Explicating an English Speaking Examination: Challenges and Considerations

Patricia Vella Briffa and Daniel Xerri
patricia.vella-briffa@um.edu.mt
daniel.xerri@um.edu.mt

Abstract: The assessment of speaking has become increasingly important in most language learning programmes. The Maltese post-secondary context saw the revision of the MATSEC Advanced English syllabus and the subsequent introduction of a speaking component as part of this high-stakes examination. The component's outline in the syllabus provided a degree of latitude but also had certain limitations; however, it was immediately clear that it required further refinement to guarantee a more rigorous speaking assessment. This article discusses the initiative undertaken by a group of lecturers of English at the University of Malta Junior College to explicate the speaking examination outlined in the syllabus. It illustrates how these lecturers positioned themselves as test developers in order to elaborate on the 'what' and 'how' of testing second language speaking at post-secondary level. In discussing various challenges and considerations, this article aims to illustrate some means of enhancing the high-stakes testing of speaking proficiency.

Keywords: high-stakes testing, language assessment, Advanced Level English, testing speaking proficiency, validity, reliability

The advent of Communicative Language Teaching in the 1970s greatly increased the significance accorded to the teaching and subsequent assessment of second language speaking skills. Since then, various examination bodies around the world have been engaged in developing effective ways of assessing such skills. In Malta, speaking has been assessed at SEC level for the past 20 years and, since 2011, it has also featured in the national benchmark exercise carried out at the

end of primary education. However, it was only in 2013 that students finishing their post-secondary education and sitting for the Advanced Level examination in English were expected to demonstrate spoken proficiency. This article discusses how a team of lecturers at the University of Malta Junior College collaborated in explicating a speaking component forming part of the Advanced Level English examination.

Speaking is most often acknowledged to be the most difficult of the four language skills, precisely because of its very complex nature. John Field considers it to be 'one of the most complex and demanding of all human mental operations'.[1] Speaking requires that learners not only know how to produce aspects of language, such as grammar, pronunciation and vocabulary, but that they also understand the sociolinguistic and pragmatic conventions associated with speech production. This complexity is further compounded by other dimensions of speaking, including talk as interaction, as transaction, and as performance.[2] All these instances confirm 'the cognitive demands imposed by a speaking task' given that 'most speech takes place under time pressure [and] most forms of speaking are reciprocal'.[3] Such demands contribute to the difficulty of developing spoken proficiency in a second language.

Speaking is considered to be a complex process and this complexity affects the means generally devised to assess this skill. The complex nature of speaking needs to be accounted for when developing valid and reliable means of assessment.

> In the testing of speaking, the interaction between participants (pairs of candidates in some tasks, interlocutor and test-taker in others), is acknowledged to be complex, made more difficult still by the interplay of variables like gender, status, cultural background, peer familiarity, the linguistic proficiency level of the test-taker and of any partner.[4]

1 John Field, 'Cognitive Validity', in *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, ed. by Lynda Taylor (Cambridge, 2011), 70.

2 Jack C. Richards, *Teaching Listening and Speaking: From Theory to Practice* (Cambridge, 2008).

3 Field, 97.

4 Charles Alderson and Jayanti Banerjee, 'Language Testing and Assessment (Part 2)', *Language Teaching*, 35 (2002), 101.

As a consequence, Arthur Hughes maintains that 'The accurate measurement of oral ability is not easy'.[5] In fact, when compared 'with the other skills, speaking is the most difficult to assess'.[6] Magdalena Aleksandrzak claims that 'Testing the oral proficiency of foreign language students is a complex task which may cause considerable problems at any stage of the process.'[7] Difficulties arise from the very early stages of determining what is to be assessed, to the ensuing stages of task selection and preparation, to the later stages of deciding on the form of assessment, and to the final stages of administering the selected test. It is precisely a discussion of the challenges and considerations underlying a locally administered high-stakes speaking test that concern the present paper.

Assessing speaking is challenging, primarily because of the nature of what should qualify as fluency in this skill. Field claims that 'Fluency is a notoriously slippery concept, and attempts to define it have caused much controversy over the years'.[8] Sari Luoma argues that 'The narrowest definitions only include a few features ... whereas the broadest uses are virtually synonymous with "speaking proficiency".'[9] In her opinion, speaking assessment needs to be seen as a series of activities that 'involve speakers in using language for the purpose of achieving a particular goal or objective in a particular speaking situation. The emphasis in this definition is on goal-oriented language use.'[10] Such a definition implies that careful attention needs to be paid to what candidates are asked to do in a speaking test. Moreover, the design of such a test needs to be driven by a consideration of the goals that candidates are expected to achieve in the course of the test.

Validity and reliability

Any form of language testing, understood as 'instruments used to measure language ability or aptitude', has to primarily address the challenge of determining validity and reliability, two different yet complementary

5 Arthur Hughes, *Testing for Language Teachers* (Cambridge, 2003), 134.
6 Suwandi dan Taufiqulloh, 'Designing Speaking Test', *Eksplanasi*, 4 (2009), 183.
7 Magdalena Aleksandrzak, 'Problems and Challenges in Teaching and Learning Speaking at Advanced Level', *Glottodidactica*, 37 (2011), 46.
8 Field, 82.
9 Sari Luoma, *Assessing Speaking* (Cambridge, 2004), 88.
10 Ibid., 31.

aspects of a test.[11] Ensuring a test's reliability is in itself a measure of validity.

Test developers need to confirm that a test possesses various kinds of validity. In the context of speaking assessment, a thorough knowledge of what is meant to be tested is necessary, if construct validity is to be ensured. According to Luoma, test developers need to be able to 'define the kind of speaking they want to test in a particular context ... develop tasks and rating criteria that test this ... and make sure that the testing and rating processes actually follow the stated plans'.[12] As Cyril J. Weir points out, this knowledge of the test construct in the early stages of test development is essential given that

> The more fully we are able to describe the construct we are attempting to measure at the *a priori* stage the more meaningful might be the statistical procedures contributing to construct validation that can subsequently be applied to the results of the test.[13]

More specifically is the consideration given to content validity, which ensures 'the basis for making a principled selection of elements for inclusion in the test'.[14] An informed selection of content will render a more accurate measurement of the language skill and will guarantee a more positive backwash effect. Another consideration is directed towards whether the test items selected are in effect a fair representation of what is meant to have been acquired in respect of the language skill being tested. Field argues for cognitive validity since this ensures that 'the task, test content and prevailing conditions' are, in fact, the same as those required outside the testing context.[15] Another aspect of validity concerns the candidates' 'familiarity with the criteria that will be used when assessing performance'.[16] This information is usually provided in a test manual.

11 Wei Li, 'Validity Considerations in Designing and Oral Test', *Journal of Language Teaching and Research*, 2 (2011), 267.

12 Luoma, 28.

13 Cyril J. Weir, *Language Testing and Validation: An Evidence-based Approach* (Basingstoke, 2005), 18.

14 Hughes, 27.

15 Field, 66.

16 Evelina Galaczi and Angela ffrench, 'Context Validity', in *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, ed. by Lynda Taylor (Cambridge, 2011), 131.

Equally important is test developers' consideration of reliability, which Wei Li defines as 'consistency in scores regardless of when and how many times a particular test is taken'.[17] Reliability can be enhanced by providing candidates with more than one task during the test, devising clear and unequivocal rubrics, and ensuring that the candidates and examiners are familiar with the test tasks and procedures. To this effect Field advocates 'feedback from piloting and past administrations'.[18] A further aspect of test reliability pertains to scoring consistency, which can be secured through examiner training and the use of a calibrated rating scale.

Practical considerations

Further to issues of validity and reliability, this paper also explores practical considerations of test conduct in order to address the full spectrum of concerns involved in the testing of speaking. Byron Gong asserts that 'a reliable and valid oral English test is also connected with how the speaking test is conducted'.[19] In fact, 'Besides validity and reliability, practicality is often quoted as the third consideration in test design.'[20] Therefore, aspects pertaining to test content, test organization, and test delivery are discussed hereunder.

The time allotted to the tasks forming part of a test is of primary concern, especially since 'The time variable is critical in information processing theories of speech production.'[21] Several benefits may result if planning time is given its due consideration in a speaking test. According to Field, 'Pre-planning time clearly assists *conceptualisation* ... increasing the likelihood of utterances that are carefully formed syntactically and of

17 Li, 268.

18 Field, 67.

19 Byron Gong, 'Considerations of Conducting Spoken English Tests for Advanced College Students', paper presented at the 36th International Association for Educational Assessment Conference, Bangkok, (22–27 August 2010) <http://selectscore.com/fullpaper/119.pdf> [accessed 20 October 2013], 7.

20 Li, 268.

21 Catherine Elder and Gillian Wigglesworth, 'An Investigation of the Effectiveness and Validity of Planning Time in Part 2 of the IELTS Speaking Test', in *IELTS Research Reports: Volume 6*, (IELTS Australia and British Council, 2006) <http://www.ielts.org/pdf/Vol6_Report1.pdf> [accessed 20 October 2013], 4.

precision in the choice of lexis.'[22] Such planning time allows for ideas to be generated, to be better expressed, and to be logically set, ultimately leading to a more accurate and fluent output.

Secondly, topic choice in a speaking test is of prime importance as knowledge of content will affect the output of each task. The topics selected for every task need to be appropriate to the candidates' level, within their age and experience, free of cultural bias, not distressing or offensive, of an appropriate cognitive level, familiar yet not too familiar, and not specialized or technical.[23] Ensuring that the chosen topics meet as many of these criteria as possible will allow candidates to achieve better results in their spoken production.

A third consideration is task purpose, which informs both the candidates' and the examiners' behaviour. Task purpose 'will facilitate goal setting and monitoring [and] has implications for the scoring validity of a task/test'.[24] Task purpose, conveyed through clearly stated instructions in the rubrics, serves to 'direct the examinees to the specifics of a task'.[25] According to Kenneth Wolf and Ellen Stevens it also helps examiners 'to anchor judgements'.[26] It is complemented by the use of an interlocutor frame, which helps 'to ensure standardisation across speaking tests and to guide and constrain examiners so that the candidates' experiences are fair and equal and the examiners' contributions are controlled'.[27] Consistency is much more likely to be achieved if the task purpose is embedded in the test's scripted rubrics and instructions.

Advanced English Speaking Examination

In 2010, the syllabus panel of the Advanced Level English Examination launched a new syllabus that was meant to implement a number of changes in the content and structure of the exam. One of the most innovative aspects of this syllabus consisted of the introduction of a speaking examination that aimed to address the long-existing lacuna regarding the assessment of

22 Field, 102.
23 Galaczi and ffrench, 150.
24 Ibid., 124.
25 Luoma, 53.
26 Kenneth Wolf and Ellen Stevens, 'The Role of Rubrics in Advancing and Assessing Student Learning', *The Journal of Effective Teaching*, 7 (2007), 12.
27 Galaczi and ffrench, 169.

speaking proficiency at post-secondary level. Its first sitting was set for May 2013. The 15-minute exam was to carry 6 percent of the global mark and this attested to the fact that lecturers and students were now expected to give speaking an adequate amount of importance as part of classroom practice. The exam adopted the format of a one-to-one setting involving the examiner and the candidate and it was structured in three parts:

1. An informal interview intended as a conversation starter, where the examiner will ask basic questions about topics such as work, study, leisure, and career plans;
2. A conversation initiated by the interlocutor, based on a prompt such as a photograph or other image that is presented to the candidate at this point in the interview;
3. A presentation expressed as a long turn by the candidate based on a question selected by the candidate from a list of five presented to her/him some minutes before entering the interview room. The set of five questions will reflect five of the following topics: lifestyle, music, sport, religion, relationships, international news, environment, war, education and entertainment.[28]

The other details specified by the syllabus concerned the marks and approximate time of each part:

- Part 1 is a guided examiner-to-candidate conversation (about 3 minutes – 4 marks);
- Part 2 is a guided examiner-to-candidate conversation (about 4 minutes – 6 marks);
- Part 3 is a candidate-to-examiner monologue (about 3 minutes – 8 marks).[29]

The syllabus expected candidates to demonstrate 'competence in handling the English language accurately, with minimal errors in grammar and spelling' and to be able 'to communicate ideas effectively and through coherent logic and structure in expression and argumentation'.[30] It considered it 'desirable' that candidates should 'demonstrate an evolved proficiency in speaking and listening skills' and for this reason the exam 'serves as a measure of the candidates' ability to speak and converse in English'.[31]

28 MATSEC, *AM Syllabus (2013): English* (Malta: MATSEC, 2010), 6.
29 Ibid., 9–10.
30 Ibid., 2.
31 Ibid., 6.

The kind of proficiency desired by the syllabus is in line with the Common European Framework's description of a C2 user of language. At C2 level, a speaker:

- shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning and also has a good command of idiomatic expressions and colloquialisms;
- maintains consistent grammatical control of complex language;
- can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty;
- can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly; can interweave his/her contribution into the joint discourse with fully natural turn-taking, referencing, allusion-making, etc.;
- can create coherent and cohesive discourse, making full and appropriate use of a variety of organizational patterns and a wide range of connectors and other cohesive devices.[32]

Such a high level of spoken proficiency would enable students to engage in a variety of exchanges, ranging from the very informal, such as being able to discuss issues of general interest, to the more formal, such as seminar and tutorial discussions, and giving presentations and answering subject-related questions in depth. Moreover, at this level of language learning, speakers would be familiar with the vocabulary relevant to general interactional and social language, and the lexis of comparing, describing, expressing opinions, agreeing, disagreeing, suggesting, speculating, evaluating, etc. As C2 users of English, they would be able to use a range of appropriate vocabulary with flexibility when giving and exchanging views on familiar, unfamiliar, and abstract topics. Most of the above resonates with what the Advanced English Speaking Examination requires candidates to be able to do. Even prior to the introduction of this speaking examination, students studying English at Advanced Level were still indirectly expected to show mastery of those features typical of proficient spoken language use. It can therefore be argued that the exam's introduction facilitated the process of concretizing what students were reasonably expected to achieve at the end of a two-year course at a post-secondary institution like Junior College.

32 Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (Strasbourg, 2001) <http://www.coe.int/t/dg4/linguistic/Source/Framework_en.pdf> [accessed 20 October 2013], 28.

Explicating the Speaking Examination

The introduction of a speaking component in a high-stakes national examination at such an advanced level of language learning was well received. This innovation was seen to serve the purpose of gauging the spoken proficiency of second language learners before progressing to higher education or employment on the basis of a qualification in English that is intrinsically meant to attest to a high level of language proficiency.

Once the syllabus was published, it was immediately clear that the newly introduced speaking examination required a further effort before it could become a valid and reliable form of assessment. The syllabus description consisted of an outline that lacked any 'connections between planning, instruction, and assessment'.[33] Consequently, this gap generated a degree of uncertainty amongst lecturers, mainly in terms of how to approach task-planning and candidate preparation for the new speaking component.

A group of lecturers within the Department of English at Junior College therefore took the initiative to address this limitation by collaborating on the task of writing detailed test specifications and sample test materials for the new speaking exam. By positioning themselves as test developers they undertook the challenging task of further explicating the speaking exam description found in the syllabus while working within the parameters set by MATSEC with regards to number of tasks set, time allotted, and weighting of marks. Only in this manner would this high-stakes speaking examination constitute a valid and reliable form of assessment of the speaking skills developed at this level of language learning.

The writing of detailed test specifications involved elaborating on the outline of the speaking component provided in the syllabus. The writing of these specifications necessitated thorough knowledge and experience of the student cohort, the educational context, and the level of speaking proficiency to be tested. This led to considerations with regards to aspects of test content and procedures, as well as the development of a set of criteria for assessment that had to be tried and tested several times. Decisions undertaken during this rigorous exercise were compiled in a manual that was meant to accompany the new speaking component.

33 Helenrose Fives and Nicole Di-Donato-Barnes, 'Classroom Test Construction: The Power of a Table of Specifications', *Practical Assessment, Research and Evaluation*, 18 (2013), 6.

The manual would not only provide lecturers and students with detailed information about the component but it would also serve as a tool that lecturers in similar educational contexts could use to enhance their assessment literacy with regards to the testing of speaking.

The specifications in the manual elaborated more fully on the three tasks outlined in the syllabus. Three tasks, progressing from a conversational format to picture interpretation and a brief discussion onto a final task of presenting one's ideas on a topic, were considered suitable to elicit as full a range as possible of speaking skills. The tasks would engage the candidate long enough for the time allotted and would not be unduly taxing. This progression of task engagement is in keeping with research findings, indicating that ideally the 'ordering of tasks ... follows a logical order from relatively structured and supported interaction under the direct control of the examiner involving topics of immediate personal relevance to more open-ended discussion with less examiner control involving more general topics'.[34] In this manner, context validity is ensured as 'At the lower levels, the discourse mode is primarily description and exposition, whereas at the higher levels argument and persuasion are also elicited.'[35] Hence the tasks test the speaking skills that candidates would typically employ in the context of an informal conversation, participation in a subject seminar, and the delivery of a presentation on a topic of specific interest. Such speaking skills are best assessed by means of interviews, live monologues, and discussions: these being some of the most valid and reliable means of spoken test types.[36] Context validity was further ensured by implementing both 'planning and non-planning conditions' in the selected tasks, as reflected in real-life speaking contexts ranging from spontaneous conversation to more formal monologues.[37]

The logistics of the Advanced English Speaking Examination demanded that a direct face-to-face format was to be adopted throughout the entire speaking exam. Furthermore, it was to be an examiner-candidate exchange, which raised the issue of 'the degree of interlocutor support'.[38] In view of the fact that the interlocutor would also assume the role of an examiner, it was considered necessary to script the interlocutor's turns

34 Field, 133.
35 Ibid., 144.
36 Scott Thornbury, *How to Teach Speaking* (Harlow, 2007), 125–6.
37 Galaczi and ffrench, 135.
38 Glenn Fulcher, *Testing Second Language Speaking* (Harlow, 2003), 19.

'through the use of an "interlocutor frame"'.[39] This would 'ensure that all test events conform to the original test design so that all test-takers participate in essentially the same event'.[40] The use of an interlocutor frame is advisable given that it sees examiners being 'trained to conduct speaking tests according to a standardized prescribed role' in order to further ensure scoring validity.[41] Furthermore, the length of each interlocutor turn was also stipulated. These measures were introduced to standardize the examiners' practices and enhance test reliability.

As already shown above, the three parts outlined in the syllabus were revisited and further explicated by a group of Junior College lecturers, who were ultimately responsible for compiling the manual for the Advanced English Speaking Examination. The next three sections demonstrate how the syllabus description of each part was explicated by means of more detailed specifications and sample test materials. These sections include a discussion of the considerations involved at every stage.

Part 1

An informal interview intended as a conversation starter, where the examiner will ask basic questions about topics such as work, study, leisure, and career plans.[42]

The syllabus intended this first task to be a means of easing the candidate into an exchange with the examiner. In the explication this was further defined as a conversational exchange to be always initiated by the examiner and to consist of a prescribed number of questions and prompts meant as a springboard for the candidate to demonstrate their conversational ability. The 'basic questions' were further specified to mean a set of six 'wh'-questions.[43] According to Barbara A. Fox and Sandra A. Thompson 'wh'-questions are especially suited to serve both as '"Specifying questions" [that] seek particular pieces of information' and as '"Telling

39 Galaczi and ffrench, 170.

40 Barry O'Sullivan and Yang Lu, 'The Impact on Candidate Language of Examiner Deviation from a Set Interlocutor Frame in the IELTS Speaking Test', in *IELTS Research Reports: Volume 6* (IELTS Australia and British Council, 2006) <http://www.ielts.org/PDF/Vol6_Report4.pdf> [accessed 20 October 2013], 4.

41 Galaczi and ffrench,167.

42 MATSEC, 6.

43 Ibid.

questions" [that] seek extended responses' aimed at generating a thread of exchanges with the examiner.[44] The questions were to be primarily about the candidates and areas of personal interest and experience that they would feel confident enough to answer. In fact, Helen Gaylard and Allan Ramsay maintain that when responding to a 'wh'-question, the participant in the conversation, in this case the candidate, always has some knowledge of the description of the entity embodied in the question word itself.[45]

Part 2

A conversation initiated by the interlocutor, based on a prompt such as a photograph or other image that is presented to the candidate at this point in the interview.[46]

In the explication the second task was further defined as an exchange, this time purposely based on a visual, which could be a picture or a series of pictures depicting aspects of a topic. It was deemed important that careful consideration should be paid to the selection of visuals. This would ensure that the items, people, or events depicted were clear and familiar, that the visual was not too detailed but busy enough to generate an exchange, and that the lexical items, grammatical structures, and language functions were within the cohort's range, especially since such visual prompts 'play a large role in the development of the student's mental model'.[47] Indeed, such use of visual prompts helps to stimulate the candidate's vocabulary bank.

A further specification was introduced that would permit the candidate to examine the visual prompt carefully before engaging in its description. Trialling determined that each candidate should be given approximately 30

44 Barbara A. Fox and Sandra A. Thompson, 'Responses to Wh-Questions in English Conversation', *Research on Language and Social Interaction*, 43 (2010), 135–6.

45 Helen Gaylard and Allan Ramsay, 'Relevant Answers to Wh-Questions', *Journal of Logic, Language and Information*, 13 (2004), 173–86.

46 MATSEC, 6.

47 Victoria Crisp and Ezekiel Sweiry, 'Can a Picture Ruin a Thousand Words? Physical Aspects of the Way Exam Questions are Laid Out and the Impact of Changing Them', paper presented at the British Educational Research Association Conference, Edinburgh (10–13 September 2003) <http://www.cambridgeassessment.org.uk/Images/109709-can-a-picture-ruin-a-thousand-words-physical-aspects-of-the-way-exam-questions-are-laid-out-and-the-impact-of-changing-them..pdf> [accessed 20 October 2013], 3.

seconds to examine the visual before being asked to embark on a description of what they could see. This planning time was considered sufficient for the candidate to be able to retrieve and generate the lexicon necessary for the ensuing brief description of the visual. The candidate, again after the trialling of sample test materials, was expected to describe the visual for approximately one minute. This would give the candidate sufficient time to describe the picture without being too taxing a task. Yet another specification introduced in the explication of Part Two was the inclusion of a set of four questions in order to elicit the candidate's opinions about a number of things related to the topic at hand. These questions were not meant to have one correct answer but required the candidate to express more in-depth views and opinions about the topic. Such open-ended questions would assist the candidate to explain thoughts and ideas as well as to connect and organize them in a logical and coherent manner.

Part 3

A presentation expressed as a long turn by the candidate based on a question selected by the candidate from a list of five presented to her/him some minutes before entering the interview room.[48]

The explication of this final task further specified that it should consist of a brief presentation or an 'extended individual performance' by the candidate on a topic or a statement.[49] Since the candidate would be selecting the topic of the presentation, a list of statements would be formulated for the candidate's consideration. A set of five statements was considered sufficient to provide the candidate with a varied selection that at the same time would not cause anxiety by being too broad. Moreover, providing a choice for the brief presentation would help boost the candidates' confidence in relation to delivery. The presentation was reserved for the end of the speaking examination as it is much more demanding for candidates, in that it requires them to generate a number of ideas and expand on these coherently and sequentially.

Consequently, a further specification was added stipulating that candidates should be allowed a brief period of preparation time in which

48 MATSEC, 6.
49 Field, 106.

to engage in macro-planning. Evelina Galaczi and Angela ffrench maintain that 'The inclusion of planning time ... has to be considered within the context of practicality.'[50] Given that the syllabus provided some leeway by using the term 'about' to specify the amount of time that each part should take, it was decided to incorporate planning time into the exam.[51] It was also specified that a paper and pencil were to be provided to candidates for them to be able to formulate thoughts and ideas as well as to use these notes as a form of guidance in the delivery of their presentation. Planning time was deemed to be particularly necessary for candidates in the local context, especially since as learners 'acquire greater knowledge of L2 and (above all) greater automaticity of lexical retrieval and speech assembly, they are able to make increasingly effective use of any time allowed'.[52] The planning time was set at two minutes, which was considered sufficient for candidates to recollect their thoughts before delivering a three-minute presentation. Field argues that 'There must be a cut-off point ... after which additional time is unhelpful and may even lead to second thoughts and a blurring of conceptual and linguistic targets.'[53] Once again, the time limit was only decided upon after the sample test materials had been adequately trialled.

Rating scale

The group of lecturers responsible for explicating the speaking component also engaged in the challenging task of developing a rating scale for the three different parts comprising the speaking examination. The rating scale would allow examiners to standardize their assessment of candidates' speaking performance in terms of both fluency and accuracy. This entailed the application of 'a multi-faceted scale, each component of which adds to an overall score'.[54]

A set of four criteria constituted the basis of this rating scale: fluency and coherence, pronunciation, vocabulary, and grammar. The rating scale consisted of 36 descriptors that were meant to comprehensively describe

50 Galaczi and ffrench, 135.
51 MATSEC, 6–7.
52 Field,103.
53 Ibid.
54 Barry O'Sullivan, 'Notes on Assessing Speaking' (Cornell University, 2008) <http://www.lrc.cornell.edu/events/past/2008–2009/papers08/osull1.pdf> [accessed 20 October 2013], 22.

a range of levels of spoken proficiency. The four criteria were given equal weighting as they were considered equally important aspects of speaking proficiency. Though admittedly much more time-consuming than a holistic rating scale, this analytic scale was considered the best way of gauging the spoken proficiency of candidates sitting for such a high-stakes examination. The merits of an analytic rating scale are attested to by the literature on second-language testing, which highlights the better guidance provided to interlocutor-examiners in terms of assessing the strengths and weaknesses of candidates' performance, hence ensuring a stronger measure of scoring validity in the long run.[55]

Trialling

The entire task of explicating the new Advanced English Speaking Examination was carried out over a three-year period. The tasks, rubrics, interlocutor frame, timing, and other elements developed as part of this explication were subjected to adequate trialling in order to ensure validity and reliability as well as to factor in each of the variables discussed in the previous sections. The information gathered as part of this trialling process facilitated the compilation of an authoritative test manual for use by all stakeholders.

Trialling served the purpose of creating examination tasks that would be 'accessible to the target candidature and which meet the requirements of the purpose stated in the exam specifications'.[56] Moreover, trialling enabled the group of lecturers acting as test developers to verify whether

1. the task has provided sufficient stimulus … to allow the candidates to fully engage with it and display their language ability;
2. there has been any misunderstanding or misrepresentation of the task;

55 The merits of analytic rating scales for assessing speaking are discussed in the following: Yasuyo Sawaki, 'Construct Validation of Analytic Rating Scales in a Speaking Assessment: Reporting a Score Profile and a Composite', *Language Testing*, 24 (2007), 355–90; Xiaoming Xi, 'Evaluating Analytic Scoring for the TOEFL Academic Speaking Test (TAST) for Operational Use', *Language Testing*, 24 (2007), 251–86; Lynda Taylor and Evelina Galaczi, 'Scoring Validity', in *Examining Speaking: Research and Practice in Assessing Second Language Speaking*, ed. by Lynda Taylor (Cambridge, 2011), 171–233; Luu Trong Tuan, 'Teaching and Assessing Speaking Performance through Analytic Scoring Approach', *Theory and Practice in Language Studies*, 2 (2012), 673–9.

56 Galaczi and ffrench, 127.

3. there is any ambiguity or lack of clarity in the wording of the rubric or in the visual.[57]

Trialling helped indicate that the task rubrics required the most modification, especially when considering that the operative language was the candidates' second language throughout. It was imperative that the rubrics were as clear and unambiguous as possible.

Conclusion

The first sitting of the Advanced English Speaking Examination took place in May 2013 and assessed the spoken proficiency of 508 candidates. The initiative to explicate this new speaking examination was particularly crucial given the lack of detail in the original syllabus outline. In the Maltese post-secondary context, the benefits were manifold and far-reaching. Explicating the speaking examination provided a 'match' between the teaching and testing of second-language speaking.[58] It also ensured the three fundamental criteria of validity, reliability and practicality. Only a rigorous consideration of all aspects of 'what' to test and 'how' to test can truly enhance high-stakes testing and provide candidates with a fair assessment of their speaking proficiency. By means of their work, the lecturers responsible for explicating the Advanced English Speaking Examination helped to achieve this aim.

Acknowledgements

The explication of the Advanced English Speaking Examination was conducted together with our colleagues Clyde Borg, Andrew Farrugia, Joseph Gerardi, and Odette Vassallo.

57 Ibid.

58 Michael Bay-Borelli, Christine Rozunick, Walter D. Way, and Eric Weisman, 'Considerations for Developing Test Specifications for Common Core Assessments' (Pearson, 2012) <http://images.pearsonassessments.com/images/tmrs/tmrs_rg/Common_Core_Test_Specifications12-01-10.pdf> [accessed 20 October 2013],12.