Is the devil you know better? Testwiseness and eliciting evidence of interactional competence in familiar versus unfamiliar triadic speaking tasks

Nicholas Glasson
Centre for Research in English Language Learning and Assessment, University of Bedfordshire, UK.

Research has proven the impact testwiseness and test preparation can have on speaking test discourse (Lam, 2015; Luk, 2010; Hüttner, 2014). The question is perhaps not whether assessment leads to contrived interaction but how interactions may differ between a more familiar task (face-to-face Cambridge B2 First Speaking) and an entirely new online task. This conversation analytic case study of a group of three in two different tasks (one face-to-face, one online) sought to explore how testwiseness is implicated in candidate talk and to explicate differences in what a more familiar and a completely new task elicited. The analytic focus was on the conversational object "I agree" but includes reference to non-verbal behaviours, turn-taking practices and the relationship between agreement and topic shift. The findings indicate a contrast in terms of closure and progressivity – that is, an observable difference in the closure of topics and the forward movement of conversation – across the two tasks. The data illustrates how the same candidates performed very differently in two tasks in the space of 40 minutes. In so doing, it raises questions about broadening the potential assessments have to target interactional competence.

Key words: Speaking assessment, Conversation Analysis, Interactional Competence

Introduction

Messick (1982) recognised that test preparation could involve the development of skills with the objective of scoring well but that are potentially construct-irrelevant. In

Email address for correspondence: nicholas.glasson@cambridge.org

© The Author(s) 2022. This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International License, which permits the user to copy, distribute, and transmit the work provided that the original authors and source are credited. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

the context of speaking assessment, this might take the form of rehearsed or contrived performances that, while superficially meeting criteria for interactional competence, are actually quite different from target domain talk-in-interaction. Such issues of construct-irrelevant testwiseness, as a by-product of candidate preparation, are a fundamental concern since they could ultimately "undermine the credibility of test scores as a measure of the targeted ability" (Knoch et al., 2020, p. 552).

As Burton (2020) reminded us, testing has always been beset by a concern regarding the "relationship between test-focused curricula and outcomes" (p. 2). Numerous studies have touched upon the impact "testwiseness" (Winke & Lim, 2017) can have on score gains via test-takers using strategic behaviours. Testwiseness is understood here to mean candidate familiarity with the task via some form of prior training. While not wholly antithetical to testing a construct, test preparedness in the form of memorisation and drilling can artificially enhance scores (Xie, 2013). It is also acknowledged that cram schools are a lucrative industry and that the capacity for live task prompts to be shared online is also utilised to coach learners (Yu et al., 2017). The impact of this was seen in the context of the American Council on the Teaching of Foreign Language's (ACTFL) oral proficiency interview assessment, where test-takers gave "memorised responses that did not address the question and were not adapted for the audience" (Cox, 2017, p. 104). Moving from more obvious instances of testwiseness leading to construct-irrelevance, it is valuable perhaps to consider the more subtle ways in which genuine (i.e., not consciously contrived) candidate discourse may differ from a more to a less familiar task.

Literature review

Test familiarity and testwiseness

Speaking tests where two or more candidates interact have been found to elicit a broader range of interactional features when compared to examiner-led interviews (Brooks, 2009; Lazaraton, 2002) and "provide opportunities for students to demonstrate not only their linguistic competence, but also their interactional abilities to relate to each other" (Gan et al. 2008, p. 331). While such tests provide broader sampling of talk-in-interaction than asymmetrical interviews (Kormos, 1999), there is evidence they are still essentially institutional in a conversation analytic sense in that

"the setting imposes constraints on what can be done and how" (Liddicoat, 2011, p. 345).

This orientation towards institutional rather than social-interactional goals is evident in terms of candidates needing to give "floor time" to peers (Brooks, 2009), a focus on task achievement over discussion (Gan, 2010), of "talking to score" (Luk, 2010), of ignoring disagreement (Hüttner, 2014) or even producing rehearsed performances (Lam, 2015). This institutionalising can be seen as a by-product of not just standardised test delivery but also a candidate's desire to score well even if this means the talk produced is contrived.

Research has found test-takers may let interactional problems pass, dealing with disagreements superficially (Hüttner, 2014; Luk, 2010; Firth, 1996). Firth (1996) noted how L2 speakers "imbue talk with an orderly and normal appearance, in the face of extraordinary, deviant, and sometimes abnormal linguistic behaviour" (p. 237). This is due to the orientations of test takers, in that they are mutually supportive, and this is often a by-product of practising together (Luk, 2010; Lam, 2015; Hüttner, 2014). For example, Hüttner found that disaffiliation (e.g., an observable difference of opinion) can co-exist alongside superficial linguistic alignment (e.g., agreement tokens). Hüttner hypothesises that candidates deem it more important in this context to show affiliation than to resolve disagreements. The performance is what matters, which leads candidates to "talk to score" (Luk, 2010). This phenomenon has been identified in other contexts. Stokoe (2013) describes how trainee police officers' role play actions were "done elaborately such that an examiner can be sure not to 'miss' an officer's skill or, perhaps more simply, adherence to the practice taught in training" (p. 181). There are matters at stake in assessed interactions and people rightly want to look their best.

At its most extreme, testwiseness can result in rehearsed, contrived interactions. Luk (2010) describes candidate discourse as "ritualized, contrived and colluded" (p. 25). She concludes that candidates showed a strong desire to "maintain the impression of being effective interlocutors for scoring purposes rather than for authentic communication" (p. 25). This manifested itself in a lack of other-speaker conversational repair (as in Hüttner, 2014) and using response tokens like "I agree with you" with little substantive elaboration of peers' points. As Luk points out, such

turn-entry devices are often used mechanically to gain speakership, facilitating a longer turn at speech that was prepared and not dependent on any active recipiency as listener (Lam, 2021). In a similar setting, Lam (2015) observed that assessed interaction "is not students' in situ execution of interactional competence in L2, but a canned product of students' execution of the competence prior to the assessed interaction in L1 during pre-task planning" (p. 56).

At its less extreme, testwiseness can still be seen to subvert known orders of interaction in everyday life. Lam (2018) highlights the centrality of responses contingent upon previous speakers within conceptions of interactional competence – specifically that agreement or disagreement is "accounted for" in a conversation analytic sense. As Lam points out, accounts are often provided for dispreferred (i.e., socially-awkward) actions like refusing an invitation, the same is not true of agreements which tend to be treated expeditiously. Both Lam's 2015 and 2018 studies however indicated how candidates felt the need to account for (i.e., rationalise) agreement.

Regarding Cambridge B2 First Speaking, research suggests speakers show an orientation towards a "commitment to maximising display of skills and minimising display of weaknesses" (Spence-Brown, 2001, p. 477). As Lazaraton and Davis (2008) observe in their conversation analytic study of the B2 First test: "just looking at the interaction in transcribed form, one gets the sense of watching a well-played tennis match, with each player hitting the ball to the other until the umpire calls time" (p. 321). Given the relation between testwiseness and language elicited, the B2 First Speaking Part 3 interactive task was a useful control in piloting a new online task where testwiseness was not a factor. The question addressed in this study is not whether assessment leads to contrived interaction or not, but how interactions may differ between a more familiar task (face-to-face B2 First Speaking) and an entirely new online task.

Online speaking tests and language elicited

Research into online interactive tests of speaking has focused on score comparability, feasibility and linguistic function analysis. While not a primary focus, research thus far can also give us clues as to the nature of online test interactions. For example, Jin and Zhang (2016), which focused on online dyads, noted that overlapping talk-in-interaction negatively impacted scores as it was incomprehensible. This prevalence of

overlap in turn-taking is at least in part attributable to the fact the test was audio-only without visual cues to rely on. This underscores the importance of embodied conduct in the online mode. Jin and Zhang (2016) note "misinterpretation of turn-closing signals" (p. 78) as a result.

In terms of interactional features, studies have suggested some differences between modes. For example, a higher frequency of clarification requests and asking questions (Nakatsuhara et al., 2017; Jin & Zhang, 2016) and greater use of conversational management items like "you first" (Jin & Zhang, 2016). The online mode was found to make turn-initiating and closing moves harder to accomplish which led to slower speaker transition and mechanical turn-taking (Nakatsuhara et al., 2017; Jin & Zhang, 2016; Ockey et al., 2017).

Where the visual channel was unavailable, talk was more like non-face-to-face interactions (Jin & Zhang, 2016; Ockey et al., 2017). Ockey et al. (2017) opted to use avatars (in a 'library' Virtual Learning Environment setting) in place of real video images but acknowledge that the use of avatars may have led to a reduced sense of "being there" – their rationale behind the use of avatars was to reduce anxiety and "bias for best" (Swain, 1984). Plough et al. (2018) note the importance of social presence, arguing that the online "medium impedes each participant's feeling that the other is 'real' and their ability to accurately read their interlocutor's behaviour" (p. 439). They argue the online mode may lead to "qualitatively different interactional behaviours" (p. 439) which will require a re-examination of what constitutes interactional competence. Similarly, Nakatsuhara et al. (2021) make the case for "studies that go beyond a comparison between the face-to-face and video-conference (VC) modes of a speaking test to an investigation of the VC mode in its own right" (p. 382). So, there is a gap here in what qualitatively differentiates online interactional competence from existing evidence that this study aims to contribute to.

Groups of three as an interactional environment

Studies of candidate group interactions go as far back as Folland & Robertson (1976) but research on grouped interactions grew post-2000 (e.g., Ockey, 2001; Nunn, 2000). These were often focused on candidate characteristics like extroversion. He and Dai (2006) did explore use of interactive language functions in triads but focused on a limited range of interactional features, dictated by external criteria.

More recent CA studies of grouped candidate interactions (3's or 4's) have provided insights into emergent topical development in conversation (Gan et al., 2008), how such interactional features may differ across proficiency levels (Gan, 2010) and how group size and test-taker characteristics may impact on performance (Nakatsuhara, 2011). Interestingly, Nakatsuhara concludes that "if test designers are interested in minimizing the effect of extraversion-level variables on test-takers' performance, groups of three would appear to be more appropriate than groups of four" (p. 504). Groups of three candidate interactions evidenced a much greater "collaborative atmosphere" and sense of solidarity when compared to groups of four candidates which showed evidence of "avoidance behaviour" and "mechanical turn-taking" (Nakatsuhara, 2011, p. 494). These criticisms of a four-way candidate interaction reflect the phenomenon of "schisming" (Egbert, 1997) whereby multiparty groupings are seen to break into smaller dyads.

Recent conversation analytic work shows how in triads there is often a dyadic participation framework in operation where two speakers of the three are engaged in talk and the third speaker is often implicitly on the conversational side-lines. In the context of assessing interactional competence, one could argue that "including a third person in a conversation is an ongoing achievement, accomplished through the use of micro-level interactional practices" (Stivers, 2021, p. 17). This makes triads a ripe context for the assessment of interactional competence as it has the potential to show not only the ability of an individual to participate in dyadic interaction but also their ability to shift into and out of dyadic talk.

Closure and progressivity in talk-in-interaction

The analysis presented here focuses on agreement sequences and, pursuant to this, it is useful to introduce two concepts from CA studies: closure and progressivity. The closure of a conversational sequence presents a specific interactional problem for speakers in that disengagement must be mutually agreed between speakers. It must also ensure that "all participants to the conversation have had the opportunity to talk about all of the things which need to be dealt with" (Liddicoat, 2011, p. 294). To disengage from a conversational topic or a conversational exchange entirely, speakers must achieve this action co-operatively through co-constructed actions.

Closing sequence example (Liddicoat, 2011, p. 295)

```
Emma:    So we'll do it at 7.30
Sue:     It'll be fun
Emma:    Alright,
Sue:     Okay
Emma:    Bye=
Sue:     =Bye
```

In the example above, we can observe how Emma and Sue have to jointly arrive at a point where they can say goodbye. Imagine, for instance, how disjunctive or disaffiliative it would be for Sue to say "Bye" in any of her previous turns. In essence, in order to complete any social action via talk, there must be co-constructed agreement. Such agreements can be viewed as "closure-implicative" (Schegloff, 2007) to the topic or sequence at hand, in effect talk is produced that "closes down a topic in some way" and that "provides for the possibility that there is nothing further to be said on the current topic" (Liddicoat, 2011, p. 298). In the context of L2 assessed interactions, it is useful to consider how such actions are achieved, since the closure of a topic is fundamental apparatus for the achievement of topic shifts (Jefferson, 1984) and more broadly, communicating meaningfully. The second conversation analytic feature relevant to this research is that of progressivity.

Progressivity example (Stivers & Robinson, 2006, p. 367)

```
Mom:     What're you doing Dad? ((Dad is at phone))
         ((0.7 silence in which Boy gazes at Dad; Dad is leaning over
         phone))
Boy:     He's listening
```

In the example above we can see how a child provides an answer to mom's question even where they are not the stated addressee of the question. As conversation analytic studies have shown, we often show a preference for any answer over a non-answer, for progress over inaction. In contexts like the one above there is evidence for a "systematic preference for the provision of an answer over the preference for the selected next speaker to respond" (Stivers & Robinson, 2006, p. 386). As Schegloff (2007) notes, such sequences are the apparatus for getting activities accomplished and to delay accomplishment of an activity is seen as "dispreferred" in conversation analytic terminology, meaning it is an action that threatens co-operation, social ease and affiliation between speakers. The preference for progressivity (i.e., maintaining momentum in social actions) is seen not only in who provides answers but how we skilfully initiate and respond to trouble sources via repair in conversation. We do this habitually as the meaningful continuance of turn-taking itself rests upon it.

Research focus

This study was focused on exploring how triadic candidate talk differed between a more familiar face-to-face task (B2 First Speaking Part 3) and a new online task test-takers had never seen before. Given the relative lack of conversation analytic study of online test-taker interactions, coupled with the unique interactional dynamic a triadic discussion presents, it was a rich context in which to explore how task familiarity plays out in terms of talk elicited.

Methods

Conversation Analysis

Recordings of face-to-face and online interactions were transcribed according to conversation analytic (CA) conventions (Jefferson, 2004), following similar studies which utilise images alongside the verbatim transcript (Mondada, 2014; 2018). The overriding aim of CA is to uncover the orderliness of talk-in-interaction, to focus neither on the individual nor broader contextual factors but the "procedural infrastructure of situated interaction" (Schegloff, 1992, p. 1338). As Lazaraton explains:

> "The basic question facing the analyst is, 'Why this now?' instead of 'that' or 'later'. The analyst attempts to model the procedures and expectations employed by the participants by proceeding as the talk does: on a turn-by-turn basis" (Lazaraton, 2002, p. 31).

Order is produced by the speakers themselves and it is the analyst's role to uncover "the structures, the machinery, the organized practices, the formal procedures, the ways in which order is produced" (Psathas, 1995, p. 3). CA can be viewed as "a process more than a product, because it grows out of the transcription and repeated and prolonged examination of materials" (Lazaraton, 2002, p. 31). CA can be focused on single case analyses (i.e., the descriptive analysis of individual transcripts) and not primarily concerned with "statistical analyses of large data aggregates" (Lazaraton, 2002, p. 31).

For the purposes of framing analytic findings, an object-focused approach (see Firth, 1996) is initially adopted whereby a particular conversational item (in this case the phrase "I agree") which was present in both task settings is described. Object-focused studies have been used a great deal in conversation analytic research to focus on the interactional purpose of lexico-grammatical constructions – an oft-referenced example is Heritage's work on the object "Oh" as a change-of-state marker (see Heritage, 1984). The question in this study is what "I agree" does in interactional terms. Beyond the target object, the analytic focus includes reference to non-verbal behaviours (specifically eye-gaze and hand gesture – as these were particularly involved in how the face-to-face task interactions were realised), turn-taking practices and the relationship between how agreement is conducted with its broader implications for topic shift. In this way, one can view the initial target object as a key for unlocking insights into broader contrasts between the differing interactional ecologies of the two tasks in terms of closure and progressivity. For transcription conventions, see Appendix 1.

Participants

A convenience sampling approach was taken with school/university/work contacts known to the researcher as the main source for participant recruitment. In total, 21 participants (assessed as within B1-C1 levels[2]) were used for detailed analysis in this small-scale pilot study (11 females, 10 males). They were mostly[3] English language students attending classes in Cambridge (UK) during November/December 2019 – just prior to the global outbreak of Covid-19. Of the 21 participants, 7 self-reported that they knew nothing about the B2 First Speaking test. However, of these 7 participants, 4 were enrolled in a class preparing them to take either B2 First or C1 Advanced Speaking tests – and the teachers had done some speaking exam practice.

One participant was 17 years old, four were 18-20 years old, ten were between 21-30 years old, five were between 31-40 years old and one participant was over 40 years old. Participant first languages included Japanese, Italian, German, Spanish, Portuguese,

[2] All participants in the pilot were assessed by a certificated Cambridge examiner from the recordings post-trial.
[3] One group of three was from a US university trialled in the US, another two groups were comprised of L2 English speakers already working in the Cambridge area (i.e., that were not actively studying English).

Dutch, Slovak, Polish, Chinese, Korean, Turkish, French and Arabic. One participant reported only 1-2 years spent learning English, four 3-4 years, three 5-6 years, four 7-8 years and nine participants reported having 9 or more years' worth of formal English tuition.

In total, 98 minutes of data was obtained across face-to-face and online tasks. All participants gave informed consent for the use of their anonymised data.

Details of Group 2 used for explication in this article are provided in Table 1. In terms of B2 First Speaking test familiarity, all three candidates were enrolled on a C1 advanced preparation course, so had had exposure to the Part 3 task design and been taught how to manage the Part 3 task interaction (confirmed in the post-trial interview).

Table 1. Pilot group 2 background information

Participant	Gender	Age	First language	Time learning English	Teacher-assessed level[4]
A	Female	21-25 years old	Spanish	More than 10 years	C1
P	Female	18-20 years old	German	More than 10 years	C1
I	Female	17 years old	Dutch	9-10 years	C1

While only Group 2 is explored in this article, the interactional features relating to agreement, closure and progressivity across the two tasks/modes were corroborated by the data across the pilot recordings.

Instruments

The pre-trial candidate survey

Before trialling, candidates were asked to complete a short survey covering basic participant information (age, gender) and aspects like familiarity with B2 First Speaking, test-taking plans, self-estimated CEFR level and frequency of video-conferencing use. The pre-trial candidate survey items are provided in Appendix 2.

[4] This was later corroborated by a multiple-marking activity with experienced Cambridge examiners and MFRM analysis (not reported in this paper) as part of a larger PhD study.

The control task (B2 First Speaking – Part 3)

The face-to-face examiner-mediated task was the B2 First Speaking Part 3 mind-map task. It is designed to cover performance sampling from CEFR B1 to C1 levels. The Part 3 task is one of four tasks which comprise the Cambridge B2 First Speaking exam.

Table 2. Cambridge B2 First Speaking test parts and timing

	Interaction pattern	**Discourse features**	**Pair**	**Group of three**
Part 1	Interlocutor asks the candidates questions to elicit personal information.	General interaction and social language. Responding to questions. Expanding on responses.	2 minutes	3 minutes
Part 2	Interlocutor delegates an individual task to each candidate with a subsequent brief response question for the other candidate.	Sustaining a long turn. Managing discourse: coherence; organisation of language and ideas; accuracy and appropriacy of linguistic resources; clarity of message	4 minutes	6 minutes
Part 3	Interlocutor delegates collaborative discussion and decision-making tasks to candidates.	Sustaining an interaction. Initiating and responding appropriately. Negotiating.	4 minutes	5 minutes
Part 4	Interlocutor leads a discussion with candidates.	Responding appropriately. Developing topics.	4 minutes	6 minutes
Total timing			14 minutes	20 minutes

The Part 3 task is the peer-peer discussion phase of the test where the examiner delegates a collaborative discussion and decision-making task to the candidates. The input is a mind map diagram with a central question (Figure 1) which is delivered with matching verbal instructions (Figure 2).

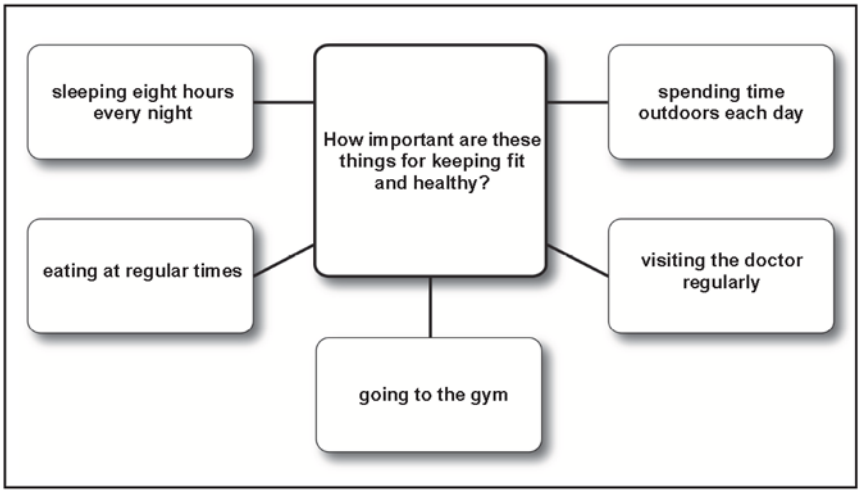


Figure 1. The B2 First Speaking Part 3 task sheet

21 Keeping fit

Part 3 4 minutes (5 minutes for groups of three)
Part 4 4 minutes (6 minutes for groups of three)

Part 3

Interlocutor Now, I'd like you to talk about something together for about two minutes *(3 minutes for groups of three).*

Here are some things people often do to keep fit and healthy and a question for you to discuss. First you have some time to look at the task.

*Place **Part 3** booklet, open at **Task 21**, in front of the candidates. Allow 15 seconds.*

Now, talk to each other about **how important these things are for keeping fit and healthy.**

Candidates *2 minutes (3 minutes for groups of three)* .....

Interlocutor Thank you. Now you have about a minute to decide **which two are most important for keeping fit in the long term.**

Candidates *1 minute (for pairs and groups of three)* .....

Interlocutor Thank you. (Can I have the booklet, please?) *Retrieve **Part 3** booklet.*

Figure 2. The B2 First Speaking Part 3 interlocutor frame

Functions the task is designed to elicit include sustaining an interaction, exchanging ideas, expressing and justifying opinions, agreeing and/or disagreeing, suggesting, speculating, evaluating and reaching a decision through negotiation. The basic process of what happens is summarised below:

> *Part 3*
>
> *This part is divided into two parts and the interlocutor asks you and your partner to talk together in both. In the first part, you will be given five written prompts and asked to discuss a question. For example, you might*

> *be asked to discuss things that might attract tourists to a town. The interlocutor does not join in the conversation. After 2 (3 minutes for groups of three) minutes, the interlocutor will give you one more minute to make a decision together which is related to what you have been discussing.*
>
> (Source: https://www.cambridgeenglish.org/images/173977-cambridge-english-first-speaking-test-examiner-comments.pdf)

The task used in this study was taken from live B2 First materials and had been through all the standard quality checks before being potentially selectable for live use[5]. Information about the B2 First exam can be found online (https://www.cambridgeenglish.org/exams-and-tests/first/exam-format/). B2 First Speaking has been used in several prior CA studies (e.g., Galaczi, 2004; 2008; 2014) – making it an appropriate control task in this context.

The new online task

The online task used a bespoke teleconferencing platform for an extended collaborative speaking task (10 minutes of peer-peer discussion).

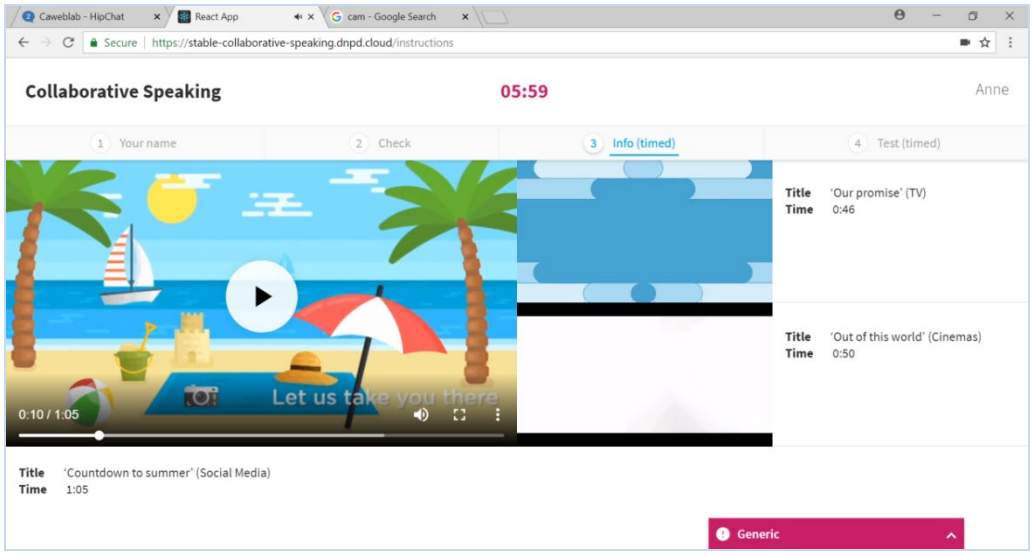


Figure 3. Screen excerpt of initial preparation screen with three-video input

Before the task begins, all three candidates are given 10 minutes to engage with a specific set of inputs (i.e., three different adverts for a drink). The videos contain text and music but not voiceovers. They are given the task of appraising each video. Once

[5] Its only reason for withdrawal from live use was that it covered a topic area already represented in the bank of potential test materials.

the preparation time ends, the group move into a shared teleconferencing space. Candidates then work together to determine which short advert is the most suitable based on the instructions they have.

As a prototype, there was no clearly defined specification, but the task was developed as an initial exploration of how the revised CEFR descriptors for online interaction might be assessed. The updated CEFR (Council of Europe, 2020) contains several useful points of reference for online spoken interaction (Table 3). Of particular relevance to this task perhaps are "instances of simultaneous (real time) and consecutive interaction"; "participation in sustained interaction with one or more interlocutors"; "reactions to embedded media"; "participation in collaborative project work" and "dealing with communication problems."

Table 3. Relevant descriptor areas for online speaking (Council of Europe, 2020)

Online conversation and discussion	**Goal-oriented online transactions and collaboration**
• instances of simultaneous (real time) and consecutive interaction, the latter giving time to prepare a draft and/or consult aids • participation in sustained interaction with one or more interlocutors • composing posts and contributions for others to respond to • comments (e.g., evaluative) on posts, comments, and contributions of others • reactions to embedded media • the ability to include symbols, images, and other codes for making the message convey tone, stress, and prosody, but also the affective/emotional side, irony etc.	• purchasing goods and services online • engaging in transactions requiring negotiation of conditions, in a service as well as client role • participating in collaborative project work • dealing with communication problems

There is a "shift from simple social exchanges and personal news towards a broader range of competences encompassing professional and educational discursive interaction at the C levels" (Council of Europe, 2020, p. 96) with an assumption speakers will be able to deal with real-time interaction from B1+ level upwards. Candidates were told before the first task was trialled that they would be assessed in both tasks. The main differences between the two tasks are summarised in Table 4.

Table 4. Summary of the main differences between the two tasks

	New online task	B2 First Speaking task
Test-takers' pre-existing familiarity with task	No familiarity with task or task training	Some familiarity with task type and prior training
Examiner presence/intervention	None (task was fully automated)	Examiner sets up and closes task, present throughout.
Within-task individual preparation time	10 minutes	15 seconds
Pre-task input	Scenario-based instructions and three short adverts	None
Within-discussion prompting (e.g., task sheets)	None	Mind-map diagram with ideas
Allowance of candidate's own notes	Yes	No
Time allocated for candidate-candidate discussion	10 minutes	4 minutes (3 minutes discussion phase, 1 minute decision phase)
Goal-orientation of task	Negotiation towards a decision	Negotiation towards a decision
Number of test-takers	3	3

The post-trial interview frame

For participants who were available, 30-minute group interviews were held. The interviews used a pre-determined frame (see Appendix 3). Interviews were held after the researcher had completed an initial conversation analysis of the samples – the rationale for this was so that the group interview could begin with general reflections on the two task experiences and then be focused on the specific phenomena of closure and progressivity identified. This approach was adopted as a complementary methodological addition to the conversation analysis (Pomerantz, 2005) to provide participant responses to the analysis as corroboration for analytic output by the researcher.

Procedure

A basic outline of the piloting procedure is provided in Figure 4 below. The pre-trial surveys were administered immediately prior to the first task being trialled. Trials of the F2F and OL tasks were counterbalanced so that one task was not taken first every time. Tasks were trialled on the same day – groups would do the first task and then, after a 15-minute break, they would do the second task. The online task was recorded within the online task platform. Participants were provided high-quality headsets to wear to ensure clarity of audio input/output. The B2 First Part 3 task was recorded

with a digital camera positioned on a tripod to the side of the examiner. The sound pick-up on the camera was sufficient to negate the need for participants to wear individual microphones. The researcher then conducted conversation analysis of the recorded data over the course of several months. After this conversation analytic phase, post-trial interviews were conducted with a small subset of six available participants.

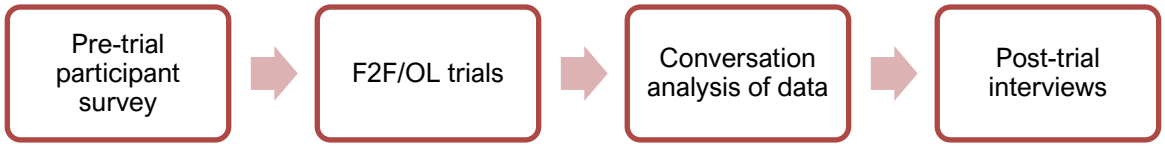


Figure 4. Outline of trial procedure

This data set was taken from an initial piloting stage of a much larger PhD research project that involves data drawn from 72 test-takers' online triadic interactions.

Findings

Object-focused analysis: "I agree"

Extract 1 is taken from the discussion phase of the B2 First task where the three speakers in Group 2 are being asked to talk about different ways of keeping fit and healthy.

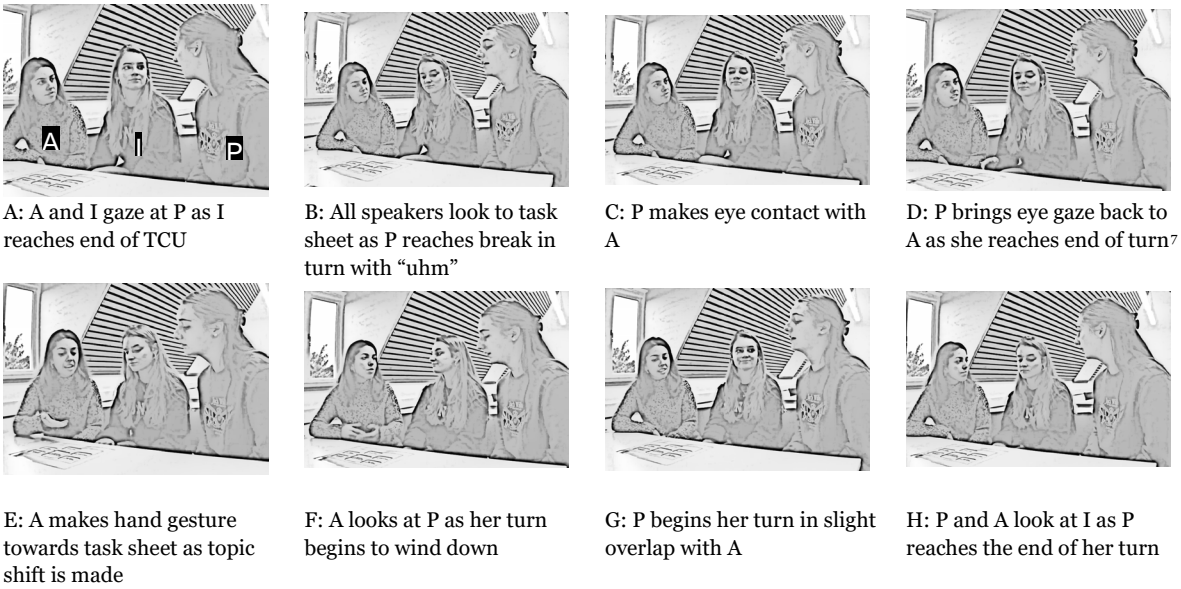
F2F Group 2 – Extract 1 (discussion phase)

```
1    EX:   Now (.) talk to each other about how important these things are
2          for keeping fit and healthy.
3    I:    Um: (.) I think that sleeping eight hours at night is very
4          important? because if you don't sleep enough (.) you can't
5          really function throughout the day?
6          so (.) you need that sleep?[A]
7→   P:    Yes I agree with you u:hm[B] I think that uhm if you[C] don't
8          have enough sleep you (.) like (.) can't go through your energy
9          resources[D]
10→  A:    Yes I can totally see your point and (.) like (.) helps you
11         during the day to focus on what you need to do and to have
12         enough energy to go throughout the day and (0.2) also maybe
13         eating at regular times[E] would be convenient to keep up your
14         health (.) maybe you have a meal plan or just the order of the
15         meals? during the day? already helps you to get enough energy
16         and to keep healthy instead[F] of being eating around all day=
17   I:    = [°Mmhmm°
18   P:    = [[G]Yes and gives your body (.) like (.) rhythm (.) you can
19         like refer to in those (.) like (.) so things are normal and
20         doesn't have to worry at the moment[H]=
21   I:    =yes and I think...
```

This sequence occurs at the opening of the three-minute discussion phase. We can see that the object "I agree" is in turn-initial position (line 7). Sequentially, it forms a second-pair part (SPP) to I's turn in which she has chosen an idea from the task sheet and provided an account for her choice – the intonation contour of which is rising in each turn-constructional unit (TCU)[6] as the turn reaches a transition relevance place. The object "I agree" is accompanied by an account-like rationale in the form of "I think that if you don't have enough sleep you like can't go through your energy resources". Between the object and the account there are hesitation tokens ("uhm"), suggesting a break in progressivity by the speaker (this occurs in P's turn in line 7). Such breaks in the progress of talk are noteworthy as they are a signal that some form of additional action is underway. In line 7, the "uhm" sounds P uses are effectively placeholders which signal to the other speakers she is not done yet. Sequentially, they maintain speakership and allow the current speaker to deliver another turn constructional unit (TCU). This can be seen in how P uses specific sounds, pausing or objects to create points of maximal incompleteness which break flow but keep control: "I think that *uhm* if you don't have enough sleep you *(.) like (.)* can't go". It is unusual and difficult for speaker change to occur at such points.

Figure 5 indicates the way in which all three speakers used gaze and gesture to achieve highly co-ordinated turn-taking. Images A and D for example show how gaze is used with precision timing to subtly project who the next potential speaker is (Goodwin, 1979, 1981). Image B shows how when P reaches this break in contiguity (i.e., to add an account for agreement) all three candidates are focused on the task sheet – so, while this is arguably a break from achieving progress in the action underway (i.e., agreement), it is not viewed as dispreferred or accountable by A or I, whose gazes are also directed at the task sheet. Image B indicates how a potential point of speaker change is actively being passed up by A and I – suggesting an expectation of an elaborated turn and an orientation towards such breaks in the flow of reaching an agreement. Indeed, it is only later in image D that we see eye gaze and body turn signaling that at the next transition-relevant place, a next speaker will be allocated.

[6] A TCU is a term used in Conversation Analysis to describe a discrete unit or increment of speech – it may be a single-word utterance ("What?") or something more complex ("When I got there (.) I was totally confused").



A: A and I gaze at P as I reaches end of TCU

B: All speakers look to task sheet as P reaches break in turn with "uhm"

C: P makes eye contact with A

D: P brings eye gaze back to A as she reaches end of turn[7]

E: A makes hand gesture towards task sheet as topic shift is made

F: A looks at P as her turn begins to wind down

G: P begins her turn in slight overlap with A

H: P and A look at I as P reaches the end of her turn

Figure 5. Embodied conduct as turn management (Extract 1)

Similarly, in line 10, A's turn begins with an upgraded form of "I agree" (Pomerantz, 1984) with "Yes I can totally see your point", before some micro-pauses lead into a similar accounting for agreement ("helps you during the day" etc.). In Extract 1 one can view lines 3-6 as a first pair part (FPP) which then leads to a subsequent second pair part (SPP) in lines 7-9 and secondary SPP in lines 10-12 before a topic shift is initiated in lines 12-13 ("also maybe eating at regular times"). This topic shift occurs mid-utterance and in a similar way to I's initial opening, has a rising intonation contour in the case of specific TCUs (e.g., "during the day?"), A then self-selects to continue speaking, providing an additional increment which warrants the suggestion being made. The topic shift in lines 12-13 is disjunctive (Jefferson, 1984) in nature and is preceded by notable pausing. Image E in Figure 5 shows how A uses hand gesture to help signal the topic shift and perhaps as a means to warrant its invoking (i.e., as part of the task).

This is followed by no-gap-no-overlap responses from I and P, with P extending their turn, providing a possible collaborative completion[8] in lines 18-20 of A's final turn in line 16. Notably, in Extract 1 in lines 7, 10, 18 and 21 we see that a turn with the agreement token "yes" in initial position yields the floor to the speaker, even where

[7] Interestingly, I is seen nodding agreement towards P just before this point but then withdraws eye contact as P brings her gaze back to A (as shown in image D).

[8] P's turn follows on quickly from the previous turn, while this could be seen as collaborative completion of a turn it is structurally more akin to an increment i.e., an additional turn constructional unit which is free-standing but links back to another, immediately-prior turn by another speaker. Grammatically it does not 'finish' or 'complete' the prior turn; it just adds to it.

there is some initial overlap or latched turns it is a prevalent mechanism for gaining the next available right to speak within the group. In Extract 1 the speakers tended to design their turns perhaps with this in mind – the initial "yes" gains you the floor, the "I agree" or equivalent phrasing achieves a sense of affiliation and the account for agreement either rephrases or expands on the original FPP.

Such turn structures can be seen in different ways. Heritage (1988) notes how the designing of turns in this way attends to another's viewpoint before turning the focus on their own perspective. In this sense the "yes, I agree" is other-attentive, the element which follows is more self-attentive. Alternatively, one could view these turns as assessment-like in that they perform agreement ostensibly via appraisals of the ideas provided. For example, A in line 10, provides agreement ("Yes I can totally see your point") and then continues ("and (.) like (.) helps you during the …"). What A provides is an extension of P's previous point about having enough energy. While such actions can be seen as topic extension or elaboration, structurally they perform a role not unlike "same assessments" which preface disagreement or, at least, re-specification of agreement (Pomerantz, 1984), for example:

JG: II.1.-15 (Taken from Pomerantz, 1984: 75)

```
1   C:    ... you've really both basically honestly gone your own ways.
2   D:    Essentially, except we've hadda good relationship et home.
```

As Pomerantz's example shows, one of the ways speakers can achieve partial agreement is by shifting the referent - in this case D's specifying that the relationship at home was contrary to the critique allows D to agree. This shifting of referents is a powerful strategy for maintaining an agreement trajectory even where disagreement is ostensibly part of what is happening. In the same way, one could view the candidates' uses of "yes I agree" plus expansions etc. as them putting their own spin on the idea as a means to both maintain agreement and also showcase some language. This can be observed in Extract 1, line 18 where P adds to A's ideas – it is not that the agreement is weak per se but that the same process of specifying a referent is present.

This could make the agreements in Extract 1 seem superficial or even mechanistic in character but if one views them as being like same-assessment agreement actions then perhaps it is not so much that agreement is accounted for but simply that participants in such test tasks have the dual interactive imperative to maintain an affiliative stance

while also creating breaks in the progressivity of talk to meet the needs of the assessment task.

In this way, via a choreographed turn-taking allocation system using eye gaze and gesture, P, A and I show an orientation towards equality of opportunity within the confines of the face-to-face task. It is not the agreement which is accountable as it is unlikely such actions are dispreferred – agreement is characteristically a preferred option which requires no accounting for and thus does not lead to an expanded sequence of turns. It is more that each speaker is expected to play their part and to do so requires avoiding sequence closure while maintaining a discussion. Even in the final turn of the decision phase of the B2 First task, Group 2 exhibited the same interactional practices:

F2F Group 2 – Extract 2 (final utterance of decision phase)

```
1→ I:    Yeah I agree with both you guys. Eating also is very
2        important (.) because of eating (.) you get to do this stuff
3        and the combination of sleeping enough and eating in a healthy
4        way (.) I think that's very good.
5  EX:   Thank you. That is the end of the test.
```

By contrast, decision-making in the online task was very different in character, as shown in Extract 3.0 below.

OL Group 2 – Extract 3.0

```
1   P:    So (.) the third one is completely out of-,
2   A:    [Yes.
3   I:    [Yes.
4→  P:    Yes okay (.) so we all agree.
```

Extract 3.0 is a short sequence focused on discounting the third video (Out of this world). We can see in line 4 that the "we all agree" is not a prelude to anything further and is used to close the sequence which began in line 1. In lines 2 and 3, there is no attempt made by A or I to re-specify the referent or "account" for their "yes". Agreement is a given, and decisions are reached - in short. In the online task for Group 2 the object "I agree" occurred once (it was used three times face-to-face), which is interesting given that talking time was significantly longer (9+ minutes online as opposed to 4 minutes total face-to-face). This single instance is shown in Extract 3.1 below.

OL Group 2 – Extract 3.1

```
1    P:    so (.) I'd say we go with the first ad and then cross
2          out the slides and therefore put the content of the second one
3          in=
4    I:    =yes=
5    P:    =not all of them but just like a hint then make it more like
6          suitable like the form of the first one not like a power point
7          presentation more like a movie to watch
8→   I:    I[agree with you per- like (.) to full extent=
9    A:     [perfect
10         [=okay
11→  I:    [=and cancel the third one completely like(.)=
12   A:    =yes=
13   I:    =no heheh (.)[we are not going to do that
14   A:                [no the third one is OUT heheh
```

In Extract 3.1 the group is discussing how to improve one of the adverts they have seen and P is explaining that the first video is the better option with changes – with a view to reaching a consensus over which video should be selected for the advertising campaign. This is in the final stages of their discussion (one minute before the talking time ends). Here we see the object "I agree" (line 8) used in a very different manner to that of the face-to-face task. It is notably different in that it does not have a rationale or account attached to it; instead, it is accompanied by an upgrade "to full extent" and sequentially the turn constructional unit "I agree with you … to full extent" can be seen as an upgrade (Pomerantz, 1984) on "yes" I gives in line 4.

Extract 3.1 exemplifies a feature of the online talk-in-interaction elicited by the video-input task in that there were no instances where agreement is oriented to as accountable (i.e., requiring explanation) or re-specifiable (i.e., requiring redefinition). In contrast to the face-to-face extracts, the "I agree" is much more closure-relevant (Schegloff, 2007) in the sense that it is oriented to by speakers here as potentiating a shift of topic. This is evidenced in the way in line 11, a continuation of I's utterance in line 8, I moves on to "and cancel the third one completely". This leads, notably, to further unaccounted-for agreement in lines 12-14 on the topic of the third video, which nobody likes (as shown in A's use of stress on "OUT" in line 14).

Whereas "yes I agree" (or a semantic equivalent) was used in turn-initial position to do the preliminary work of what is often a much longer turn or series of TCUs in the face-to-face task, in the online task these objects are used in a sequence closing manner. These turns are often minimal and "do not project any further talk" (Liddicoat, 2011, p. 187); they are closure implicative (Schegloff, 2007) in relation to the topic or idea being discussed. In Extract 3.1, I's "yes" in line 4, A's "perfect" in line

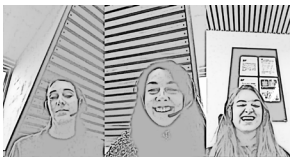
9 and "okay" in line 10 and "yes" in line 12 are all indications more generally of an orientation by speakers not to explain agreements but to seek a natural and expedient closure of the topic under discussion (of which Extract 3.0 above is an extreme case).

More generally, the online interactive environment is shown here to be one of "absolute adjacency" (Jefferson, 1986) where there is no beat of silence between turns (e.g., Extract 3.1, lines 11-13). Where overlaps occur, they are quickly resolved. While instances of overlapped speech can potentially lead to misunderstandings, Extract 3.1 shows how I and A can still understand one another in that the joke and laughter is visibly shared in lines 13-14. These two turns at talk are also a sign of the heightened sense of affiliation one sees in the online interaction more generally. A's turn in line 14 is a completion of I's "no" in the immediately prior line. Such completions are indicators of sophisticated and coordinated behaviors, reliant on close recipient monitoring and a clear understanding of what a turn in progress is projecting (Bolden, 2003). In contrast to Extract 1 lines 18-20, the collaborative turn-taking in Extract 3.1, lines 13-14, shows a grammatical and semantic completion of I's turn which is indicative of heightened engagement.

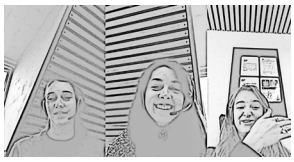
It is useful to note differences in embodied conduct, too. Extract 3.2 leads immediately from where Extract 3.1 stopped and is a continuation of the decision-making sequence but, crucially, starts after the point at which the three speakers have decided which video to use.

OL Group 2 – Extract 3.2

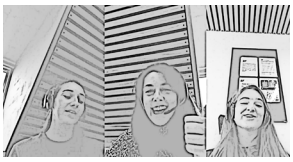
```
15  I:    [heheh yesA
16  P:    [it's not us
17  I:    er:: noB [heh heh
18  P:             [heh heh
19  A:    so:: I guess we came to an agreeˈmentˌ with that yeah?
20        That's greatC
21  P:    [°yes (inaudible)°
22  I:    [yes we agree:::d [that's gu:::d hehehD
23  A:                      [yay! heheh
```



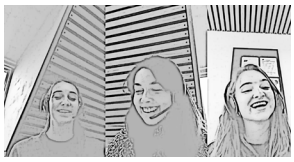
A: A and I laugh as P says "It's not us"



B: I makes a 'swipe left' gesture as she says "er no"



C: A gives a 'thumbs up'



D: P and A smile as I joins in joke "that's gu:::d"

Figure 6. Embodied conduct in the online task

In Extract 3.2 I and A use hand gesture as a way to amplify what is being said (Image B and C in Figure 6). I uses a swipe left gesture to make a joke about the unsuitability of the third video while A holds a thumbs up gesture for several seconds to make a light-hearted comment on the groups surprising ability to come to an agreement. The key thing to note, in comparison with the face-to-face embodied conduct shown in Extract 1, is that the speakers orient to being visible to their co-interactants and the affordances (Hutchby, 2001) of the medium (i.e., that hand gestures are a viable means to convey meaning alongside the verbal). It is interesting to consider the impact of delimited or obfuscated eye-gaze has in the online mode, given the relative importance it seemed to play in Extract 1 in coordinating speaker actions and turn-taking and what we know of eye-gaze's importance to talk-in-interaction in terms of next-speaker selection (Rossano, 2012; Hjulstad, 2016). For Group 2, we can see how the teleconferenced mode typically made all three speakers more tuned in throughout the talk (e.g., Extract 3.2) when compared to the choreography of Extract 1. Yet, at the same time, they did not have the interactional affordance of being able to direct eye-gaze to coordinate next-speaker selection. This constraint in online interactions may well be why there was a greater need to tune in more.

Notably, groups were asked to speak for significantly longer in the online task and invariably they did so. This begs the question: how was this achieved? Disjunctive topic shifts were a feature of the face-to-face task and, while a necessary expedient in a shorter task in order to cover the points in the input, they effectively limit the nature of the talk more broadly. In the online task, topic shifts were more "stepwise" in character (Jefferson, 1984). Extract 4 is part of a discussion which emerged out of all three speakers querying the fact "100's of free flights" were offered as prizes for a supposedly eco-friendly company. As we pick up the discussion, A, I and P are discussing other incentives the company could offer in their advert that better suit this environmental ethos.

OL Group 2 – Extract 4 (discussion of eco-friendly video (Our promise))

```
1   P:    Yeah (.) instead of (.) like oh flying around the world (.)
2         we could do like (.) uhm=
3   I:    =Yeah=
4   P:    =A picnic in the park and that would be (.) like- (.) and then
5         we have our drink with us in the picnic
6   A:    Ye::s. [Yeah (.) that would be good
7   I:           [°or train?°
8   A:             [What sorry?
9   I:    Yes or:  [(.)on the bike? Or train journeys?
10        Or something with (.) cycling maybe?
```

```
11  A:    Mhmmm (.) Or maybe backpacking? hhhh
12  I:    Y[e::s
13  P:     [Yeah.=
14  I:    =Good one.=
15  P:    =So=
16  A:    =That could be a good one (.) you need your sugar and you
17        need your energy
18  I:    [Mhmm
19  A:    [to go backpacking heheh=
20  I:    =And maybe? also show something like a schoo:l? and it's like
21        their la:st days? (.) all stuff like that because like
22  A:    [Yeah
23  I:    [Summer is close and everything but people are still in offices
24        and still in school
25  A:    So (.) what we could probably do is keep this summery vibe, but
26        in-
27  I:    [Mmhmm.
28  A:    [But instead of selling u:hm (.) well not selling but offering
29        a trip or a long-term trip just showing that the summer starts
30        (.) whenever you take one of our soft drinks
31  P:    [°Yeah°
32  I:    [↑Mmm↓  tha:ts a good one
33  A:    You know, (.) maybe they could work with that?
34  I:    Mmhmm=
35  A:    =What do you think,
36  P:    Yeah uhm I'd say it's a goo- great idea I would say uhm
37        we should include real-life pictures?
```

Again, one can observe the rapid and frequent speaker changes, the frequent but quickly resolved overlaps. In line 8, there is also an other-initiated repair operation ("What sorry?") unlike anything seen in the face-to-face data for this group. This also provides evidence of how the turn-taking system can resolve issues of overlap specifically in the online interactional environment, overlaps potentially caused by latency (Seuren et al., 2021).

We can note the frequent speaker changes are linked to each speaker turn providing incrementally more to the previous speaker turn (e.g., lines 3-14). In contrast to the longer turns seen in the face-to-face task, these turns are shorter and far more embedded in the sequence of turns in which they arrive. The overarching trajectory of the talk is on arriving at a conversational destination rather than the less focused talk seen in Extract 1 – there is a sense of progressivity (Stivers & Robinson, 2006) being maintained. Agreements are closure-implicative and do not rely on re-specification of a referent. There is a greater sense of other-attentiveness in the way suggestions are try-marked (as in lines 7, 9 and 11). Appraisals and assessments are succinct (e.g., "good one" in line 14).

All such features lead to a more natural interaction where stepwise topic shifts are seen to occur. Up to line 19 of Extract 4 the discussion has been focused on different ideas for the prizes which could be given away. In lines 20-21 and 23-24 I begins to shift the focus to an ancillary matter (i.e., the people in the advert[9]). A then picks up this idea "keep this summery vibe" and extends it – this turn effectively stabilises the topic shift. In lines 25-30 A completes the topical pivot away from ideas for the giveaways to a more general approach to the advert ("summer starts whenever you take one of our soft drinks"). This is met with agreement by I and P; A recycles this confirmation request in line 35 ("What do you think?") to which P provides an upgraded assessment ("it's a goo- great idea") before another ancillary matter is raised ("real-life pictures").

Elicitation of opinion in line 35 is sequentially and functionally much more natural than in face-to-face data in that, while it is still focused on bringing a speaker in from the cold, the focus is on appraising an idea that is the product of their peers, which is very different from an idea presented on a mind-map, that does not rely on having monitored the trajectory of talk. Thus, P's turn is important as it indicates that while she has not been as active in the talk in this sequence, she has been monitoring it and can integrate her projected next topic into the pre-existing sequence of turns without disjunction or disruption (i.e., having real images not artwork, to support the summer vibe idea). Also note its framing as a question (i.e., try-marked), again showing an other-attentiveness which is more prevalent in online data. Topics are not initiated in a blunt or disjunctive fashion; they are ancillary matters (i.e., connected to prior focus) which stabilise within a sequence of turns and across speakers.

Post-trial interviews

Findings were reflected in candidate post-trial interviews. A, P and I were asked to reflect on the two interactions, face-to-face and online:

> "Although it (the online task) is an examined task, you're having a conversation of three and the examiner is not there - you kind of forget what they tell you in the exam training that *always say why* at least in my case it was that."
>
> Group 2 – Candidate A

[9] This is a subtler move than may appear to the reader because in the video the people were shown on holiday as part of the plug for free holidays to be won – the prizes and the actions shown were intertwined.

A's comments above are notably close to observations made in prior research in terms of performed talk (Stokoe, 2013; Hüttner, 2014; Luk, 2010; Lam, 2015). As P highlights, there is a desire to qualify agreements face-to-face as a means to an end and topics require less interactional work to establish when they are omnipresent:

> "Yeah, I feel that when you're talking to the examiner you wanna impress them by using a lot of words and showing what you can do – so you try to talk as much as possible and giving a reason just makes sense then cos you can pick out another topic again without having to think about something new."
>
> Group 2 - Candidate P

P also explained that in training for the face-to-face task she had developed a habit of avoiding disagreement because "we would never be able to finish" the task (i.e., cover all five ideas). Again, this fits with CA findings presented here in that agreement with qualification is an expedient yet sufficient means to impress for the purposes of assessment while avoiding a potentially less controllable (and disaffiliative) sequence-expanding disagreement.

Discussion and implications

This study explored the nature of the differences in the interactions generated in the two task settings and more broadly to investigate how online affordances may be exploited to create a less testwise interaction (i.e., where the interactional agenda is not pre-determined by the task's omnipresence).

Turn-taking in the face-to-face B2 First task was highly choreographed and part of a precision-timed ecology of embodied actions (notably eye-gaze). This was combined with longer turns at speech. Agreement is often accompanied by the current speaker re-specifying a referent in a manner not dissimilar to other conditional agreement phenomena (Pomerantz, 1984). While there is an attentiveness to other-speaker utterances, topic shifts are disjunctive in character (perhaps deliberately, to be visible for assessment).

In contrast, the online task elicited swifter speaker change with shorter turns and more evident no-gap, no-overlap transitions between speakers. While overlaps were more present online, these were quickly dealt with and by their very existence led to repair

initiations which are an important facet of interactional competence. Agreement in the online task was swift and led to the closure of one topic and the opening of another. Speakers did not try to prolong deliberations. This, in turn, allowed for a more exploratory discourse (Mercer, 2002) containing more subtle topic shifts and talk that was more focused on getting things done. The evidence here suggests a far greater speaker focus on topic closure to achieve progressivity in the online interaction than the more familiar, more controlled task.

Gesture, while only fleetingly covered, was different in the two modes. In the face-to-face mode, hand gestures were deictic in nature (i.e., to point to the shared task resource). In the online mode, they were more emblematic in nature with a direct semantic relation to the talk (e.g., thumbs up and swipe left). This suggests that mode is influential in the non-verbal behaviours tasks may elicit, which in turn would need to be recognised in any expansions or elaborations of the interactional competence construct where NVBs feature. Similarly, eye-gaze behaviours were evident in face-to-face interaction as a means to coordinate turn-taking that was not replicable online.

Albeit small in scale, this qualitative study raises interesting questions regarding the way in which online speaking assessments might be utilised to further promote a sociolinguistic-interactional construct where hitherto monological and semi-direct testing has predominated. Such tests are often underpinned by an individual psycholinguistic construct which primarily sees talk-in-interaction as a validity threat as much as a validity asset - due to the variability it naturally brings into the speech samples elicited. This conception of speaking focuses "little attention to situational and social context of language use" (Roever & Kasper, 2018, p. 332). Online tasks, like the one used in this study, have the potential to rebalance this current state of high-stakes speaking assessments.

The challenge for test developers is the extent to which we can harness new task designs (be they online or face-to-face) to capture less testwise samples in the future - to better represent a sociolinguistic-interactional construct of speaking. Arguably, and in line with previous studies, one means to achieve this could be the adoption of triads. The manner in which candidates were able to bring either themselves or others into the dyadic participatory framework of the interaction (as in Extract 4, lines 35-36) suggests triads can elicit a wider range of interactional skills. Entry into this dyadic

participatory framework in the online task required interactional work for it to come off; it was not merely a case of waiting for one's turn.

The relative familiarity of the candidates with the two tasks led to qualitatively different interactions. At the risk of corrupting a well-worn interactional metaphor, what we have seen here is the difference between a grand slam tennis final and an exhibition match. Echoing the findings of Youn (2020), these examples of test-taker interactions "illustrate how learners' orientation to the contexts and contingencies result in a different degree of interactional fluency" (Youn, 2020, p. 109).

As ever, it is a case of balancing reliability with validity. It is not the contention of this paper that candidates should be hurled unwittingly into unfamiliar tasks for the sake of validity. It is more that task designs like a mind map present a pre-packaged list of points for discussion which potentially diminishes the scope of the task to elicit stepwise conversational moves and the kind of progressivity seen in everyday interaction. This is perhaps particularly true where candidates are trained in how to manage and apportion the topical content of tasks across turns and speakers. In essence, this concern for reliability eats away somewhat at the scope for candidates to create their own locally-relevant points and to weave them into the flow of the interaction. All this is not to say that one task is inherently better than another. However, it does reflect arguments for a renewed focus on the central importance of interactional competence to the assessment of speaking (Salaberry & Burch, 2021).

Limitations

This research was conducted with a small convenience sample, so the generalisability of findings is limited. It is also worth remembering that trials were conducted immediately prior to the global pandemic, as such data may represent a snapshot of pre-covid interaction that is different from how candidates behave post-pandemic. The two tasks were very different so it is reasonable to assume they would elicit different interactional samples (the focus in this paper is not that they wouldn't differ, but to explore how they were different). With specific reference to testwiseness, there is no knowing what well-trained candidates may do interactionally with this kind of online task and it is acknowledged that if somebody wants to contrive interaction, they will find a way. A key point in this regard is that this contrivance becomes far harder

to hide in a conversation where topical points or shifts emerge within talk itself. In terms of examiner presence or absence, preparation time, task duration – the researcher accepts these are all variables that could impact language elicited and further studies should explore these in a more controlled fashion. This was not a comparability study. It was an exploratory study of the kind argued for in recent literature (e.g., Nakatsuhara et al., 2021).

Conclusion

Testwiseness has the effect of creating a self-fulfilling prophecy in terms of language sampling for speaking assessment – as candidates become better able to "talk to score" (Luk, 2010) it impedes meaningful inferences about interactional competence. Arguably, task performances like those on the B2 First task here might be viewed as unnecessarily limiting or controlling in terms of what they enable candidates to do interactionally. While this enhances comparability of task outputs across samples, it does so at the expense of potentially sampling much more context-sensitive phenomena like the ability to move stepwise through topics or to mimic the progressivity that prevails in talk-in-interaction in real life. As shown, tasks can afford test-takers greater scope "to use the target language to talk about what the participants themselves make locally relevant, rather than to the (still omni-relevant) activity of the language assessment task" (Burch & Kley, 2020, p. 49).

Candidate orientation to task and setting cannot be underestimated in terms of the influence it has on language samples obtained – and this is no less true of online interactions as they are explored in future. This study is hopefully a useful contribution to future task design initiatives, in understanding the affordances of the online setting and discussions of the threat "testwiseness" presents to meaningful assessment of interactional competence.

Ethical approval

This research was approved by the ethics committee of the author's post-graduate research institute in the UK, following all legal considerations (e.g., full DBS checks) and in line with BAAL and ILTA ethical guidelines.

References

Bolden, G. (2003). Multiple modalities in collaborative turn sequences, *Gesture*, *3*(2), 187-212. https://doi.org/10.1075/gest.3.2.04bol

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, *26*(3), 341–366. https://doi.org/10.1177/0265532209104666

Burch, A. R. & Kley, K. (2020). Assessing Interactional Competence: The role of intersubjectivity in a paired-speaking assessment task. *Papers in Language Testing and Assessment*, *9*(1), 25-63. https://arts.unimelb.edu.au/__data/assets/pdf_file/0010/3545371/2020-91_2_Burch-and-Kley.pdf

Burton, J. D. (2020). "How Scripted is this Going to Be?" Raters' Views of Authenticity in Speaking-performance Tests, *Language Assessment Quarterly*, *17*(3), 244-261. https://doi.org/10.1080/15434303.2020.1754829

Council of Europe (2020) *Common European Framework of Reference for Languages: Companion Volume with New Descriptors*. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Cox, T. (2017). Understanding intermediate-level speakers' strengths and weaknesses: An examination of OPIc tests from Korean learners of English. *Foreign Language Annals*, *50*(1), 84–113. https://doi.org/10.1111/flan.12258

Egbert, M. M. (1997). Schisming: The collaborative transformation from a single conversation to multiple conversation. *Research on Language and Social Interaction*, *30*(1), 1–51. https://doi.org/10.1207/s15327973rlsi3001_1

Folland, D. & Robertson, D. (1976). Towards objectivity in group oral testing. *ELT Journal*, *30*, 156–167. https://doi.org/10.1093/elt/XXX.2.156

Firth, A. (1996). The discursive accomplishment of normality: On 'lingua franca' English and conversation analysis. *Journal of Pragmatics*, *26*, 237-259. https://doi.org/10.1016/0378-2166(96)00014-8

Galaczi, E. (2004) Peer–peer interaction in a speaking test: The case of the First Certificate in English, New York: Columbia University. Unpublished doctoral dissertation.

Galaczi, E. (2008) Peer–peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly*, *5*(2): 89-119. https://doi.org/10.1080/15434300801934702

Galaczi, E. (2014) Interactional competence across proficiency levels: how do learners manage interaction in paired speaking tests? *Applied Linguistics*, *35*(5), 553–574. https://doi.org/10.1093/applin/amt017

Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students, *Language Testing*, *27*(4), 585-602. https://doi.org/10.1177%2F0265532210364049

Gan, Z., Davison, C., & Hamp-Lyons, L. (2008). Topic negotiation in peer group oral assessment situations: a conversation analytic approach, *Applied Linguistics*, *30*(3), 315-334. https://doi.org/10.1093/applin/amn035

Goodwin, C. (1979). 'The interactive construction of a sentence in natural conversation' in G Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology*, pp. 97-121.

Goodwin, C. (1981). *Conversational Organisation: Interactions between Speakers and Hearers*, Academic Press.

He, L., & Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing*, *23*(3), 370–401. https://doi.org/10.1191%2F0265532206lt333oa

Heritage, J., C. (1984). "A change-of-state token and aspects of its sequential placement", *In Structures of Social Action: Studies in Conversation Analysis (J. Maxwell Atkinson, John Heritage, eds.)*, Cambridge, U.K., Cambridge University Press, pp. 299–345.

Heritage, J. C. (1988). Explanations as accounts: A conversation analytic perspective. In C. Antaki (Ed.), *Understanding everyday explanation: A casebook of methods*, 127-144. Sage.

Hjulstad, J. (2016). Practices of organizing built space in videoconference mediated interactions, *Research on Language and Social Interaction*, *49*, 325–341. https://doi.org/10.1080/08351813.2016.1199087

Hutchby, I. (2001). *Conversation and Technology: From the Telephone to the Internet*, Polity Press.

Hüttner, J. (2014). Agreeing to disagree: 'doing disagreement' in assessed oral L2 interactions, *Classroom Discourse*, *5*(2), 194-215. https://doi.org/10.1080/19463014.2014.893897

Jefferson, G. (1984). On stepwise transition from talk about a trouble to inappropriately next-positioned matters in J. M. Atkinson and J. Heritage (Eds): *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, 191-222.

Jefferson, G. (1986). Notes on 'latency' in overlap, *Human Studies 9*, 153-183. https://www.jstor.org/stable/20008965

Jefferson, G. (2004). 'Glossary of transcript symbols with an introduction' in G. H. Lerner (ed.): *Conversation Analysis: Studies from the First Generation*. John Benjamins, 13–31.

Jin, Y. & Zhang, L. (2016). The Impact of Test Mode on the Use of Communication Strategies in Paired Discussion. In: Yu G., Jin Y. (Eds) *Assessing Chinese Learners of English*, Palgrave Macmillan.

Knoch, U., Huisman, A., Elder, C., Kong, X., & McKenna, A. (2020). Drawing on repeat test takers to study test preparation practices and their links to score gains. *Language Testing*, *37*(4), 550–572. https://doi.org/10.1177/0265532220927407

Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: a conversation analysis of role-plays and non-scripted interviews in language exams. *Language Testing*, *16*(2), 163-188. https://doi.org/10.1177%2F026553229901600203

Lam, D. M. K. (2015). Contriving authentic interaction: Task implementation and engagement in school-based speaking assessment in Hong Kong. In G. Yu & Y. Jin (Eds) *Assessing Chinese learners of English: Language constructs, consequences and conundrums* (pp. 38-60), Palgrave Macmillan.

Lam, D. M. K. (2018). What counts as "responding"? Contingency on previous speaker contribution as a feature of interactional competence. *Language Testing*, *35*(3), 377–401. https://doi.org/10.1177/0265532218758126

Lam, D. M. K. (2021). Don't Turn a Deaf Ear: A Case for Assessing Interactive Listening. *Applied Linguistics*, *42*(4), 740–764, https://doi.org/10.1093/applin/amaa064

Lazaraton, A. (2002). *Studies in Language Testing 14: A Qualitative Approach to the Validation of Oral Language Tests*. Cambridge University Press.

Lazaraton, A., & Davis, L. (2008) A microanalytic perspective on discourse, proficiency, and identity in paired oral assessment. *Language Assessment Quarterly*, *5*(4), 313-335. https://doi.org/10.1080/15434300802457513

Liddicoat, A. J. (2011). *An Introduction to Conversation Analysis (Second Edition)*, Continuum.

Luk, J. (2010). Talking to Score: Impression Management in L2 Oral Assessment and the Co-Construction of a Test Discourse Genre. *Language Assessment Quarterly*, *7*(1), 25-53. https://doi.org/10.1080/15434300903473997

Mercer, N. (2002). *Words and Minds*. Routledge.

Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice, *Educational Psychologist*, *17*(2), 67-91. https://doi.org/10.1080/00461528209529246

Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, *65*, 137-156. https://doi.org/10.1016/j.pragma.2014.04.004

Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality. *Research on Language and Social Interaction*, *51*(1), 85-106. https://doi.org/10.1080/08351813.2018.1413878

Nakatsuhara, F. (2011). Effects of test-taker characteristics and the number of participants in group oral tests. *Language Testing*, *28*(4), 483-508. https://doi.org/10.1177%2F0265532211398110

Nakatsuhara, F., Inoue, C., Berry, V., and Galaczi, E. (2017). Exploring the use of video-conferencing technology in the assessment of spoken language: A mixed-methods study. *Language Assessment Quarterly*, *14*(1), 1–18. https://doi.org/10.1080/15434303.2016.1263637

Nakatsuhara, F., Inoue, C., Berry, V. & Galaczi, E. (2021). Video-conferencing speaking tests: do they measure the same construct as face-to-face tests? *Assessment in Education: Principles, Policy & Practice*, *28*(4), 369-388. https://doi.org/10.1080/0969594X.2021.1951163

Nunn, R. (2000). Designing rating scales for small-group interaction. *English Language Teaching Journal*, *54*(2), 169–178. https://doi.org/10.1093/elt/54.2.169

Ockey, G. J. (2001) Is the oral interview superior to the group oral? *Working Papers, International University of Japan*, 11, 22–40.

Ockey, G. J., Gu, L. and Keehner, M. (2017) Web-Based Virtual Environments for Facilitating Assessment of L2 Oral Communication Ability, *Language Assessment Quarterly*, *14*(4), 346-359. https://doi.org/10.1080/15434303.2017.1400036

Plough, I., Banerjee, J., Iwashita, N. (2018). Interactional competence: Genie out of the bottle. *Language Testing*, *35*(3), 427-445. https://doi.org/10.1177%2F0265532218772325

Pomerantz, A. (1984). Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 57-101). Cambridge University Press.

Pomerantz, A. (2005). "Using participants' video stimulated comments to complement analyses of interactional practices." In H. te Molder & J. Potter (Eds.) *Talk and cognition: Discourse, mind and social interaction* (pp. 93-113). Cambridge University Press.

Psathas, G. (1995). *Qualitative research methods, Vol. 35. Conversation analysis: The study of talk-in-interaction*. Sage Publications, Inc.

Roever, C. & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing*, *35*(3), 331-355. https://doi.org/10.1177%2F0265532218758128

Rossano, F. (2012). Gaze in Conversation. In *The Handbook of Conversation Analysis* (eds J. Sidnell and T. Stivers).

Salaberry, M. R. & Burch, A. R. (Eds.) (2021). *Assessing Speaking in Context: Expanding the Construct and its Applications*. Multilingual Matters.

Schegloff, E. A. (1992). *Introduction to Harvey Sacks* in G. Jefferson (ed.) *Lectures on conversation*, ix-lxii, Blackwell.

Schegloff, E. A. (2007). Sequence Organization in Interaction: A Primer in Conversation Analysis. *Sequence Organization in Interaction: A Primer in Conversation Analysis I*.

Seuren, L. M., Wherton, J., Greenhalgh, T. and Shaw, S. E. (2021) Whose turn is it anyway? Latency and the organization of turn-taking in video-mediated interaction. *Journal of Pragmatics*, 172, 63-78. https://doi.org/10.1016/j.pragma.2020.11.005

Spence-Brown, R. (2001). The eye of the beholder: Authenticity in an embedded assessment task. *Language Testing*, *18*(4), 463–481. https://doi.org/10.1177%2F026553220101800408

Stivers, T. (2021). Is Conversation Built for Two? The Partitioning of Social Interaction. *Research on Language and Social Interaction*, *54*(1), 1-19. https://doi.org/10.1080/08351813.2020.1864158

Stivers, T. & Robinson, J. D. (2006). A preference for progressivity in interaction. *Language in Society*, *35*, 367-392. https://doi:10.1017/S0047404506060179

Stokoe, E. (2013). The (In)Authenticity of Simulated Talk: Comparing Role-Played and Actual Interaction and the Implications for Communication Training. *Research on Language and Social Interaction*, *46*(2), 165-185. https://doi.org/10.1080/08351813.2013.780341

Swain, M. (1984). Large-scale communicative language testing: A case study. In S. J. Savingnon, & M. S. Berns (Eds.), *Initiatives in communicative language teaching*. Reading, MA: Addison Wesley.

Winke, P. & Lim, H. (2017). The effects of test preparation on second-language listening test performance. *Language Assessment Quarterly*, *14*(4), 380–397. https://doi.org/10.1080/15434303.2017.1399396

Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, *10*(2),196–218. https://doi.org/10.1080/15434303.2012.721423

Youn, S. J. (2020). Pragmatic variables in role-play design for the context validity of assessing interactional competence. *Papers in Language Testing and Assessment*, *9*(1), 95-125. http://www.altaanz.org/uploads/5/9/0/8/5908292/2020_9_1__4_youn.pdf

Yu, G., He, L., Rae-Dickens, P., Kiely, R., Lu, Y., Zhang, J., & Zhang, Y. (2017). Preparing for the speaking tasks of the TOEFL iBT® test: An investigation of the journeys of Chinese test takers. *TOEFL iBT Research Report, 28*. https://doi.org/10.1002/ets2.12145

Appendix 1

Symbol / annotation	Explanation
,	A comma indicates a continuing intonation contour
?	A question mark indicates a rising intonation
.	A full stop indicates falling or terminal intonation
–	A dash indicates a cut off word or sound (i.e. due to speaker withdrawing or lowering volume of speech)
A: word= B: =word	Equals symbols are used to indicate where two turns are 'latched' (i.e. they come one after the other with no noticeable beat of silence, no-gap, no-overlap)
(0.3)	A timed pause, the decimal is intended to show tenths of a second duration (e.g. a pause marked (1.2) = 1 second and two tenths of a second).
(.)	An un-timed pause, usually a micro-pause
(word)	Bracketed words indicate where a word is the analyst's best guess at what the item is due to unclear sound or articulation
WORD	Where a word appears in all capitals it indicates louder than surrounding talk (i.e., in the context of the interaction)
°word°	Degree symbols are used to indicate speech which is noticeably quieter than surrounding talk
word	Underlining indicates where a particular syllable or item is given noticeable stress
>word<	Narrowing chevrons indicate speech which is faster than surrounding talk
<word>	Widening chevrons indicate speech which is slower than surrounding talk
wo:::rd	Colon symbols indicate a stretched sound, the more colons the more stretched.
[word [word	Brackets indicate the onset of overlapping talk
↑word↓	Arrows indicate audible changes in pitch
hhh	The use of 'h' on its own indicate audible out-breaths, the number of letters indicating length/duration
.hhh	The use of 'h' fronted by a full stop indicate audible in-breaths, the number of letters indicating length/duration
heheh	This type is used to indicate laughter
…	Three dots indicate that the remainder of the turn at speech have been omitted by the analyst
word^{A}	The use of capitalised A, B, C etc. as superscript within the transcript are used to indicate specific snapshots of non-verbal/embodied conduct which coincide with the verbal transcript.
	Embodied actions are shown in an adjacent table, where relevant. This approach has been adopted to disrupt the reading of the transcript as little as possible while still giving adequate room for inclusion of non-verbal phenomena. Images are 'bleached' to maintain an indication of conduct while protecting individuals' identities.

Appendix 2

What is your first name?

Which school are you at?

- [Participants were provided a list of schools involved]

Gender

- Male
- Female
- Prefer not to say

Age

- Under 18 years old (Note: consent must be confirmed by parent/guardian below)
- 18-20 years old
- 21-25 years old
- 26-30 years old
- 31-35 years old
- 36-40 years old
- 41-45 years old
- 46-50 years old
- Over 50 years old

What is your first language?

How long have you been learning English?

- I have never had formal language teaching
- Less than one year
- 1-2 years
- 3-4 years
- 5-6 years
- 7-8 years
- 9-10 years
- More than 10 years

Are you planning to take an exam in the future?

- Yes
- No
- Maybe

If you answered 'yes' please put the name of the exam and when you plan to take it below.

How familiar are you with the Cambridge First Speaking test?

- I know nothing about this test
- I've heard about it, but that's all
- I know a little bit about the test
- I'm somewhat familiar with the test
- I'm very familiar with the test

How often do you use teleconferencing software like Skype?

- Very frequently

- Frequently
- Occasionally
- Rarely
- Very Rarely
- Never

If you use teleconferencing software, what do you use it for?

- Social (friends)
- Social (family)
- Work
- Study
- Other (please specify below)

If you selected 'other' please describe how you use teleconferencing software below.

How much time do you usually spend using the internet each day?

- Less than 1 hour a day
- 1-2 hours a day
- 3-4 hours a day
- 5-6 hours a day
- More than 6 hours a day

Do you have a laptop computer available to you where you live currently?

- Yes
- No

Does it have internet access?

- Yes
- No

Appendix 3

Interviewer frame

Opening discussion

Interviewer: Okay, so in the trials you did two tasks. One was an online task where you talked for almost ten minutes without an examiner. The other task was a more standard face-to-face mind map task – in this case from B2 First Part 3 – which was a three-minute discussion followed by a one-minute decision-making phase.

In both cases, we recorded what happened and before we look at some observations I've made, I want to ask for your general impression.

Looking back, do you feel the two tasks led you to talk to each other in a different way?

Back-up: Did you talk differently in the two tasks? Why?
Extension: How was it different?

Participants: [Each is asked to respond in turn]

Interviewer: And, of the two tasks, which did you prefer? Why?

Participants: [Each is asked to respond in turn]

At this point the interviewer takes a more reactive role, for example, if training is highlighted as a feature of why talk was different the interviewer should ask the other participants if this is also their impression etc.

Specific focus on features found

Interviewer: Okay, so since the recordings were made I have been viewing them and I have found some interesting things in the recordings. Here is a transcript of your talk in the face-to-face task [interviewer shows excerpt and describes feature of interest].

My question to you is why you think the talk is this way?
Note: Interviewer may directly contrast with online talk examples, as appropriate.

Participants: [Each is asked to respond in turn]

Closing

Interviewer: And, before we stop, is there anything else you would like to raise or say about your experience of doing these two tasks?

Participants: [Each is asked to respond in turn]

Interviewer: Thank you for all your help with this research both today and in the trialling.