Language assessment literacy for language learning-oriented assessment

Liz Hamp-Lyons

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

This paper reflects on the findings of a small-scale and exploratory study which attempted to explore whether and how learning-oriented assessment opportunities might be revealed in, or inserted into formal speaking tests, order to provide language assessment literacy opportunities for language teachers teaching in test preparation courses as well as teachers training to become speaking test raters. Hamp-Lyons and Green (2014) closely studied a set of authentic speaking test video samples from the Cambridge: First (First Certificate of English) speaking test, in order to learn whether, and where, learning-oriented behaviours could be encouraged or added to interlocutors' behaviours, without disrupting the required reliability and validity of the test. We paid particular attention to some basic components of effective interaction that we would want an examiner or interlocutor to exhibit if they seek to encourage interactive responses from test candidates: body language (in particular eye contact; intonation, pacing and pausing); management of turn-taking and elicitation of candidate-candidate interaction. We call this shift in focus to view tests as learning opportunities *learning-oriented language assessment (LOLA).*

Keywords: language assessment literacy; learning-oriented language assessment; speaking assessment

Recent developments in language testing and educational measurement have embraced formative assessment and broader concepts of *learning-oriented assessment* [LOA] (Carless 2003, 2007). These developments all involve cognitive engagement by both learners and teachers, and an emphasis on strategies to support *assessment for learning* [AfL] (Black & Wiliam, 1998; Stiggins *et al,* 2004; Wiliam & Thompson, 2007). In a recent study, Tony Green and I elaborated principles of what we have called '*learning-oriented* **language** *assessment*' (LOLA), and we reached the view that the factors affecting a 'learning orientation' in any assessment have as much to do with the beliefs and principles of

Email address for correspondence: liz.hamp-lyons@beds.ac.uk

teaching as they do with principles in testing and assessment. We chose a particularly challenging context in which to explore LOLA. We chose to explore ways in which we might support teachers and test developers to bring more learning opportunities into large-scale formal tests. With funding from Cambridge Assessment, we conducted a research project that aimed to explore whether, and how, greater pedagogical validity might be brought into such tests and associated materials. We chose to look specifically at the speaking test within the Cambridge: First (formerly known as FCE), an international English language test aimed at the CEFR B2 level.

Principles of learning-oriented assessment

Although formative assessment and the cluster of assessment types usually associated with it are not completely distinct from the kinds of assessment usually grouped together under the term summative assessment, we have found it useful to identify principles and processes that are central to modern forms of formative assessment and link these specifically to learning-oriented assessment: in our case, to learning-oriented **language** assessment.

Carless, Joughin, Liu *et al*. (2006) suggest that effective learning-oriented assessments are characterised by three essential features:

1. tasks are designed to stimulate effective learning;
2. learners are actively engaged in evaluating the quality of their own performance and that of their peers;
3. feedback is timely and shows learners how to act to progress their work, supporting current and future learning.

All three of these key aspects of learning-oriented assessments are very complex, and until recently have not been part of the (rather sparse) resources for language teachers wishing to increase their assessment literacy. In fact it is only recently that most language teachers have begun to see assessment as a part of the responsibility of the teacher rather than depending on large-sale tests to drive their assessment information. Therefore, drawing on the body of work briefly referred to above and on our own experiences as educators, test developers and trainers of test developers and users, we began by elaborating on these three foundational principles and placing them at the core of our definition of learning-oriented language assessment. We (Hamp-Lyons and Green 2014) then added two further key elements:

4. teacher questioning (Black, Harrison et al., 2003), and

5. scaffolding of performance (Shepard, 2005; Andrews, Clarke & Callahan, 2008).

1: Tasks that engage students in learning

Assessment tasks can never fully replicate target language use in naturally occurring situations, partly because test takers are always aware that they are being assessed, and partly because target language use domains can never be fully defined or assessed. However, assessment tasks should lead students to engage in processes that develop their learning, and help them to build the skills they will require if they are to perform effectively in the real world. Language assessment tasks should emulate the kinds of tasks that language users engage in beyond the classroom or examination hall, that is, tasks that are marked by complexity and a degree of unpredictability. This means assessment tasks should be designed deliberately to be learning tasks. In learning-oriented assessment the task is key, because "the overarching question is always 'How does this assessment practice support learning?" (Carless, Joughin, Liu *et al*, 2006, p. 9). If language assessment is to be language learning/language learner-oriented, assessment tasks must be selected and constructed so that they become learning tasks: the relevant skills need to be built into the assessment literacy materials for LOLA.

2: Learner engagement, peer- and self-assessment

'Sharing success criteria' is one of the key practices in assessment for learning, and it also increases learner motivation (Mak, forthcoming). Carless, Joughin, Liu *et al*. (2006) see awareness of the criteria for judging performance as helping students to close the gap between their current performance and the desired standard. Providing rating criteria and samples of performance at different levels for teachers and learners in preparing for a test gives teachers and learners the opportunity to build their understanding of the required performance standards and to take concrete steps towards progress. Teaching learners how to use assessment criteria as scaffolds to self-assess or comment on their peers' work stimulates the active engagement of learners in thinking about their own learning. Underpinning such teaching is a deeper reflection on the meaning and value of specific assessment criteria than many teachers have been prepared for in their initial training any 'language testing' course they may have had and calls for specific teacher assessment literacy materials.

3: Feedback and feed-forward

Feedback is a key route for a learner to understand what s/he has done well or poorly (and perhaps why); but when done well feedback can also be a mechanism for promoting learning. Feedback provides learners with guidance on what they can do to change their

performance in order to improve it on a new task: this is known as *feed-forward* (Duncan, 2007). Feedback fosters 'current' learning because it helps learners to build their understanding of concepts. Effective feedback becomes feed-forward when it supports the learner not only to improve on the immediate task, but also to generalize the feedback for application in other similar tasks—and to new tasks and new learning contexts. Face-to-face speaking tests have the potential to provide learners with at least two forms of feedback: immediate feedback ('feed-up') from the examiner embedded in interaction (in the form of prompting or error correction or similar interventions), and retrospective feedback provided after the event in the form of a test score or report on performance. The first of these can help learners to modify and improve performance and to learn how to perform better during the test event. Although the provision of retrospective feedback on speaking test performance is quite common, little research has looked at the use of immediate feedback practices during a speaking assessment event. In any test event, of course, it is difficult or impossible to integrate feed-forward in any planned way. However, as conceptualization and understanding of feed-forward grows, it should become possible to incorporate this into language teacher assessment literacy training and resources.

4: Questioning Processes and Practices

'Questioning' in formative assessment usually relates to how tasks are represented and negotiated with learners. But in speaking assessment (as in classrooms) the tasks in the formal sense of 'items' offered to learners are also potentially stimuli for students/test-takers to display their full range of ability beyond the narrow categories normally captured in rating scales. From the latter perspective, questioning should elicit the test candidate's best performance across all domains of language. Furthermore, in LOLA, good questioning by teachers (and assessors) is of course important, but learner questioning, in a wide range of modes and processes, is at least equally important. With encouragement, learners can use questions to back-channel, request clarification, elicit external feedback, plan learning strategies, self-assess, and use questioning as a spur to critical thinking (Hamp-Lyons & Tavares, 2011). Clearly, this complex aspect of language assessment calls for explicit assessment literacy support for teachers attempting the learning-oriented use of questioning.

5: Scaffolding

Scaffolding is a deliberate educational process by which skills or materials are introduced gradually to learners, and is a constant and essential feature of language teaching and learning. Questioning may be seen as one component of scaffolding. Effective scaffolding depends on teachers' skilled observation of learning and (at least in the case of language

learning) on their ability to work contingently and in real time with learners as they engage in language activity. Well done, scaffolding is very fruitful; however, concerns are often raised about 'unintentional scaffolding', the moments that seem to slip in to speaking tests, usually without the assessor or interlocutor being aware of them (Galaczi, 2008). This is a complex area of teacher assessment literacy, and seems to suggest that teachers, and teachers as interlocutors on speaking tests, need to become more aware of what scaffolding **is** before they could be expected to avoid it.

The LOLA model

Figure 1 shows the foundational principles of learning-oriented assessment (LOA) and the key processes or strategies through which teachers and assessors can put LOA into practice, laid out beside their 'cultural equivalents' in large-scale testing. The model is conceptual: note that it does not suggest a contrast between summative and formative assessment in the sense that those terms are usually applied (see Hamp-Lyons, 2007 p. 494). Rather, it lays out the extreme points of the learner-oriented *vs.* authority-oriented continuum (here we use 'continuum' only metaphorically since there appears to be no research to identify the ends and parameters of a continuum in fact).

We believe that the special conditions of *learning-oriented* language assessment require an elaboration of this model, one that acknowledges the importance of interaction, as described by Young (2000), who uses the term 'interactional competence'. We therefore added a further row, to suggest the continuum between a language assessment that, through applying, eliciting and probing interactive competence, is in intention inherently construct representative, and one at the further end of the continuum is neither interactive nor learning-oriented: this is shown in Figure 1a.

Learning-oriented assessment	⇔	Large-scale testing
Learning-focused tasks		**Judgement-focused tasks**
Scaffolded task completion		Support not permitted
Interactive/exploratory questioning		Interlocutor/assessor questioning
Learner-involved assessment		**Learner-excluded assessment**
Self-assessment		Authority assessment
Peer evaluation		Authority assessment
Learning-focused feedback		**Judgement-focused feedback**
Interlocutor scaffolding		Summary decision reporting
Immediate feedback		Delayed or no feedback
Focus on feed-forward		No feed-forward
Assessment is contingent		**Assessment is absolute**

Figure 1. An emerging model of learning-oriented assessment and its contrasts with large-scale testing

The centrality of interaction in speaking assessment

Although interaction has not itself been a core principle either in assessment for learning or in LOA, this is because most research has been carried out in subject areas other than language learning. In the context of **speaking** assessment, interaction can be considered both a key skill for language learners and a key context in which there should be fertile ground for the development of LOA opportunities, processes and practices – and assessment. Because the skills that underpin effective interaction are vital to successful communication and achievement of social and professional goals, we must make interaction a focus—of test items and task design, of the teaching in courses preparing for speaking exams, of interlocutor/assessor speech during the test event, as well as between interlocutor and candidate(s) and between candidates. The role and use of questioning, and the effective use of scaffolding are vital components of LOA as we conceive of it, and they are also aspects of a speaking test event that are most likely to demand or lead to the use of candidates' interaction skills.

Learning-oriented language assessment is inherently interactive		Large-scale testing under- represents the constructs of language

Figure 1a. Acknowledging the importance of interaction

When <u>all</u> elements of this model are present we may refer to the resulting forms of assessment as *learning-oriented language assessments.*

The Study

1. The FCE Speaking Component

Candidates take the speaking component of the FCE in pairs or groups of three; however, throughout the test they are assessed on their individual performance and not in relation to each other. They are awarded marks by two examiners: an assessor and an interlocutor. The interlocutor awards a mark for the performance as a whole, using the Global Achievement scale, while the assessor awards marks for five individual criteria: Grammatical Resource; Lexical Resource; Discourse Management; Pronunciation; Interactive Communication.

The test lasts approximately 14 minutes, and the task types that occur in the test are Part 1 short exchanges with the interlocutor and with the other candidate; Part 2 a 1-minute 'long turn'; Part 3 a collaborative task involving the two candidates; Part 4 a discussion.

The design of the test means that assessors can focus on different criteria in different parts of the test; but each student receives only one score on each criterion/domain, and not separate scores on each test part.

2. What do interlocutors do?

Clearly, as in most paired speaking tests, most sections of the Cambridge: First Speaking test are dominated by interlocutor questioning. Little guidance is offered in teacher support materials on questioning processes other than a focus on the types of question that will appear in each part of the test and advice on interactive questioning as a strategy to demonstrate engagement to satisfy the 'Interactive Communication' criterion.

When Young and Milanovich (1992) looked at the FCE interview, they found a very unbalanced interviewer-candidate relationship: the interviewer displayed the features of power and the interviewee(s) displayed submissiveness. Goal-setting and expectations were very different for the interviewer versus the candidate, and questioning was driven by assumptions of fact-based answers. These findings and other Cambridge research led to a number of studies of interview discourse in the Cambridge Speaking tests and a decision to move to a more structured format.

Lazaraton (2002) discussing her studies of Cambridge oral proficiency interviews, specifically the CASE, had reported that interlocutors provided accommodations to candidates that were not specified by the instructions to and training of examiners/interlocutors. Over a period of years, these issues were pursued within Cambridge, first with 'main suite' exams and then with IELTS, and, as explained by Brown & Taylor (2006) led to the implementation of more rigorous rater training and the development of a more structured interlocutor frame in order to ensure that there would be less variation in the interviewing technique of examiners. However, research on the effects of accommodations by speaking test raters on student test-takers have shown mixed results (Ross 1992; O'Sullivan & Lu (2006). To my knowledge, no studies reporting assessment literacy training for new or experienced raters have been done.

3. Is there a language learning orientation in FCE Speaking?

To study this question, we studied four FCE Speaking test standardising samples and six other samples, and paid most attention to the standardising samples since they are taken to represent the official Cambridge view of what an FCE Speaking test event should look

and sound like. 2   We drew on cognitive validity theory (Weir 2005), conversation analysis (Seedhouse, 2005) and models of interactional competence (Young, 2011). Nakatsuhara (2008) identified substantial differences in interviewer interactional behavior, and Nakatsuhara (2009) found a range of interactional competence: scaffolding behaviour, use of body language, negotiation of meaning; topic personalisation; types and frequency of sequence-openers; joint utterance completion, avoidance behaviour, involvement of jokes and turn-taking.

Bringing these aspects together with our focus on looking for a language-learning orientation in these formal test events, Hamp-Lyons and Green (2014) intensively studied the samples we had been given and came up with the following key categories of interlocutor behavior which seem the most supportive of what we have called '*learning-oriented language assessment*' (LOLA):

- body language including eye contact
- topic maintenance and management
- intonation, pacing, pausing
- clarification strategies (or absence of)
- error correction or repair
- management of turn-taking in paired section
- elicitation of candidate-candidate interaction

In this paper I look at several examples of interlocutor behavior, trying to identify LOLA-promoting and LOLA-potential as well as LOA-discouraging behaviours, with a view to building teacher/interlocutor assessment literacy materials from the data. Where it was also possible (depending on time of shift of camera angle) to confidently relate candidate verbal or non-verbal response to the immediately-preceding interlocutor behavior, relevant comments have been included. Green (this issue) discusses LOLA promoting behaviours and potential strategies for teachers and learners in the test preparation materials.

Learning orientation in Part 1: Interview

In Part 1, candidates are introduced to the interlocutor and assessor, and are asked simple questions intended in part to put them at ease. The assessment focus during Part 1 is on general interactional and social language: ability to provide information about oneself and to offer personal opinions on a range of topics. Although this section is described as

---

2 Our samples were provided by Cambridge ESOL. However, examples are available on the internet: the sample here, on YouTube, is an official Cambridge ESOL video:
https://www.youtube.com/watch?v=EdeZp0n0JHw

"short exchanges with the interlocutor and with the other candidate", we did not observe any cases of interaction with the other candidate in the 10 sample interviews provided.

Eye contact

Often, when conversation data is transcribed, body language is not transcribed, or only limited body language data is recorded. Language testing is turning its attention to language as a whole structure of meaning over multiple channels, and eye contact, proximity behaviour and so on are increasingly of interest. Research increasingly supports the claim that "gaze functions to provide information, regulate interaction, express intimacy, exercise social control, and facilitate service and task goals." (Kleinke, 1986, p. 78), and business negotiation manuals always stress the importance of eye contact in establishing solidarity. Research into the effects of eye contact on the elaboration of speech in oral interviews would be valuable, perhaps most importantly with lower proficiency candidates.

Because of the importance of questioning as an LOA-promoting process, and indeed of questioning in the classroom, we might expect this to be a fairly natural shared experience for candidates and interlocutors – and in some cases it is. But there are fairly frequent instances of the interlocutor failing to make eye contact with a candidate, or making eye contact at an inappropriate moment:

Extract1 1: FCE Interview A:

> Interlocutor (Int): (eyes down) First of all (makes eye contact towards both candidates) we'd like to know something about you, so I'm (eyes down) going to ask you (glances up) some questions (eyes down) about yourselves (3 sec pause/silence, eyes still down)… (looking up) where are (looking at S1) you from .. (name)?

Unfortunately, we do not see the candidates' reaction while the interlocutor is speaking: but the moments of eye contact are out of phase with the meaning of what Int is saying: when someone tells us they are going to ask us a question (especially in a high-stakes testing context) our attention is fully on them and we are expecting them to look at us. In this case, Int looks *down* when saying "questions" and there is (what feels like a long-ish) pause before she looks up quite suddenly and asks (with gaze apparently toward the two candidates in general, although our angle of sight is very poor) and only after "are" does she say "you" while noticeably turning her head to look at <u>one</u> candidate. This (what felt to us) inappropriate eye contact practice was repeated at times throughout this interview and in two others where we saw this examiner. This is not *learning-oriented assessment* in

action nor any precursor of it. The need to remain close to the frame is evident in the interlocutor's body language, specifically eye contact and gaze patterns, during Part 1.

Topic initiation and maintenance

As Young (2011) indicates, one of the essential discursive practices of effective interaction is topic management: that is, the ability to identify and follow a topic; to notice and react to topic shifts; and to appropriately shift and/or nominate new topics. Young notes that '[t]opic management includes preferences for certain topics over others and decisions as to who has the right to introduce a given topic, how long a topic persists in discourse, and who has the right to change the topic' (p. 8). In OPIs topic management can be a problem area because the interlocutor-candidate(s) encounter is neither natural nor equal. Not surprisingly, in all the standardizing samples we observed some instances of sudden topic shift, or failure to maintain topic, by the interlocutor: for example:

Extract 2: FCE Interview A:

> S1: I like my Korean food and … (0.50-1.03--very hard to understand pronunciation: looks like S2 is not understanding)

[*During this 13 second turn there is no clarification request, request for repetition, or error correction, from I*]

> Int: (smiles) 'kay (draws breath creating a pause: renews eye contact) Do you come from a large or small family?

The interlocutor can be forgiven for not understanding S1; but the rejection of all of the previous turn and the immediate move to a new question with the same candidate is quite the opposite of learning-oriented. This unannounced topic shift would be difficult to manage in normal conversation between speakers of the same language: in discussion between speakers of 3 different languages, and especially immediately after a failed attempt at interaction, it is quite marked. S1's confusion is evident in the very short following turn and S1's eye movement, looking up and right as if for external input:

> S1: er… er… I'm from.. small family (looks into distance (… nods) … (looks at I)

However, the following interlocutor handles the difficulty of topic progression and of eye contact in this artificial context effectively:

Extract 3: FCE Interview C:

> Int: Looking at S2) So (S2'name) (looks down a second) What do people do in their **free** time (looks down, then up).. where you live?

> S2: (1.12-1.27): We have a very big park [15 sec turn]…

> Int: ah **hah** (looks down then up) (S1's name), what do you do enjoy.. um (glances down) (hand gesture) doing *in the evenings*? (this phrase is spoken faster and with rise-fall on *evenings*)

Int's insertion of "in the evenings" in place of "in … free time" into the frame, coupled with a slight stress and tone emphasis, resulted in noticing and comprehension, and a fruitful turn, from S1. Whether or not this would be considered deviation from the frame is unreported. In the next turn, Int needs to at the same time shift the topic and address the other speaker: she begins by turning her body slightly to S2 and naming her:

> Int: (Name)… What did you do last weekend?

> S2: I went to London.. because I'd like to visit the… Tower of London (continues)

This interlocutor's use of gaze and the slight pause before addressing the question to S2 seems to be effective in priming the candidate for a turn, as the candidate speaks for 23 seconds.

Intonation, pacing, pausing

In considering why, in addition to the problems with eye contact and topic management in Interview A, Int A conveyed a sense of being less oriented to the candidate than Int C we noted rather fast speech patterns, and pauses that appeared to be 'interlocutor-oriented', by which we mean that they occurred when the Int needed time to think or make strategic decisions on how to go forward. In contrast we noted a slightly slower pace of speech from Int C, and very short but observable pauses in putting questions to candidates, which we suggest were more candidate-oriented, suggesting an inclination towards a learning orientation.

Learning orientation in Part 2: Individual 'long turn'

Part 2 consists of an explanation and directions by the interlocutor, and a one-minute turn by one candidate to describe a visual, followed by a shorter turn (about 20secs) from the second candidate giving their own opinion. In the samples we viewed there was almost no occurrence of eye contact between the two candidates, and no case where the Int managed to facilitate a S-to-S exchange. In fact in some cases the speaking candidate remained eyes/head down throughout her/his long turn. One sample was remarkable in

that both candidates kept their voices completely flat throughout their long turns (even though one of them made quite a clever joke) and nor did they make eye contact with the Int.

In our Part 2 video samples we saw no occurrences of **clarification strategies,** very minimal **self-repair,** and no occurrences of **Int error correction.** There was almost no opportunity for Int-S interaction because of the structure of this part: instructions were delivered as a monologue which closed with 'OK?" or 'All right?" from the interlocutor. Much of the time the camera was directed at the pictures and not at Int or the Ss, so we could not observe interlocutor **eye contact**. **Pacing** seemed a little slower than in Part 1 in most cases, but we did not use a time counter.

In terms of the principles of LOLA it would seem that Part 2 offers opportunities for initial scaffolding, reformulation or narrowing the task by the Int to provide the S with enough grounding to get started on the intended task. Whether that would be judged to be too much or inappropriate support in the speaking test context would depend on the construct behind the test design that has led to candidates being given such a task. However, it seemed evident to us that the requirement of control over the input and interaction with the candidates results in much narrower ground for language learning-oriented strategies for the interlocutor

Learning orientation in Part 3: Conversation

Brooks (2009) found that "when test-takers interacted with other students in the paired test, the interaction was much more complex and revealed the co-construction of a more linguistically demanding performance than did the interaction between examiners and students. The paired testing format resulted in more interaction, negotiation of meaning, consideration of the interlocutor and more complex output." (p. 341). In line with Brooks, in Part 3 we saw more linguistic range and more interactional strategies, as the extract from this interview illustrates:

Extract 4: FCE Interview D: The task is to choose from a selection of pictures of holidays which ones would be best.

S2: So "C", what do you think about we visit Europe by train?

S1: By train? Um.. it's a good... it's a good transport, train.. because it's quite safe.. you don't have to take care where you are going… and you are very quickly from one city to the other one… and you can (mumbles)

S1:   [yeah…

S2**:**   [… different cities

S1: (loudly) and you can go with your friends

S2: yes, yes

S1: and I think for young people you have (hand movement) very.. cheap prices… to go by train around Europe

S2: yes, yes

S1: hm hmm

S2: … and very safe

S1: yeah

This interaction looks and sounds quite like a conversation despite the inevitable artificial context of the interview room. The candidates make eye contact, turning towards each other as they speak, their turns overlap, and they look at each other while listening. There are plenty of non-verbal as well as verbal markers of agreement and solidarity: nodding, smiling, hand gestures. The intonation and pacing are natural for NNS-NNS conversation at this level. Each candidate contributes new information with some degree of enthusiasm in their intonation. However, the next extract from a different video sample shows something different:

Extract 5: FCE Interview B: The task is to choose from a selection of pictures which three jobs are the best.

> S1: "Ye-es… I agree with you (flat)… but then um.. I think this job (points) would be quite interesting and er nice to hear about and er… to have it (downtone).
>
> S2: er.. I disagree with you (flat) because if you do that job for many… er … for long time, at last you 'll be bored I think (flat)
>
> S1: (flat) Ye..es.. -and you will (slightly lifts the tone and speeds delivery) be fat also. *(S2 looks at her and laughs: she has not looked up, but smiles when she hears his chuckle)*

These Ss have clearly prepared thoroughly for the test and know exactly what to do. (This was the case to some extent with all the samples we saw.) They begin the task without interaction over organization or strategy. They each in turn use appropriate, formulaic

turn-taking language, agreeing then disagreeing with each other. They carry through the task exactly as instructed. However, their voices are flat almost throughout and they make very little eye contact.

In terms of **management of turn-taking**, in Part 3 this is all done by the candidates themselves, although it seems to result in stilted turn exchanges. Despite the existence of a substantial body of research on conversational turn-taking, much more and closer research within the context of language tests is needed before it would be possible to identify causal factors of the differences between the two interactions discussed above. However, it seems evident that at a modest level there is a learning orientation in the way the candidates in Interview D are involved in their interaction that is absent in Interview B. Is it language that is being learned, or accommodation? Does the difference affect scores? We don't know—but surely we should try to find out. There are many tools of discourse analysis and conversation analysis as well as of lesson planning in teacher education that could be brought to bear on activities to go beyond teaching the 'rules' of turn-taking and towards helping learners see what makes real conversation and help prospective candidates to develop strategies they can apply within the test itself. This suggestion moves towards the idea, often raised recently in professional conferences, that there should be assessment literacy materials available for learners/test-takers as well as for teacher and test raters/interlocutors.

Learning orientation in Part 4: Discussion

At the time of this study[3], the free Cambridge English teacher support resources offered teachers this task to prepare their learners for *Cambridge: First* speaking:

> 'Imagine you are going to interview a famous explorer. Think of five questions you could ask. Your questions should encourage the explorer to speak as much as possible. Write your questions in the spaces below.'

The stated intention here is to show "how open questions can help to elicit more ideas than yes/no questions". However, in all ten samples we viewed, the interlocutors approached the 'discussion' task as a topic continuation from Part 3, using a series of questions addressed to one or other of the candidates, as this extract illustrates:

Extract 6: FCE INTERVIEW B:

> Int: Do you think it's useful for students to hear about people's jobs?

---

[3] The Part 4 section has since been updated

S1: It's very useful. And it (have?) "becaw"-become compulsory subject to school—(louder) so that children have to be informed for what job they have to follow er.. what jobs er lot of money.. so they can leave (live) their lives good and nice (falling tone)

Int: Thank you (looks down, then up) (Name), how well do schools prepare young people for work?

S2: I think very good (makes eye contact, nods) we can have all the information we want… and erm.. it's very good.. because students must know a lot of things about jobs… and I think er.. from school we learn all the things we want (nods)… and we need (falling tone)

Int: (Name) how difficult is it for young people to find work in your country?

S1: erm.. I find that the students and…the young people in general … it's difficult to find a job and not so ..diff--more difficult to find a job that… interesting

These two students by this point were a little more animated than in Part 3 and there seemed to be opportunities for Int to take a step back and monitor the discussion between the two of them, 'lubricating' it with back-channelling. But as the extract shows, this did not happen. Probably as a consequence, we did not see candidates "show how (they can use) open questions… to elicit more ideas than yes/no questions". There are no occurrences of **candidate-candidate questioning** or other interaction during Part 4 in these samples; we did not see any attempts by interlocutors to elicit inter-candidate interaction, nor did we see any three-way interaction. In order to fully understand why this is the case, we would need to see interlocutor/assessor training materials and to participate in or observe a training session/programme. However, it is clear that initiating somewhat spontaneous 'discussion' effectively enough to set it going and then let it run on its own steam requires interlocutors who have the requisite skills and confidence to facilitate that type of peer-to-peer conversation. This should be within the reach of many FCE candidates: LOLA-focussed teacher materials and training sessions could be very valuable here.

In the next Part 4 extract, we see a student who seems to be below or borderline to the required standards, and Int is probing for further evidence, giving the impression of a less authentic-style 'discussion':

Extract 6: FCE interview A

Int: Do you like the idea of these changes, or do you think the lake should be kept as it is now?

S1: ummmmm…. Can you repeat?

Int: (exact repetition)

S1: I think if we keep as now is better than.. change… m…

Int: S2?

S2: I think some change can be done but not.. too much

Int: Why do you think people like to spend their free time in the countryside? (S2)?

S2: Because it's quiet and you can relax yourself and have er and stay in the .. nature because sometimes people needs.. who living in the city they forget about the beauty of the countryside and sometimes you need to go there to remind you ..of so many beautiful things

Int: S1??

S1: I think just-the most important thing.. wants to be (mumbles) … because many people live in.. live in cities and they can't be the (?) because they are very busy and they have to work every day and maybe they go to holiday to the.. to the countryside

(Int: yes)

Int: How important is it for children to experience life in the country… (S1)?

S1:I think it's very important for children because… mmm… er… yes.. if er because some, many children they are living in city and they are very [selpish?] and they er have a chance to live in countryside they can be less er selpish.. so I think (smiles) it's good for children

Int: mm (nods)… Why might people who live in the countryside want to visit the city? (**now** looks at and names S1)

S1: It's like same thing. e.. if I live in city maybe I want go in country but.. maybe if … their life is too relax maybe er.. maybe their life is too relax maybe they want.. visit.. but maybe some day maybe they don't like city

Int: m hm.. S2?

S2: Ye I agree with (S1)

This was the only instance of a student verbalising a request for clarification in the samples we viewed. Perhaps the combination of two choices—do you like X or do you think Y? with the immediate qualification "should be kept as it is now?" was too grammatically complex for this candidate. After the exact repetition, S1 manages a limited answer: Int's strategy of then turning to S2 and asking the same question (which was not otherwise in this Int's behaviour) is reminiscent of classroom teacher data showing that when a teacher does not back-channel to a student's response but repeats the question to another student, this means the 'answer' was 'wrong'; and that students recognize this. This set of turns seems not only not to be a discussion, but instead to be another questioning session, with the questioning intensively focused on S1, the weaker of the two. Throughout this interview S2 has been part-turned towards S1; the Int's face is mobile and empathetic when S1 is speaking. This was clearly a LOLA opportunity: the Int could have drawn S2 into questioning S1 herself. Swain, Brooks & Tocalli-Beller (2002) describe classroom contexts in which peer-peer support interactions between students at the upper limit of their ZPD as having the potential to raise the performance of the participants. Viewing these instances of Part 4 made us wonder what the real purpose of this Part is and if it does function to probe a different aspect of candidates' speaking performance, as intended. Part 4 seems to offer the best opportunity in this test for application of the LOLA principles above. Deliberately building in interaction through training interlocutors, and providing focused teacher support materials for teachers teaching test prep classes should lead to multiple opportunities for candidates to show what they can do with the spoken language—making this a more *learning-oriented language assessment.*

Implications to be drawn for LOLA

1. Learning-oriented task design and use

As Saville (2009) has argued, a learning orientation needs to be intentionally designed *into* an assessment. Looking at the structure of the test with its four parts, we can see that efforts have been made to keep learning in mind as well as exam success, and we can see some small windows of possibility for a LOLA 'mind set' to begin to seed itself into the test. But there is no deliberate aim towards scaffolding for task completion, and tasks seem not to be designed primarily as learning opportunities but as language display. Moving in that direction would require a programme of research-informed development

built around a willingness to consider changing the way the construct of 'speaking' is currently conceived in the test; test developers/item writers would need re-orientation to help them understand (a) the goals of moving towards LOLA and (b) what these goals look like in practice. With this underpinning, materials to help teachers how to implement LOLA principles in their own classroom test prep tasks would become feasible.

2. Learner-involved assessment: Peer and self-assessment opportunities

There is a rich and substantial literature on peer feedback and self-assessment, and we need to discover the extent to which, in the context of classrooms where Cambridge English Language Assessment examinations are being prepared for, there is evidence of learning being treated as a shared enterprise. LOA elements such as learner autonomy and learning support from peers need to exist in parallel with the collection, selection and submission of evidence of learning. There are indications that Cambridge English Language Assessment is aware this: for example, Galaczi (2013) says: "…our understanding of interactional competence both in the classroom and as a construct underlying tests and assessment scales needs to broaden to include not just interactional features such as topic development organization, **but also listener support strategies and turn-taking management**" (p. 553). These features of dialogue are complex and have been found in a number of studies to be salient to scores. In these videoed speaking test events we observed two moments when a candidate gave small body language signals or under the breath murmurs suggesting an urge to help their co-candidate, but in neither case did they do so. This is, of course, quite unlike co-construction as frequently found in conversation and dyadic interaction (Norton, 2013). Inserting peer- and self-assessment strategies such as listener support into the classroom will begin to provide learners with the tools to engage with each other over something significant—their own performance, and to practice turn-taking, questioning, repair and other important interactional skills in a personally relevant and 'authentic' context. Similarly, finding ways to provide peer feedback during the test, and encourage candidates to self-assess and repair in real time if they can, would make the test more authentic to everyday NNS-NNS or NNS-NS interaction; and would also make the test itself a learning-oriented experience.

3. Learning-focused feedback

The absence of all kinds of feedback in these data was striking, especially when looking for strategies to help teachers and future speaking assessment interlocutors (and trainers) to build some LOLA into the task they have. Examples of interlocutor scaffolding to support learners to complete task are absent in the small dataset we studied.

We did observe cases where an interlocutor smiled more genuinely/enthusiastically when a candidate said something particularly interesting or amusing, and we may have missed instances because of the use of only one camera in the data. We also observed instances where (to our eyes/ears) the interlocutor revealed a negative response to a candidate turn with a downtoned 'hm', for example. If this is feedback, it is of the most limited kind (and even this is discouraged by the interlocutor frame). Whether this back-channelling was noted by the candidates is a different question.

4. Focus on feed-forward

This is the LOA principle that seems to be most problematic for the Main Suite exams at present, or at least for the FCE as we have observed it. The suggestion made above, about finding ways to bring peer feedback and self- assessment into the test, apply here also: however, we have not observed any instances of either behaviour in this small study. A step further forward towards a learning orientation in the design of the speaking tests would be the inclusion of opportunities/encouragement for self-repair. The key tools of feed-forward are **scaffolding** and **interactive/exploratory questioning**.

Scaffolding

Possibly the slight slowing of pace by two interlocutors when directly addressing a question or an instruction to (a) candidate(s) is a form of accommodation, but can't be considered scaffolding. In Interview B the conversation flags and is lackluster for half the allotted time of Part 3. Perhaps some contingent scaffolding could have 'jump-started' further ideas and interaction and provided a richer language sample for assessment.

Interactive/exploratory questioning

In the turns of interlocutors, the dominant forms are direct instructions and direct questions. In contrast, opportunities for candidates to question their interlocutor do not occur. Some of the more complex questions posed by interlocutors closely resemble essay tasks, e.g. 'Would you rather live in a small town or a large city?' Even simple 'why' and 'how' questions would allow for more open responses. Currently, however, there is very little evidence to be found for an intention towards a learning orientation in these FCE data.

Discussion

The limited data we have studied closely has shown a (fairly narrow) range of forms of interaction in the various phases of the test, but essentially no scaffolding and

predominantly rote questioning throughout. As Carless, Joughin, Liu *et al* (2006) showed, opportunities often exist but may necessitate a shift of paradigm before those opportunities can be perceived and then acted upon. It should also be noted that tests from other testing bodies display quite similar patterns and a common practice of discouraging truly interactive exchange between interlocutor and test-taker(s). What hope is there, then, for LOLA to enter into speaking assessment contexts?

As O'Sullivan & Lu (2006) point out, the strongest factor in explaining candidate performance is the interlocutor. Since the interlocutor manages the interview throughout, we might well hold him or her responsible for the absence of learning opportunities. However, we have clearly seen in the recorded samples that these interlocutors are tied to a restrictive interlocutor frame that prevents them from using the majority of their skills of 'normal' conversation and even their classroom interaction skills. Our reading of the relevant research, both that linked to research on Cambridge tests and more widely, suggests that there are arguments for and against a rigid interlocutor frame. But to move towards speaking assessments that are language learning-oriented, interlocutors will need a professional development programme that will help them expand their questioning repertoire to include LOLA strategies. It is not usual to talk about 'language assessment literacy' in the context of training test raters, but work in this area is now emerging. In the specific context of Cambridge Assessment speaking tests, mention should be made of recent webinars conducted by Hamilton (2012) and Hamilton and Jones (2013) which have reached out to teachers to introduce ideas of learning-oriented assessment to teachers (and researchers) who are particularly interested. Although the Cambridge model is much more test-oriented than the model Hamp-Lyons and Green (2014) have proposed, it is a step in the right direction.

What can teachers do?

When importance is placed on examination success, this implies pressure on teachers to limit content to reflect test task characteristics. In the case of the FCE Speaking paper, this would appear to involve influences on such classroom activities as timing – the use of brief timed discussions, short impromptu presentations – input – use of paired photographs – and topics – predetermined, non-specialist, inoffensive and internationally accessible: i.e., on the demands of the test tasks rather than on authentic tasks tied to locally relevant learning goals. It is expected that there will be an influence from the test criteria onto the kinds of language and interaction strategies that will be valued and encouraged in the classroom. Again, there is much in the Cambridge English assessment criteria that is essential for communication beyond the test event. Accurate use of a range of language, organisation of ideas, clear pronunciation and effective turn taking are all prerequisites for effective communication. On the other hand, there are

aspects of the test interaction that may not generalise equally well. Learners are likely to learn little from the test about the pragmatics of service encounters or personal tutorials or other real-world interactions. The scripted interviewer role restricts the kinds of help that the test taker can request and limits responsive listening on the part of the examiner. These elements, which are not significant in genuine interaction, are areas that could well be revised or replaced with consciously language learning-oriented tasks and criteria.

Clearly, introducing a genuine learning orientation into a speaking exams is difficult and probably implies a long-term agenda with several smaller stages of change along the way, but it seems clear that the role of the interlocutor is absolutely key. Following are a few ideas for ways to bring a greater LOLA emphasis into contexts such as the large-scale speaking test this study has discussed.

Redefining the role of the Interlocutor

There may be a valuable role for judiciously-timed scaffolding in the warm-up stages of speaking tests, through back-channelling, contingent questioning and sometimes more expert diagnostic probing of the kinds that are for many teachers a near-natural skill. For example, in these Cambridge: First samples, we observed opportunities for interlocutor-candidate(s) in Part 1 through better topic management by interlocutors. Part 2 (the 'long turn') is the most artificial stage of the test: it shows fewest problems in interlocutor behaviours, but it offers least in the way of opportunities for LOA. Task re-design could support interlocutors in making clear to candidates that something more than desultory description is expected, and this could be brought home to them by probing questions, since genuine questioning is a key element in LOLA. Part 3, the ''collaborative' task, has considerable potential for a learning orientation to be designed-in, and could be made somewhat authentic given its relative similarity to a group project in typical undergraduate and graduate university courses; Part 4 is labeled 'discussion' but at best it significantly under-represents the construct of a discussion. Taking the interlocutor out of the frame would free up the potential for candidates' co-construction of language to solve a problem or perform a task and would thus provide a different perspective on candidates' skills.

Create clear opportunities for learner involvement

One of the defining features of learning-oriented assessment is its insistence on involving learners in the process of assessment. Learners should be helped to build an understanding of the criteria by which performance is judged, develop, through supported practice, the ability to assess their own work and the work of others, and take lessons from these processes to improve their own performance. The LOLA principles

suggest that candidate-focused criteria and materials for understanding performances will be valuable and should include opportunities for learners to practice self-assessing and obtaining feedback—there is an important role here for the greater exploitation of many of the new developments in educational technology. Giving and getting peer feedback could be a part of the speaking test itself, and would bring out different facets of performance.

Some conclusions

This study has found that the key to successful LOLA in practice will be the training and monitoring of the performance of interlocutors. This is a key element for LOLA as a newly-emerging area of language assessment literacy, since we have only recently begun to understood just how critical the interlocutor role in speaking assessments. The other essential component is to provide opportunities for teachers who prepare students for this or other speaking tests to attend seminars/webinars on LOLA in practice, and for materials on test agency websites to go through regular updating to enable teachers to upgrade their own LOLA skills and methods.

References

Andrews, S., M. Clarke & C. Callahan. (2008, December). *Integrating assessment and learning in task design through activity theory.* Downloaded from http://web.hku.hk/~sbapro/doc/SBA_and_task_design-Dec_2008.ppt

Black, P., C. Harrison, C. Lee, B. Marshall & D. Wiliam. (2003). *Assessment for Learning: Putting It into Practice* (Buckingham, U.K: Open University Press.

Black, P. & D. Wiliam. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*, 1: 7-74.

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

Brown, A. & L. Taylor. (2006, November). A worldwide survey of examiners' views and experience of the revised IELTS Speaking Test. Cambridge Research Notes, 2: 14-18.

Carless, D. (2003, November). Learning-oriented assessment. In *Evaluation and Assessment Conference, University of South Australia, Adelaide* (Vol. 25).

Carless, D. (2007). Learning-oriented assessment: conceptual basis and practical implications, *Innovations in Education and Teaching International, 44,* 1: 57-66.

Carless, D., Joughin, G., Liu, N. F. and associates (2006). *How assessment supports learning: learning-oriented assessment in Action.* Hong Kong, Hong Kong University Press.

Duncan, N. (2007). 'Feed-forward': Improving students' use of tutors' comments. *Assessment & Evaluation in Higher Education*, *32*, 3: 271-283.

Galaczi, E. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in English. *Language Assessment Quarterly, 5*, 2: 89-119.

Galaczi, E. (2013). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking tests? *Applied Linguistics 34,* 5: 553-574.

Hamilton, M. (2012). *Learning Oriented Assessment: Integrating assessment with learning.* Cambridge Assessment webinar for teachers.

Hamilton, M. & N. Jones. (2013). How can assessment support learning? A learning-oriented approach. Downloaded 1 December 2016 from https://www.youtube.com/watch?v=UtSeNH9PvHw

Hamp-Lyons, L. (2007). The impact of testing practices on teaching. In *International handbook of English language teaching* (pp. 487-504). Springer.

Hamp-Lyons, L. & A. Green. (2014). Applying a concept model of learning-oriented language assessment to a large-scale speaking test. Report submitted to Cambridge Assessment (January).

Hamp-Lyons, L. & N. Tavares. (2011). Interactive assessment: A dialogic and collaborative approach to assessing learners' oral language. In D. Tsagari & I. Csepes (Eds.) *Classroom-based language assessment* (pp. 29-46). Frankfurt: Peter Lang.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin, 100*, 1: 78-100.

Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests.* SiLT 14. Cambridge: UCLES/CUP.

Mak, P. (forthcoming). Exploring EFL Elementary Students' Views about a Writing Feedback Innovation.

Nakatsuhara, F. (2008). Inter-interviewer variation in oral interview tests. *ELT Journal 62, 3: 266-275.*

Nakatsuhara, F. (2009). *Conversational styles in group oral interaction: How is the conversation co-created*? Unpublished PhD: University of Exeter.

Norton, J. (2013). Performing identities in speaking tests: co-construction revisited. *LAQ 10*, 3: 309-330.

O'Sullivan, B. & Lu, Y. (2006). The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. In P. McGovern and S. Walsh (Eds.) *IELTS Research Reports, Vol. 6*: 91-117.

Ross, S. (1992). Accommodative questions in oral proficiency interviews. Language Testing 9, 2: 173-186.

Saville, N. (2009). Developing a model for investigating the impact of language assessment. *Cambridge Research Notes 42*: 24-33.

Seedhouse, P. (2005). Conversation analysis and language learning. *Language Teaching, 38:* 165-187.

Shepard, L. (2005, November). Linking Formative Assessment to Scaffolding. *Educational Leadership:* 66-70.

Stiggins, R., J. Arter, J. Chappuis & S. Chappuis (2004). *Classroom Assessment for Student Learning: Doing It Right, Using It Well*. Assessment Training Institute/ETS.

Swain, M., L. Brooks & A. Tocalli-Beller (2002). Peer-Peer Dialogue as a Means of Second Language Learning. *ARAL*: 171-185.

Weir, C. (2005).*Language testing and validation: an evidence-based approach.* Basingstoke: Palgrave Macmillan.

Wiliam, D. (2006a). *Assessment for learning: why, what and how?* Paper presented to the Cambridge Assessment Network. Downloaded from http://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/Cambridge%20AfL%20keynote.doc

Wiliam, D., & Thompson, M. (2007) Integrating assessment with instruction: what will it take to make it work? In C. A. Dwyer (Ed.) *The future of assessment: shaping teaching and learning* (pp. 53-82). Mahwah, NJ: Lawrence Erlbaum Associates.

Young, R. F. (2000). Interactional competence: Challenges for validity. Paper presented at a joint LTRC/AAAL symposium on "Interdisciplinary Interfaces with Language Testing". (Available online as ED 444 361).

Young, R. F. (2011). Interactional competence in language learning, teaching, and testing. *Handbook of research in second language teaching and learning*, *2*, 426-443. Ed. E. Hinkel. Routledge.

Young, R. & M. Milanovic. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition, 14*, 4: 403-424.