



# Enabling interaction: A study of examiner behaviour in an interactive speaking test

A mixed methods study of the Trinity  
GESE Interactive Task

Richard Harris

# Graded Examinations in Spoken English (GESE)

- One examiner, one candidate
- GESE Grades 7 and 8 – CEFR B2
- Three tasks
  - Topic Discussion
  - Interactive Task
  - Conversation
- Interactive task generates candidate-led interaction

# Purpose of the study

- To investigate what strategies are used by examiners in the GESE Interactive Task
- To investigate whether the strategies found have an effect on objective measures of candidate performance

# Corpus transcripts



- Complete transcriptions of Interactive Tasks extracted from the corpus
- Examined linguistic and para-linguistic behaviour of both examiners and candidates
- 54 transcripts
- 37 Interactive Prompts
- 10 nationalities

# Methodology

- Examiners
  - Grounded theory developed a taxonomy of strategies that examiners use to allow candidates to take the lead in interaction
  - Identified 8 strategies
- Candidates
  - Quantitative analysis of the candidates' responses to the different strategies
  - Fluency
  - Lexis
  - Grammar

# Examiner strategies

- ACCEPTING
- HOLDING
- INAUTHENTICITY
- PARSIMONY
- REFORMULATING
- REJECTING
- RESOLVING
- STEERING

# Examiner strategies

- ACCEPTING
  - HOLDING
  - INAUTHENTICITY
  - PARSIMONY
  - REFORMULATING
  - REJECTING
  - RESOLVING
  - STEERING
- Behaviour such as
    - Clarifying
    - Confirming
    - Repeating
    - Restating the original dilemma

# Example - REFORMULATING

*My nephew wants to work in the music business when he leaves school, but his parents think he should go to university. I'm not sure I agree with them.*

<S> and er if he er can find a good job to to go to live to live to have a good life why not? </u>

<E> but he </u>

<S> do you agree? </u>

<E> he doesn't want to go to university he wants to leave school at sixteen </u>



# Examiner strategies

- ACCEPTING
- HOLDING
- INAUTHENTICITY
- PARSIMONY
- REFORMULATING
- REJECTING
- RESOLVING
- STEERING

# Examiner strategies

- ACCEPTING
  - HOLDING
  - INAUTHENTICITY
  - PARSIMONY
  - REFORMULATING
  - REJECTING
  - RESOLVING
  - STEERING
- A neutral response that maintains the dialogue but doesn't appear to make demands upon its direction
  - May include information intended to add grist to the mill of the candidate's questioning

# Examiner strategies

- ACCEPTING
  - HOLDING
  - INAUTHENTICITY
  - PARSIMONY
  - REFORMULATING
  - REJECTING
  - RESOLVING
  - STEERING
- Overtly asking for solutions
  - Overtly or more subtly steering a candidate towards an outcome apparently preferred by the examiner
  - Proffering own solutions
  - Raising the stakes
  - Pseudo-passivity

# Example – HOLDING

*My teenage nephew has decided that he wants to take up boxing as a way to get fit. I'm not sure it's such a great idea.*

<E> I I don't know really I think erm perhaps he's got some new friends who are older than him and and they are really into boxing  
</u>

<S> mm </u>

<E> and so maybe it's because of his friends you know </u>

<S> mm </u>

<E> and they're they're not nice boys they're bad boys you know </u>

# Example – HOLDING

*My teenage nephew has decided that he wants to take up boxing as a way to get fit. I'm not sure it's such a great idea.*

<E> I I don't know really I think erm perhaps he's got some new friends who are older than him and and they are really into boxing  
</u>

<S> mm </u>

<E> and so maybe it's because of his friends you know </u>

<S> mm </u>

<E> and they're they're not nice boys they're bad boys you know </u>

# Example – STEERING

*My teenage nephew has decided that he wants to take up boxing as a way to get fit. I'm not sure it's such a great idea.*

*<E> I I don't know really I think erm perhaps he's got some new friends who are older than him and and they are really into boxing </u>*

*<S> mm </u>*

*<E> and so maybe it's because of his friends you know </u>*

*<S> mm </u>*

*<E> and they're they're not nice boys they're bad boys you know </u>*

# Examiner strategies

- ACCEPTING
- HOLDING
- INAUTHENTICITY
- PARSIMONY
- REFORMULATING
- REJECTING
- RESOLVING
- STEERING

# Examiner strategies

- ACCEPTING
  - HOLDING
  - INAUTHENTICITY
  - PARSIMONY
  - REFORMULATING
  - REJECTING
  - RESOLVING
  - STEERING
- Giving limited information in response to a request (in an apparently unforthcoming way)
  - Giving a minimal response to prompt the candidate to contribute / question more



# Example - PARSIMONY

*My nephew used to dress very well, but now he's totally changed his appearance and I'm not sure now what to think about it.*

<S> yes so er nowa= he has just changed or want to change </u>

<E> no he has changed </u>

<S> has changed </u>

<E> totally changed </u>

<S> totally changed </u>

<E> yeah his appearance yes </u>

<S> okay well it's a pity </u>

<E> mm </u>

<S> no </u>

<E> mm </u>

# What follows the strategy? Performance

- Extracted all candidate responses that followed the strategies
- Across 8 strategies, 886 examples of performance
- Created three responses
  - Original response (backchannels removed)
  - Response with pauses removed
  - Pruned response (after Iwashita, Brown, McNamara & O'Hagan, 2008)

# What follows the strategy? Performance

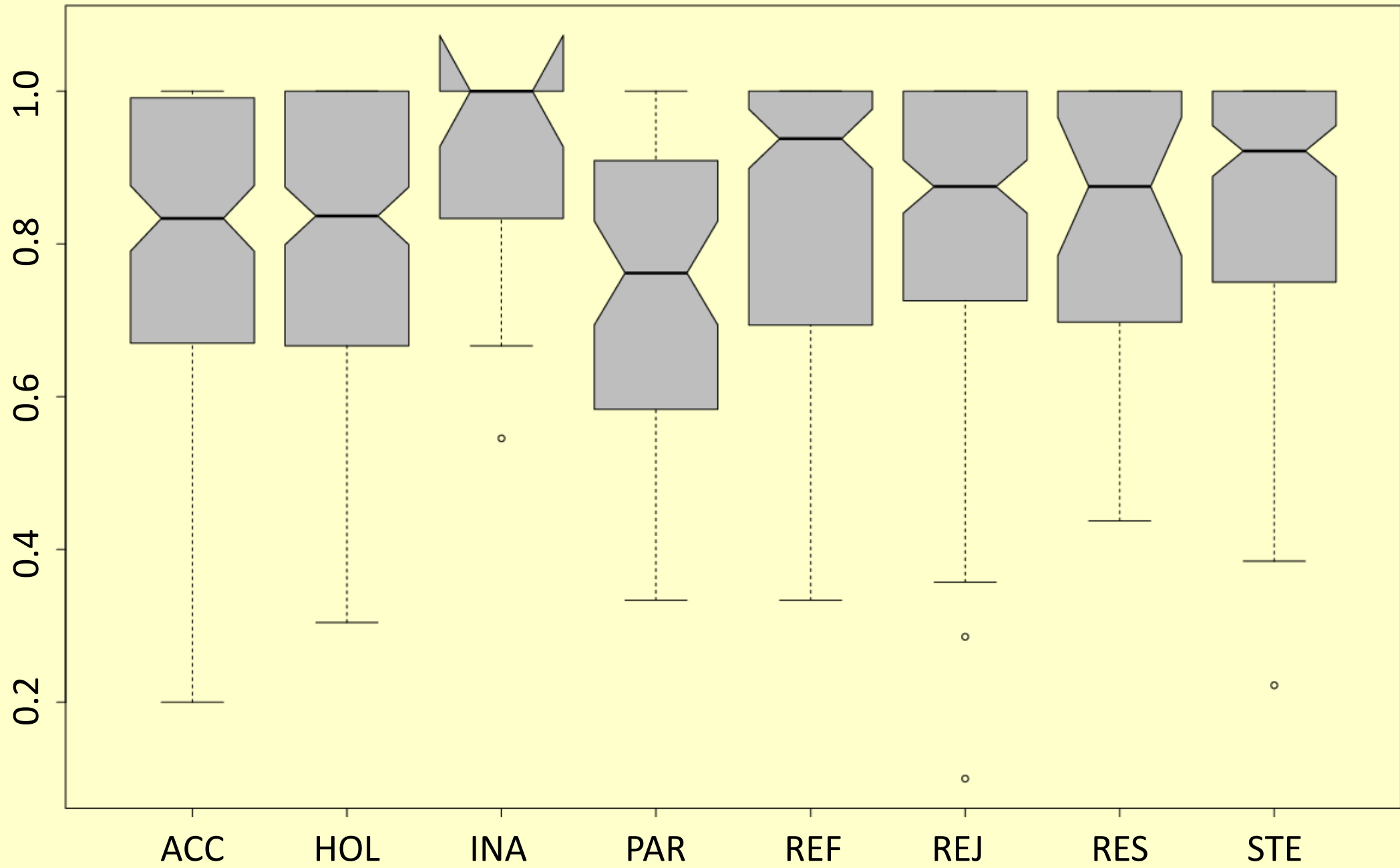
- Fluency
  - Length of response
  - Reformulation ratio
  - Pause ratio
- Lexis
  - TTR
  - COCA Range
  - COCA Frequency
- Grammar
  - Grammar errors
  - Complex t-unit ratio
  - Complex nominals per t-unit

(after Iwashita et al., 2008)

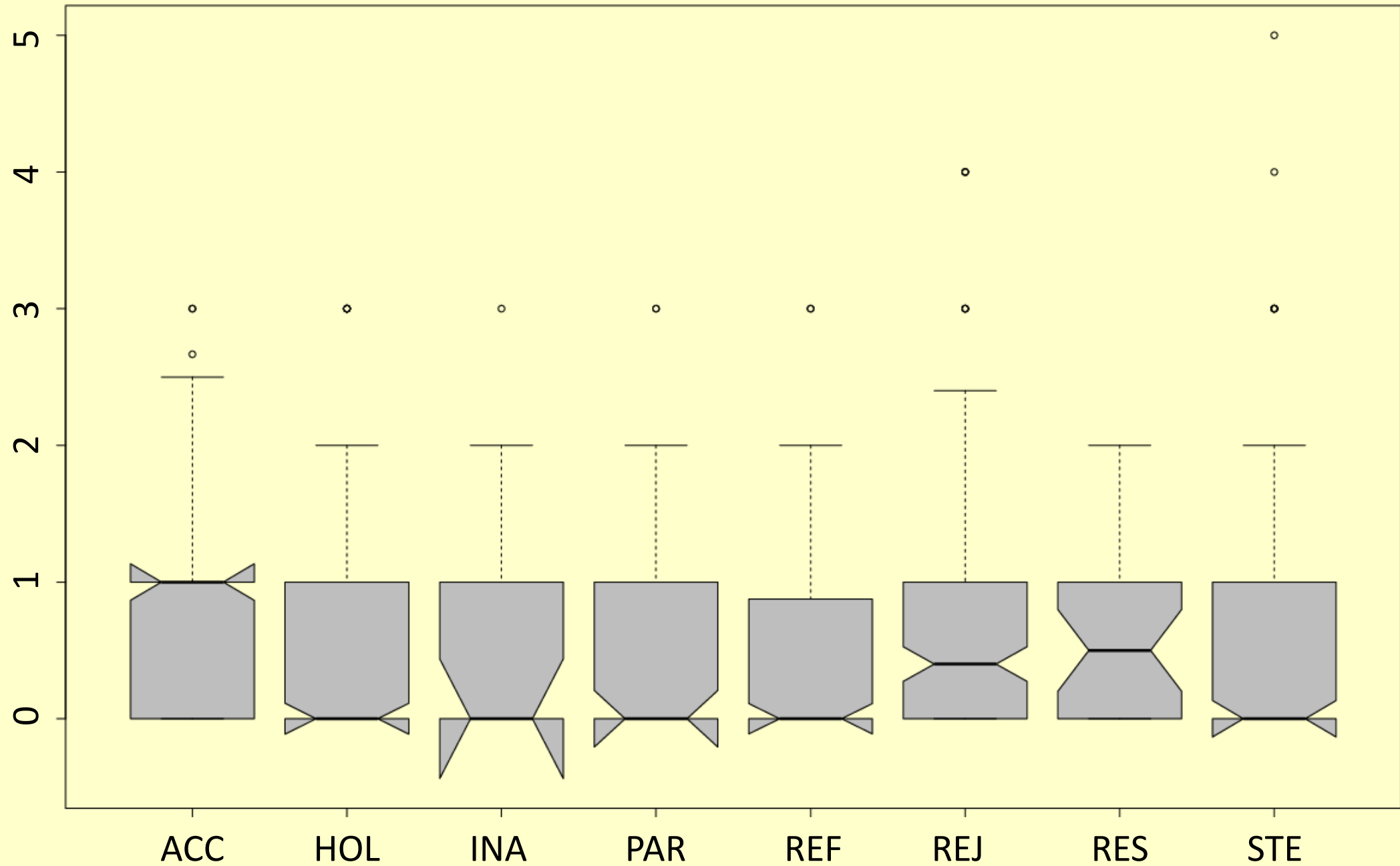
The logo for SiNLP is a rounded rectangular button with a gradient from dark orange to light orange, featuring the text "SiNLP" in white.The logo for TAASSC is a blue rounded rectangular button with a slight shadow, featuring the text "TAASSC" in orange.The logo for GAMET is a purple rounded rectangular button with a slight shadow, featuring the text "GAMET" in orange.The logo for TAALES is a green rounded rectangular button with a slight shadow, featuring the text "TAALES" in grey.

<https://www.linguisticanalysisistools.org/tools.html>

Reformulation Ratio



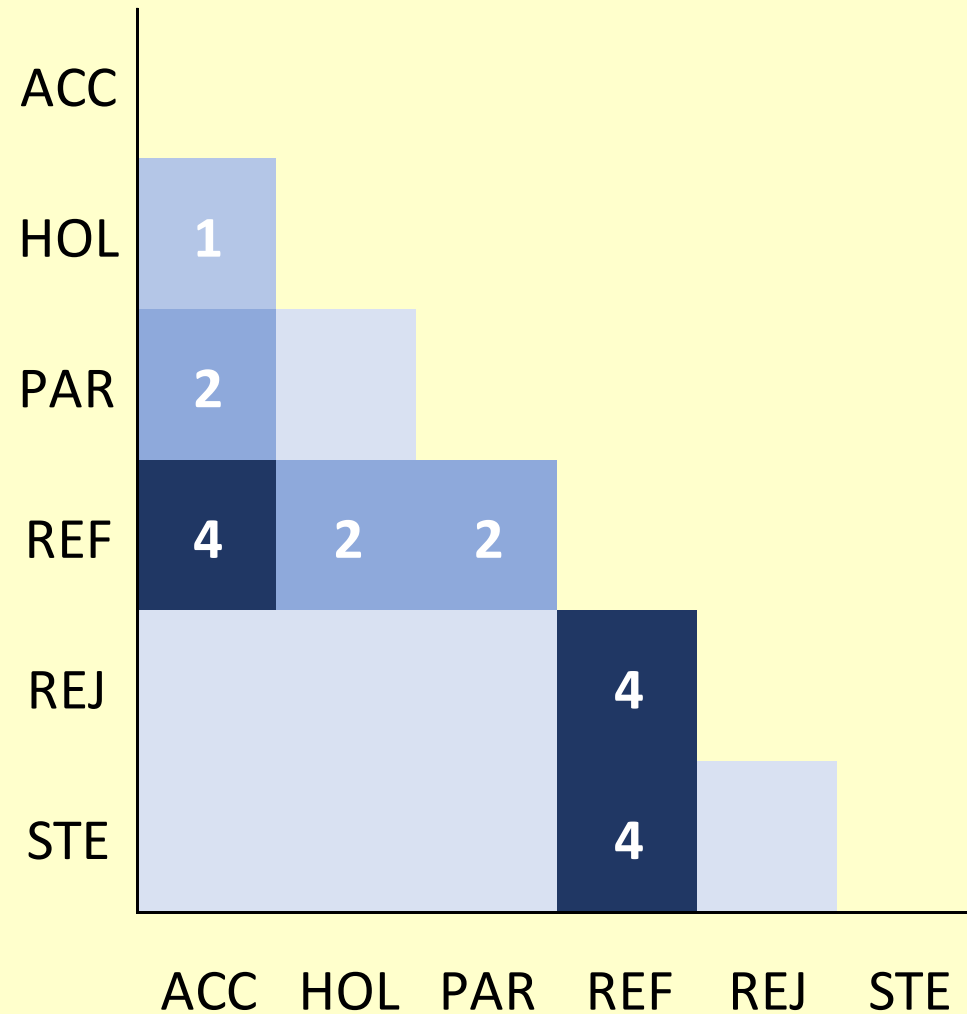
Complex Nominals per T-Unit



# Omnibus Kruskal-Wallis H Tests

	$\chi^2$	<i>d.f.</i>	<i>n</i>	<i>p-value</i>	Significant Pairwise Comparisons						
Length	34.458	5	508	0.000 (sig.)		REF-HOL	REF-STE	REF-REJ	REF-ACC		
Reform Ratio	26.465	5	508	0.000 (sig.)	PAR-REF	HOL-REF					
Pause Ratio	18.784	5	508	0.002 (sig.)	PAR-REF		STE-REF				
TTR	28.591	5	508	0.000 (sig.)				REJ-REF	ACC-REF		
COCA Range	7.239	5	508	0.203 (n.s.)							
COCA Freq	8.763	5	508	0.119 (n.s.)							
Error	11.295	5	508	0.046 (sig.)							
Complex T	34.101	5	508	0.000 (sig.)			REF-STE	REF-REJ	REF-ACC	HOL-ACC	PAR-ACC
Complex Nom	23.693	5	508	0.000 (sig.)			REF-STE	REF-REJ	REF-ACC		PAR-ACC

# Significant Pairwise Comparisons



# Post-hoc Mann-Whitney U Tests with mean ranks and effect sizes

Measure	Strategy 1	Mean rank	Strategy 2	Mean rank	Z	n	p	r	R <sup>2</sup>	Effect size (Cohen, 1988)
Length	REF	102.88	HOL	78.12	-3.199	180	0.001	0.24	0.06	small
Length	REF	79.53	STE	101.47	-2.842	180	0.004	0.21	0.04	small
Length	REF	72.58	REJ	108.42	-4.631	180	0.000	0.35	0.12	medium
Length	REF	70.32	ACC	110.68	-5.213	180	0.000	0.39	0.15	medium
Reform Ratio	PAR	57.78	REF	85.27	-3.990	148	0.000	0.33	0.11	medium
Reform Ratio	HOL	75.73	REF	105.27	-3.950	180	0.000	0.29	0.09	small
Pause Ratio	PAR	59.46	REF	84.19	-3.654	148	0.000	0.30	0.09	medium
Pause Ratio	STE	78.98	REF	102.02	-3.165	180	0.002	0.24	0.06	small
TTR	REJ	75.41	REF	105.59	-4.608	180	0.000	0.34	0.12	medium
TTR	ACC	76.28	REF	104.72	-4.343	180	0.000	0.32	0.10	medium
Complex T	REF	80.12	STE	100.88	-3.400	180	0.001	0.25	0.06	small
Complex T	REF	77.79	REJ	103.21	-4.028	180	0.000	0.30	0.09	medium
Complex T	REF	74.48	ACC	106.52	-4.925	180	0.000	0.37	0.13	medium
Complex T	HOL	77.87	ACC	103.13	-3.767	180	0.000	0.28	0.08	small
Complex T	PAR	63.85	ACC	81.36	-2.743	148	0.006	0.23	0.05	small
Complex Nom	REF	80.29	STE	100.71	-2.968	180	0.003	0.22	0.05	small
Complex Nom	REF	79.21	REJ	101.79	-3.241	180	0.001	0.24	0.06	small
Complex Nom	REF	74.53	ACC	106.47	-4.484	180	0.000	0.33	0.11	medium
Complex Nom	PAR	63.05	ACC	81.88	-2.787	148	0.005	0.23	0.05	small



# Post-hoc Mann-Whitney U Tests with mean ranks and effect sizes

Measure	Strategy 1	Mean rank	Strategy 2	Mean rank	Z	n	p	r	R <sup>2</sup>	Effect size (Cohen, 1988)
Length	REF	102.88	HOL	78.12	-3.199	180	0.001	0.24	0.06	small
Length	REF	79.53	STE	101.47	-2.842	180	0.004	0.21	0.04	small
Length	REF	72.58	REJ	108.42	-4.631	180	0.000	0.35	0.12	medium
Length	REF	70.32	ACC	110.68	-5.213	180	0.000	0.39	0.15	medium
<b>Reform Ratio</b>	<b>PAR</b>	<b>57.78</b>	<b>REF</b>	<b>85.27</b>	-3.990	148	0.000	<b>0.33</b>	0.11	<b>medium</b>
<b>Reform Ratio</b>	<b>HOL</b>	<b>75.73</b>	<b>REF</b>	<b>105.27</b>	-3.950	180	0.000	<b>0.29</b>	0.09	<b>small</b>
Pause Ratio	PAR	59.46	REF	84.19	-3.654	148	0.000	0.30	0.09	medium
Pause Ratio	STE	78.98	REF	102.02	-3.165	180	0.002	0.24	0.06	small
TTR	REJ	75.41	REF	105.59	-4.608	180	0.000	0.34	0.12	medium
TTR	ACC	76.28	REF	104.72	-4.343	180	0.000	0.32	0.10	medium
Complex T	REF	80.12	STE	100.88	-3.400	180	0.001	0.25	0.06	small
Complex T	REF	77.79	REJ	103.21	-4.028	180	0.000	0.30	0.09	medium
Complex T	REF	74.48	ACC	106.52	-4.925	180	0.000	0.37	0.13	medium
Complex T	HOL	77.87	ACC	103.13	-3.767	180	0.000	0.28	0.08	small
Complex T	PAR	63.85	ACC	81.36	-2.743	148	0.006	0.23	0.05	small
Complex Nom	REF	80.29	STE	100.71	-2.968	180	0.003	0.22	0.05	small
Complex Nom	REF	79.21	REJ	101.79	-3.241	180	0.001	0.24	0.06	small
Complex Nom	REF	74.53	ACC	106.47	-4.484	180	0.000	0.33	0.11	medium
Complex Nom	PAR	63.05	ACC	81.88	-2.787	148	0.005	0.23	0.05	small

# Post-hoc Mann-Whitney U Tests with mean ranks and effect sizes

Measure	Strategy 1	Mean rank	Strategy 2	Mean rank	Z	n	p	r	R <sup>2</sup>	Effect size (Cohen, 1988)
Length	REF	102.88	HOL	78.12	-3.199	180	0.001	0.24	0.06	small
Length	REF	79.53	STE	101.47	-2.842	180	0.004	0.21	0.04	small
Length	REF	72.58	REJ	108.42	-4.631	180	0.000	0.35	0.12	medium
Length	REF	70.32	ACC	110.68	-5.213	180	0.000	0.39	0.15	medium
Reform Ratio	PAR	57.78	REF	85.27	-3.990	148	0.000	0.33	0.11	medium
Reform Ratio	HOL	75.73	REF	105.27	-3.950	180	0.000	0.29	0.09	small
Pause Ratio	PAR	59.46	REF	84.19	-3.654	148	0.000	0.30	0.09	medium
Pause Ratio	STE	78.98	REF	102.02	-3.165	180	0.002	0.24	0.06	small
TTR	REJ	75.41	REF	105.59	-4.608	180	0.000	0.34	0.12	medium
TTR	ACC	76.28	REF	104.72	-4.343	180	0.000	0.32	0.10	medium
Complex T	REF	80.12	STE	100.88	-3.400	180	0.001	0.25	0.06	small
Complex T	REF	77.79	REJ	103.21	-4.028	180	0.000	0.30	0.09	medium
Complex T	REF	74.48	ACC	106.52	-4.925	180	0.000	0.37	0.13	medium
Complex T	HOL	77.87	ACC	103.13	-3.767	180	0.000	0.28	0.08	small
Complex T	PAR	63.85	ACC	81.36	-2.743	148	0.006	0.23	0.05	small
<b>Complex Nom</b>	<b>REF</b>	<b>80.29</b>	<b>STE</b>	<b>100.71</b>	-2.968	180	0.003	<b>0.22</b>	0.05	<b>small</b>
<b>Complex Nom</b>	<b>REF</b>	<b>79.21</b>	<b>REJ</b>	<b>101.79</b>	-3.241	180	0.001	<b>0.24</b>	0.06	<b>small</b>
<b>Complex Nom</b>	<b>REF</b>	<b>74.53</b>	<b>ACC</b>	<b>106.47</b>	-4.484	180	0.000	<b>0.33</b>	0.11	<b>medium</b>
<b>Complex Nom</b>	<b>PAR</b>	<b>63.05</b>	<b>ACC</b>	<b>81.88</b>	-2.787	148	0.005	<b>0.23</b>	0.05	<b>small</b>

# Discussion

- HOLDING, PARSIMONY and STEERING elicit **less fluent responses** than REFORMULATING.
  - The goal of reformulating is to remove uncertainty
  - Holding increases cognitive load
  - Parsimony introduces uncertainty in an interactional context
  - Steering may introduce tension

# Discussion

- ACCEPTING, REJECTING and STEERING elicit **responses of greater grammatical complexity** than REFORMULATING and PARSIMONY (as measured by complex nominals per t-unit).
  - Accepting, rejecting and steering generate the need for justification, questioning, or elaboration on the part of the candidate.
  - Parsimony is intended to leave space for candidates' contributions.
  - Space is not what is expected in an interactional context.

# Summary

- Grounded theory to explore spoken interaction
- Empirical quantitative analysis of candidate performance
- Examiner behaviour has an effect on candidate performance
- Improved understanding of the scope of the effect in an interactional context
- Future directions?

Thank you for listening

# References

- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1–25.
- Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of Memory and Language*, 75, 104–116.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.
- Crossley, S. A., Bustamante, A., & Bradfield, F. (2018). GAMET: A tool for automatically assessing grammar and mechanic errors in learner corpora. In *14th American Association for Corpus Linguistics (AACL) Conference*.
- Davies, M. (2008). The corpus of contemporary American English (COCA): 400+ million words, 1990-present.
- Gablasova, D., Brezina, V., & McEnery, A. (2009). The Trinity Lancaster Corpus: Development, Description and Application. *International Journal of Learner Corpus Research*, 5(2), 126–160.
- Glaser, B., & Strauss, A. L. (1973). *The discovery of grounded theory strategies for qualitative research* (1st pbk. ed., Observations (Chicago, Ill.)). Chicago: Aldine Pub.
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24–49. <https://doi.org/10.1093/applin/amm017>
- Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication. Retrieved from [http://scholarworks.gsu.edu/alesl\\_diss/35](http://scholarworks.gsu.edu/alesl_diss/35)
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786.
- Kyle, Kristopher, Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior Research Methods*, 50(3), 1030–1046.
- Trinity College London. (2018). *Exam Information : Graded Examinations in Spoken English ( GESE )*. Trinity College London. Retrieved from [www.trinitycollege.com](http://www.trinitycollege.com)
- Wessa, P. (2017). Notched Boxplots (v1.0.10) in Free Statistics Software (v1.2.1), Office for Research Development and Education. Retrieved July 22, 2019, from [http://www.wessa.net/rwasp\\_notchedbox1.wasp/](http://www.wessa.net/rwasp_notchedbox1.wasp/)

Questions?

[richard.harris@trinitycollege.com](mailto:richard.harris@trinitycollege.com)