

Exploring performance across two delivery
modes for the same L2 speaking test:
face-to-face and computer delivery using
Zoom

Fumiyo Nakatsuhara, Chihiro Inoue and Lynda Taylor
(CRELLA, University of Bedfordshire)

Vivien Berry (British Council)

Evelina Galaczi (Cambridge English Language Assessment)



RESEARCH BACKGROUND

Background

Different speaking test delivery modes

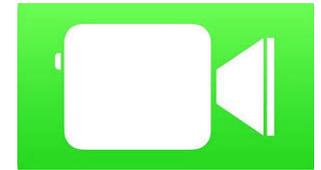
- **f2f mode:** Cambridge Main Suite exams, IELTS
- **Computer-delivered mode (administered by computer only):** TOEFL, PTE
- **Both modes:** BULATS

What about a **computer-delivered mode administered by an examiner** on the other side of the screen, to combine advantages of both these modes?

- **Computer-delivered mode (administered by examiner):** Computer-based Standard Speaking Test (ALC, 1999 – but no longer available)

Background (cont.)

- Rapid advances in online video communication technology:
e.g. Skype, Facetime, Zoom



Zoom: online video conferencing program (<http://www.zoom.us>), which offers high definition video conferencing and desktop sharing



Rationale

- Ensuring accessibility and fairness in remote areas where frequent delivery of f2f tests is not possible
- Making use of recent, advances in technology to offer computer-delivered tests without losing interactional features / responding to the broadening construct of speaking (increasing use of online video communication)

Aims of the Project

To compare performance features across f2f and Zoom-delivery modes with regard to:

- 1. *test-takers' linguistic output and scores* on the 2 modes and *their perceptions* of the 2 modes**
- 2. *examiners' test management and rating behaviours* across the 2 modes, including *their perceptions* of the 2 delivery conditions**

Research Questions

- **RQ1a:** Are there any differences in **linguistic output, specifically types of language function**, elicited from test-takers under f2f and Zoom-delivery conditions?
- **RQ1b:** Are there any differences in **test-takers' scores** between f2f and Zoom-delivery conditions?
- **RQ1c:** What are **test-takers' perceptions** of taking the test under f2f and Zoom-delivery conditions?

- **RQ2a:** Are there any differences in **examiners' test administration behaviour (i.e. as interlocutor)** under f2f and Zoom-delivery conditions?
- **RQ2b:** Are there any differences in **examiners' rating behaviour** when they assess test-takers under f2f and Zoom-delivery conditions?
- **RQ2c:** What are **examiners' perceptions** of examining under f2f and Zoom-delivery conditions?

RESEARCH DESIGN

- mixed methods approach -

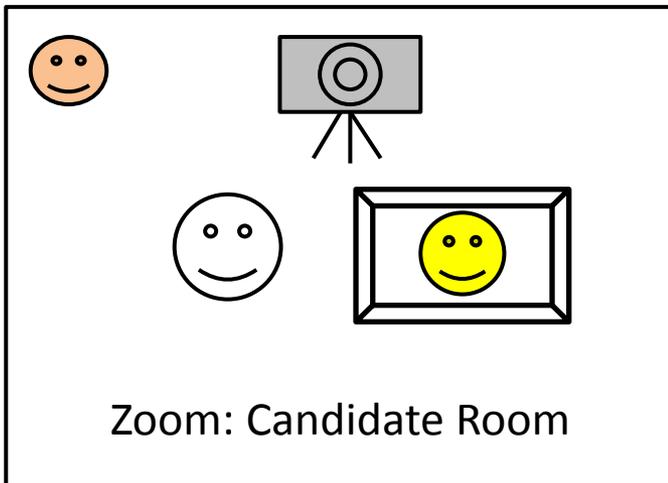
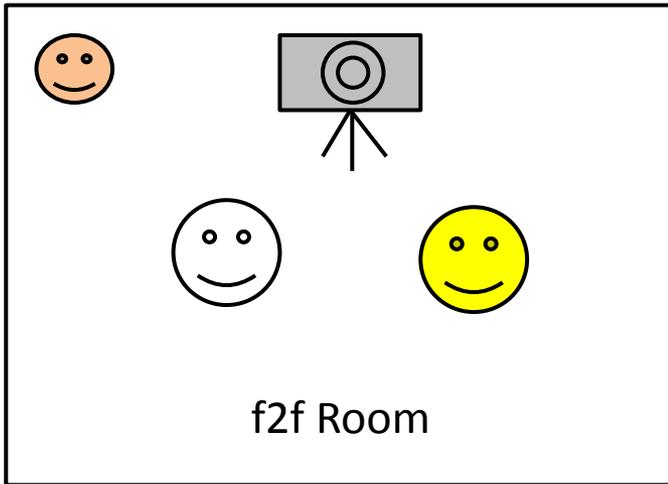
Data collection

- **Test-takers:** 32 test-takers (CEFR B1 – C1) taking both f2f and Zoom-delivered tests
- **Examiners:** 4 trained examiners (Examiners A – D)
- **Test structure:**

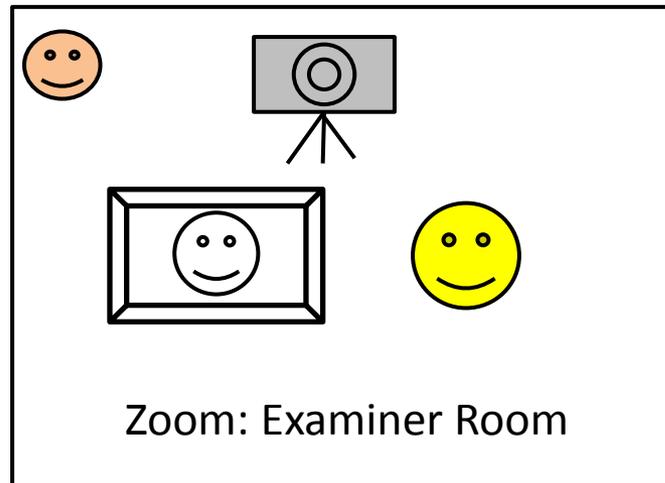
Part 1	Intro. and interview	4-5 mins.	2 test versions
Part 2	Individual long turn	3-4 mins.	- Travelling
Part 3	Two-way discussion	4-5 mins.	- Success

- **Counter-balanced:** delivery mode x test version
- **Rating criteria:** 1) Fluency and coherence; 2) Lexical resource; 3) Grammatical range and accuracy; 4) Pronunciation

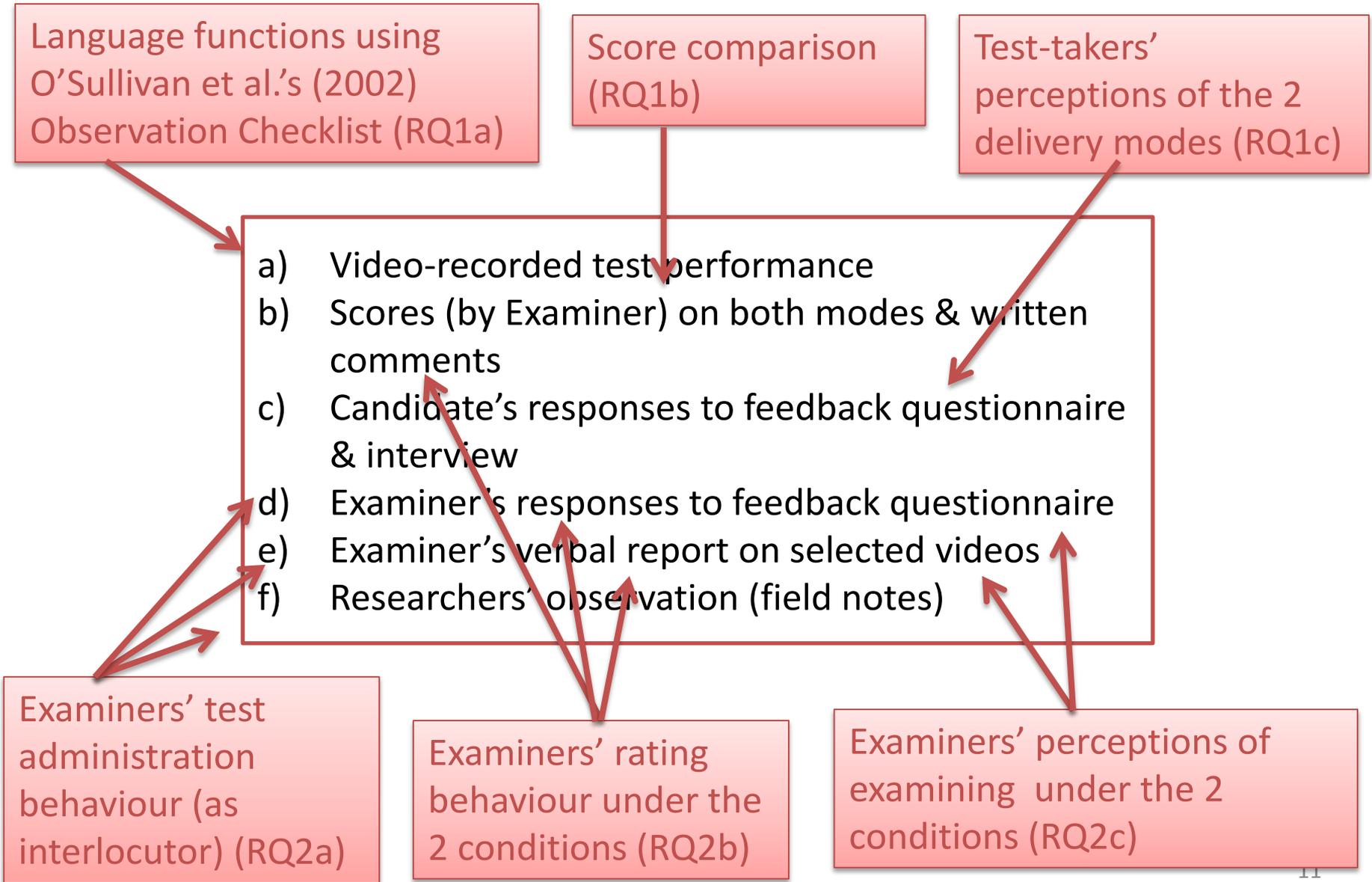
Gathered Data



- a) Video-recorded test performances
- b) Scores (by Examiner) on both modes & written comments
- c) Candidates' responses to feedback questionnaire & interview
- d) Examiners' responses to feedback questionnaire
- e) Examiners' verbal reports on selected videos
- f) Researchers' observations (field notes)



Data Analysis



RESULTS

Language Functions (RQ1a)

- 5 functions that test-takers used sig. differently under the 2 conditions.
 - Asking for clarification (Parts 1 & 3) Zoom > f2f
 - Elaborating (Part 3) Zoom > f2f
 - Comparing (Part 3) f2f > Zoom
 - Suggesting (Part 3) f2f > Zoom
 - Modifying (Part 3) f2f > Zoom
- No sig. differences in Part 2

Test scores (RQ1b)

Category	Test mode	Mean (SD)	t (df = 31)	Sig
Fluency	f2f	6.594 (1.043)	1.791	.083
	Zoom	6.406 (1.160)		
Lexis	f2f	6.750 (1.047)	.297	.768
	Zoom	6.719 (1.143)		
Grammar	f2f	6.625 (1.008)	.000	1.000
	Zoom	6.625 (1.100)		
Pronunciation	f2f	6.688 (.780)	1.717	.096
	Zoom	6.531 (.879)		
Overall (mean)	f2f	6.644 (.829)	1.459	.155
	Zoom	6.570 (.982)		
Overall (rounded)	f2f	6.547 (.883)	1.044	.305
	Zoom	6.469 (.991)		

Test-taker questionnaire & interview (RQ1c)

	Test mode	Mean (SD)	t (df = 31)	Sig
Understand the examiner?	f2f	4.72 (.46)	6.18	.000
	Zoom	3.72 (.77)		
Taking the test was (1. V difficult – 5. V easy)	f2f	3.84 (.85)	3.65	.001
	Zoom	3.13 (.83)		

	f2f	Zoom	No diff
Which test made you more nervous?	28.1%	46.9%	25.0%
Which test was more difficult?	12.5%	65.6%	21.9%
Which test gave you more opportunity to speak English?	50.0%	18.8%	31.3%
Which test did you prefer?	84.4%	9.4%	6.3%

Selected comments from candidates



Having good eye contact and seeing the examiner's facial expressions and postures in the f2f mode helped me to understand him better. (S06)

I'm shy and I always become nervous when I speak to a person f2f. Instead, the computer test made me more relaxed. (S07)



I felt that I was able to speak more f2f, because I was able to understand when to speak and chip in. (S17)

I preferred the f2f mode. The examiner gave me lively encouragement. (S10)



The quality of sound in Zoom wasn't always good. (S24)



On the computer, it doesn't matter what I say as the interviewer is not in front of me. I felt freer and was able to speak more. Perhaps I made more mistakes, though. (S05)



The both modes were the same. It's still speaking to someone. (S20)



Examiners' test administration behaviour (RQ2a)

Interactional behaviour:

- role and frequency of examiner response tokens, e.g. nodding, back-channelling
- rate and articulation of examiner speech
- effect of examiner intonation
- use of gestures by examiners (and awareness of gestures used by candidates)
- issues related to turn-taking management
- requests for clarification

Issues specific to Zoom mode:

- negative effects of delayed video transmission
- the way the candidate can impact on the sound quality
- need to control the direction of the interview

Examiners' rating behaviour (RQ2b)

- With Zoom, Examiner D couldn't really hear S20's **word endings or other micro-phonologic features**. He didn't want to give the benefit of the doubt, so scored 5.0 for Pronunciation and it could have been even 4.0 (6.0 in f2f mode); he gave 5.0 for Grammar (6.0 in f2f mode)
- Examiner B wasn't sure whether she needed extra efforts due to S07's pronunciation or due to the technology. Also, **visual information is very important to rate Pronunciation**, but due to an unclear view of the candidate and her picture and sound not exactly synchronised, she felt that she perhaps overrated S07's pronunciation under the zoom condition (Zoom at 6.0 and f2f at 5.0)

Examiners' perceptions of two modes (RQ2c)

(1) Ease of administration

	Mode	Min	Max	Mean (SD)
Comfortable in overall administration	f2f	4	5	4.75 (0.50)
	Zoom	2	5	3.75 (1.26)
Ease of administering Part 1	f2f	4	5	4.75 (0.50)
	Zoom	2	5	4.00 (1.41)
Ease of administering Part 2	f2f	4	5	4.75 (0.50)
	Zoom	3	5	4.00 (0.82)
Ease of administering Part 3	f2f	2	5	4.00 (1.41)
	Zoom	2	5	4.00 (1.41)
Ease of administering interlocutor frame	f2f	3	5	4.50 (1.00)
	Zoom	2	5	3.75 (1.89)

Examiners' perceptions of two modes (RQ2c)

(2) Ease of rating

	Mode	Min	Max	Mean (SD)
Comfortable overall in rating performance	f2f	3	5	4.50 (1.00)
	Zoom	1	5	3.50 (1.73)
Ease of applying Fluency and Coherence scale	f2f	4	5	4.75 (0.50)
	Zoom	1	5	3.75 (1.89)
Ease of applying Lexical Resource scale	f2f	4	5	4.75 (0.50)
	Zoom	3	5	4.25 (0.96)
Ease of applying Grammatical Range and Accuracy scale	f2f	4	5	4.75 (0.50)
	Zoom	2	5	4.00 (1.41)
Ease of applying Pronunciation scale	f2f	4	5	4.75 (0.50)
	Zoom	1	5	4.00 (1.41)
Confidence in accuracy of rating	f2f	3	5	4.00 (0.82)
	Zoom	1	5	3.50 (1.73)

Examiners' perceptions of two modes (RQ2c)

(3) Comparisons of two modes

	f2f	COMPUTER	NO DIFFERENCE
Which mode of speaking test did you feel more comfortable with?	1	2	1
Which mode of speaking test did you feel was easier for you to <u>administer</u> ?	2	0	2
Which mode of speaking test did you feel was easier for you to <u>rate</u> ?	3	1	0
Which mode of speaking test do you think gave a better chance for the test-taker to demonstrate their level of English language proficiency?	1	1	2
Which speaking test did you prefer?	2	1	1

CONCLUSION

Main Findings

- The two modes were generally comparable in terms of scores
- Some differences in elicited language functions and examiner behaviour
- Examiner comments suggest that exposure to computer-based testing of speaking encourages familiarity and confidence with that mode, and this can presumably be enhanced through appropriate training and more experience

Ways Forward

- Examining further some Zoom sessions which had much better internet connections than others, since this would allow us to have baseline data of Zoom working at its best at this point
- Re-analysing all audio recordings by presenting them randomly to two or more experienced raters. This would allow for a more thorough statistical analysis of differences in scores awarded on the two different modes of oral tests
- Examining audio/video data would also allow a focus on a range of conversational features, e.g. length of turn, turn interruptions/overlaps, gaps between turns
- Addressing the need for explicit examiner and candidate training if the introduction of computer-based oral testing is to be considered in the future

Thank you!

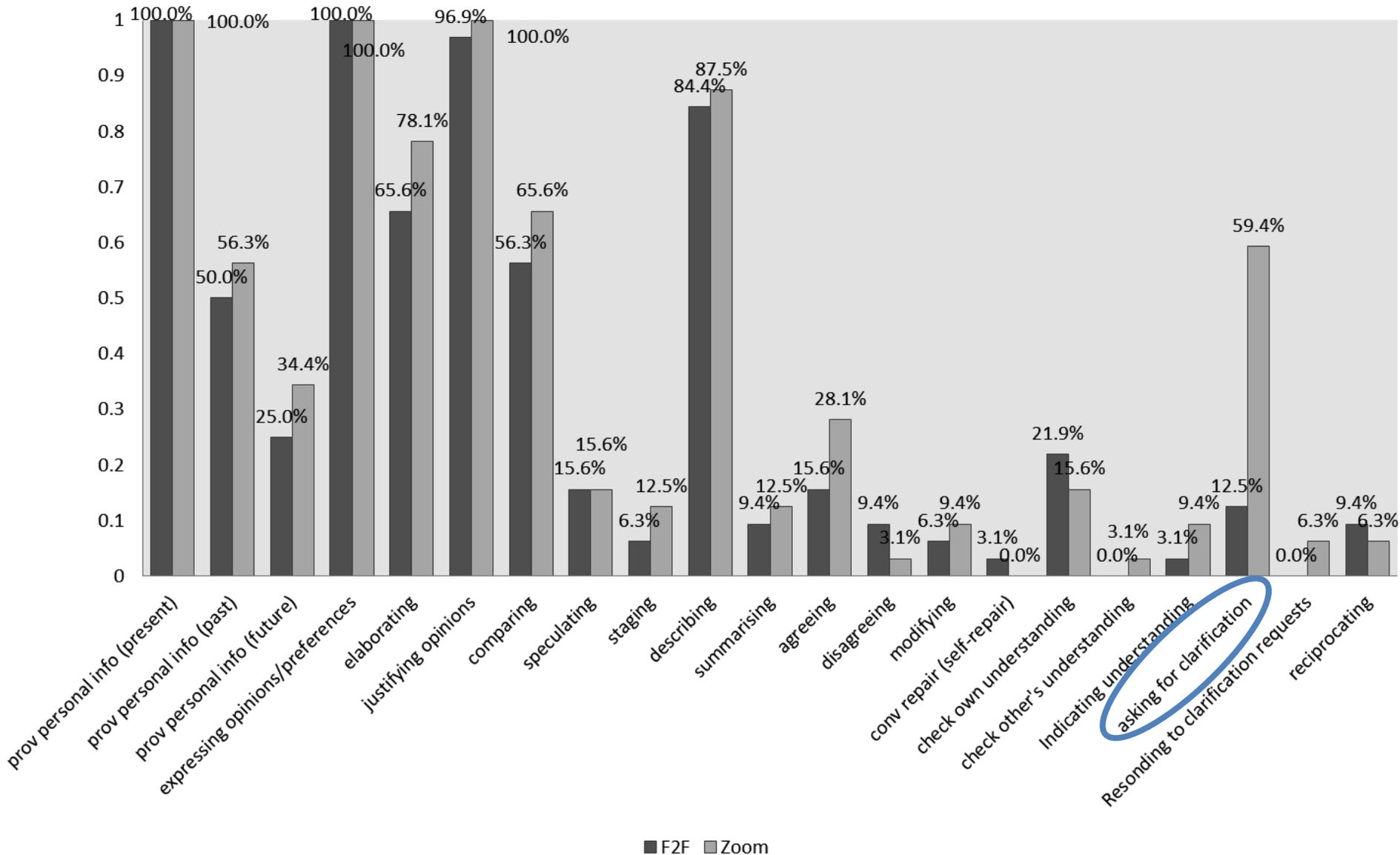
The researchers gratefully acknowledge support from:

- **Jermaine Prince** (IT specialist, British Council)
- **Michal Mazanek & Ivo Vicente** (IELTS, Ealing, Hammersmith & West London College)

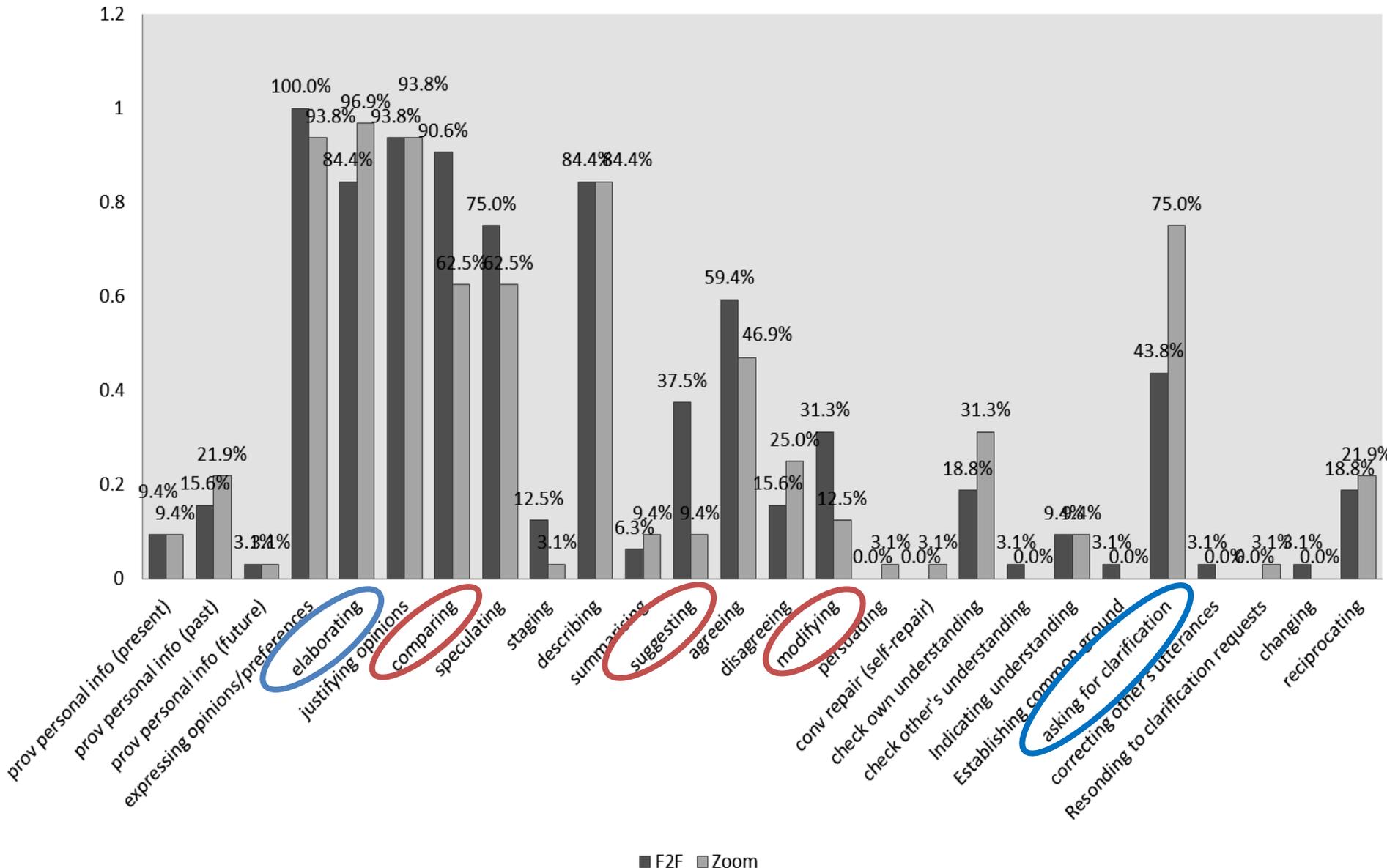


CAMBRIDGE ENGLISH
Language Assessment
Part of the University of Cambridge

Part 1 (Interview)



Part 3 (Discussion)



Selected notes from examiners' verbal reports (1)

Examiner D tended to nod more on f2f mode to facilitate the candidate, but on computer mode, he didn't do it very much due to the delays in video transmission

Examiner A nodded much more on computer; she said this might have been because the candidate asked for repetition; she was afraid that he might think she's not getting what he said if she doesn't show understanding

All 4 examiners reported that they needed to articulate each word more slowly and to slow down his speech on Zoom-delivery mode

Examiner D uses 'Why?' with an intonation which doesn't make it sound like interrogation – which is easy to convey in f2f, but with computer, due to sound quality and transmission speed, he's not sure if that subtlety is conveyed

Selected notes from examiners' verbal reports (2)

Under the f2f condition, Examiner A pointed out sub-questions on the prompt card using her finger to facilitate more speech, but couldn't do so under Zoom condition

Examiner C commented: she cannot see the candidate's hands on computer; therefore there's limited information available. Cannot get the same rapport on computer as f2f interviews

Examiner D doesn't think he can use gestures on computer as effectively as he could under the f2f condition (e.g. leaning slightly back to show that it's the candidate's turn to speak)

Examiner C commented that turn-taking was different due to the sound quality and slow communication; because turn-taking is slower, a smaller amount of language (evidence) is elicited during the time allowed.

Examiners' rating behaviour (RQ2b)

Examiner D tended to nod more on f2f mode to facilitate the candidate, but on computer mode, he didn't do it very much due to the delays in video transmission

Examiner A nodded much more on computer; she said this might have been because the candidate asked for repetition; she was afraid that he might think she's not getting what he said if she doesn't show understanding

All 4 examiners reported that they needed to articulate each word more slowly and to slow down his speech on Zoom-delivery mode

Examiner D uses 'Why?' with an intonation which doesn't make it sound like interrogation – which is easy to convey in f2f, but with computer, due to sound quality and transmission speed, he's not sure if that subtlety is conveyed