

Effects of extended planning time in a listening-into-speaking task: Some preliminary observations

Chihiro Inoue & Daniel Lam
CRELLA, University of Bedfordshire, UK

Background

- A growing body of research focused on integrated tasks in the field of language testing (e.g. Brown et al., 2005; Lee, 2006; Sawaki et al., 2009, Chan, 2013, Rukthong, 2016)
- Studies on integrated *speaking* tasks still account for a relatively small percentage
- Several studies examined the constructs of listening-into-speaking tasks by analysing the scores (e.g. Lee, 2006; Sawaki et al., 2009), expert judgements (Brown et al., 2005) and elicited performances (Brown et al., 2005; Frost et al., 2011), but very few examined the cognitive processes

TOEFL iBT lecture listening-into-speaking task

- Cognitive processes and skills that are required to successfully complete this type of tasks are highly relevant to what students do when they want to check understanding and ask questions after a lecture, report and discuss the lecture content with a tutor or study group members afterwards
- In real-life, students will take time to plan what they say/report
- Is 20 seconds enough for such planning?

Effects of short planning time

Study by	Planning	Key findings
Wigglesworth (1997)	None 1 min	Planning only helped the more highly proficient learners on complexity and accuracy on more cognitively demanding tasks
Mehnert (1998)	None 1 min 5 min 10 min	<ul style="list-style-type: none"> • Fluency and lexical density improved with each increase of planning time • Accuracy increased only with 1 min • Complexity was higher only with 10 min
Elder & Iwashita (2005)	1.25 min 4.25 min	No significant effects on CAF or candidate perceptions of task
Elder & Wigglesworth (2006)	1 min 2 min	No significant impact on scores (of IELTS Part 2 long-turn task) or on CAF
Li, Chen & Li (2015)	None 30 sec 1 min 2 min 3 min 5 min	<ul style="list-style-type: none"> • 30 sec was insufficient for any improvement; even detrimental to fluency compared with no planning • 1-min was the threshold that led to significant improvement on accuracy • Longer planning time produced gradually more accurate but not steadily more fluent or complex utterances; the degrees of improvement decreased • 5 min had diminishing effects • Candidate preferred 1-3 min planning time

Use of planning time x levels

- Ortega (2005) examined the process of planning through a retrospective method and found that the low-level participants often reported using retrieval strategies while advanced learners used more rehearsal strategies while planning
- Planning would be beneficial for higher level participants in improving the formal aspects of performance, and lower level participants may need to allocate more attention and time to prioritize planning the content and lexis (Wigglesworth ,1997; Ortega, 1995; 1999)

Research Questions

1. Do candidates plan for language and mentally rehearse their speech with the current 20-second planning time provided in the TOEFL iBT Integrated Speaking (Lecture and Question) tasks?
2. Do candidates **start planning for language and mentally rehearsing their speech with a longer length of planning time (to be piloted)**, especially at higher levels of proficiency?
3. Do the longer length of planning time (to be piloted) lead to **better performance in terms of the linguistic measures of CAF and integration of the lecture input?**
4. Do the longer length of planning time (to be piloted) lead to higher scores in the analytic rating categories of TOEFL iBT Integrated Speaking Rubrics?

Pilot study

10 pilot participants (5 higher, 5 lower) performing 5 tasks with 5 lengths of planning time (20 sec, 60 sec, 90 sec, 120 sec, 180 sec)

ID	L1	Test score
PS01	Portuguese	IELTS Speaking (mock)* 6.0
PS02	Italian	IELTS Speaking (mock)* 7.0
PS03	Bengali	IELTS Speaking 6.0 (Overall 7.0)
PS04	Polish	IELTS Speaking (mock)* 6.5
PS05	Polish	CAE 70/100 (speaking marked as 'exceptional')
PS06	Urdu	IELTS Speaking 5.5 (Overall 5.5)
PS07	Mandarin	IELTS Speaking 5.5 (Overall 5.0)
PS08	Cantonese	IELTS Speaking 5.0 (Overall 5.0)
PS09	Arabic	IELTS Speaking 5.5 (Overall 4.5)
PS10	Arabic	IELTS Speaking 5.5 (Overall 5.0)

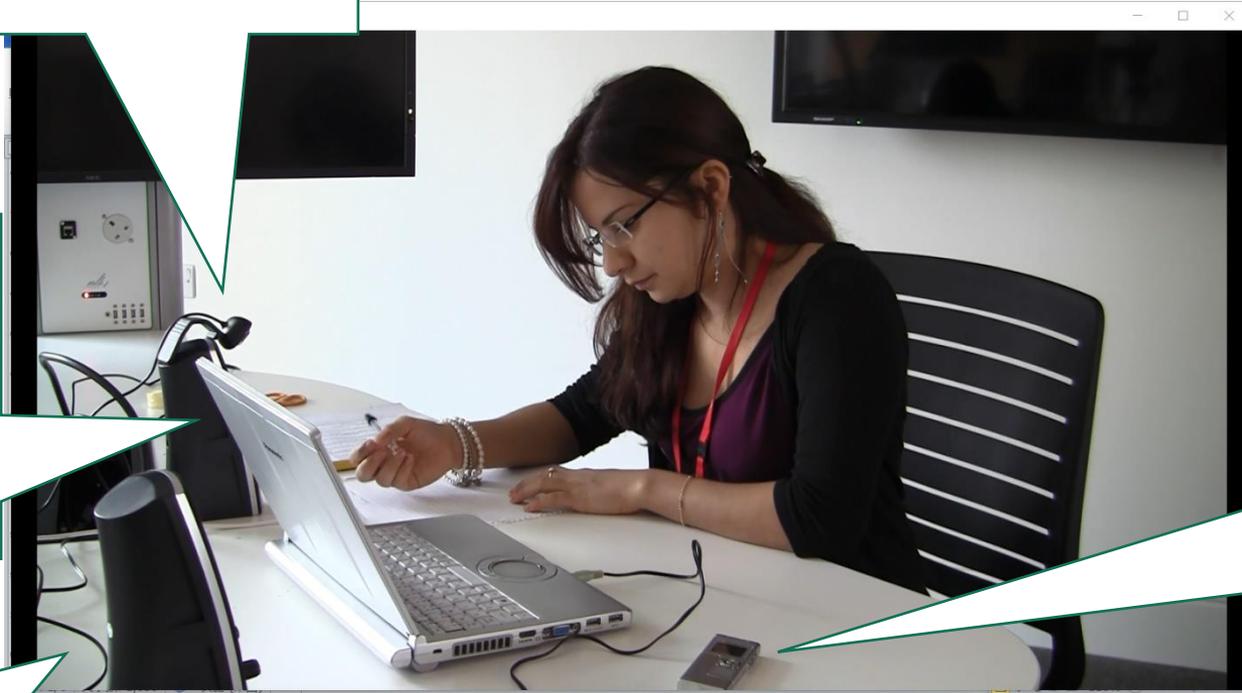
Note. *A mock IELTS exam was administered by a certified IELTS Speaking examiner as part of the preparation course.

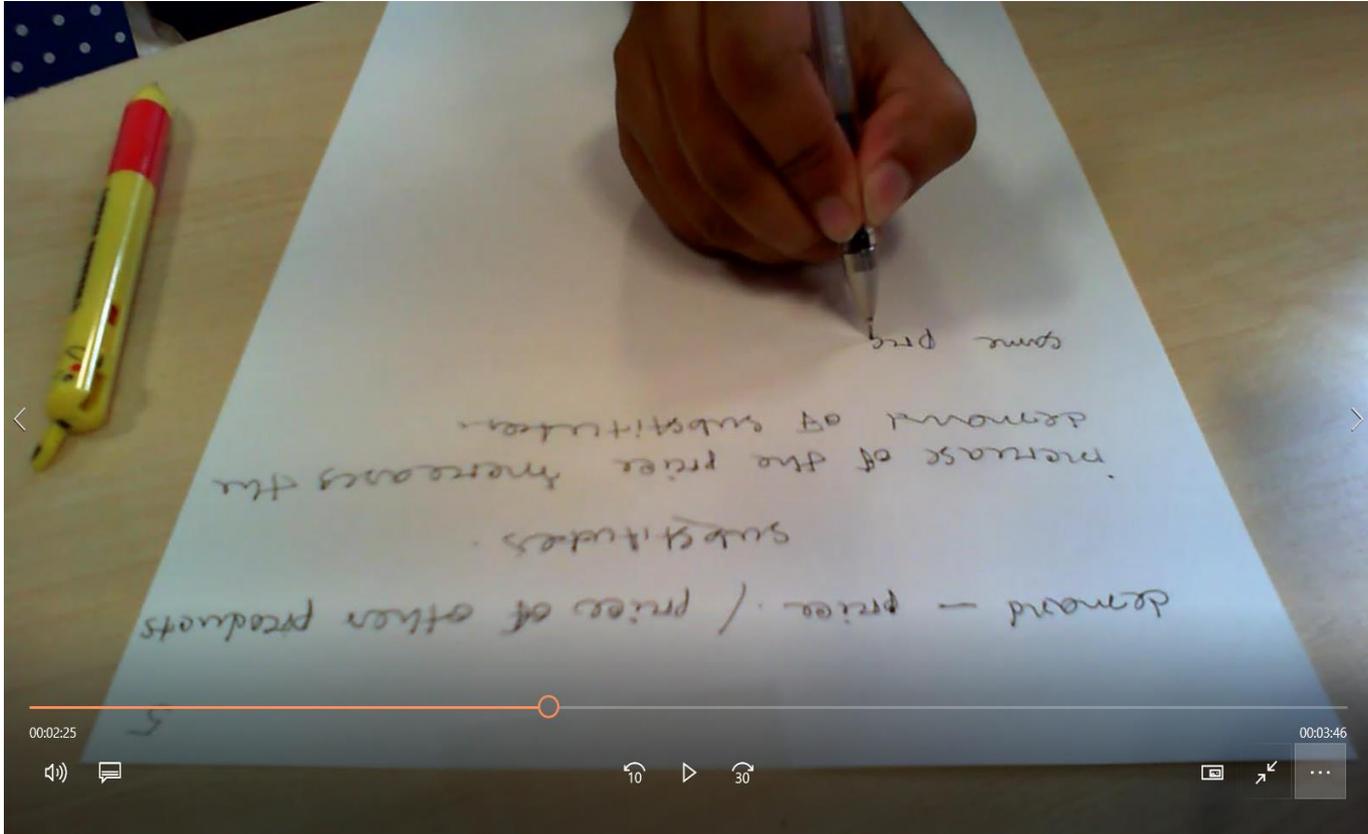
A webcam put on a speaker in front of participant's writing hand in order to capture note-taking behavior

Laptop with prompts in front of participant

One speaker on each side of laptop

Audio recorder to run throughout session





Variables for analysis

Variables	Details
Lexical complexity, syntactic complexity, accuracy and fluency (CAF)	<ul style="list-style-type: none">• Syllables per minute (Kormos & Denes, 2004)• Errors per 100 words (Mehnert, 1998; Inoue, 2016)• Weighted errors per clause ratio (Foster & Wigglesworth, 2016)• D value (VOCD) (Malvern & Richards, 1997)• Measure of textual diversity (MTLD) (McCarthy, 2005)• Percentages of words found in BNC-COCA vocabulary list (Nation & Webb, 2011; using VocabProfile (Cobb, n.d.)• Subordinate clauses per AS-unit (Crookes, 1989; Mehnert, 1998)• Number of verb elements per AS-unit (Nakatsuhara & Field, 2012)
Idea units	Idea units presented in the lecture recordings and reproduced in the spoken performances (Frost et al., 2011)
Reported behavior during task performance	Coded using thematic analysis (i.e. Swain et al., 2009)

Procedure for Idea Unit analysis

1. Segmenting *Idea Units* (IU)
 - Lecture listening input
 - Candidates' task performances
 2. Coding for lecture content reproduction
 - Types
 - Accuracy
 3. Counting the frequencies for types and accuracy of the lecture content reproduction
- Following Frost, Elder, and Wigglesworth (2011)

Lecture content reproduction: Types

1. **IU: Replication** of **individual** idea units from lecture input
2. **CIU: Combination or condensation** of two or more idea units from lecture input
3. **MP: Macro-propositions** – making generalizations or inferences of one or more idea units from the lecture input
4. **Others** – points not mentioned in the lecture input; participants' own views; IUs carrying a structuring discourse or rhetorical function.
5. **Repeat** – Ideas already mentioned in the previous IUs

Accuracy of reproduction

A. **Accurate** reproduction (A)

B. **Inaccurate** reproduction (I)

- referred to as 'distortions' in Frost et al. (2011)

C. **Partially accurate** reproduction (P)

- Capturing the general topic/idea but not specific details or relations between ideas
- Some details are accurate while others are inaccurate
- Reproducing key words but not accurately representing the idea(s)

- Coded as distortions in Frost et al. (2011)
- But potentially capture a more nuanced pattern of content reproduction (by lower proficiency candidates) (?)

Frequencies (type and accuracy)

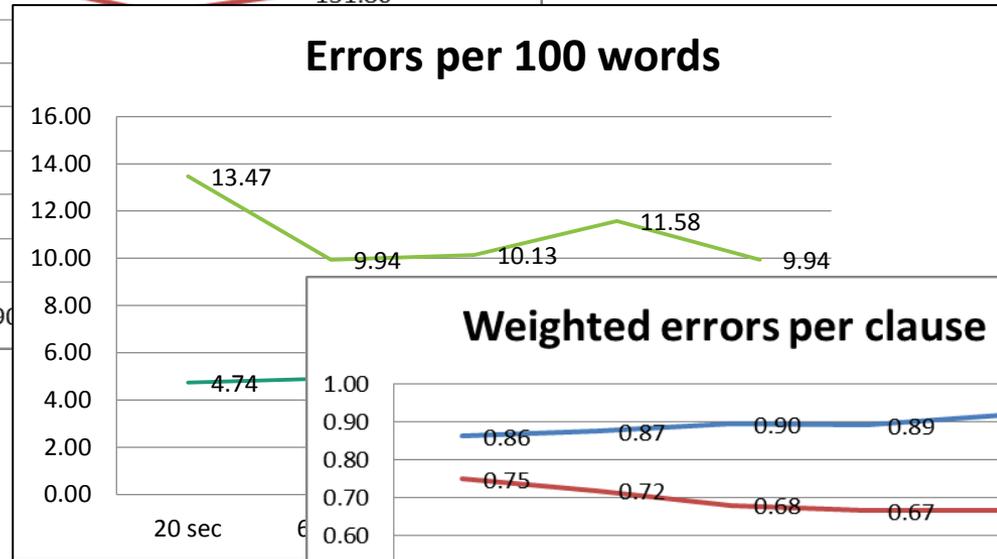
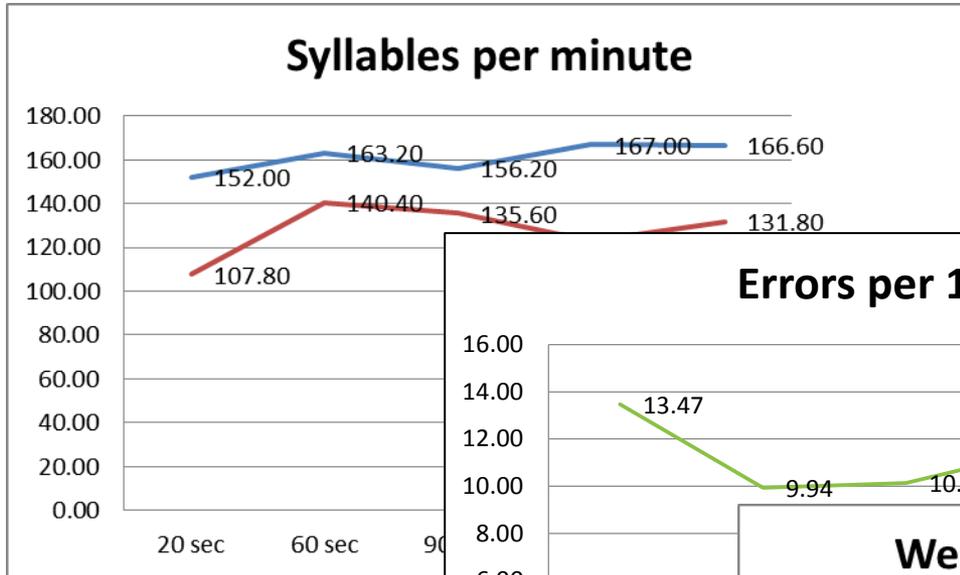
- **For each planning time length...**
- **Accuracy:** Total number of a) accurate, b) inaccurate, and c) partially accurate reproductions
- **Type:** Total number of IIU, CIU, MP reproduced
- **Accuracy + Type:** Number of a) accurate, b) inaccurate, and c) partially accurate reproductions of IIU, CIU, MP
- Length adjusted (number / total IUs reproduced)
- High group vs. low group

Results

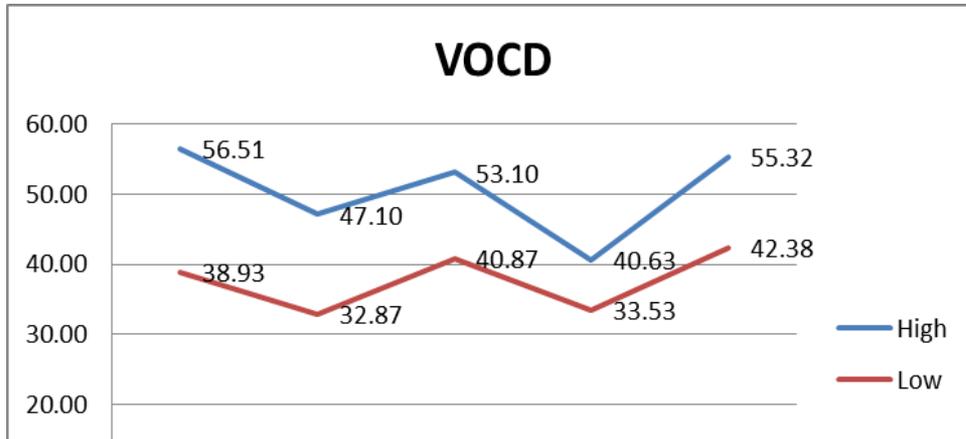
Amount of production (no. of words)

		20 sec	60 sec	90 sec	120 sec	180 sec
High	Mean	117.60	124.80	110.80	118.00	117.20
	SD	27.83	18.16	17.71	42.94	33.20
Low	Mean	71.80	96.80	92.40	89.40	88.20
	SD	9.93	38.70	34.53	40.65	25.80

CAF results: fluency & accuracy

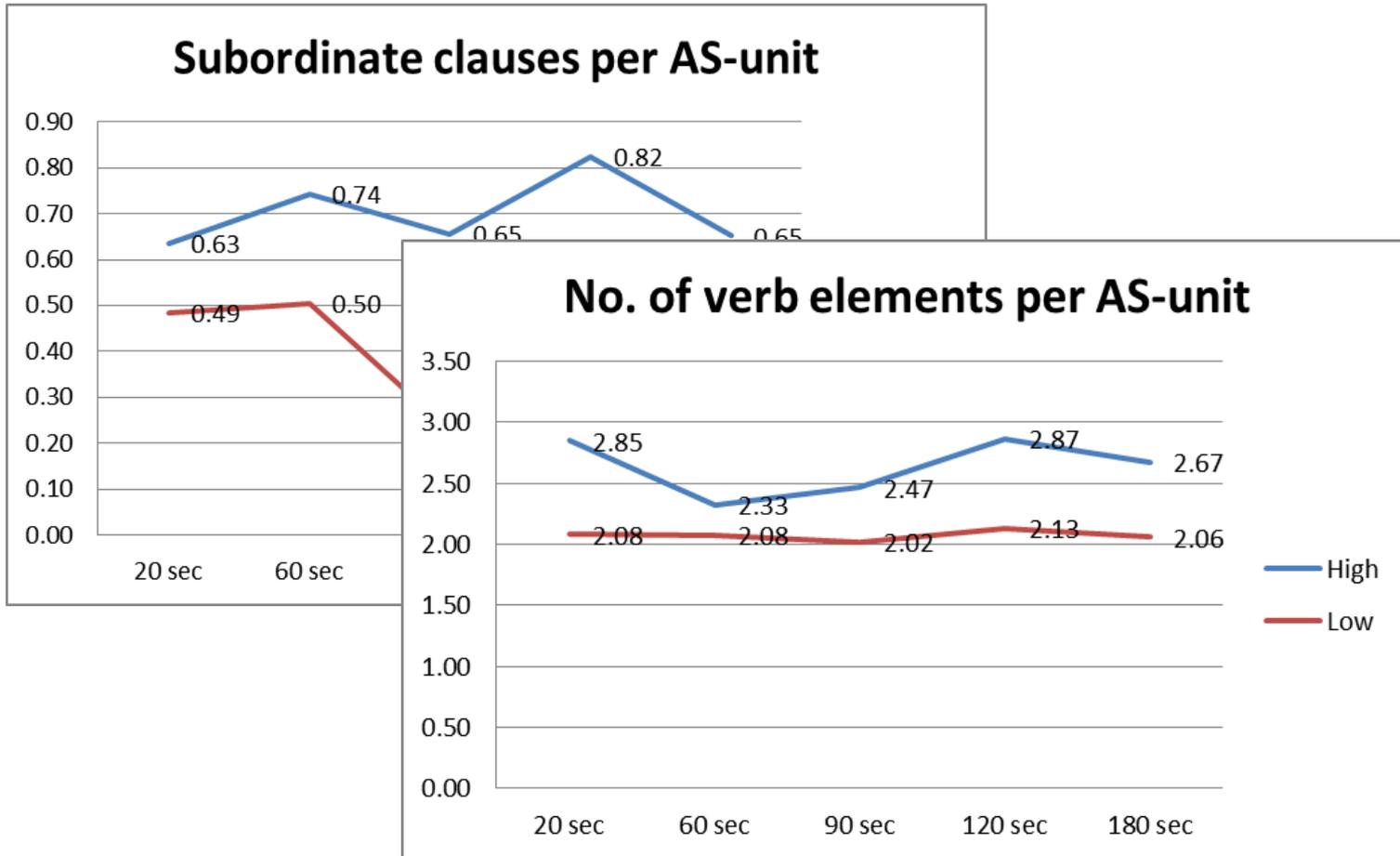


CAF results: lexical complexity



	High					Low				
	20 sec	60 sec	90 sec	120 sec	180 sec	20 sec	60 sec	90 sec	120 sec	180 sec
K-1	83.07	90.60	75.92	73.85	81.92	85.78	90.52	80.71	77.40	87.71
K-2	7.59	4.10	8.63	14.95	8.43	5.16	6.17	7.12	12.03	7.07
K-3	4.18	3.12	12.63	5.90	2.06	3.43	2.91	10.53	7.89	1.91
K-4	5.16	0.62	2.14	2.58	3.82	5.64		1.52	2.26	2.22
K-5		1.02		0.69	1.91			1.27		
K-6		2.48		1.19	0.79		1.56	0.68		0.70
K-7				2.74	2.30				2.19	1.27
K-18					1.81					1.82
Off list	0.00	0.00	0.85	1.15	0.00			0.68		

CAF results: syntactic complexity

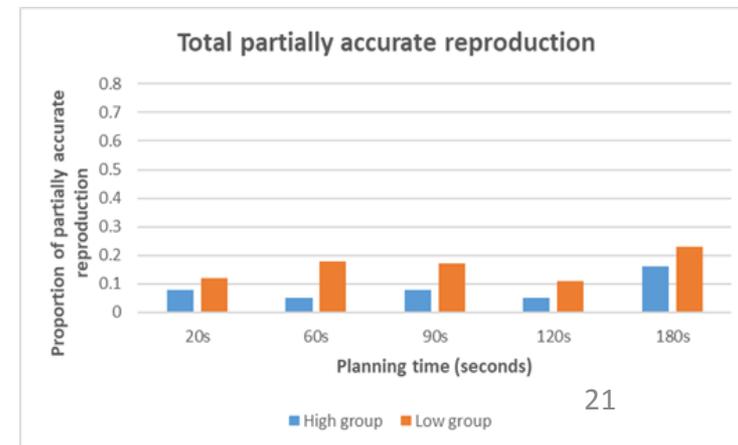
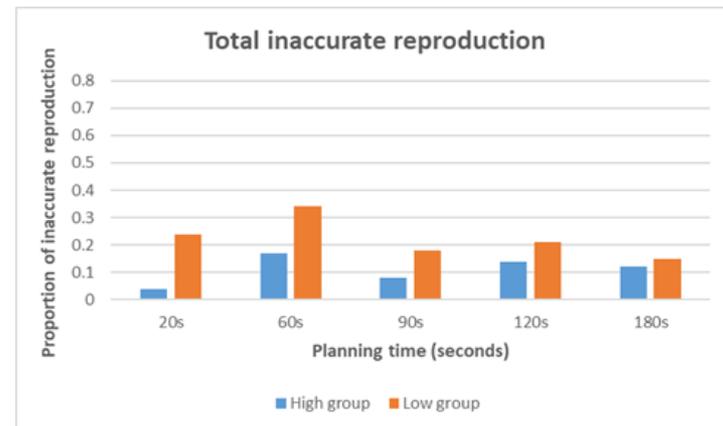
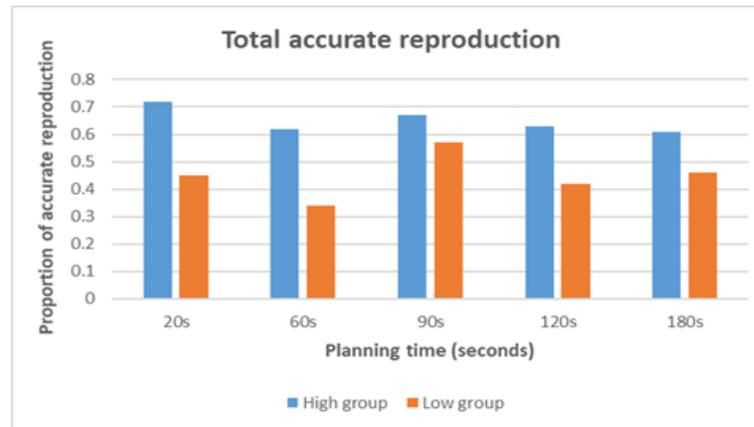


From stimulated recalls...

Candidate	Plan for language	Mentally rehearse
PS01	90, 120, 180	
PS02	60, 90	
PS03	60, 180	60, 180
PS04	All	
PS05		
PS06	20, 90, 120, 180	120, 180
PS07	90	
PS08	60, 90, 180	90
PS09	60, 90, 120, 180	
PS10	All	

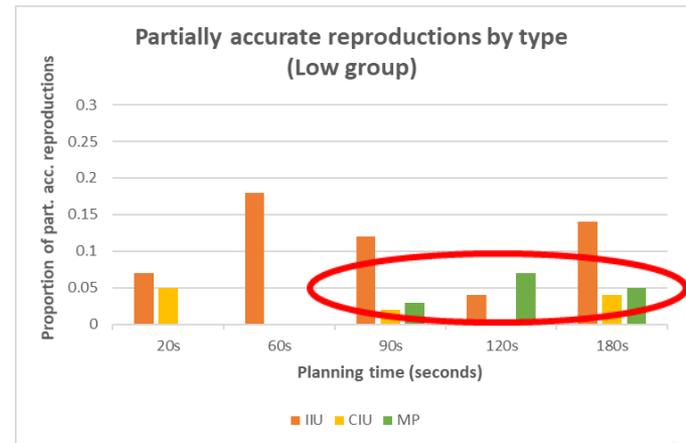
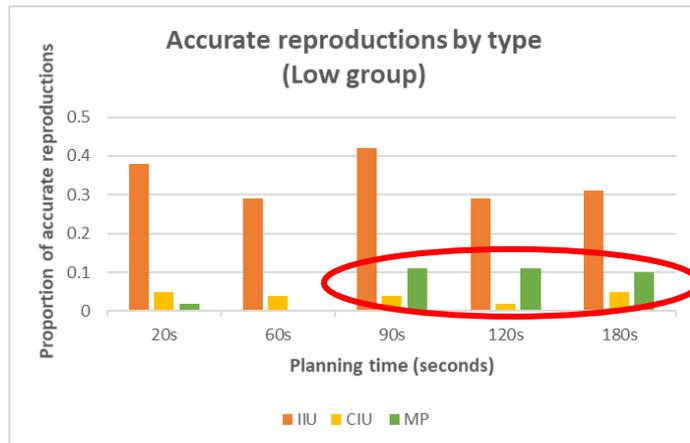
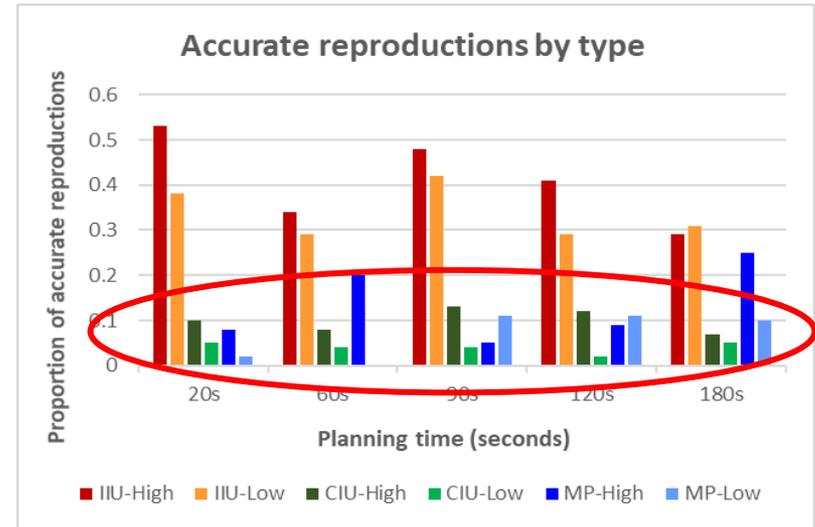
Idea units: Accuracy (overall)

- High group
 - higher number of A
- Low group
 - higher number of I & P
- No clear relationship with planning time length



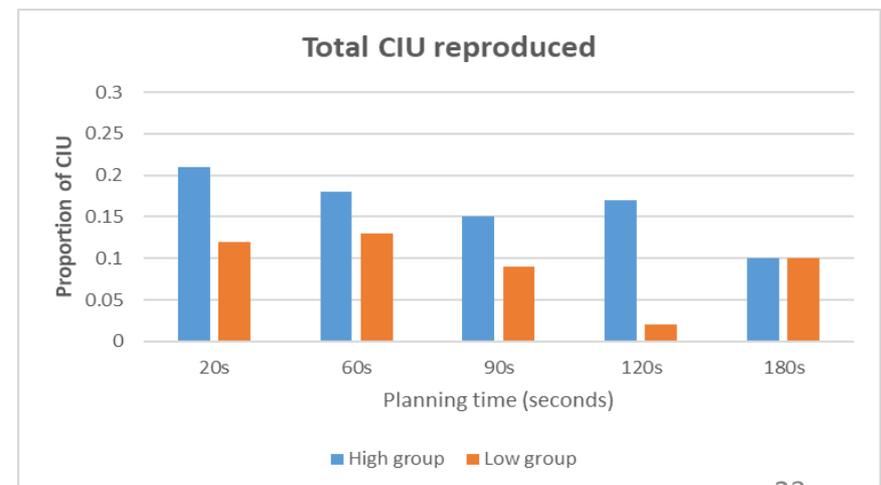
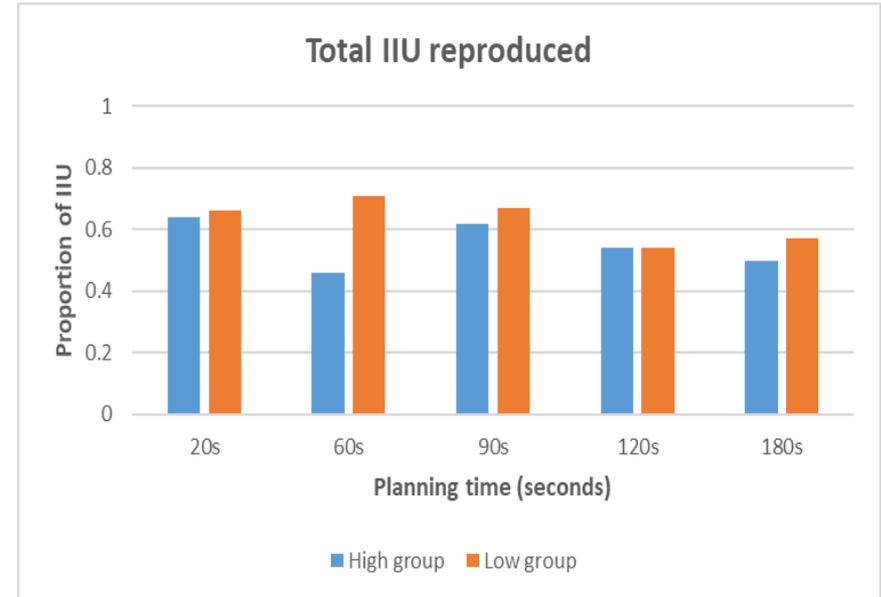
Idea units: Accuracy (by type)

- More CIU and MP than in Frost et al. (2011)
- No clear patterning with planning time length, but among low group...
 - Small proportions of MPs at 90s, 120s, 180s
 - ~ zero at 20s and 60s



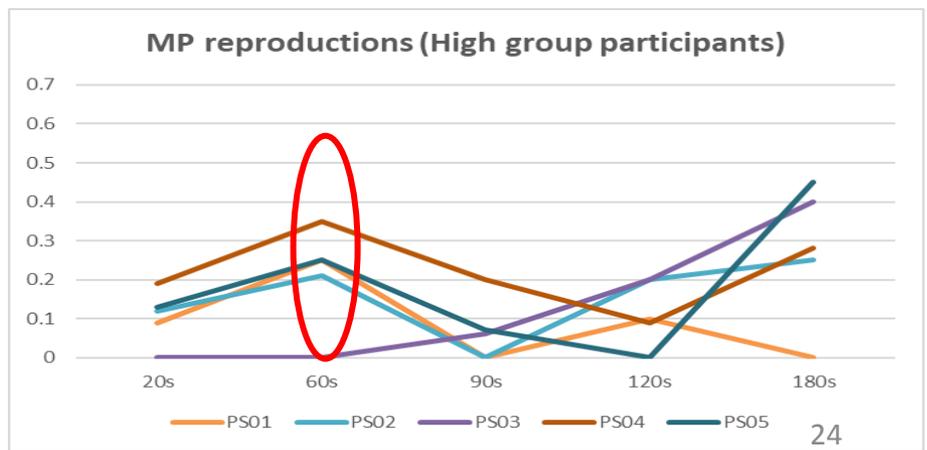
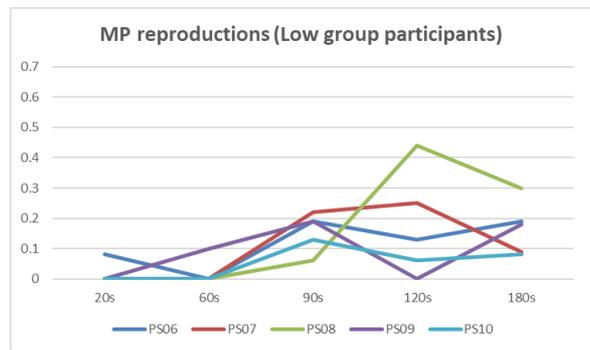
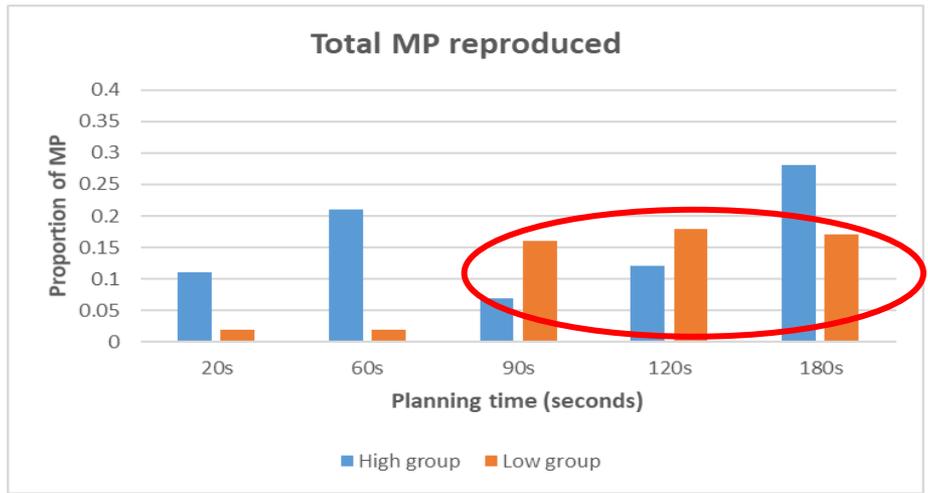
Idea units: Total reproductions (IIU and CIU)

- No clear patterning with planning time length for IIU and CIU



Idea units: Total reproductions (MP)

- Low group:
 - Attempted more MP reproductions at 90s, 120s, 180s
- High group:
 - 4 of 5 participants showed an increase of MPs at 60s
 - No further increase at 90s and 120s
 - 4 of 5 had highest MPs at 180s



Planning time and lecture content reproduction: Some speculations...

- Longer planning time → revisit the lecture content
 - Esp. for lower proficiency candidates
 - **Compensating for limited comprehension** while listening
 - Reported behavior during planning time
 - ‘reviewing notes to make sense of their own notes’
 - ‘trying to remember what was said in the lecture’
- Longer planning time → **make generalizations or inferences** from key points in the lecture
 - rather than simply reporting individual ideas or summarizing them
 - But...benefit less clear among higher proficiency candidates

Discussion

Summary of results:

Linguistic performance, content reproduction, and reported planning behavior

Aspect of performance	Pattern
Fluency	Might increase with 60-second planning time
Accuracy	Might increase slightly as planning time increases
Lexical complexity	Might not change across planning time lengths; no differential effects for high/low proficiency groups
Syntactic complexity	Might not change across planning time lengths; no differential effects for high/low proficiency groups
Lecture content reproduction	More MPs attempted at 60s for high group; More MPs attempted at 90s+ for low group
Reported planning behavior	Planning for language and mental rehearsal might be reported from 60 or 90 seconds

Discussion

- Lexical and syntactic complexity not influenced by planning time?
 - Perhaps participants focused more on reviewing notes and organizing thoughts, rather than on using more sophisticated words or more complex structures
- Little effect on reproduction of lecture content (for high group)?
 - An important mediating factor: how much the candidate understood while listening

Looking forward...

Implications for the main study

- Larger sample (n = 70)
 - 35 higher proficiency, 35 lower proficiency
- Two planning time conditions
 - Original planning time: 20s
 - Extended planning time: 90s (TBC)
 - Counter-balanced with task/topic
- Performances scored by trained ETS raters
 - Holistic AND analytic scores
- Examining reported planning behavior
 - Stimulated recall (n = 16)
 - Questionnaire/checklist (n = 54)

References

- Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks* (TOEFL Monograph Series MS-29). Princeton, NJ: Educational Testing Service.
- Chan, S. (2013). *Establishing the validity of reading-into-writing test tasks for the UK academic context*. (Doctoral dissertation). Retrieved from University of Bedfordshire Repository at <http://uobrep.openrepository.com/uobrep/handle/10547/312629>
- Cobb, T. (n.d.). Web Vocabprofile. Accessed from <http://www.lextutor.ca/vp/>
- Crookes, G. (1989). Planning and interlanguage variability. *Studies in Second Language Acquisition*, 11, 367-383.
- Elder, C. & Iwashita, N. Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp.219-238). Amsterdam: John Benjamins.
- Elder, C. & Wigglesworth, G. (2006). An investigation of the effectiveness and validity of planning time in Part 2 of the IELTS Speaking Test. *IELTS Research Report*, 6, 1-28.
- Foster, P. & Wigglesworth, G. (2016). Capturing accuracy in second language performance: The case for a Weighted Clause Ratio. *Annual Review of Applied Linguistics*, 36, 98–116.
- Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test-takers' oral performances. *Language Testing*, 29(3), 345-369.
- Inoue, C. (2016). A comparative study of the variables used to measure syntactic complexity and accuracy in task-based research. *Language Learning Journal*, 44(4), 487-505.
- Jarvis, S. (2002). Short texts, best-fitting curves and new measures of lexical diversity. *Language Testing*, 19(1), 57-84.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System*, 32, 145-164.
- Lee, Y.-W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independent tasks. *Language Testing*, 23(2), 131-166.

- Li, L., Chen, J., & Sun, L. (2015). The effects of different lengths of pretask planning time on L2 learners' oral test performance. *TESOL Quarterly*, 49(1), 38-66.
- Malvern, D. & Richards, B. (2012). Measure of lexical richness. *The Encyclopedia of Applied Linguistics* (pp.3622-3627).
- Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition*, 20, 83-108.
- Nakatsuhara, F. & Field, J. (2012). *An investigation of examiner interventions in relation to the listening demands that they make on candidates in the GESE exams*. Unpublished project report submitted to Trinity College London.
- Ortega, L. 1995. *The effects of planning in L2 Spanish oral narratives* (Research Note No. 15). Honolulu: University of Hawai'i, Second Language Teaching and Curriculum Centre.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21(1), 109-148.
- Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pre-task planning. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp.77-109). Amsterdam: John Benjamins.
- Rukthong, A. (2016). *Investigating the listening construct underlying listening-to-summarize tasks*. (Doctoral dissertation). Retrieved from Lancaster University Research Directory at <http://eprints.lancs.ac.uk/78054/1/2016ARukthongPhD>
- Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL Internet-based test. *Language Testing*, 26(1), 5-30.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14, 101-122.
- Wigglesworth, G., & Elder, C. (2010). An investigation of the effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly*, 7, 1-24.