

Learning-oriented Language Test Preparation Materials: A contradiction in terms?

Anthony Green
University of Bedfordshire

The impact of the use of assessment on teaching and learning is increasingly regarded as a key concern in evaluating assessment use. Realising intended forms of impact relies on more than the design of an assessment: account must also be taken of the ways in which teachers, learners and others understand the demands of the assessment and incorporate these into their practice. The measures that testing agencies take to present and explicate their tests to teachers and other stakeholders therefore play an important role in promoting intended impact and mitigating unintended, negative impact. Materials that support teachers in preparing learners to take tests (such as descriptions of the test, preparation materials and teacher training resources) play an important role in communicating the test providers' intentions. In this study, these support materials are analysed. The selected materials, provided to teachers by Cambridge English Language Assessment, go with the Speaking component of a major international test of general English proficiency: *Cambridge English: First*. The study addresses how these materials might embody or reflect learning-oriented assessment principles of task authenticity, learner engagement and feedback within a coherent systemic theory of action, reconciling formative and summative assessment functions to the benefit of learning.

Keywords: Impact; test preparation; alignment; washback

1. Learning-Oriented Language Assessment

Learning-oriented approaches to assessment (LoA) are often contrasted with the rigidity and constraints of standardised testing (Turner & Purpura, 2016). Learning-oriented approaches open a space for researchers, teachers and teacher trainers to re-conceptualise classroom assessment within a separate paradigm: an interactive, formative tool for the improvement of learning, rather than a means of measuring, ranking and selecting learners (Scarino, 2013). It is a key principle that LoA discourages the tendency for teachers to 'model their own classroom tests after the

highly limiting and less-than-ideal tasks found on typical standardized tests' (Pellegrino, 2014). While it has been argued that classroom assessments should be judged against different quality criteria from standardised tests (Brookhart, 2003), advocates of LoA have argued that all assessments, including those used for summative purposes, should be evaluated according to how successful they are in promoting learning. Carless, Joughin and Liu (2006) have suggested that three features are indicative of such a learning orientation:

- 'assessment tasks as learning tasks': tasks that afford opportunities for learning because they invoke real life activities and are completed using 'the knowledge and skills that the course requires'
- learners evaluating their own work and that of their peers
- feedback that involves learners using information from assessment to progress their learning (pp.9-10)

The first of these learning-oriented principles, task authenticity, is clearly compatible with the language testing orthodoxy that tests of communicative language ability should reflect the characteristics of tasks in the target language use domain to which test performance is intended to generalise (Bachman & Palmer, 2010). The second learning-oriented principle referred to by Carless et al. (2006), related to, but distinct from task authenticity, relates to learner agency: self-regulation, interdependence and the potential for these to cultivate lifelong learning. Boud (2000) and Boud and Soler (2016) have used the term 'sustainable assessment' to reflect 'the need for all assessment practices to equip learners for the challenges of learning and practice they will face once their current episode of learning is complete' (p. 401). To sustain this, Boud and Soler (2016) suggested that responsibility for the learning process and for judging its success needs to be devolved to learners themselves. Strategies suggested for promoting self-regulation and interdependence through classroom assessment include the sharing of success criteria and the use of peer- and self-assessment. Following Ramaprasad (1983), the third defining principle listed by Carless Joughin and Liu (2006) – feedback – is conceived in the LoA literature not merely as the provision of comments on learner performance, but as a regulatory mechanism: the use of the observable results of learning (changes in performance over time) to modify the processes of teaching and learning (Boud & Molloy, 2013). Feedback in this sense has only occurred when teachers and learners demonstrate that they are able to use insights from assessment to improve performance and to progress learning.

One large international provider of English language tests, Cambridge English Language Assessment (CELA), has taken up learning-oriented assessment as a strategy for both promoting and measuring learning: 'The term LOA is chosen to emphasise that all levels of assessment can and should contribute to both the effectiveness of learning and the reliable evaluation of outcomes' (www.cambridgeenglish.org/research-and-validation/fitness-for-purpose/loa, see

also Jones & Saville, 2016). This raises the question of whether and to what extent the traditional standardised tests in current use meet these criteria. Are they, or can they become, learning oriented? Hamp-Lyons and Green (2014) reviewed CELA's current practices in relation to LoA principles. Elsewhere in this issue, Hamp-Lyons reports on the learning-oriented assessment potential of existing speaking tests. In this paper, I look at how the tests are presented to teachers through the materials provided to support preparation and consider the extent to which these support a learning orientation as conceived in the LoA literature.

2. Test preparation materials

From a psychometric perspective, test preparation has been regarded by test developers as something of a dark art. It is seen as a threat to test validity, sometimes of questionable ethicality, seeking to exploit features of test design to inflate scores without accompanying gains in proficiency. In other words, it is (if successful in boosting scores) a source of construct irrelevant variance, or 'test score pollution' (Haladyna, Nolen & Haas, 1991). On the other hand, from the perspective of test impact, it is understood that tests need to operate in concert with the curriculum if beneficial effects on teaching and learning are to be achieved (Brown, 2008). Taking a liberal view of validity, one that includes test consequences within validity, Haertel (2013) suggested that it should include consideration of what he termed 'messaging effects, whereby testing is used to shape perceptions or understandings... in ways that do not depend directly on the actual scores' (p.5). Taking a similar stance in setting an agenda for validity research into classroom-based assessment (CBA) in language education, Turner (2012) included a focus on curriculum alignment. She posed the question, 'How do teachers mediate between CBA and preparing students for external tests? What are the underlying commonalities needed to create coherence across the different assessment components in an education system?' (p.75). In conceptualising and evaluating the potential impact of assessment use, the means by which a test is presented and marketed to teachers are arguably at least as important as the qualities of the assessment instrument itself (Green, 2007).

In responding to Turner's (2012) question on behalf of CELA, Jones and Saville (2016) highlighted alignment through shared construct definition based on the levels of the Common European Framework of Reference for Languages (Council of Europe 2001). They suggested that in a learning-oriented language assessment system, the interactional authenticity of tasks is an important consideration. Successful task performance (in class and on the test) should entail the knowledge, skills and abilities required for effective social action in the target language use domain. Both learning tasks and assessment tasks should therefore relate to, and be modelled on, the kinds of tasks that language users engage in beyond the classroom or examination hall.

In addition, Jones and Saville (2016) argued that successful alignment of content needs to be underpinned by a shared interpretation of constructs by all participants in the education system: 'it is essential that the constructs which the tasks implement and make observable are understood in the same way by schools, assessment bodies, governments and society at large.' (p.99). Alignment in language education can only be effected to the extent that educators can find sufficient consensus on the nature of language as a skill and the objectives of language learning. Testing agencies that seek to engender intended impacts cannot remain aloof from the classroom, but must find effective means of communication and points of compatibility with educators on the aims and processes of language learning. In short, 'effective communication and appropriate support for teachers is a prerequisite for achieving intended impacts.' (Saville, 2009, p. 227).

A number of studies in language education, mostly framed as 'washback' research, have investigated score outcomes to establish whether test preparation has been successful in boosting learners' scores (see for example, Elder & O'Loughlin, 2003; Green, 2007; Liu, 2014; Xie, 2013), generally finding only limited evidence for the comparative effectiveness of test preparation programmes. Other studies have used surveys or qualitative interview and observation-based methods to explore processes: how teachers and learners adjust their practices to meet the perceived demands of a test (see for example, Green, 2006; Mickan & Motteram, 2009; Wall & Horak 2011; Gan, 2016). A few studies have considered the impact of tests on materials used in preparing learners for tests. Saville & Hawkey (2004), for example, described the development of an instrument for the analysis of textbook materials intended to gauge the extent to which a test had influenced these. However, none have given sustained attention to how language test providers communicate what they regard as appropriate forms of test preparation or steps that they take to encourage these. Indeed, it is a striking feature of most washback research that the intentions of test providers regarding preparation practices are vague or unstated.

Saville (2009) suggested that testing programmes intending to improve student learning should set out a "theory of action" (p.251). A term widely used in the field of policy evaluation, a theory of action is a plausible and verifiable chain of inter-related steps that makes it apparent how any intended outcomes, such as improvements in teaching and learning, are expected to be brought about through a policy intervention (such as the introduction or revision of an assessment system). As an example of a theory of action, Bennett (2010) describes how the introduction of a Cognitively Based Assessment to US schools includes professional development for teachers, which is expected to lead to the formation of teacher communities for reflecting on the use of the assessment, in turn leading to improved teacher understanding of the content domain, resulting in improved learning outcomes for students (the ultimate impact sought through the introduction of the assessment). Saville (2009) suggests that a

useful theory of action for test reform initiatives should address both intended and potential unintended effects of the assessment system and set out the means by which intended impacts can be promoted and damaging effects avoided or mitigated. In arguing that test development processes should embed strategies to promote intended consequences, or to achieve ‘impact-by-design’, Saville (2009) suggested that, among other strategies, ‘communication of key concepts (e.g. construct related features of the tests) [to teachers] needs to be addressed as an ongoing feature of the [examination] system’ (p. 263). The theory of action underlying CELA’s impact strategy (Saville, 2009; Jones & Saville, 2016) includes communication about the intended uses of a test, information on what it is intended to measure and advice on effective test preparation practices.

Test providers typically offer teachers and learners information and materials designed to help them to prepare. These materials convey both explicit and implicit messages about the testing programme, its relationship to local curricula and the test preparation practices that the providers consider appropriate. Where the improvement of learning is the priority, as Moss (2016) has observed, there is a need for test providers to ‘support educators in connecting test-based information to their practice to explain outcomes, frame questions or problems and explore solutions.’ (p. 237). Relating testing to learning and teaching is particularly challenging when tests are used on a global scale and so must seek compatibility with multiple education systems. Although the CELA agenda for impact research reported in publications such as Hawkey (2006) has noted professional support for teachers as an important factor, this has not been a major focus in studies conducted to date. This paper examines teacher support materials associated with the *Cambridge English: First Speaking* test. It addresses the question: How do CELA teacher support materials transmit the learning-oriented assessment principles of i) task authenticity, ii) learner engagement with assessment criteria and processes and iii) feedback?

3. Methods

The study focussed on the Speaking paper of *Cambridge English: First*, a well-established, international test associated with large numbers of preparation programmes. This study involved a document analysis of test preparation material provided by CELA and directed primarily at teachers. Although it should be noted that documents are not the only resource available (for example, CELA also organises an extensive worldwide programme of seminars for teachers), these materials are very widely obtainable and it was anticipated that they would articulate the test developer’s perspective on test preparation and set out options for the integration or embedding of CELA examinations within language learning programmes.

An initial review of the material directed at teachers through CELA websites was used to select texts that appeared representative of the CELA approach. The chosen documents were then systematically reviewed. Evidence was sought for the extent to which these resources, explicitly or implicitly, referred to and invoked the three LoA criteria of i) task authenticity, ii) learner engagement with assessment criteria and processes and iii) feedback. This analysis provides the basis for an appraisal of whether the current teacher support materials communicate a learning-oriented approach and whether there are opportunities for strengthening the role of LoA in this area.

CELA resource materials for teachers extend from an open forum in which teachers may share ideas and activities for test preparation to subscription courses for teachers, including courses centred on preparation for specific examinations. The Cambridge English website also lists 'official' test preparation textbooks such as Capel and Sharp (2014). Although the inclusion of teacher posts and the listing of third party materials may suggest approval on the part of CELA, this does not necessarily imply that the authors share the test developer's standpoint on appropriate test preparation. As the purpose of the study was to consider the guidance offered directly by the test developer, such third-party materials were excluded.

Within the support material for *Cambridge English: First* offered to teachers by CELA, three types were identified through this initial review.

1. The first type of resource familiarized teachers with the basic characteristics of the examination. Available both online and in print, the default source for familiarisation is the *Handbook for Teachers for Cambridge English: First (FCE)*. The 2012 edition (UCLES 2012) (referred to as 'the handbook') was current at the time of the study.
2. The second type of resource was material for teachers to use in the classroom to help them to familiarize learners with test demands and to prepare them for taking the test. Free resources for teachers preparing learners for *Cambridge English: First* were offered on two websites operated by CELA: www.teachers.cambridgeesol.org and teachingsupport.cambridgeenglish.org. These two sites have now been superseded by a third, www.cambridgeenglish.org/teaching-english, which provides much of the same material.
3. The third source of support consisted of teacher training material, directed at teacher professional development. Relevant to preparation for FCE, there are subscription courses hosted on the Cambridge English Teacher website (www.cambridgeenglishteacher.org) directed at teachers preparing students for *Cambridge English: First*, or at introducing teachers to the CELA approach to assessment.

3.1. The Handbook for Teachers

Handbooks, freely available either in print or by download from the CELA website, provide essential information about the purpose of the relevant examination, an overview of the content and sets of sample material for each of the papers. As such, they are key documents for teachers seeking help with preparation for CELA examinations.

For *Cambridge English: First*, there are four papers: *Reading and Use of English*, *Writing*, *Listening* and *Speaking*. For *Speaking*, the handbook included an overview of the paper and a basic description of each test part. The overview covered *timing*, *number of parts*, *interaction patterns*, *task types* (the people and type of interaction involved: conversation, discussion, collaboration) and *marks* (test takers are scored on their performance across all four parts). For each of the four parts the *timing*, *communicative focus for assessment* and *task type and format* were described. A full sample *Speaking* paper, was provided including the frame (or outline script) for the interlocutor and all input materials required for the test takers. Rating scales were also included – both an analytic scale used by the assessor (the examiner who observes, but does not participate in the interaction) and a holistic scale for use by the interlocutor (the examiner who interacts with the test takers). A glossary in the handbook explained or elaborated some of the terms used in the scales.

3.2. Free Teacher Support web resources

www.teachers.cambridgeesol.org provided a free source of information and resources for teachers preparing students to take CELA examinations. It offered downloadable materials for use in the classroom including sample papers, copies of the test rating scales, videos of test performances and advice on test preparation. These downloadable resources were divided on the website into two sets: 'created by teachers' (33 documents linked to *Cambridge English: First*) and 'created by Cambridge English' (18 documents linked to *Cambridge English: First*). In addition to sample test papers, the documents created by Cambridge English included lesson plans focused on different Parts of the test. A resource titled *FCE Speaking Part 4 Discussion*, was selected for inclusion in this study as, following an initial review, it was judged by the researchers to be representative of the material on offer. This took the form of a lesson plan with instructions for teachers and exercises for students. One of these exercises asked: 'How much do you know about Part 4 of the Speaking test? Read these statements below. One of them is false. Which one?' Another part of the lesson plan suggested that teachers should invite their students to '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 plan called for students to compare the questions they had produced with examples from Part 4 of the Speaking paper. According to the lesson

plan, this was intended to show learners how open questions can help to elicit more ideas than yes/no questions. This activity was followed by student practice in interviewing each other. The stated aims of the lesson were to 'give students an overview of Part 4 of [the Speaking paper], and to help them understand how to respond to open questions.'

Reflecting forthcoming changes to the examination, a second 'Teaching resources' site offering 'Teaching Support for 2015 exam updates' was located at teachingsupport.cambridgeenglish.org. This lacked the teacher-created material, but also provided similar information about the exams, sample papers and classroom activities as well as a link to www.teachers.cambridgeesol.org. One document that could be downloaded freely from this site was chosen as representative of the resources available. This provided a more general set of 'Teaching Tips from Cambridge Examiners' on all four papers on the *Cambridge English: First* examination. The section on the Speaking paper provided a brief summary of the format of each Part followed by numbered tips for teachers: 29 in all. Each tip was a paragraph in length. They included suggestions such as the following:

- (Part 1) Encourage your students to look for opportunities to socialise with English speakers. In class, they could role-play social occasions in which they meet new people, e.g. parties, train journeys, starting a new job. This will give them the opportunity to practise a range of topics for this part of the test.
- (Part 2) Students often find it useful to observe a good model answer given by a more advanced learner of English or by the teacher.
- (Part 3) Remind your students to make positive contributions to move the discussion forward. They should be encouraged to respond to each other's contributions by agreeing, disagreeing and questioning each other, rather than just giving information about the task.
- (Part 4) Encourage your students to give full answers to the questions asked. They can do this by keeping useful question words in their heads, e.g. 'Why?', 'How?', 'When?', 'Where?'. If, when answering a question, students also respond to related question words like these, they will give full contributions.

3.3. Subscription resources: Cambridge English Teacher

To establish the extent to which support materials foster learning-oriented assessment principles, courses on the *Cambridge English Teacher* website presented by CELA and Cambridge University Press were also included. One of the twenty courses on offer at the time of the study was titled *Understanding Assessment*, while six others focussed on preparation for specific CELA examinations (e.g. *How to teach Cambridge English: Young Learners*). Other courses focus on topics such as teaching language skills (Writing, Listening etc.) or teaching specific groups of learners (Primary learners, teenage learners). Two courses were chosen for review because they appeared to be the most

relevant to preparing learners for the *Cambridge English: First Speaking* paper: *Understanding Assessment* and *How to teach Cambridge English: First (FCE)*.

3.3.1. Subscription resources (1): *Cambridge English Teacher Understanding Assessment*

Understanding Assessment promised general guidance for teachers on how they might implement assessment in their classrooms and how CELA tests might support them in doing so. The course included the following ten units: Unit 1: Purposes of and approaches to assessment; Unit 2: Key principles of testing; Unit 3: Language levels and the CEFR [Common European Framework of Reference for Languages (Council of Europe 2001)]; Unit 4: Task types; Unit 5: Testing vocabulary and grammar; Unit 6: Testing reading; Unit 7: Testing listening; Unit 8: Testing writing; Unit 9: Testing speaking; Unit 10: Evaluating tests and whole test design.

3.3.2. Subscription resources (2): *Cambridge English Teacher How to Teach Cambridge English: First*

How to Teach Cambridge English: First familiarised the participants (practising language teachers) with the format and content of the test and offered advice on effective preparation strategies. Five of the ten units were devoted to guidance on preparing tasks to test grammar and vocabulary, reading, listening, writing and speaking with further units on task types and whole test design. Units 9 and 10 of the ten-unit course covered the Speaking paper. They examined what the test takers are required to do in each test part, the language that the tasks are designed to elicit and the roles of the two examiners. The assessment criteria were presented and explained and there was guidance for teachers on preparation activities, including lesson plans. There were video samples of test taker performance to view and evaluate.

Examples of tasks were drawn from CELA examinations: Unit 9 on testing speaking credited the 2012 *Cambridge English: Key for Schools*, *Cambridge English: First for Schools* and *Cambridge English: Advanced* handbooks for teachers. The unit objectives included the following:

‘By the end of Unit 9 you will

- be aware of the practical arrangements that have to be taken into consideration for speaking tests
- understand the skills involved in speaking and the importance of selecting which skills to test
- be aware of the need to select speaking tasks carefully
- be aware of the need for clear assessment criteria and rating scales for speaking’.

4. Results and Discussion

4.1. Learning-Oriented Principle 1: task authenticity

The principle of task authenticity involves connecting assessment tasks to language use in the world beyond the classroom. The *Cambridge English: First* handbook made claims regarding the authenticity of the test tasks: ‘Cambridge English exams... are based on realistic tasks and situations so that preparing for their exam gives learners real-life language skills’ – and suggested that they ‘encourage positive learning experiences, and seek to achieve a positive impact on teaching wherever possible’ (p. 2). Three apparently distinct populations of test taker were envisaged. ‘learners who want to... start working in an English-speaking environment; study at an upper-intermediate level, such as foundation or pathway courses; live independently in an English-speaking country’ (p. 3). This diversity of purpose was not further explained and it was not made clear how the test tasks connected to use of language relevant to these domains. Such information might help teachers to make the connection between the assessment and real-world language use: the learning-oriented assessment requirement that assessment should share the goals of the programme, and that the programme should prepare learners to use language for practical purposes in the real world.

Some of the preparation activities on offer in the *Teaching Tips from Cambridge Examiners* document from www.teachers.cambridgeesol.org invoked a world beyond the test. This was most notable for Part 1 (introductions and exchange of personal information) and Part 4 (discussion). In preparing for Part 1, it was suggested that learners should mingle and greet each other. In preparing for Part 4 they could nominate topics that interested them for group discussion. As the test tasks closely resemble familiar forms of social interaction – meeting and greeting, informal discussions – it is not difficult to imagine learners engaging in very similar ways with friends and acquaintances outside the classroom. In this sense, the material did, in accordance with LoA principles, connect the test tasks to real-world language use.

The relationships between Part 2 and Part 3 and their real-world parallels are more opaque and this is perhaps why the practice tasks – e.g. cut out and compare two photos from magazines, choose the most important event in history – often appeared less natural. If the test tasks are emblematic of a wider universe of speech genres, Part 3 and (to an even greater extent) Part 2 appear to represent more educational and professional forms of speech – presentations and negotiations – where topics in target language use situations are likely to involve more specialised forms of knowledge. Reproducing such tasks in a generally accessible test is more challenging and so it may be inevitable that the test tasks appear less true to life.

What was missing from the free teacher resources was any indication of how the tasks mapped onto broader learning objectives or longer term language learning needs rather than the immediate imperative to pass the examination. This has implications for the sustainability of assessment: How might a teacher preparing students for further education or for work or career enhancement (testing purposes indicated in the specifications) exploit the test tasks in ways that link to these target domains? Although the *Understanding Assessment* course advised teachers preparing for tests to relate the skills tested in the exam to the learners' own needs and interests, it offered no guidance on using assessment to discover more about learners' needs or any practical illustration of how tested skills could be connected to longer-term learning goals.

The *Understanding Assessment* course appeared to take it for granted that the assessment of speaking involves the use of face to face speaking tests (as in *Cambridge English: First*) and options for assessment formats were all presented within this restriction. Variations were suggested (individual, paired or grouped test takers; spoken, visual, audio-visual or written input), but alternatives such as the use of presentations to an audience, observation of spontaneous class interaction or even computer or tape mediated speaking tests were not discussed. Advice was provided in Unit 9.5 on the need to match task types to test purpose: 'we need to specify exactly what we want to test', although for speaking there was no explicit reference to the concept of authenticity or how this should be maximised in a communicative language assessment approach, or in LoA.

It was appropriate that the word *testing* was used in the course unit titles as the emphasis was very clearly on designing test tasks with very little consideration given to alternative forms of assessment. In fact, the term *assessment* was used interchangeably with testing throughout the course and there was no discussion of any distinction between the two. Course participants were informed in Unit 1.4, for example that 'formative assessment is testing which takes place during, rather than at the end of, a course of study.' This does not reflect the much broader interpretation of the term in the LoA literature.

On the *How to Teach Cambridge English: First* course, beyond some generic features of effective communication ('things that effective speakers do') referenced in Unit 9, there was no discussion of the extent to which the test tasks might, in keeping with learning-oriented principles, replicate or represent real-life language use. Although advice was given (mainly in Sections 9.8 to 9.13 and 10.8 to 10.13) on teaching learners strategies that might help them to cope with test demands, the focus of the course was circumscribed by the scope of the test tasks: common topics for discussion, strategies for interaction, gambits for turn taking and organising short monologues. It was not made clear how participating in preparing for test tasks and training in these techniques might connect to longer-term language learning goals or how teachers

might most appropriately integrate test preparation into a locally developed curriculum.

4.2. Learning-Oriented Principle II: learner engagement and self-regulation

In LoA, the principle of learner engagement involves learners building an understanding of what is meant by successful language use through the sharing of success criteria and the use of peer- and self-assessment. The assessment criteria were published in the handbook and appeared on the free websites as downloadable resources. It was suggested that learners should become familiar with them and there were exercises on the websites that might help with the process. There were even some suggestions that they could be used for self- or peer-assessment. For example, in the resource titled *Practising First Speaking Part 2*:

‘students could get into the habit of timing themselves talking about any topic for one minute. They should record themselves, if possible, and listen to themselves. They should particularly consider how well they organise their talk and how they could improve this... Students can use the recordings to analyse their strengths and weaknesses, relating them to the assessment criteria used in the First: Speaking test.’

However, this activity was envisaged as a self-study option and there was no advice on training learners to carry out self-evaluation.

A number of other examples of student engagement in assessment appeared in the free teacher resource materials. In *Practising First Speaking Part 3*, students were asked to reflect on their performance using a checklist – Did you ask your partner(s) for their opinion? Did you dominate the exchange? Did you speak too little? However, it could not be claimed that self- and peer-assessment were a consistent and central element in the materials provided for download. They did not appear regularly, they had no place among the *Teaching Tips from Cambridge Examiners* and they were not presented as an essential element in student learning.

Although Section 2.1 of the Introduction to the *Understanding Assessment* course suggested that assessment could be carried out ‘by national and international organisations, by teachers, by learners, by learners and teachers collaboratively, or by learners in collaboration with each other’, the only example of self-assessment by learners came in Section 3.3 on the use of the CEFR self-assessment grids to arrive at an overall judgement on one’s own level of proficiency. The use of self-assessment to reflect on the performance of specific tasks was not touched on. Similarly, peer-assessment was not further addressed in the *Understanding Assessment* course. Section 2.5 on Impact provided advice on promoting learning within test preparation courses, but did not suggest the use of self- or peer assessment in this context.

Although the use of peer and self-assessment was not given a prominent role, there were elements in the course that could help teachers to understand assessment criteria and so might also help in building this awareness among learners as well. In addition to the CEFR Self-Assessment grid, Section 3.4 presented the CEFR table of 'Qualitative aspects of spoken language use' (Council of Europe 2001, pp.28-9) for use in assessing two sample performances illustrative of different CEFR levels. Section 9.6 presented CELA assessment criteria – with another performance sample for rating – and it was suggested that awareness of the criteria 'can give learners, teachers and schools information to help them plan further learning and teaching'. However, there was no elaboration showing how this might be managed in class.

Participants in the *How to Teach Cambridge English: First* course were directed to the *Specifications and Sample Papers* booklet in the free teacher support resources as part of a process of familiarisation with the test. Review tasks invited participants to match test parts to the kinds of language elicited and to match assessment criteria to their definitions. In Unit 9, participants were presented with a list of 'things that effective speakers do'. This included such features as 'using a range or variety of grammar and vocabulary'; 'not hesitating too much'; 'showing that you are listening by responding to your partner'; 'using word and sentence stress appropriately' and 'talking for appropriate amounts of time, from short answers to longer discussions, depending on the situation'. This list served as an introduction to the assessment criteria and the scales used by the two examiners during the speaking test, which were presented in Section 9.4. In this section of the course, Tasks 4 and 5 asked participants to match the 'things that effective speakers do' to the four assessment criteria of *Grammar and Vocabulary*, *Discourse Management*, *Pronunciation* and *Interactive Communication*. There were opportunities to watch test takers performing on each test part and in Section 10.5, there were explanations of how the test takers' performances were evaluated in relation to the assessment criteria.

Such material should provide participating teachers with a clear understanding of how test performance is assessed. No doubt, teachers could use similar techniques to help their students to understand the criteria. Raising learners' awareness of the assessment criteria as a basis for feedback was advocated in Section 9.11, but no explicit suggestions were made about how to conduct this awareness raising.

Some use of peer assessment in class was clearly envisaged. In Task 14, teaching strategies characterised as 'effective' included 'demonstrate or play recordings of short/unclear answers and interesting/natural answers and get learners to pick out the differences'. The lesson plan in 9.12 included opportunities for learners to identify areas of strength and weakness in learner responses to the teacher's questions. Suggested activities sometimes involved learners as assessors of performance – whether this took the form of responses modelled by the teacher, presented through recorded samples or carried out in class by their peers. On the other hand, there were

no statements about the value of peer or self-assessment in promoting learning and it was not given a central role in classroom activities. Section 9.8 suggested that during test practice students should be grouped in threes with one student acting as the observing examiner while the others acted as test takers, but it was not suggested that the student in the examiner role should award scores or comment on the other learners' performances.

Overall, the support material did include elements of learner engagement in assessment, but this was not the systematic feature that would be expected in a LoA approach. Learner self-regulation through self- and peer-assessment did not regularly feature in the teacher support material. Although there were occasional suggestions for self-assessment activities, these were not the central focus in any of the lesson plans. Success criteria, in the form of the test rating scales, were shared with teachers, but there was limited guidance on using these with learners.

4.3. Learning-Oriented Principle III: feedback

The feedback principle involves the use of assessment evidence to progress learning. In the handbook, feedback was mentioned only in relation to comments on performance on practice tests. It was not included among the *Teaching Tips from Cambridge Examiners* resource and, like self- and peer-assessment, appeared only intermittently in the lesson plans on offer from the free teacher resource websites. Following discussions in both *Practising First Speaking Part 3* and *Practising First Speaking Part 4* the notes for the teacher suggested eliciting feedback from the students on others' performances. On the other hand, there was little indication of the importance placed in learning-oriented assessment on the need for opportunities to act on the comments provided.

The introduction to the *Understanding Assessment* course described how large scale test providers were 'working towards ensuring that tests are designed so that they can be used formatively wherever possible'. The following strategies were said to embody this development:

- 'produce tests that have a positive washback/impact on learners and teachers
- provide more detailed feedback in the form of exam reports and graphical profiles of candidates' abilities
- provide online and other support for teachers/candidates
- use portfolios and continuous assessment
- conduct washback/impact studies'

Although portfolios and continuous assessment were not subsequently addressed in the course, issues of feedback and impact were covered in later units. In Section 2.5 on test preparation, teachers were advised to 'first say what learners did well, then what they can improve on and finally try to end with something positive and encouraging'

and to 'give positive feedback on test preparation tasks and the test itself. For example, put ticks and encouraging comments next to what is correct, not only crosses next to what is incorrect'. These recommendations fit with learning-oriented feedback recommendations for commentary that focusses on task performance and the gap between current and intended performance. On the other hand (in common with the error coding system described in Unit 8: Writing), they would seem to fall short of the learning-oriented requirement to close that gap between current and intended performance by providing learners with opportunities to use the information provided to modify and improve task performance.

In Unit 9, course participants were invited to watch a speaking test and prepare feedback for the test taker. They were then invited to view an examiner providing detailed evaluative comments to the test taker on her performance (although no such comments are given following an official test). Although these comments were given after the test event, they were certainly comprehensive and did include specific suggestions to improve performance. Unfortunately, participants did not see the feedback loop completed. There was no evidence of the use of the examiner's comments by the test taker resulting in improved performance. This is partly because, contrary to the dialogic feedback model advocated by learning-oriented assessment proponents, the comments were delivered by the assessor in the form of a monologue. The test taker was not (visibly) given opportunities to reformulate or self-correct in ways that might help her to learn.

In short, the advice on feedback in the *Understanding Assessment* course centred on providing information to learners and test takers either in the form of detailed and informative score reports from test providers or corrections and comments from teachers. However, in learning-oriented assessment, feedback is understood to be most effective in driving learning if there is evidence of improved performance. The course did not offer guidance on effective strategies for bringing this about.

The *How to teach Cambridge English: First* course did give advice on providing feedback on learner performance. For example, advice to learners connected to some useful expressions was included in Sections 9.8, 9.9, 10.8 and 10.9 and these could be used in feedback to learners on their performance on practice tasks. On the other hand, integrating feedback into the teaching and learning cycle was not a course focus and there was no advice on how to encourage learners to seek out feedback and use it to improve performance.

Some of the suggestions on test taking strategies might constitute good general advice on interactive language use:

- Part 1: 'If you are not sure of the question, ask for help' – 'sorry, can you repeat?'

- Part 2: 'Organise [your] answer so that it is easy for the examiner to understand what [you] are saying'
- Part 3: 'Make sure you are collaborating by asking your partner's opinion and showing that you're listening to them'

Other elements seemed limited to the specific demands imposed by the test situation:

- Part 1: 'Say more than just "yes" or "no" even if it is a yes / no question'
- Part 2: Compare photos rather than describe them separately' – 'both/ neither of them...'
- Part 3: 'After talking about all or most of the pictures, move on to discuss the second question'

There was also advice for teachers on interventions that might help learners to overcome common weaknesses. Again some of these suggestions might be helpful in many language learning contexts – training learners to cope with a lack of vocabulary by paraphrasing, setting up role play situations where learners would naturally meet each other and exchange social and personal information. Others were clearly tied to the specific demands of the test and so could threaten to 'narrow the curriculum': addressing a lack of ideas by brainstorming the kinds of questions or topics that might come up in Part 1 with some vocabulary suggestions for each. The advice included warnings against misguided test preparation practices such as asking learners to write down and learn answers to common questions or giving extensive practice in describing individual pictures (rather than comparing pairs of pictures).

It is noteworthy that there was no information provided on how to use official *Cambridge English: First* results to drive learning forward: the formative use of summative test results. The course appeared to be limited to preparation for the examination rather than its effective integration into teaching/learning processes as envisaged in the kind of learning-oriented assessment cycle described by Jones and Saville (2016). There was no mention of how the tests might be used to modify future learning objectives or of the scope for alignment between *Cambridge English: First* as an external measure and locally relevant objectives.

5. Conclusions

This paper has explored the extent to which the presentation of the *Cambridge English: First* speaking test to teachers articulates the potential for alignment between external tests and local classroom practices in accord with learning-oriented principles. The literature relating to learning-oriented assessment suggested that this might be accomplished by drawing attention to the relationships between test tasks, local curricula and practical language use; by engaging learners in setting goals and

monitoring their progress in relation to the standards embodied by *Cambridge English: First*; and by harnessing feedback processes to improve learning.

The inclusion of a Speaking paper in *Cambridge English: First* is a valuable spur to the teaching of spoken language skills. The four Parts of the test imply a degree of variety in the test preparation classroom that should assist learning: variety that might not be found in preparation for tests that are limited to a single examiner: single test taker interview format. The volume and variety of resources offered by CELA to support preparation conveys a commitment to making connections between the examination and the language classroom. On the other hand, the controlled nature of the test and the fixed nature of the four Parts exert a certain pressure on teachers to constrain what is taught. There must always be a risk that any examination may come to eclipse longer-term learning goals especially in contexts (such as formal schooling) where such longer term goals are not yet clearly in focus.

As yet, the presentation of the test appears to do relatively little to communicate a learning-oriented vision for test preparation. The tendency noted in the literature on washback for language teachers to base preparation activities on the demands of the test tasks rather than on locally relevant learning goals is not seriously challenged by the CELA teacher support materials. On the other hand, it must be acknowledged that this paper is limited by its reliance on documentary evidence. The uses made and practical effects of the presentation of CELA examinations on teaching and learning in local contexts remains to be investigated.

In the context of a test of general proficiency, it may be unclear whether or how learners will progress to using English for travel, further education, employment or to meet other personal or social needs. In such circumstances there would seem to be a particular risk that satisfying the requirements of the examination may become the overriding objective for learning. When the examination is complete, the goal is met and language learning can end. This impulse could be offset by guidance for teachers on interpreting test preparation in relation to potential domains of use beyond the classroom and the examination hall. Although the *Cambridge English: First* test tasks are said to be realistic, there is little explanation in the resources of what realities they represent or of how they are designed to capture the characteristics of language use in relevant target domains.

If 'good learning requires learners to become autonomous and capable of self-direction' (Jones & Saville, 2016, p. 42), the resources for teachers offered little guidance on how this could be fostered through preparing for *Cambridge English: First*. The resources offered no systematic advice on formative processes or on how test results could be used 'not as a final summative judgement, but rather as feedback to guide further learning.' (Jones & Saville, 2016, p. 86). In sum, the teaching resources did not appear to encourage teachers to seek or build coherence between CELA examinations

and the multiple local educational contexts in which the examinations are used. Encouragingly, since the completion of this study, CELA has begun to enhance its support for teachers by including more guidance on LoA: a course titled *Learning Oriented Assessment* is now available through the Cambridge English Teacher website.

One promising way forward suggested by Jones and Saville (2016) is to explore the meaning of alignment between tests and the CEFR (Council of Europe 2001). Although the linking of language tests, including CELA examinations, has involved relating them to CEFR proficiency levels, a richer form of linking is exemplified by the European Language Portfolio (ELP). The ELP, developed in tandem with the CEFR, is intended to support autonomous, self-regulated lifelong learning within a 'pedagogical culture in which learning, teaching, and assessment are naturally integrated with each other' (Little & Erickson, 2015, p. 125). It shares and manifests the conception of language learners in the CEFR as social agents developing 'a communicative competence to which all knowledge and experience of language contributes and in which languages interrelate and interact' (Council of Europe, 2001, p. 4). In this, it is an attractive tool for constructive alignment that embodies learning-oriented principles. It may help to clarify how external tests connect to personal language learning goals. However, as Little and Erickson (2015) acknowledged, to date 'the ELP has failed to gain significant purchase in any of the Council of Europe's member states.' (p. 126). It remains to be seen whether more effective integration with external testing might influence its future endorsement.

6. Acknowledgements

The research reported in this paper was funded with support from Cambridge English Language Assessment.

7. References

- Bachman, L. F. & Palmer, A. (2010). *Language Assessment in Practice: Developing language assessments and justifying their use in the real world*. Oxford: Oxford University Press.
- Bennett, R. (2010). Cognitively Based Assessment of, for, and as Learning (CBAL): A Preliminary Theory of Action for Summative and Formative Assessment. *Measurement*, 8, 70–91.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698-712.

- Boud, D. & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413.
- Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, 22(4), 5-12.
- Brown, J.D. (2008) Testing-Context Analysis: Assessment is Just Another Part of Language Curriculum Development, *Language Assessment Quarterly*, 5:4, 275-312.
- Capel, A. & Sharp, W. (2014). *Objective First. 4th Edition*. Cambridge: Cambridge English.
- Carless, D., Joughin, G., & Liu, N. F. (2006). How assessment supports learning: Learning-oriented assessment in action, Hong Kong: Hong Kong University Press.
- Cheng, L. (2005). *Changing language testing through language teaching*. (Studies in Language Testing Vol. 21). Cambridge: UCLES & Cambridge University Press.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Elder, C., & O'Loughlin, K. (2003). Investigating the relationship between intensive English language study and band score gain on IELTS. In *IELTS Research Reports*, 4, (pp. 207–254). Canberra: IELTS Australia.
- Gan, Z. (2016). Understanding Oral Proficiency Task Difficulty: An Activity Theory Perspective. In *Contemporary Second Language Assessment: Contemporary Applied Linguistics*, 4, (pp.147-164) London: Bloomsbury.
- Green, A.B. (2006). Washback to the learner: Learner and teacher perspectives on IELTS preparation course expectations and outcomes. *Assessing Writing*, 11(2), 113-134.
- Green, A.B. (2007) *IELTS Washback in Context: Preparation for academic writing in higher education*. (Studies in Language Testing Vol. 25). Cambridge: UCLES & Cambridge University Press.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20(5), 2-7).
- Hamp-Lyons, L. & Green, A.B. (2014). Applying a concept model of learning-oriented language assessment to a large-scale speaking test. Unpublished project report to Cambridge English Language Assessment.
- Haertel, E. (2013). How is testing supposed to improve schooling?. *Measurement: interdisciplinary research and perspectives*, 11(1-2), 1-18.
- Hawkey, R. (2006). *Impact Theory and Practice*. (Studies in Language Testing Vol. 24). Cambridge: UCLES & Cambridge University Press.

- Jones, N., & Saville, N. (2016). *Learning Oriented Assessment: A systemic approach* (Studies in Language Testing Vol. 45). Cambridge: UCLES & Cambridge University Press.
- Little, D., & Erickson, G. (2015). Learner Identity, Learner Agency, and the Assessment of Language Proficiency: Some Reflections Prompted by the Common European Framework of Reference for Languages. *Annual Review of Applied Linguistics*, 35, 120-139.
- Liu, O. (2014). Investigating the relationship between test preparation and TOEFL iBT performance. TOEFL Research Report No. RR-14-15. Princeton, NJ: Educational testing Service.
- Mickan, P. & Motteram, J. (2009). The preparation practices of IELTS candidates: Case studies. In IELTS Research Reports, 10. (pp. 223-262). Canberra: IELTS Australia.
- Moss, P.A. (2016) Shifting the focus of validity for test use, *Assessment in Education: Principles, Policy & Practice*, 23(2), 236-251
- Pellegrino, J. W. (2014). Assessment as a positive influence on 21st century teaching and learning: A systems approach to progress. *Psicología Educativa*, 20(2), 65-77.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4-13.
- Saville, N.D. (2009). Developing a Model for Investigating the Impact of Language Assessment within Educational Contexts by a Public Examination Provider. Unpublished PhD Thesis. University of Bedfordshire, UK.
- Saville, N.D. & Hawkey, R.A. (2004). The IELTS Impact Study: Investigating Washback on Teaching Materials. In L. Cheng, Y. Watanabe & A. Curtis (Eds.), *Washback in Language Testing: Research Context and Methods* (pp. 73-96). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Scarino, A. (2013). Language assessment literacy as self-awareness: Understanding the role of interpretation in assessment and in teacher learning. *Language Testing*, 30(3), 309-327.
- Turner, E. (2012). Classroom assessment. In Fulcher, G. & Davidson, F. (eds.), *Handbook of Language Testing*. Abingdon, Oxon.: Routledge, pp. 65-78.
- Turner, C. E. & Purpura, J. E. (2016). Learning-oriented assessment in the classroom. In Tsagari, D., & Banerjee, J. (Eds.). (2016). *Handbook of Second Language Assessment*. Berlin: DeGruyter Mouton.
- UCLES (University of Cambridge Local Examinations Syndicate) (2012). *Cambridge English First: First Certificate in English (FCE): Handbook for teachers*. Cambridge: UCLES. Retrieved December 6, 2012 from www.teachers.cambridgeesol.org/ts/digital-Assets/117578_Cambridge_English_First_FCE_Handbook.pdf.

- Wall, D., & Horak, T. (2011). The impact of changes in the TOEFL exam on teaching in a sample of countries in Europe: Phase 3 and 4. (TOEFL Research Report RR-11-41). Princeton, NJ: Educational Testing Service.
- Xie, Q. (2013). Does test preparation work? Implications for score validity. *Language Assessment Quarterly*, 10(2), 196-218.