BRITISH COUNCIL

ENGLISH LANGUAGE
ASSESSMENT RESEARCH GROUP

EXAMINING TEST-TAKER PROCESSES AND STRATEGIES AND STAKEHOLDER PERCEPTIONS OF RELEVANCE OF THE APTIS FOR TEACHERS SPEAKING TEST IN THE AUSTRALIAN CONTEXT

AR-G/2020/1

Sally O'Hagan and Kellie Frost
Language Testing Research Centre, The University of Melbourne

ARAGs RESEARCH REPORTS ONLINE
SERIES EDITORS: RICHARD SPIBY AND CAROLYN WESTBROOK
ISSN 2057-5203
© BRITISH COUNCIL 2020

ABSTRACT

This study investigated the relevance of the Aptis for Teachers Speaking test to the assessment of teachers' English speaking skills in the Australian context. It examined the speaking processes and strategies elicited by the Aptis for Teachers Speaking test tasks and compared these with stakeholder perceptions of classroom oral communication demands.

Strong English speaking skills are considered central to effective teaching in Australian classrooms, and the importance of oral communication and interaction is emphasised in professional standards for teachers. This study investigated the potential of Aptis for Teachers (Speaking) to provide a more authentic representation of the construct of speaking ability in Australian classrooms compared with general proficiency tests, which are commonly used to assess teachers' English speaking ability. The aims of the study were to examine the speaking processes and strategies engaged by the Aptis for Teachers Speaking tasks, and to explore test-taker and teacher educator perceptions of the appropriateness of the tasks as measures of classroom-ready speaking skills in the Australian context. The project provides important empirical evidence relevant to validity claims in relation to the speaking component of Aptis for Teachers, as well as insights into key stakeholder perceptions of the test.

Authors

Sally O'Hagan is a Research Fellow in the Language Testing Research Centre at the University of Melbourne where she works on test development, program evaluation, test validation, and standard-setting research. Her work spans a range of contexts including the K–12 sector, pre-university language preparation, post-entry assessment in higher education, and language in the professions, including teaching and healthcare. Sally also teaches into the Applied Linguistics program at the University of Melbourne, and formerly taught general and academic English to adults and younger learners. Her doctoral research examined assessor behaviour and judgement in the context of discipline-based academic writing tasks. Sally's work has appeared in *Language Testing*, *Studies in Higher Education*, *Journal of Advanced Nursing* and in books published by Peter Lang and Springer.

Kellie Frost is a lecturer in English for Academic Purposes (EAP) and Applied Linguistics at the University of Melbourne. Her research interests are in the areas of EAP, English in the workplace, and language testing and immigration policy. She previously worked in the Language Testing Research Centre at the University of Melbourne, where she led a range of language test development, consulting and research projects. Her PhD, completed in 2018, examined the impact of the use of English test scores in selection processes for skilled migration in Australia. Kellie's has published articles in *Language Testing* and chapters in a number of research volumes, including the *Oxford Handbook of Language Policy and Planning*, winner of the 2019 British Association of Applied Linguistics book prize. She has held research grants from the British Council, the Educational Testing Service, and Pearson Education.

ACKNOWLEDGEMENTS

The authors acknowledge the generous support of the British Council for funding this study and providing guidance throughout the project. We are also grateful to Cathie Elder for her input into the development of the aims and scope of the project; to Josh Clothier for his assistance with protocol development, data collection, transcription and data preparation; and to Kathryn Weller for her assistance with coding.

CONTENTS

1. BACKGROUND **5**

1.1 Rationale for the study 5
1.2 Teacher language proficiency 6
1.3 Test-taker processes and strategies 6

2. THE PRESENT STUDY **7**

2.1 Participants 7
2.2 Instruments 9
2.2.1 Aptis for Teachers Speaking test 9
2.2.2 Background questionnaires 11
2.2.3 Verbal report protocol 11
2.2.4 Interview protocol 11
2.3 Data collection procedures 12
2.4 Data analysis 13
2.5 Verbal report coding scheme 13

3. RESULTS **14**

3.1 Test-taker processes and strategies engaged by Aptis for Teachers Speaking tasks 14
3.1.1 Orientation to task 16
3.1.1.1 Generating ideas 16
3.1.1.2 Selecting ideas 16
3.1.1.3 Reflecting on the question 17
3.1.1.4 Linking questions and image 17
3.1.2 Communication 18
3.1.2.1 Avoiding 18
3.1.2.2 Approximating, Paraphrasing to avoid repetition, Word choice 18
3.1.2.3 Speeding delivery, Expanding the message to fill time 19
3.1.2.4 Expanding the message to improve response 19
3.1.2.5 Organising and structuring response 19
3.1.2.6 Slowing delivery, Relating task to experience, Referring to notes 20
3.1.3 Cognition 21
3.1.3.1 Engaging with stimulus images 21
3.1.3.2 Using imagery 22
3.1.3.3 Notetaking (Part 4) 22
3.1.3.4 Using L1 23
3.1.4 Metacognition 23
3.1.4.1 Self-correcting linguistic error, Evaluating performance, Identifying communication issues 23
3.1.4.2 Reflecting on orientation, Planning, Monitoring time 24
3.1.4.3 Reflecting on picture stimulus, Reacting to test format 25
3.1.5 Affect 26
3.2 Stakeholder perceptions of Aptis for Teachers Speaking tasks 26
3.2.1 Oral communication demands of the classroom 27
3.2.1.1 Explanations and instructions 27
3.2.1.2 Classroom management 27
3.2.1.3 Dialogic teaching 28
3.2.1.4 Giving feedback 29
3.2.1.5 Modelling language 29
3.2.1.6 Differentiation 30
3.2.1.7 Vocabulary 30
3.2.2 Relevance of Aptis for Teachers Speaking tasks to classroom communication 31
3.2.3 Aptis for Teachers Speaking test compared with current approved tests 33

4. SUMMARY AND DISCUSSION OF FINDINGS **37**

5. CONCLUSION **40**

5.1 Limitations 40

5.2 Future research and recommendations 41

REFERENCES **42**

APPENDIX A: Sample speaking questions (screenshots) **44**

APPENDIX B: Background questionnaire (educators) **46**

APPENDIX C: Background questionnaire (trainees) **47**

APPENDIX D: Verbal report protocol **48**

APPENDIX E: Interview protocol (educators) **50**

APPENDIX F: Interview protocol (trainees) **51**

APPENDIX G: Verbal report coding categories with examples **52**

List of tables

Table 1: Test-taker (trainee teacher) participants 8

Table 2: Teacher educator participants 9

Table 3: Overview of Aptis for Teachers Speaking test: parts 2, 3 and 4 10

Table 4: Verbal report coding scheme 13

Table 5: Frequencies of reported test-taker behaviours 15

List of figures

Figure 1: Overview of research design 12

1. BACKGROUND

1.1 Rationale for the study

Recent research into employment outcomes for graduating international students in Australia suggests a misalignment between the communication skills they possess and the expectations of employers across various professions, including teaching (for example, Arkoudis et al., 2009). This has led to increased scrutiny of universities' efforts to develop the English language skills of international students over the duration of their courses (IEAC, 2013), as well as a progressive tightening of the English language requirements for teacher registration (VIT, 2015).

Increasingly restrictive English language criteria has, in turn, led to questions concerning the adequacy and appropriateness of existing language tests as gatekeeping devices in this domain (Murray, Cross & Cruickshank, 2014). Some have criticised the use of general proficiency or academic English tests, such as IELTS, for professional registration as inappropriate, given that this purpose is beyond that for which such tests were designed (e.g. Hall, 2010). Meanwhile, the use of specific purpose tests has also been controversial: specific purpose test constructs extend beyond language abilities, yet test requirements only apply to speakers of English as an additional language (EAL), thus raising issues of fairness (Murray, Riazi & Cross, 2012).

Aptis for Teachers is not currently one of the English language tests accepted for the purpose of teacher registration in Australia. The test, however, has the potential to provide a more authentic representation of the construct of speaking ability in Australian classroom contexts than a general proficiency test, since it is designed for use in assessing the English proficiency of teachers. According to the test developer, the content of Aptis for Teachers "relates specifically to teachers and questions tap into themes and scenarios that teachers come across every day" (British Council, n.d.). At the same time as the Aptis test content is teaching focused, the assessment criteria are language proficiency-based (www.britishcouncil.org/aptis), which means it is not a specific purpose test, in the strictest sense. In this way, the fairness issues raised by specific purpose test constructs, which elicit communication skills that English as a first language speakers may not possess, are potentially mitigated. Aptis for Teachers may also represent a useful tool for Australian universities offering pre-service teacher training. Its content relevance may help promote positive washback, thereby promoting the language-readiness of graduating teachers for professional registration and classroom practice.

Teachers' oral communication skills, as discussed further below, are considered central to effective teaching in Australian classrooms, and the importance of spoken interaction is emphasised in professional standards for teachers and the associated registration requirements. The aim of the current study is to examine the speaking processes and strategies elicited by the Aptis for Teachers Speaking tasks, and to explore test-taker and teacher educator perceptions of the appropriateness of the tasks as measures of classroom-ready speaking skills in the Australian context. By investigating test-taker behaviour and different stakeholder perceptions, the study will contribute empirical evidence relevant to validity claims for the appropriateness of the tasks and underlying test construct of the Aptis for Teachers Speaking test. The research will also contribute to a better understanding of the oral communication demands faced by EAL trainee teachers.

1.2 Teacher language proficiency

Since 2011, professional standards for teachers in Australia have been set out in a national framework known as the Australian Professional Standards for Teachers (APST), which specifies the requirements for registration and permission to practice in Australia, including English language proficiency requirements. The APST emphasises the importance of talk- and interaction-based teaching and learning in classrooms, with oral skills highly valued and considered central to effective teaching (VIT, 2009).

General proficiency language tests, such as IELTS, are increasingly used as gatekeeping tools for professional registration and practice in Australia. The appropriateness of using general proficiency English language tests in these contexts has been questioned by many (see for example, Hall, 2010), as mentioned above. Existing theoretical understandings of speaking processes and strategies highlight the importance of context and communication purpose in influencing oral production (Bachman & Palmer, 1996; Bygate, 1987; Fulcher, 2003; Luoma, 2004). Further, there is a body of research that supports the specificity of the structure of classroom discourses compared to other language use domains (for example, Sinclair & Coulthard, 1992).

The question of how the construct of teacher language proficiency should be defined, however, remains controversial. As Elder and Kim (2013) suggest, there is compelling evidence to support the notion that the language proficiency required by teachers extends beyond general or academic language proficiency. Equally, however, the roles of teachers vary widely according to subject area, classroom context, and school environment, among other things, and as such, the communication demands of the profession are also likely to be highly diversified.

Aptis for Teachers is a variant of a general proficiency test, with test content tailored to the teaching domain (www.britishcouncil.org/aptis). Thus, in light of the complex issues surrounding conceptualising teacher language proficiency for testing purposes, the extent to which Aptis for Teachers can provide a valid measure of the language proficiency demands relevant to EAL teachers in the Australian domain is worthy of investigation. Furthermore, stakeholder perceptions of the test and its suitability for the Australian context, in particular the views of test-takers and educators, provide an additional and important source of evidence in the validity argument for Aptis.

1.3 Test-taker processes and strategies

Investigation of test-taker cognition in language testing research has been used to establish the extent of correspondence between the intended test construct and the processes and strategies engaged by test-takers during task performance, with sufficient congruence suggesting that test tasks are relevant to real-life language demands of the target domain (Purpura, 2014). Verbal protocol methodologies are typically employed in such studies (for example, Barkaoui, 2015, 2016; Plakans, 2008). In the case of second language speaking tests, because verbal reporting techniques would necessarily intrude into oral production for the task, to avoid disrupting performance, retrospective, rather than current, reporting methods are used in studies of test-taker cognition (Cohen, 2012; Cohen & Olshtain, 1993). Specifically, stimulated recalls are an established method for investigating strategic behaviours on speaking tasks (for example, Swain et al., 2009). Previous studies have defined test-taker processes and strategies in various ways, reflecting a lack of consensus over definitions and operationalisation. The present study draws upon the definition of strategic behaviours in Swain et al. (2009) which conceptualises strategies as "the deliberate thoughts and behaviors used to manage or carry out cognitive processes with the goal of successful test performance" (p.2). This definition spotlights the relationship between strategies and cognitive processes (strategies being used to regulate and/or reflect upon cognitive processes), as well as the goal-directed nature of strategy use. It also conceives of strategies as behaviours (thoughts, actions) that are consciously engaged in by test-takers and, hence, accessible to self-reporting methodologies, including retrospective recall.

Previous studies of test-taker processes and strategies in language testing have sought to identify the types of processes and strategies engaged in, and to explore the relationships amongst strategy use and such factors as proficiency level, task type, delivery mode, and test performance (e.g. Barkaoui, 2016; Phakiti, 2003; Plakans, 2009; Purpura 1998; Yoshida-Morise,1998). This study contributes evidence of the processes and strategies engaged by the Aptis for Teachers Speaking tasks, as well as domain-insider (teacher educator and trainee) perceptions of the test tasks and their relationship to the real-world communicative demands of the classroom.

2. THE PRESENT STUDY

Given the existing emphasis on teachers' oral communication skills, the study focuses on the speaking test component of Aptis for Teachers. Aptis for Teachers (Speaking) is a computer-mediated test, which takes 12 minutes under normal test conditions. Details of the test format and structure are given in Section 2.2.1. The study addresses the following research questions.

1. What are the speaking processes and strategies engaged by test-takers in response to the Aptis for Teachers Speaking tasks?
2. What are test-takers' and teacher educators' perceptions of the relevance and suitability of the Aptis for Teachers Speaking tasks as measures of the classroom readiness of trainee teachers' oral skills?

For the first research question, verbal reports were used to investigate the processes and strategies engaged in by test-takers in response to the Aptis for Teachers Speaking tasks. For the second question, interviews with trainee teachers and teacher educators were used to explore the appropriateness of the tasks as measures of classroom readiness of EAL trainee teachers' speaking skills in the Australian context.

2.1 Participants

Two groups of stakeholder participants were recruited via the researchers' existing contacts within university teacher training courses in Melbourne:

- test takers: EAL background trainee teachers in, or approaching, the final semester of a two-year graduate level pre-service teaching course (Master of Teaching) at a university in Melbourne
- teacher educators: lecturers in a graduate level pre-service teaching course at a university in Melbourne.

Across the two groups, there were a total of 15 participants: ten trainee teachers and five teacher educators. The trainees were recruited via an email invitation circulated to two student cohorts in the Master of Teaching program – primary and secondary. Respondents were screened for suitability, with only EAL background trainees in, or approaching their final semester selected to participate.
The trainee participants were being prepared for generalist teaching, as well as disciplinary specialisations, for primary or secondary level. The disciplinary specialisations included mathematics, sciences, humanities, EAL, and languages other than English (LOTE). All had already undertaken three or more periods (each of 10–20 days) of supervised teaching practice on school placements for the practicum component of their training. They were aged from 21 to 40 years and were from a range of L1 backgrounds: six were L1 speakers of Chinese (including Cantonese, Mandarin, Shanghainese), and the remainder, Indonesian, Turkish, Kannada, or Hindi. A summary of the trainees' background characteristics is given in Table 1.

Table 1: Test-taker (trainee teacher) participants

ID No.	L1	Gender	Age	Master of Teaching – Primary (P) or Secondary (S)	Disciplinary specialisation (if chosen)	Aptis Speaking test form
T-1	Chinese	F	21–30	P	Unknown	B
T-2	Chinese	F	21–30	P	LOTE	A
T-3	Indonesian	F	21–30	S	Humanities, Media	B
T-4	Turkish	F	31–40	S	Mathematics, Physics	B
T-5	Chinese	M	21–30	S	EAL, Japanese	B
T-6	Chinese	F	21–30	S	Biology, Science	B
T-7	Chinese	M	31–40	S	Chemistry, Biology	A
T-8	Kannada	F	21–30	P	Mathematics	A
T-9	Hindi	F	31–40	P	Mathematics	A
T-10	Chinese	F	21–30	S	LOTE, Biology	A

The educators were recruited via an email invitation sent to the researchers' existing contacts, some of whom forwarded the invitation to selected colleagues in the Master of Teaching program. In keeping with the scope of the study, only a small number of participants were sought; therefore, the first five educators who volunteered were recruited. The educators were lecturers and/or clinical specialists in a Master of Teaching program. The clinical specialist's role is to support the development of clinical teaching skills through seminars and classroom observations and by acting as a resource for trainee teachers during their teaching placements in schools. In addition to the educators' current roles in pre-service teacher education, all were experienced classroom teachers at primary and/or secondary school level in the Australian context (with from 5 to 37 years of experience), across a range of subject areas including mathematics, humanities, media, drama, EAL, and LOTE. Three educators had additional experience in post-secondary/adult education, and four had also taught in schools and post-secondary contexts outside Australia (from 1 to 9 years). A summary of the educator' background characteristics is given in Table 2.

Table 2: Teacher educator participants

ID No.	Gender	Age	Highest qualification (teaching)	Current role (Master of Teaching program)	Classroom experience: - years - level (P, S, P-S)* - specialisation	Teaching experience outside Australia: - years - level (P, S, P-S) - LOI**
Ed-1	M	61+	Master of Education	Lecturer and clinical specialist	32 S mathematics	2.5 S LOI: English
Ed-2	M	51–60	Master of Education	Lecturer	35 S English, LOTE, drama	1 P-S LOI: English
Ed-3	M	51–60	Master of Education	Clinical specialist	6 S media, drama, English	none
Ed-4	M	31–40	Master of Teaching	Clinical specialist	5 P, S, P-S history, humanities, English	2 S LOI: English
Ed-5	F	61+	Master of Education & Master of Literacy	Lecturer and clinical specialist	37 P, S, P-S EAL, LOTE	9 LOI: English

* P – primary; S – secondary; P-S – post-secondary/adult education
** LOI – primary language of instruction

2.2 Instruments

The following instruments were used for data collection: Aptis for Teachers Speaking test (two parallel forms), questionnaires, interview and verbal report protocols.

2.2.1 Aptis for Teachers Speaking test

Two parallel forms of the Aptis for Teachers Speaking test were used in the study (labelled as forms 'A' and 'B' for the purposes of the study). The test has the following four parts:

1. Part 1 consists of three questions on personal topics about the test-taker and their interests. Questions are given one at a time and test-takers have 30 seconds to respond to each question.

2. Part 2 consists of a photographic image of a school-related topic (depicting a classroom scenario, for example) and three questions. The test-taker is first asked to describe the image (question 1) and is then asked two questions about the topic. In this part, the test-taker is required to describe, explain, give reasons and express their opinion. Questions are given one at a time and the test-taker has 45 seconds to respond to each question.

3. Part 3 consists of two photographic images of a school-related topic and three questions. The test-taker is first asked to compare the two images (question 1) and is then asked two questions about the topic. In this part, the test-taker is required to describe, compare, explain, give reasons. Questions are given one at a time and the test-taker has 45 seconds to respond to each question.

4. Part 4 consists of a photographic image of a school-related topic and three questions related to the topic. In this part, the test-taker is required to describe, explain and to discuss (giving reasons for an opinion, or weighing two aspects, such as advantages and disadvantages). The three questions are provided at the same time. The test-taker is given one minute to prepare a response to the three questions and may make notes during this time, followed by two minutes' speaking time.

As laid out in the developer's technical report (O'Sullivan, 2015), Aptis test tasks include the following performance parameters, specified in terms of the language of input and output (expected response): **channel**, **discourse mode**, **text length**, **topic familiarity**, **writer/speaker relationship**, **nature of information**. In the case of the teacher variant of the speaking test, described above, the input **channel** includes photographic images and aural/written on-screen questions; **discourse mode** relates to rhetorical purpose and structure (describe, give reasons, compare, and so on); and **text length** is specified in speaking time (displayed in an on-screen timer). **Topic familiarity**, intended to afford test-takers the opportunity to demonstrate their best performance (O'Sullivan, 2015), is achieved with the use of teaching-related input and questions in parts 2-4. The **writer/speaker relationship** is not explicitly defined for the speaking task; however, it is clear from the input and format that the (automated) voice delivering the test questions is intended to be envisaged as an audience. Finally, the **nature of information**, primarily the complexity and degree of concreteness/ abstraction of the expected output, is varied progressively within each test part, as well as across the test parts. The questions within each part move from less, to more complex in terms of the functional output required (thus, for example, from description to opinion in part 2, and from description to discussion in part 3). Likewise, the test parts are arranged in order of increasing difficulty, moving from personal topics (part 1), through topics with a concrete relation to the photographs (parts 2 and 3), to more abstract domain-related topics (Part 4). In the final test part, the test-taker is also given a 'longer turn' to speak (two minutes). To illustrate the appearance and format of the speaking tasks in the online environment, sample tasks extracted from the *Aptis Candidate Guide* are provided in Appendix A (note the themes and topics of the sample tasks shown are not indicative of the teacher variant).

As indicated in the above description of the four parts of the test, Part 1 is focused on personal topics, while Parts 2, 3 and 4 are focused on school-related topics (with questions referring to the test-taker's own classroom experience included). As it is these test parts that are tailored to the domain of teaching, Parts 2, 3 and 4, as summarised in Table 3, were the focus of the present study.

Table 3: Overview of Aptis for Teachers Speaking test: parts 2, 3 and 4

	Input	Output (functions)	Time allowance
Part 2	1 photograph 3 questions (one by one)	Describe, explain, give reasons, opinion	Response time: 45 seconds per question
Part 3	2 photographs 3 questions (one by one)	Describe, compare, explain, give reasons	Response time: 45 seconds per question
Part 4	1 photograph 3 questions (at once)	Describe, explain, give reasons, discuss	Preparation with note-taking: 1 minute Response time: 2 minutes (longer turn)

Aptis for Teachers is normally administered by computer via an automated online system which regulates the timing and delivery of the speaking task i.e. instructions and questions (on-screen and audio), picture stimuli, preparation time (part 4) and response times. To collect data with the educators, the test forms were converted to PDF documents and provided to the educators in hard copy. For the trainees, it was necessary to administer the test off-line to enable the test to be paused between Parts 2, 3 and 4 for the collection of verbal reports. To do this, the test forms were replicated as PowerPoint slideshows with the standardised format (timing, presentation, visuals, audio) preserved within each of the test parts. A 'practice test' form was also created for use with trainees for the purposes of test familiarisation and verbal report training. The practice test consisted of one question extracted from part 2 of the test. This was administered in the same PowerPoint format as described above.

2.2.2 Background questionnaires

Questionnaires were used to gather background information about the participants. For educators, the items were about age, gender, qualifications and experience in teacher education, classroom experience, familiarity with English language requirements for teacher registration, and familiarity with Aptis for Teachers (see Appendix B). For trainees, the items were about age, gender, language and educational background, and teaching experience (see Appendix C).

2.2.3 Verbal report protocol

A protocol for conducting the verbal reporting sessions with trainees was developed with a researcher script to ensure procedures were consistent with all participants. The protocol included instructions for administering the speaking test, for training participants to provide verbal reports, and for prompting participants appropriately during their verbal reports, i.e. using minimal, non-leading formulations to elicit further verbalisation if required (see Appendix D). The design of the stimulated recall method for verbal reporting used in the study was informed by a review of relevant literature (including Cohen, 2000; Gass & Mackey, 2000; Jourdenais, 2001), and drew on the approach described in Swain et al. (2009) and the procedures recommended by Green (1998).

Prior to the main study, a pilot study was conducted with three test-taker participants and the results used to refine the final verbal report protocol. To ensure procedures for administering the speaking test were consistent, sections of the researcher script were modified to elaborate on how to run and pause the slideshow format of the test which had been prepared for the purposes of the study. In the interests of managing the time commitment for participants so they would not become fatigued, the following procedural changes were also introduced to the protocol: the number of test questions used for familiarisation and training for participants was reduced from three to one; in the main part of the session directly following training, to maintain the integrity of the test structure, participants took all four parts of the test but only provided verbal reports for the domain-relevant parts of the test, i.e. Parts 2, 3 and 4.

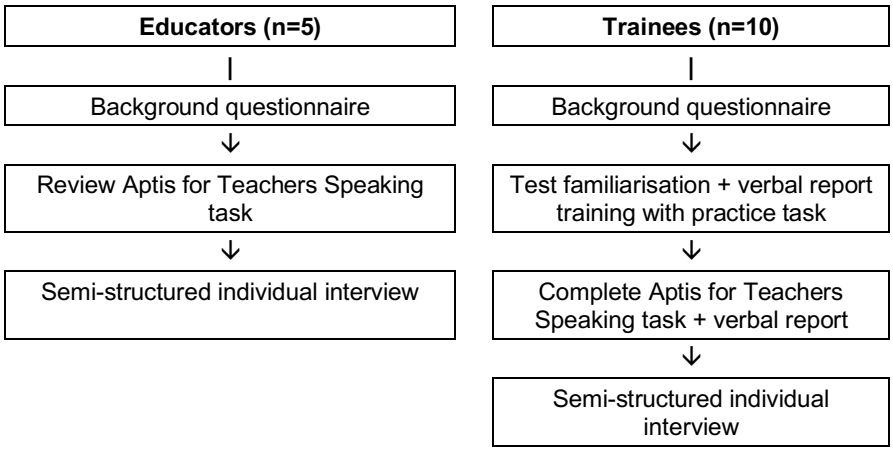
2.2.4 Interview protocol

A semi-structured interview protocol was used to elicit participants' views of the Aptis for Teachers Speaking tasks in relation to the oral communication demands of the classroom and in relation to the speaking skills valued and addressed in pre-service teacher training. They were also asked to consider the speaking tasks in relation to tests currently approved for teacher registration purposes in Australia. A semi-structured design was used to enable in-depth exploration of key issues as they arose during the interviews. The protocol was developed in two versions: one tailored to educators (see Appendix E) and one to trainees (see Appendix F). Prior to the main data collection, pilot interviews were conducted with two participants, one from each stakeholder group, and refinements made to the respective versions of the interview protocol to ensure the wording of the questions was clear.

2.3 Data collection procedures

An overview of the research design is given in Figure 1. Data collection proceeded after obtaining ethics approval from the University of Melbourne Human Ethics Advisory Group. Each participant first received a written overview of the study and its aims. Then, as outlined in the figure, the educators and trainees completed a background questionnaire (see Section 2.2.2 above).

Figure 1: Overview of research design



The educators then individually reviewed two parallel forms of the Aptis for Teachers Speaking test (forms 'A' and 'B'). As already noted, the test materials were provided to the educators as hardcopy PDF documents. This was done to enable them to examine the test materials at their convenience. Note, however, they had been carefully briefed to ensure they had a clear understanding of the automated, computer-based mode of delivery. After reviewing the test materials, the educators took part in an individual interview with one of the researchers, using the protocol described above (Section 2.2.4). The interviews, which were audio-recorded for later transcription and analysis, ranged from 30–60 minutes in duration.

For trainees, data was collected in individual verbal report sessions conducted by a trained research assistant. A short 'practice test' (see Section 2.2.1) was first administered to familiarise trainees with the test and for verbal report training purposes. Once the training was completed, the trainees took one of two parallel forms of the test ('A' or 'B'). The two forms were randomly allocated, while ensuring both forms were used in equal numbers. The test was delivered in the slideshow format described above (Section 2.2.3), allowing the test to be paused at the end of each part. The trainees' test performances were video-recorded for use as recall stimuli for the verbal reports. Trainees took all four parts of the test (to preserve the integrity of the test-taking experience) and produced a verbal report immediately after completing each of Parts 2, 3 and 4, i.e., the parts of the test with domain-relevant content (see Section 2.2.1). To do this, the test was paused at the end of Part 2, and the trainee instructed to watch the video of their performance as a stimulus for recalling and verbalising what they were thinking and doing while undertaking the speaking task. While producing their verbal report, the trainees were able to pause, rewind, and continue playing the recording as they wished. The verbal report protocol is described in Section 2.2.3. The test was then resumed, and the process repeated for the remaining test parts (3 and 4). In the final part of the session, trainees took part in an individual interview according to the protocol described above (Section 2.2.4). The sessions with trainees were each approximately 1.5 hours in duration. In addition to the video recording of the test performances, the sessions were audio-recorded in their entirety and the verbal reports and interviews transcribed for analysis.

2.4 Data analysis

The interviews and verbal reports were transcribed and then subjected to thematic analysis (Miles & Huberman, 1994; Saldana 2013) using NVivo software. To provide a richer context for the transcripts during the coding process, the audio recordings of the interviews and verbal report sessions were also consulted throughout the coding process rather than relying on transcripts alone. For the interviews, thematic codes at the superordinate level were derived from the pre-structured nature of the data (i.e. from the interview questions). Emergent themes in the interview data were also used to create more detailed sub-codes, and to refine the higher level codes. For the verbal reports, thematic coding was broadly informed by the literature on processes and strategies in oral communication (Bachman & Palmer, 1996; Bygate, 1987; Cohen & Olshtain, 1993; Fulcher, 2003; Luoma, 2004), and relevant aspects of Swain et al.'s (2009) taxonomy of test-takers' reported strategic behaviours in a speaking assessment. Swain et al.'s taxonomy was drawn from accepted frameworks of speaking strategies identified in the second language acquisition and language testing literature, and thus represented a useful starting point for developing a coding scheme.

Working with preliminary coding categories derived from Swain et al.'s (2009) taxonomy, two researchers each independently coded half of the verbal report dataset before coming together to discuss and revise or add codes to reflect emerging themes. Refinements were made through further iterations of coding, discussion of issues and disagreements, and revisions to the scheme to arrive at a final coding scheme, detailed below in Section 2.5, Table 4. Using the final coding scheme, one researcher coded the full verbal report data set. To conduct an inter-coder reliability check, a trained research assistant coded a subset (20%) of the data. The subset consisted of two intact verbal reports for all three domain-relevant parts of the test (i.e. Parts 2, 3 and 4), with both test forms ('A' and 'B') represented. Inter-coder agreement was established by calculating the total number of coding agreements divided by the total number of coding decisions; the percentage agreement between the two coders was 84%. Once the coding had been completed, a tally was taken of the number of reported behaviours coded to each category. To enable comparison of the relative frequencies of reported behaviours across categories, raw frequencies were converted to percentages of the total number of coded behaviours overall.

2.5 Verbal report coding scheme

The coding scheme is shown in Table 4. As shown in the table, the categories of reported strategic behaviours, of which there are 29 in all, are organised into five superordinate categories: Orientation to task, Communication, Cognition, Metacognition, and Affect. Definitions and illustrative data extracts for each of the categories of reported behaviours in the coding scheme are given in Appendix G.

Table 4: Verbal report coding scheme

Orientation to task	**Cognition**
Generating ideas	Engaging with stimulus images
Selecting ideas	Using imagery
Reflecting on the question	Using L1
Linking questions and image (Part 4)	Notetaking (Part 4)
Communication	**Metacognition**
Avoiding	Evaluating performance
Approximating	Identifying communication issues
Slowing delivery to gain time	Monitoring time
Expanding the message to fill time	Planning
Expanding the message to improve response	Self-correcting linguistic error
Word choice	Reflecting on orientation
Relating task to experience/knowledge	Reflecting on picture stimulus
Organising and structuring response	Reacting to test format
Paraphrasing to avoid repetition	**Affect**
Speeding delivery	Justifying performance
Referring to notes (Part 4)	Reflecting on affect

3. RESULTS

The results are reported in two main sections below. In Section 3.1, the processes and strategies reported by test-takers in undertaking the Aptis for Teachers Speaking tasks are described, and their distribution in the data reported. This is followed in Section 3.2 by the findings of the interviews with trainee teachers and teacher educators.

3.1 Test-taker processes and strategies engaged by Aptis for Teachers Speaking tasks

A summary of the frequencies of occurrences of strategies and processes identified through an analysis of the verbal report data are given in Table 5. The table shows the number of instances of reported behaviours for each of the 29 coding categories individually and tallied for each of the five superordinate categories (Orientation to task, Communication, Cognition, Metacognition, and Affect). The table reports the raw frequencies 'overall', i.e. summed across the verbal reports from all 10 test-taker participants, and for each of the 10 test-takers individually, identified by test-taker IDs 'T-1' through to 'T-10'. To facilitate comparison of the relative frequencies across categories, the overall raw frequencies (n) are also shown as percentages of the total number of reported behaviours across all categories (n=364).

In terms of superordinate categories, it can be seen from Table 5 that the most frequently reported types of processes and strategies were **metacognitive** (accounting for 29.12% of the total number of reported behaviours) and **communicative** (28.30%). These were followed in frequency by **task orientation** (16.48%), **cognitive** (13.19%) and **affective** (12.91%) behaviours. As can be seen in the table, with one exception, behaviours in each superordinate category were reported by all 10 test-takers at least once (in the case of test taker T-4, no affective behaviours were reported).

Amongst the individual categories, the most widely reported processes and strategies were **Relating task to experience or knowledge** (13.46% of the total number of reported behaviours), **Identifying communication issues** (8.79%), **Monitoring time** (8.53%) and **Justifying performance** (7.14%). These were followed in frequency by **Reflecting on affect** (5.77%), **Selecting ideas** (5.49%), **Notetaking** (5.22%), **Reflecting on the question** (4.95%), **Engaging with stimulus** (4.95%), **Generating ideas** (4.67%), **Slowing delivery** (3.85%) and **Evaluating performance** (3.57%). Each of these categories of behaviours were reported by at least seven of the 10 test-takers.

Three further categories (**Reacting to test format**, **Referring to notes**, **Expanding to fill time**), respectively accounted for between 2.75% and 2.20% of all reported behaviours. Each of these categories of behaviours were reported by at least five of the 10 test-takers.

The remaining categories of behaviours respectively accounted for yet a smaller proportion of the reported behaviours (1.65% and below). Nevertheless, for the majority of categories in this group, the behaviours were reported by more than one test-taker. The least widely reported behaviours in this group (i.e., reported by only one test-taker) were **Using L1**, **Speeding delivery**, **Paraphrasing**, and **Avoiding**.

Table 5: Frequencies of reported test-taker behaviours

	Overall		T-1	T-2	T-3	T-4	T-5	T-6	T-7	T-8	T-9	T-10
	n	%	n	n	n	n	n	n	n	n	n	n
Orientation to task	**60**	**16.48**	**4**	**12**	**2**	**10**	**3**	**9**	**5**	**4**	**5**	**6**
Generating ideas	17	4.67	0	1	1	3	1	1	2	2	2	4
Linking questions and image	5	1.37	2	0	0	2	0	0	0	0	0	1
Reflecting on the question	18	4.95	0	3	0	3	2	6	2	1	0	1
Selecting ideas	20	5.49	2	8	1	2	0	2	1	1	3	0
Communication	**103**	**28.30**	**9**	**21**	**14**	**16**	**4**	**5**	**6**	**12**	**11**	**5**
Approximating	4	1.10	0	0	2	0	0	0	1	1	0	0
Avoiding	4	1.10	0	4	0	0	0	0	0	0	0	0
Expanding to fill time	8	2.20	2	0	0	2	1	1	1	1	0	0
Expanding to improve	5	1.37	0	1	2	0	1	0	0	1	0	0
Organising / structuring	6	1.65	1	1	1	1	0	1	0	0	1	0
Paraphrasing	1	0.27	0	0	0	0	0	0	0	1	0	0
Referring to notes	9	2.47	2	3	1	0	0	1	0	0	1	1
Relating task to experience	49	13.46	2	11	7	7	2	2	2	8	7	1
Slowing delivery	14	3.85	1	1	1	5	0	0	2	0	1	3
Speeding delivery	1	0.27	1	0	0	0	0	0	0	0	0	0
Word choice	2	0.55	0	0	0	1	0	0	0	0	1	0
Cognition	**48**	**13.19**	**4**	**14**	**4**	**9**	**1**	**3**	**3**	**3**	**3**	**4**
Engaging with stimulus	18	4.95	2	2	0	5	0	2	1	3	2	1
Notetaking (Part 4)	19	5.22	2	3	2	4	1	1	2	0	1	3
Using imagery	5	1.37	0	3	2	0	0	0	0	0	0	0
Using L1	6	1.65	0	6	0	0	0	0	0	0	0	0
Metacognition	**106**	**29.12**	**10**	**19**	**15**	**5**	**12**	**5**	**11**	**5**	**5**	**19**
Evaluating performance	13	3.57	1	3	1	0	1	0	2	1	2	2
Identifying comm. issues	32	8.79	2	5	8	0	4	0	1	3	0	9
Monitoring time	31	8.52	4	6	5	4	1	3	3	0	3	2
Planning	6	1.65	1	2	1	0	0	0	2	0	0	0
Reacting to test format	10	2.75	2	0	0	0	3	2	1	0	0	2
Reflecting on orientation	6	1.65	0	3	0	0	1	0	0	0	0	2
Reflecting on stimulus	5	1.37	0	0	0	1	2	0	1	1	0	0
Self-correcting	3	0.82	0	0	0	0	0	0	1	0	0	2
Affect	**47**	**12.91**	**4**	**5**	**4**	**0**	**6**	**8**	**4**	**2**	**4**	**10**
Justifying performance	26	7.14	3	2	0	0	4	3	3	2	3	6
Reflecting on affect	21	5.77	1	3	4	0	2	5	1	0	1	4
TOTAL	**364**	**100%**	**31**	**71**	**39**	**40**	**26**	**30**	**29**	**26**	**28**	**44**

In the remainder of this section, the individual processes and strategies reported by test-takers while undertaking the Aptis for Teachers Speaking tasks, organised under the five superordinate categories, are described and illustrated with extracts from the test-takers' verbal reports.

3.1.1 Orientation to task

Behaviours grouped under this superordinate category are associated with what test-takers do to orient themselves to the test task. This involves situating themselves in relation to the task, and establishing the parameters of the task and what it requires of the test-taker. In doing so, test-takers set about framing the content of their response, and in this sense are engaging in a kind of macro-planning. As can be seen in Table 5 above, four individual categories of strategies and processes are grouped here: Generating ideas, Selecting ideas, Reflecting on the question, and Linking question and image. The most frequently reported behaviours in this group were Selecting ideas (n=20), Reflecting on the question (n=18) and Generating ideas (n=17).

3.1.1.1 Generating ideas

In reporting on the process of generating ideas for a response, test-takers described drawing on their professional knowledge, from teacher training or classroom experience, as illustrated with Excerpt 1, below. Strategies used by test-takers for generating ideas also entailed situating themselves 'as teachers' in relation to the task. Test-taker T-8 (Excerpt 2), for example, recalls authentic classroom experiences. In a further example, T-5 (Excerpt 3) is oriented towards 'thinking as a teacher' rather than focusing on his 'speaking' performance.

> Excerpt 1: I was thinking about um probably, I should just say something about, about the like the teaching method or the pedagogy that I have learned. [T-10]
>
> Excerpt 2: Just thinking about my time as a teacher, and what, the groupwork or the projects that we've done together as a class in groups and what were the good things that happened out of it. [T-8]
>
> Excerpt 3: I am not really thinking whether I should perform well in my speaking, I am just thinking as a teacher [T-5]

3.1.1.2 Selecting ideas

In selecting ideas, test-takers described making choices from the ideas generated for the response (for example, T-9 in Excerpt 4). They also described the rationales for their choices, which included pedagogical concerns (as exemplified in Excerpt 5), as well as how readily they were able to recall experiences and ideas. For instance, T-2 (Excerpt 6), who identified a range of possible experiences to draw upon for their response, opted for 'the most recent' of these, while T-6 (Excerpt 7) chose an experience they remembered because it was a time when they hadn't performed well.

> Excerpt 4: So a lot of things came to mind because I've done group projects with students, and ah, I was trying to pick and choose which one to talk about. [T-9]
>
> Excerpt 5: I reckon as a teacher I would prefer to teach a class where everybody feels more together, and they would be more encouraged to participate as well. So that's why I picked the top one. [T-1]
>
> Excerpt 6: Because as a teacher you help students individually a lot. So there is a lot of moment going on. So I pick an recent one to start. [T-2]
>
> Excerpt 7: Yeah, so for the first question, 'do you remember one time that you're doing the lesson plan', um, yeah. So, I just talk about, ah, the time I remember, the first time. So, 'cos I didn't do well, that's why I still remember it. [T-6]

3.1.1.3 Reflecting on the question

Reflecting on the question included test-takers recalling, or restating individual questions within the test parts (Excerpt 8), and considering the meaning of the question and the nature of the required response. Some test-taker comments reflected difficulties they encountered understanding the task requirements. For example, T-6 (Excerpt 9), in seeking to understand what was required by the task, reflects on the wording of the question, 'describe the picture'. Other reflections included comments on the difficulty of the question type; for example, T-4 (Excerpt 10) comments on the difficulty of selecting information associated with 'why' questions. Other test-takers reflected on how readily the questions could be related to their own professional background. In Excerpt 11, for example, T-10 responds positively to the question, specifically because she can relate it to her knowledge of teaching. On the other hand, T-7 (Excerpt 12) expresses concern that a question relying on a specific aspect of domain knowledge or experience (in this case, certain types of classroom experiences) would be unanswerable for test-takers whose experiences did not match the question.

> Excerpt 8: That's when I re-read the question and it said, 'why is it good for young children to play sports?' [T-8]
>
> Excerpt 9: She say 'try to describe the picture', right? And then I'm actually not sure about the question because 'describe the picture' means to describe the things in, the objects in the picture. Or imagine what actually happened there. [T-6]
>
> Excerpt 10: I think the 'why' questions are always the most difficult questions, I guess. Because, there may be lots of reasons and you have to pick one. [T-4]
>
> Excerpt 11: And then the next question, I think the second question is quite interesting and a really good question…Sounds like I can probably express some of my ideas and maybe relate it. Relate some of this question to my teaching philosophy and what my perspective is regarding teaching. [T-10]
>
> Excerpt 12: I don't know. If there is any teacher who never had a kid…never came up to he or her. What is she or he going to do? They can't answer the question, can they? [T-7]

3.1.1.4 Linking questions and image

Linking questions and images refers to the process of establishing the relationship between the questions and picture stimulus. This was reported only in relation to Part 4 of the test. As explained in Section 2.2.1, compared with the other test parts, Part 4 is concerned with a relatively abstract domain-related topic. Further, in Part 4, the questions and expected response seem to be tied less to the stimulus image and more to the test-taker's experience in the domain. Three test-takers reported some difficulty in understanding how, if at all, the questions were connected to the picture. As indicated in Excerpt 13, T-4 concludes that they 'are not closely related'. In a similar vein, T-1 (Excerpt 14) implies that the picture was perhaps not necessary to the task. As Excerpt 15 demonstrates, T-10 felt similarly that the picture was not needed, but goes further in saying that it posed a distraction from the task of responding to the questions.

> Excerpt 13: I just first saw the picture, and then, ah heard the question, and then I tried to find out the relation between the picture and the question. And it took a little bit ah, of thinking time, because they were not closely related. [T-4]
>
> Excerpt 14: I wouldn't have known what it was until I read the question. It was just a folder or something. It wasn't obvious what it was, so I don't think the picture would've made a difference in this question. [T-1]
>
> Excerpt 15: And then I realised, oh, I don't actually need to look at the picture, and then I just read the questions. It's a bit distracting. [T-10]

3.1.2 Communication

As shown in Table 5 above, 11 different categories of communicative strategies were identified in the verbal report data: Approximating, Avoiding, Expanding to fill time, Expanding to improve, Organising and structuring, Paraphrasing, Referring to notes, Relating the task to experience, Slowing delivery, Speeding delivery, and Word choice. As can be seen from the table, by far the most commonly reported communicative strategy was Relating the task to experience (n=49); this was also reported by all test-takers in the study. This was followed in frequency by Slowing delivery (n=14), Referring to notes (n=9), Expanding to fill time (n=8), and Organising/structuring (n=6), each reported by at least six test-takers. Of the less frequently reported strategies, Expanding to improve, Approximating, and Word choice were reported by at least two test-takers. The remaining communicative strategies (Avoiding, Speeding delivery and Paraphrasing) were each reported by only one test-taker.

As will be explained in the descriptions that follow, communicative strategies were reported by test-takers in relation to different kinds of communication problems. In response to encountering specific linguistic difficulties, some of the strategies used were essentially reductive (reducing what is communicated, or avoiding the problem encountered), while others focused on achievement, or successful communication of the intended message. Reported communication strategies were also responsive to other parameters of the test task, including the time allowance; for instance, test-takers reported using strategies for extending or reducing the length of their response in light of awareness of the remaining time allowance. Strategies were also used in relation to qualitative aspects of performance, including structure/organisation of the response, and the quality of ideas. When faced with the problem of running out of ideas or not knowing what to say, in order to maintain or regain fluency, test-takers reported using strategies for gaining planning time and for accessing ideas.

3.1.2.1 Avoiding

Reported avoidance behaviours, essentially reductive in nature, included the use of repetition, message substitution, and message abandonment in order to avoid linguistic difficulties. In the first example below (Excerpt 16), T-2 describes leaving the message incomplete because she can't find an English word for what she wanted to say. In the second example (Excerpt 17), the same test-taker describes resorting to 'description' in order to avoid the language of comparison, which she was 'too scared to think about'.

> Excerpt 16: And there's one thing I wanna add but there's not a word in English I could describe it. [T-2]
>
> Excerpt 17: And then I start describing it, because I was too scared to think about um, what sentence I should use to compare. Because, um, for, I have to think about like, what structure of the sentence I should use to articulate my meaning of comparison ah, without repeating one. [T-2]

3.1.2.2 Approximating, Paraphrasing to avoid repetition, Word choice

Three types of achievement-oriented strategies – Approximating, Paraphrasing and Word choice – were identified in the verbal report data. Approximating refers to the strategy of using lexical or grammatical substitution to approximate the meaning or function of an unknown word or structure. Approximating with lexical substitution is illustrated in Excerpt 18.

> Excerpt 18: I can't think of the name of that sport, that's why I was, you know, using 'martial art'. What is it, judo? [T-7]

Paraphrasing to avoid repetition was rarely reported (one instance, by one test-taker only). Test-taker T-8 reported that she used paraphrase to avoid repeating wording she had already used in the same test part: as explained in Excerpt 19, she had already 'touched on' the same ideas when responding to a previous question.

> Excerpt 19: I touched on that in my previous response, so I was a little bit unsure of, okay, do I just repeat it, or do I rephrase it. [T-8]

The strategy labelled Word choice was also rarely reported (one instance each, by two test-takers). This strategy entailed exercising word choice with careful attention to accuracy or perceived sensitivities in interpreting what was depicted in the picture stimulus. As the data excerpts below illustrate, use of this strategy yielded more generic lexical choices: in the case of T-4 (Excerpt 20), who was unsure of the roles and professions of the people in the photograph, 'adults' (rather than 'teachers') is chosen out of a concern for accuracy; in the case of T-9 (Excerpt 21), 'person' is chosen to avoid making assumptions about the ages and roles of the individuals pictured.

> Excerpt 20: Then I changed my mind and said, 'okay, don't tell them teachers, just tell them adults'. [T-4]

> Excerpt 21: I thought it was about teaching, but I didn't want to be presumptuous about it. I thought it could've just been students working with each other, because, I couldn't tell their ages…So I say things like 'one person' and 'people' instead of 'children' and 'teacher'. [T-9]

3.1.2.3 Speeding delivery, Expanding the message to fill time

Both of the strategies, Speeding delivery and Expanding the message to fill time, were reported by test-takers in relation to awareness of the remaining time allowance for speaking (displayed in an on-screen timer). Speeding delivery, used to deal with a lack of remaining time, entailed increasing the rate of speech in order to complete the full intended response within the time allowance. Only a single instance of this strategy was reported. Conversely, the second strategy was used to deal with 'extra' time remaining. Expanding the message entailed extending the response with additional ideas, details, or embellishments. As illustrated with Excerpt 22 and Excerpt 23, the goal in doing this was to fill the remaining time, rather than to improve the quality of the response to the questions.

> Excerpt 22: And, ah, to fill up the time, you can see in the end I try to say something, you know, to make up the time, not quite relevant to a lot of things, you know. [T-7]

> Excerpt 23: And then, since I had time and I couldn't think of, ah, anything else in my part, then I focused on the picture, and just to fill the time, I said, 'oh, that teacher is doing that'. [T-6]

3.1.2.4 Expanding the message to improve response

Unlike the strategy of Expanding to fill time (described above), this strategy of expanding the message with more information, details or elaboration, was used with a concern for improving the quality of the response to the question. This is illustrated in Excerpt 24, where T-3 is motivated by wanting to increase the 'complexity' of her answer. In another example (Excerpt 25), T-8 describes elaborating response to better address the question.

> Excerpt 24: I was trying to expand on my answer…just to make my answer a little more complex. [T-3]

> Excerpt 25: I realised I wasn't just talking about children, I was talking about students in general. Yeah, um, then I was like okay, when you're young it's easier to grasp things as opposed to when you're older. [T-8]

3.1.2.5 Organising and structuring response

This reported strategy involved organising the response by considering the sequencing and linking ideas while speaking. This was done with consideration for the type of information being presented in the response. This is illustrated in Excerpt 26, for example, where T-2 describes a distinction between 'details', 'structure' and 'colour' as an organising principle. This strategy was also used with consideration for rhetorical purpose. For example, in Excerpt 27, T-3 seeks to link ideas with attention to textual form ('comparison'), while in Excerpt 28, T-9 considers her preferred argument structure when she opts to talk about 'benefits' before 'problems'.

Excerpt 26: Cos like, I sort of like, structuring like what should I say. So after saying the details thing, I can actually just go back and look at the structure and the colour. [T-2]

Excerpt 27: I was trying to connect that with the first statement that I said, and I was trying to link it with the second picture. So the form of comparison. [T-3]

Excerpt 28: As soon as I said 'there are problems' I was like, maybe I should go from the benefits side instead of saying what the problems of not doing it are, and that's when I changed that…I personally believe it's a better argument to say what's the benefits to something so that one realises this is what I get out of it. [T-9]

The final example below (Excerpt 29) shows how one test-taker reportedly used this strategy to deal with limited time for speaking: after checking the timer, T-6 describes structuring her response (that is, streamlining the structure to deal with two questions at once), specifically to enable her to complete it within the time allowance.

Excerpt 29: Yeah and see the timer. So, I think I just speed it up and then, yeah, so I kind of combine the second question with the third one. [T-6]

3.1.2.6 Slowing delivery, Relating task to experience, Referring to notes

Test-takers reported three types of strategy for either regaining or maintaining fluency in communicating their message. Slowing delivery was used to gain planning time when test-takers were having trouble thinking of what to say, or how to say it. It was achieved by slowing the rate of speech, hesitating, pausing, or repeating the question. As illustrated in the examples below, test-takers described using these behaviours to gain time for thinking of what to say next (Excerpt 30), retrieving ideas (Excerpt 31), organising the response (Excerpt 32) and making language choices to formulate the message (Excerpt 33).

Excerpt 30: I was just pausing and thinking of ways to link imagination into that. [T-3]

Excerpt 31: And the hesitation was also where I was thinking of an article that talks directly about this, and I was going 'what does it say?'. [T-9]

Excerpt 32: So I repeat the questions again, so, ah, help me organise what I'm going to say. So I repeat the question again. [T-2]

Excerpt 33: And I think I was trying to decide whether I should say 'should' or 'do'... I was thinking about like which verb I should use. [T-10]

The strategy of Relating task to experience, as already noted, was the most the dominant of all reported strategies: as reported in Table 5 above, it was identified in 49 instances and mentioned by all 10 test-takers. As illustrated in the examples below, this strategy involved drawing on professional knowledge and experience, acquired through teacher training and other relevant classroom experience, to come up with relevant ideas (Excerpt 34 and Excerpt 35), and appropriate phrases and pedagogical terminology (Excerpt 36). In the final example for this strategy (Excerpt 37), T-8, after stating that she has no teaching experience in the disciplinary specialty related to the question, describes casting further afield for ideas, drawing on what she has learned from her own reading and prior studies.

Excerpt 34: I had no idea what to say. I was, yeah, sort of 'oh no', trying to sort of remember whether I learned this back in my degree or whether I heard it. [T-3]

Excerpt 35: I just, ah, tried to remember, how I did that as a teacher, when I was working in Turkey…I just thought of what I had done. [T-4]

Excerpt 36: I used a lot of words, um I just roll out of my mouth because we've been saying it during the placement. We've been saying it during the course. Like, 'intervention'... 'go around and help them' and 'prompting questions'. [T-2]

> Excerpt 37: Again, this is not an area that I've really, delved into, so I've never been involved in any of the physical education lessons. So I was again, just trying to think about all the times I've seen the children play or work together, yeah…And I read a lot about it too, through the years that I've studied, um, even when I studied psych. [T-8]

The strategy of Referring to notes was reported in relation to undertaking Part 4 only (the part where test-takers may take notes during the allocated preparation time). Test-takers reported using this strategy when faced with not remembering what they had planned to say or when they were uncertain how to organise their response. For example, T-3 (Excerpt 38), describes how she referred to her notes to keep her response 'on track', and T-1 (Excerpt 39), to organise her response around the ordered list of keywords in her notes. In Excerpt 40, T-2 refers to her notes to help him decide which ideas to include in her response.

> Excerpt 38: So I was just really, really grateful I have the notes. Because I was starting to lose track of what I was saying, because I had dotted points on my notes, and just follow through that. [T-3]

> Excerpt 39: I was kind of like, oh, I've got lots to say, but then I don't know what to say first. So when that happens, what I usually do is look down at my notes. [T-1]

> Excerpt 40: I was thinking, should I talk about that math group, or should I talk about the inquiry? And then I looked down, I saw I write down 'inquiry' and 'website' and so that's what I'm gonna stick on with. [T-2]

While T-3 (Excerpt 38, above) was 'grateful' for having her notes to refer to, in some instances, test-takers reported that their notes were not helpful because they were somehow inadequate or incomplete. For instance, T-2 (Excerpt 41) describes how she 'looked down' at her notes for guidance, only to find that what she had hoped or expected, was not there.

> Excerpt 41: I just looked down to kinda like, there's also the dot points that I haven't got time to write down in my mind…I thought I would wrote down, but nope, I didn't. [T-2]

3.1.3 Cognition

As shown in Table 5 above, four types of cognitive strategies were identified in the verbal report data: Engaging with stimulus, Using imagery, Using L1, and Notetaking. The most frequently reported of these were Notetaking (n=19) and Engaging with stimulus (n=18). The reported cognitive strategies involved processes for engaging with the task and manipulating language in order to make sense of the input (the picture prompts and questions), retrieve ideas, and produce language for a response to the task.

3.1.3.1 Engaging with stimulus images

Test-takers described engaging with the stimulus images in various ways, including establishing what is depicted in the photograph (for example, counting how many students are pictured, as illustrated in Excerpt 42), interpreting aspects of the context and situation (for example, T-4 in Excerpt 43, makes her own assessment of the mood of the participants in the scene depicted), and scanning the images for details. Excerpt 44 illustrates another type of engagement with the stimulus image: T-2 describes the strategy of 'locating herself' within the scene depicted, and of manipulating the contents of the scene ('I changed it') to create a congruence between what she can see when she looks at the photograph, and what she can see when she is 'in' the picture.

> Excerpt 42: I was trying to count like how many people there are there in the picture, but then at first, I was gonna say seven, but then I saw obviously there's a teacher sitting there, so I just change to six students plus a teacher. [T-10]

> Excerpt 43: I focused on the faces. Ah, saw that people are having fun. Ah, the teacher was doing her job, ah, willingly, yeah, she was happy and ah trying to help, um, students. [T-4]

> Excerpt 44: I tried to put myself into the picture. So, um, if I'm in the picture, and that globe or something would be at the back of it. And then I think, no, it's a picture, and I look at their angles, I then think, oh no, the global, it's not at the back of the classroom, it's actually in the corner, so I changed it. [T-2]

Test-takers also described problems in interpreting the images or deciphering what was depicted in them. For example, in Excerpt 45, T-4 isn't able to fully understand the scenario depicted because the photograph doesn't contain enough information ('the computers were off'), while T-6 (Excerpt 46) had difficulty decoding the image, explaining that she was 'confused' about which people are teachers and which are students.

> Excerpt 45: I just focused on the ah computers I saw, and tried to guess, ah, what lesson would it be. And I couldn't, you know, infer-, the computers were off, shut down, so I couldn't, you know guess. So I just tried to guess. [T-4]

> Excerpt 46: I just trying to see what's in the first picture because for the second one, it's more obvious like with a lot of computers, but for the first one, I'm just a bit confused the person in the front. Are they actually teachers? Because this guy on the left, he looks like student as well so I'm not sure. [T-6]

3.1.3.2 Using imagery

Using imagery refers to the strategy of conjuring mental images of memories from the classroom and teacher training. Test-takers' mental images were invoked as 'pictures', 'flashbacks', or moving images ('like a video', Excerpt 47) which supported the generation or recall of ideas, and served as stimuli for language production. In some instances when test-takers used the communicative strategy of relating the task to their own experience (see Section 3.1.2, above), the cognitive strategy of using imagery was employed in conjunction with this. For example, in drawing on her professional experience for ideas, T-2 (Excerpt 48) describes 'a picture of what happened'.

> Excerpt 47: As I was talk, I had like a video playing in my head of what happened last time. So I was trying to describe like what was in my head, yeah. [T-3]

> Excerpt 48: When I describe those, it's actually the memory itself, like, the picture of what happened, the storyline of what happened in my mind. [T-2]

3.1.3.3 Notetaking (Part 4)

Notetaking was identified in the verbal report data only in relation to Part 4, in which test-takers are allowed to write notes during the preparation time. Notetaking was used in the preparation time to plan and organise the response. As a communication strategy (Section 3.1.2, above), the notes themselves were later consulted to aid the recall of ideas and organisation of the response during the speaking performance. Notetaking was underpinned by processes of ideas transformation (summarisation, selection, organisation) and language manipulation (keywords, abbreviations, symbols), and these processes can be seen reflected in the different approaches to notetaking that were reported by test-takers. These included writing down keywords or main points, and writing abbreviations or symbols to serve as mnemonic devices, as illustrated in Excerpt 49, for example. In Excerpt 50, T-9 explains how this approach was engaged in conjunction with an awareness of the time available for preparation (in which writing at length was not feasible). Test-takers also used notetaking to create a 'template' or outline which anticipated the structure of their spoken response. As illustrated in Excerpt 51, these structures were organised around the three questions in the input of Part 4. Test-takers reported adding details to their outline to guide their speaking performance, such as a 'logical' arrangement of keywords that could be referred to while speaking (Excerpt 51), or contextual information that could be included to improve communication of the message (Excerpt 52).

Excerpt 49: I just put down 'd' and it's not remember but I have thoughts in those random doodles, so those draft, so when I look at it, those thoughts just coming back…it can reminds me of what I was thinking. [T-2]

Excerpt 50: I realised I had one minute, which isn't enough to write out whole sentences, so I was trying my best to write down keywords to guide my thinking while I'm doing the talking. [T-9]

Excerpt 51: I just break it down because three questions will break it down into three parts, and each part I just write down maybe four keywords, and then yeah, just in a logical way so that later on, I can skip things if I want. [T-6]

Excerpt 52: And then I realised I should probably contextualise it. And so, in the first part I wrote, 'during last placement', just remind myself to sort of contextualise it a little bit. [T-3]

3.1.3.4 Using L1

Only one test-taker (T-2) reported using their L1 while undertaking the test tasks. This strategy comprised thinking in L1 to generate and formulate ideas, and taking notes in L1 (for test Part 4). Using L1 was engaged in conjunction with the communicative strategy of drawing on professional knowledge or experience for generating and formulating ideas (see Section 3.1.2) and seemed to depend on whether this knowledge/experience was located in an L1 or EAL context for the test taker. As T-2 describes it (Excerpt 53), she used Mandarin (L1) when she was dealing with concepts that she was most familiar with from, or had learned in an L1 context.

Excerpt 53: When it comes to this question, I want to talk about like how important sunshine is to the human body, and that area is what I learn in Mandarin. So when I structuring the dot points, they are in Mandarin. [T-2]

3.1.4 Metacognition

As shown in Table 5 above, the metacognitive strategies identified in the data were coded into eight categories. The most frequently reported of these were Identifying communication issues (n=32) and Monitoring time (n=31). These were followed in frequency by Evaluating performance (n=13), Reacting to test format (n=10), Planning (n=6), Reflecting on orientation (n=6), Reflecting on stimulus (n=5) and Self-correcting (n=3). Each of these strategies entailed conscious reflection on the test-taking process and experience, the insights from which were used by test-takers to evaluate and plan their performance in various ways. The descriptions that follow are organised around a common strategic focus or orientation.

3.1.4.1 Self-correcting linguistic error, Evaluating performance, Identifying communication issues

Three types of metacognitive strategy were focused on monitoring and evaluation of performance: Self-correcting linguistic error, Evaluating performance, and Identifying communication issues. In the first of these, test-takers reported monitoring the accuracy of their language production to self-correct grammatical or lexical errors. In the next strategy, Evaluating performance, test-takers evaluated the quality of their performance in terms of language production, including pronunciation, word choice, grammar, or fluency (as in Excerpt 54, for instance); their response to a particular question (Excerpt 55); or their overall task performance (Excerpt 56).

Excerpt 54: I think I stammered a lot there. [T-10]

Excerpt 55: I sounded quite vague, I realised, in that answer. [T-9]

Excerpt 56: So I think that I did really badly in this question and also the question afterwards. [T-5]

The strategy, Identifying communication issues, entailed awareness that a communication difficulty had been encountered. Examples of the reported difficulties include running out of ideas (for example, in Excerpt 57), lack of linguistic resources (Excerpt 58), or inability to judge audience-appropriateness of the intended message (Excerpt 59).

> Excerpt 57: I'm totally out of things to say. I had no idea what to say after that. [T-1]
>
> Excerpt 58: But for all of a sudden, how should I name that term? My mind just went blank. [T-5]
>
> Excerpt 59: I was trying to say there are students like of different races and I think, oh, probably it's not appropriate thing to say, so I just stop. [T-10]

3.1.4.2 Reflecting on orientation, Planning, Monitoring time

The next group of strategies were oriented to planning and goal setting: Reflecting on orientation, Planning, and Monitoring time. In Reflecting on orientation, test-takers reflected on their orientation to the task (described in Section 3.1.1 above) in relation to their interpretation of what the task required them to do, and in relation to their own performance goals or the assessment criteria they envisioned for the test (as illustrated in Excerpt 60 and Excerpt 61). In another example, T-5 (Excerpt 62), uncertain about the task requirements ('do you really want me to describe what the picture is?'), considers what his performance goals might be. Test-takers also pondered whether the veracity or acceptability of their ideas and opinions would matter in the assessment, as illustrated in Excerpt 63.

> Excerpt 60: I cared about the words that I was saying. 'Cos I treated it like a real test, like displaying your range of vocabulary that you have, I learnt is an important thing. [T-10]
>
> Excerpt 61: So if, like, I'm trying to thinking…should I use some fancy words to get me a higher score? [T-2]
>
> Excerpt 62: I was wondering, are you assessing my speaking performance, or do you really want me to describe what the picture is? [T-5]
>
> Excerpt 63: What if [the examiner] disagree?…I don't think I actually believing in it…It's really conflict, like, when you are asked about your opinion on something, and then you start saying those common opinion and you think, 'anyway, no that is not scientifically proved'. [T-2]

The strategy of Planning involved making advance decisions about organisation, length, structure and rhetorical purpose of the response, as illustrated in Excerpt 64, for example. Planning also included giving attention to cohesion in the response (Excerpt 65). In their use of planning strategies, test-takers took into consideration the time allowance (and monitored this, as will be described below). For example, T-2 (Excerpt 66) plans the length of her response in relation to the time allowance. Test-takers also took into consideration the available ideas; thus, to describe two images (Part 3), T-1 (Excerpt 67) describes choosing 'comparison' as the discourse mode on the basis of the ideas she was able to generate.

> Excerpt 64: I was trying to think of…whether I should summarise each of the picture first, or whether I should start comparing them. So I was like, hm, which technique should I go? [T-3]
>
> Excerpt 65: Try to pick and organise my thinking, and organise what I'm going to say…I try to answer the three questions one by one, but don't make them sound too separately, but when I'm answering the first question, I will start to answer the second question, and by connecting the language. [T-7]
>
> Excerpt 66: I want to decide whether I gonna tell the story long or short. If it's long version, there's not enough time, and I'm probably gonna tell a different story. [T-2]
>
> Excerpt 67: I didn't know what to say about them both, so instead of talking about each of them separately, I decided to go and compare them. [T-1]

The strategy of Monitoring time entailed checking the on-screen timer to keep track of the time allowance remaining for production, and for planning the response (Part 4). This strategy tended to be linked to Planning strategies (as mentioned above), and to Communicative strategies (Section 3.1.2 above). For example, T-4 (Excerpt 68) expands the message, having realised there is still more time left, while T-9 (Excerpt 69) decides to truncate the message to fit her response into the remaining time. Monitoring time was also associated with Affective strategies (described in Section 3.1.5 below) i.e., reflecting on feelings of pressure or anxiety triggered by awareness of the time, and justifying performance on the basis of not having enough time to prepare or respond.

> Excerpt 68: Then, I was constantly checking the time, too, so, um I saw that I had a lot of time, so I started to talk about the things that I saw on the table. [T-4]

> Excerpt 69: I wanted to share more on the different aspects of engagement, and then I looked at the time going, just the word engagement should be sufficient, and I moved on to disadvantages. [T-9]

3.1.4.3 Reflecting on picture stimulus, Reacting to test format

The two metacognitive strategies grouped here, Reflecting on picture stimulus and Reacting to test format, involved evaluative responses to the test content (stimulus images) and aspects of the test format. In their reflections on the stimulus photographs, some test-takers made negative remarks about the content of the photographs, including the accessibility of the content and, therefore, their value as a stimulus for ideas. Test-taker T-5 (Excerpt 70), for example, felt that the two images presented in Part 3 weren't 'very comparable', despite the task requirement for a comparison. In another example, T-4 (Excerpt 71) felt a little stymied by the image presented in Part 4 ('it limited my thinking'), explaining that the image and the questions were not complementary from her perspective. One test-taker (T-8) also reported that certain images in the test (of a popular pastime) were not very accessible to her because she lacked any specific knowledge of the particular activities depicted (Excerpt 72).

> Excerpt 70: I don't find them very comparable…Because the first one looks like a classroom and the second one looks like a computer lab. So if you give me two classroom-like pictures, it will make more sense. [T-5]

> Excerpt 71: But, sometimes planning doesn't require a computer. You just take a pen and paper and some things came to your mind and you just take notes to a paper and so, that's why it wasn't the, ah, um, it limited my thinking. [T-4: Part 4]

> Excerpt 72: When I first saw those pictures, sports isn't really an area of my interest…so my first thought wasn't very positive [T-8]

Test-taker reports of their reactions to the test format were concerned with the aural/on-screen presentation of questions in Part 4 of the test, including uncertainty about whether the questions would remain on-screen for the duration of Part 4 or would 'disappear' (Excerpt 73), and confusion about the number of questions in Part 4. This confusion stemmed from the mechanics of the on-screen text, i.e., whether the number of question marks corresponded to the number of questions (Excerpt 74). One test-taker (T-10) also commented on the alert ('beep') which sounds the onset of speaking time, stating that she would prefer an extra 'one or two second' pause between the prompt and the 'beep'.

> Excerpt 73: I straight to looking at the questions to make sure I didn't miss anything because I thought maybe the questions would disappear after she finishes. [T-1]

> Excerpt 74: Cos I'm not so sure if it's only two questions or three questions. Cos I only see two question marks. [T-5]

3.1.5 Affect

Affective strategies identified in the data were coded into two categories: Reflecting on affect and Justifying performance. In the first category, Reflecting on affect, test-takers reflected on the affective states (i.e. mood and emotions) that they experienced while undertaking the tasks. Test-takers reported feeling stressed, nervous or anxious in response to aspects of the test situation, such as the computer format (Excerpt 75), or the time constraint (Excerpt 76). While these reactions were negative, test-takers also described positive affect, such as feeling comfortable, confident or 'good' (Excerpt 77) during the test.

> Excerpt 75: I hope this, this, won't be a real thing because it's really stressing me out. Yeah. By doing it in front of the computer. [T-5]
>
> Excerpt 76: I think this one I'm just not ready. Like, I haven't think properly to answer that, so I'm bit nervous…so when I see the timer, I just get nervous. Yeah. [T-6]
>
> Excerpt 77: Probably using words like stimulus, I was feeling quite good, because it sounds quite good. [T-10]

In the second category of affective strategies, Justifying performance, test-takers reported links between affective factors, the context and situation, and the quality of their performance. Thus, to explain the quality of their performance on the task, test-takers appealed to the test format and content, and their own affective response to it. As illustrated in Excerpt 78 below, test-takers reported being hampered by a lack of time to prepare (in relation to Parts 2 and 3 of the test, as well Part 4 which includes preparation time). Other factors that test-takers felt had a negative impact on their performance included the test situation itself or 'waiting for that bell' (Excerpt 79), not being interested in or familiar with certain topics (Excerpt 80), and the aural/on-screen presentation of the questions; for example, T-10, uncertain about the number of questions (see Section 3.1.4 above), describes not having noticed one of the questions in Part 4 as accounting for her incomplete response to the task (Excerpt 81).

> Excerpt 78: If I have more time, like give me some more time, like give me some time to write down the answers, then I can probably elaborate it better. [T-5]
>
> Excerpt 79: Maybe if I wasn't waiting for that bell or something, if it wasn't test conditions or something, I'd probably be able to speak more. [T-1]
>
> Excerpt 80: So that's why I didn't have much to say, only because it isn't an area of my interest. [T-8]
>
> Excerpt 81: I didn't see the first part of the question, so I didn't write anything about the last time that I wrote a detailed lesson plan. I don't think I actually talked much about it. [T-10]

3.2 Stakeholder perceptions of Aptis for Teachers Speaking tasks

The interviews with trainee teachers and teacher educators were used to explore understandings of the speaking demands faced by teachers in Australian classrooms and, given these demands, perceptions of the suitability and relevance of Aptis for Teachers Speaking for measuring teachers' oral skills. As indicated above (Section 2.2.4), a semi-structured interview protocol was used with both stakeholder groups (see Appendices D and E). The interviews were staged around three broad themes: the oral communication demands faced by teachers, the relevance of Aptis for Teachers Speaking tasks to these demands, and how the Aptis tasks compare with the approved tests used in teacher registration processes in Australia (noting that, in the local context of the study, the point of comparison was IELTS, this being the only test approved for this purpose that was available in the

local context and that participants were familiar with). The findings are reported below under these three themes, and are illustrated with extracts from the interview transcripts, labelled with participant IDs: 'T-1' to 'T-10' for trainees, and 'Ed-1' to Ed-5' for educators.

3.2.1 Oral communication demands of the classroom

A number of speaking skills were identified as necessary for meeting the oral communication demands of the classroom: explanations and instructions, classroom management, dialogic teaching, feedback, modelling language, accommodation, and vocabulary knowledge. As will be seen below, the importance of some of these skills was considered to be contingent on particular classroom contexts. Further, while the oral communication demands of the classroom were broadly understood in terms of linguistic resources, they were also seen to have a strong cultural component.

3.2.1.1 Explanations and instructions

Giving clear explanations and instructions in the classroom was considered important by trainees and educators. Speaking clearly and slowly was described as one aspect of this, as was breaking down information into smaller, or less complex parts. In Excerpt 82 below, T-1 describes the particular importance of explaining things slowly for younger students in primary school. Giving instructions was described by one educator (Ed-2, Excerpt 83) as 'a big skill set', involving making effective choices about the sequencing and selection of information, and checking for comprehension. Depending on the disciplinary context, giving instructions clearly was also implicated in student safety. In the context of the practical science classroom, for example, T-10 stresses the importance of ensuring instructions are clear enough to be understood by the class (Excerpt 84).

> Excerpt 82: And so I feel like, um, talking really fast would also move everything faster. The kids feel like they're being rushed, and they won't feel encouraged to answer questions. So I feel like it's really important to slow everything down. [T-1]

> Excerpt 83: I think, um, anybody getting a set of instructions in an order that's going to make sense to a group of students...that's actually, it's the sequencing, it's the clarity, and not overwhelming people with too much information at once, checking along the way that – it's a big skill set, I would say. [Ed-2]

> Excerpt 84: So and also when I was doing the science placement…I did write down all the safety rules that I need to speak really carefully, sentence by sentence, and I did try to practice it. Cos it's really important. Nothing's more important than to protect the child's safety, you know…you need to speak really clearly. Even though like, you have an accent. [T-10]

3.2.1.2 Classroom management

Communication skills for effective classroom management, identified by trainees and educators, were concerned with student discipline, and the general management of behaviour in the classroom. Trainees felt that these skills were dependent on specific classroom phraseology which needed to be learned for the Australian context. For example, T-10 (Excerpt 85) describes encountering economical phrases for prompting students to action such as 'chairs up', and T-3 (Excerpt 86) cites 'eyes this way' as an example of the 'special classroom lingo' that she learned in her pre-service training in Australia.

> Excerpt 85: I think people speak in different ways in Australia and in my own home country… Yeah, like, 'put up your chairs' and, um, there are even some interesting sentences that I was trying to learn, like, 'off you go'...if you say things like this, like 'off you go' it sounds good… Even this short phrases can be tricky, like 'chairs up'. [T-10]

> Excerpt 86: I didn't have a lot of like, sort of vocabulary in classroom management…There was this guy called [name], and he would have like a lot of special classroom lingo. So instead of saying, 'shut up sit down', which is what you wanna say in the heat of the moment, but you should break it down to steps, and plan your language accordingly. So the first one is to like give a direct direction. So first is like 'eyes this way' and then…if that didn't work, you move on to stage two…there'd be a word for it…Oh, 'language of choice'. [T-3]

Further, Excerpt 86 illustrates a particular form of communication used for classroom management (directions, given in stages as needed) that may not be familiar or valued in some contexts. Indeed, one educator (Ed-2, Excerpt 87) noted that facility with forms of communication such as this, depend on knowledge of classroom cultures (rather than English proficiency):

> Excerpt 87: I think it's the intercultural factors that are the biggest challenge…The way that they've been at school as a student and learnt, and the way that they may have done some work in their home country as a teacher, is a very different way of operating than in a mainstream Australian classroom. So…how they think about oral communication and directness, and typically being direct or indirect in managing behaviour, I would say, are the kind of challenges in the mindset. [Ed-2]

More generally, cultural knowledge was regarded as necessary precondition for being able to communicate effectively in the classroom. As T-7 explains (Excerpt 88), in order to know 'how to say things', it is necessary to 'know the culture'.

> Excerpt 88: And as a foreigner you need to know the culture. You need to know the situation wheth – how you say things differently in different situation. I think ah, mainly um, it's about knowing the culture. It's more important than anything else. Because, for us, um, English is not a big problem. [T-7]

3.2.1.3 Dialogic teaching

Dialogic teaching, seen by trainees and educators as essential for effective practice in Australian classrooms, was valued for activating students' thinking by engaging them in dialogue and conversation with the teacher and with peers. The communication skills associated with this were asking 'proper' questions, facilitating 'conversation' and 'listening', as illustrated in Excerpt 89 and Excerpt 90. Listening, and giving students the time to respond to questions, was seen as a particularly important aspect of dialogic teaching in the primary school context.

> Excerpt 89: Dialogic teaching is very important. You have to know how to, ask questions. Ask proper questions. To prompt the students' thinking…So I have benefited most from our training is to know, how to have more interactions with the students, and how I could ask more questions. [T-5]

> Excerpt 90: Um, definitely those dialogic teaching skills. So, being able to converse with a group of children, and also at the same time, respond to their queries, which comes up quite often in a primary classroom…I think, more than speaking skills, it's more those other micro skills. Things like questioning, which are pretty big, because they promote higher order thinking. So, questioning, the teacher questioning the classroom, students questioning each other, students questioning the teacher. So all of this is conversation in the classroom…So again, it comes down to that dialogics. To be able to listen and respond to that child's queries. [T-8]

According to Ed-5, who drills down into the speaking skills underpinning a dialogic approach (Excerpt 91), in classroom questioning, appropriate forms need to be used (i.e. open questions), and teachers require a sufficient linguistic resource for the flexibility needed in question-answer sequences to facilitate learning. Educators considered these dialogic skills as particularly challenging for EAL background trainees. For example, Ed-3 felt that such trainees, not being equipped with experience and/or linguistic resources for dialogic teaching, were 'always talking' in the classroom instead of allowing students the space to respond (Excerpt 92). As already observed by T-8 above (Excerpt 90), Ed-3 also notes here the importance of this in the primary school context.

> Excerpt 91: We often look at the type of questioning you use, the dialogic approach. Are you using closed questions? Are we using open questions? So we actually look at how we're using questions to deepen students' understanding. So we look at language, in that way, that flexibility when students are asking questions, to get them to think about it. So that's another ah dialogic approach requires um, linguistic resource. [Ed-5]

> Excerpt 92: They're not practised enough to, to use language to its full extent and that sometimes means they're filling in gaps where they should be just waiting for students to respond. They feel they've always got to be talking…especially with primary students um, they are not allowing students to respond adequately. [Ed-3]

3.2.1.4 Giving feedback

Giving effective oral feedback to students was identified by educators as a key communication skill for the classroom. According to Ed-2 (Excerpt 93), to provide effective feedback, teachers need linguistic resources for appropriate 'phrasing' and the interactional competence to judge the timing of feedback during the 'flow' of talk.

> Excerpt 93: An important part of the oral interaction I think, that all teachers, is giving feedback… directly in the classroom obviously to students, and, how that might best be phrased. And, when it's appropriate to do so in oral interaction is another thing we take in language obviously is, um, when it's good to interrupt the flow of somebody else's speech and correct, um, or not. So those are kind of skill sets, but it's also the language to be used in those situations. [Ed-2]

3.2.1.5 Modelling language

The skills for modelling language that participants felt were important for the classroom relate to normative dimensions of language use including producing grammatically accurate language, using comprehensible pronunciation and speaking in a familiar accent. As indicated in the following illustrative extracts (Excerpt 94 and Excerpt 95), the primary school classroom was felt to be the most important context for the ability to model accurate linguistic form due to its perceived role in literacy development.

> Excerpt 94: I had first graders in semester one. And, um, they're just learning how to read certain words and say certain words, and I think it's really important to sort of, model how they're meant to say it. So you model the way their parents say it and their friends say it. [T-1]

> Excerpt 95: I think the idea of modelling, especially in primary school…there's a certain expectation to be able to model to students in a classroom of you know forty kids, where they start to pick up things, and say, hey, that's, I think that's wrong, that apostrophe goes there. [Ed-4]

Educators had divergent views on the importance modelling the norms of 'Standard English', however. This can be seen by comparing the following two extracts. In the first, Ed-3 stresses the importance of linguistic accuracy out of a concern for clear communication (Excerpt 96), while in the second, Ed-5 expresses the view that it is acceptable for the teacher's language to reflect the language use of the local context, which may reflect a linguistically diverse population of English language users (Excerpt 97).

> Excerpt 96: Overseas teachers who may not use the correct tense in English, often the meaning of what they are trying to get across is subject to confusion [Ed-3]

> Excerpt 97: Their students don't speak perfect English…[this trainee's] English is quite acceptable to students because that's what they hear [Ed-5]

3.2.1.6 Differentiation

The term, differentiation, was used by educators to refer to linguistic behaviours of communicative accommodation, such as slowing the rate of speech (as described by T-8 in Excerpt 98) or using more simple lexis and structures (T-6, Excerpt 99). Trainees and educators stressed the importance of this type of modification of language choices and style of delivery for enabling effective communication in diverse classrooms, which include students from EAL or non-English speaking backgrounds. Ed-5 (Excerpt 100) emphasised the importance of the linguistic resource needed to support differentiated language use ('to paraphrase, to rephrase, to simplify') and stressed the importance of differentiation for a range of classroom contexts, including the disciplinary diversity of primary school settings in Australia, where generalists teaching (across disciplines) is most common.

> Excerpt 98: Because I had a girl from Korea come in mid-year, and she didn't speak much English. So how do I convey what I wanted the whole class to do? So, I think, understanding that even students come from so many different backgrounds, and I think teachers, like trainee teachers need to be trained in making things as simple as possible for these children. And communicating clearly, effectively, in the most simple way. [T-8]

> Excerpt 99: Yes, I have to use very simple language…I've got a, three girls from Vietnam. Yeah, they don't speak a lot of English. [T-6]

> Excerpt 100: Sometimes they don't have the linguistic resource in which to re-use language, to paraphrase, to rephrase, to simplify. Because if we're teaching concepts, ideas, and I'm thinking in the secondary context, even in the primary context, we're teaching maths, we're teaching science, we're teaching history, we're teaching English. We've got 25 learners, and we need to differentiate our teaching, sometimes to 25 learners, and that's where I find there needs to be more development of that linguistic resource. [Ed-5]

3.2.1.7 Vocabulary

Trainees felt that a wide vocabulary knowledge was important for communicating effectively in the classroom because it enables teachers to give detailed and precise explanations of disciplinary content and, as T-9 explains (Excerpt 101), because it allows for more nuanced communication with students. As T-9 also pointed out, through modelling behaviours, teachers could also develop students' vocabulary knowledge.

> Excerpt 101: There are also different connotations to how things are said. Um, so the same thing can be said in different ways…And there are certain words that just mean differently, and to the benefit of students' confident and motivation. And it's also, um, modelling to students, so when students hear the teacher use these words, they also learn these words. [T-9]

The terminology of specific disciplines or school subjects was also identified by trainees as an important aspect of vocabulary knowledge for the classroom. Trainees mentioned the importance of being able to use, and explain the meaning of discipline specific terms, in geography (Excerpt 102) or sciences (Excerpt 103), for example.

> Excerpt 102: Subject specific, um, terminology…So I was teaching geography during my placement, and there was a lot of sort of discipline specific words that I was like, okay, I didn't know that about maps. [T-3]

> Excerpt 103: There a lot of English you don't talk in daily life when we were teaching. In science especially. Especially in chemistry and biology…The real thing is when you talk about professional English. Those science language, um, that's what matters in teaching. In my subjects. [T-7]

Note for some trainees and educators, when it comes to teaching curriculum content, 'language' may matter less than knowing the content, particularly in the context of secondary school science classrooms. For example, for Ed-1 (Excerpt 104), the ability to provide accurate and precise explanations is underpinned first and foremost by content knowledge. Similarly, T-7 (Excerpt 105) privileges accuracy of content over grammatically accuracy when it comes to communication in science classrooms.

> Excerpt 104: In secondary, I suppose, um, for better or for worse, it ah has to do oftentimes with content. Ah, for maths/science subjects and for humanities and English subjects, it's often to do with the precision involved in explaining ah content. [Ed-1]

> Excerpt 105: It is not when I use, a pro right or wrong, when I use a verb right or wrong, when I use a tense right or wrong, because the student will understand you. But if you use, um, when you're teaching, if you talk the content wrong, that's what matters. [T-7]

3.2.2 Relevance of Aptis for Teachers Speaking tasks to classroom communication

The relevance of the Aptis for Teachers Speaking tasks was evaluated by the trainees and educators in terms of content and format and, given the nature of these task dimensions, the extent to which the task output (expected responses) reflect the oral communication demands of the classroom.

The content of the Speaking tasks was perceived by trainees and educators to be highly relevant to teaching, with the picture stimuli and questions contextualized to the types of classroom scenarios that teachers would be likely to encounter in the Australian context. As illustrated in the remarks from Ed-4 (Excerpt 106) and Ed-2 (Excerpt 107), the ideas tapped into by the questions were also felt to be congruent with issues and concerns of contemporary pedagogical theory and practice.

> Excerpt 106: A good set of questions, asking for reflections, on the contemporary research and practice. [Ed-4]

> Excerpt 107: Ah, describing a picture. Um, that does kind of pry into the student's stance on differentiation, which I think is quite useful from a content point of view. [Ed-2]

However, educators also noted that the output, or expected responses to the speaking tasks did not reflect some of the important oral communication demands of the classroom; namely, those involved in classroom management, giving feedback, accommodation (differentiation) and the broader practices of the dialogic teaching approach, as described under Section 3.2.1 above. This issue was construed in terms of the test content since the speaking tasks do not actually elicit classroom genres where teacher-student interaction takes place. For example, after remarking in Excerpt 107 (above) on the value and relevance of the tasks for demonstrating knowledge of pedagogical theory, Ed-2 goes on to explain (Excerpt 108 below), that the tasks nevertheless are not suitable for eliciting the language of classroom practice, since they don't reflect the communication tasks of the classroom. Others similarly observed that the tasks required test-takers to reflect and share their ideas on pedagogical theory. While these processes were seen as having value in relation to other kinds of professional communication (such as employment interviews, or parent-teacher conferences), as illustrated in Excerpt 109, they were not seen as relevant to speaking skills for the classroom.

> Excerpt 108: But I'm not sure that that helps in any of the actual tasks that [trainee] students have to do in the classroom…That doesn't mean then that the [trainee] students have the language of instruction to do that…It's testing the candidate's ability to speak about situations in the classroom, and methodology, rather than testing their ability to use important language in the classroom and important interaction styles. [Ed-2]

> Excerpt 109: I think it's more a little talk on theory, rather than- for me, the language resource required for me isn't about how would you group students in that classroom. It's sort of, tell me about what sort of instructional language would you use? Tell me about what sort of language would you use if a student came up to you. [Ed-5]

A trainee (T-2), who similarly observed that the tasks don't elicit the language of the classroom, understood this as resulting from the 'audience' for the speaking tasks, which, as described in Excerpt 110, she envisioned to be professional colleagues or parents, rather than students in a classroom context. As another trainee put it (Excerpt 111), the tasks aren't relevant to classroom speaking because 'there are no students here'.

> Excerpt 110: So it's completely different…it's not classroom, not the language that you use with children. More like the language you use when someone else is observing your class. Like teaching fellow supervisors. Or you kinda like, imagine myself talking to parents, talking to some parents about like how do you, how do you cope with children, how do you teach students. [T-2]

> Excerpt 111: Why not because um, there are no students here…So I don't think it reflects a teacher's actual performance, regardless of the subject. [T-4]

On the whole, the monologic format of the tasks, which precludes the testing of interactive speaking skills, was regarded as a limitation of the test in terms of its relevance to speaking skills for the classroom. Furthermore, some participants saw the monologic format of the test as somewhat of a clash with their professional values when it comes to dialogic teaching in the Australian context, where classroom dialogue and conversation is considered key to effective practice. As one trainee observed (Excerpt 112), the monologic responses required by the test tasks, particularly in the longer turn for Part 4, are at odds with how she has been trained.

> Excerpt 112: They encourage you not to say too much in class, less teacher talk and more student collaboration. [T-1]

Despite these observations that the speaking tasks do not reflect classroom interaction, attitudes to the content-relevance of the tasks were generally positive, as described above. However, there was some diversity amongst stakeholders' views on whether this was necessarily appropriate for a test of teachers' speaking skills. One educator expressed a concern that the orientation of the tasks to well-known topics could allow test-wise candidates to produce rehearsed responses, thus diminishing the usefulness of scores for decision-making. For others, the content of the tasks was contentious due to perceived issues of cultural bias and test fairness. According to Ed-5, for example (Excerpt 113), while the content of the tasks was relevant to communication skills for teaching, it also seemed to be embedded in the values of a specific cultural context. At the same time, however, another educator argued conversely (Excerpt 114) that the tasks were not specific enough for Australian classroom contexts.

> Excerpt 113: A lot of the questions are based on Western educational pedagogy…
> are we testing English, or are we are we testing someone's pedagogical philosophy…
> That's what I found a little bit confronting there. [Ed-5]

> Excerpt 114: I think that, 'cos oral communication, obviously, isn't just about the words,
> it's about being able to operate appropriately in that context. And that's where in fact the
> Aptis is culturally neutral, I don't think that necessarily helps. [Ed-2]

Participants also commented on the appropriateness of the task content with regards to the level of specificity of the topics, felt by some to be too little. For example, one trainee (T-7), who was preparing for teaching in the sciences, felt that being proficient in 'science language' is what matters in his classroom; as shown in Excerpt 115, he argued that a test of teachers' speaking skills that targets grammatical accuracy, for instance, rather than disciplinary communication, is 'missing the point'.

In another example, T-10 (Excerpt 116) feels the Speaking tasks allow test-takers to demonstrate whether they can speak clearly enough, but not whether they have the ability to teach curriculum content.

> Excerpt 115: It is not when I use a pro right or wrong, when I use a verb right or wrong, when I use a tense right or wrong, because the student will understand you. But…if you talk the content wrong, that's what matters. But they never test this in this kind of English test. So I think they're sort of missing the point. [T-7]

> Excerpt 116: If you can get a really good score for this test, it means you have the ability to speak clearly and express your ideas. But, 'cos in the classroom, you are always teaching, um, your content knowledge and it depends on exactly what you are teaching, so there are many things you would never speak in a test like this. [T-10]

Finally, the timing of the tasks was remarked upon by trainees, specifically for Part 4, whose comments suggested that the time allowances had implications for the authenticity of the tasks. One trainee (T-8, Excerpt 117) remarked that the preparation time (1 minute) was out of line with the performance demands of the real world, where more time would routinely be available. Another concern raised about timing was in relation to speaking time, also for Part 4. As illustrated in Excerpt 118, some trainees felt that the complexity of the response warranted longer than the allocated speaking time of 2 minutes.

> Excerpt 117: And even as teachers, we have more time to prepare. Suppose I'm taking a lesson on health. I'd have time to prepare a lesson plan, and understand what it entails…whereas this, it was so specific, with no preparation time. [T-8]

> Excerpt 118: If it's meaning in terms of teaching skills, in terms of teaching, those time, it's just not enough to explain. [T-2]

3.2.3 Aptis for Teachers Speaking test compared with current approved tests

As already mentioned, in the local context of the study, the point of comparison for the stakeholders interviewed for this study was the speaking component of IELTS, which is the widely available approved test for demonstrating English language proficiency for teacher registration purposes in Australia. In comparing Aptis for Teachers Speaking with the speaking component of IELTS, trainees discussed their impressions of the two tests in terms of task difficulty and format, and the relevance of the content to their professional knowledge and experience. For educators, the content and the format of the two tests were also the salient points of comparison. Trainees also expressed their personal preference for taking one test over the other, typically taking into consideration which test they felt would offer them the better chance of success.

Overwhelmingly (as already seen in Section 3.2.2), the content of the Aptis Speaking tasks was felt to be highly relevant to the teaching domain. In contrast, the content of the IELTS speaking tasks was observed to be generic in nature. In general, this difference was felt to make the Aptis tasks more appropriate for measuring speaking skills of teachers although, as will be seen below, there were mixed views on this. Firstly, according to one trainee (T-9), a test geared for teachers, such as Aptis, would provide more useful information to test users; as she explains in (Excerpt 119), she felt that Aptis for Teachers test results might be preferred by employers since these would demonstrate a level of domain-relevant knowledge. Her remarks (Excerpt 120) also point to the positive washback that could be anticipated from a professionally-relevant test since it would be likely to promote test preparation practices that are relevant to the classroom.

> Excerpt 119: I would hope that if I do this and I do well on this, then I am more (liable) to be considered by schools, over candidates who have not. Because I think this would take so much more preparation and it requires one to have done group-work projects, and to have taught PE lessons. So, yeah, this this would show knowledge a lot better. [T-9]

> Excerpt 120: The IELTS is a simpler test. Not that this is hard, but this is subjective to what one says about pedagogical approaches which can change how the…teaching candidate is perceived. And I think this will take so much more preparation. So given the time I have for preparation, if I have a lot of it, I might go for that. [T-9]

Compared with the generic content of IELTS, some trainees also regarded the domain-relevance of Aptis for Teachers as preferable for an English test for teachers because they felt it would give pre-service teachers who take the test a better opportunity to demonstrate their speaking ability. As T-4 explains, for example (Excerpt 121), she would always have something to say in response to questions about teaching ('It's my profession'), whereas with IELTS, answering questions could be 'really difficult' for her if asked to speak on a topic beyond her interests.

> Excerpt 121: Because it's about teaching. It's my profession. Even though I have no idea about the question, I can answer in some way. I can find a way to answer the question. In IELTS, they may, ah, ask a question about media, and I may or may not be interested in that topic, and ah, that makes it really difficult to answer questions when you're not interested in the subject. [T-4]

Another trainee (T-10) felt similarly that the teaching-related content of Aptis would be 'a good thing' for pre-service teachers because test-takers could draw on their professional training and experience for ideas (Excerpt 122). However, as T-10 explains in Excerpt 123, she was ambivalent about her personal preference for Aptis or IELTS because she felt her existing familiarity with IELTS might give her a better chance of achieving her desired score.

> Excerpt 122: In terms of questions, of course this one asks you, like, more about your classroom and your experience as a teacher. So it probably means if you have finished a teaching course, you will have many things to talk about if you sit a test like this. That's a good thing then for the teachers to sit a test like this. [T-10]

> Excerpt 123: For myself, I'm already familiar with IELTS, so I would probably go for an IELTS, but if I like failed many times and I didn't get my ideal score, I can go for this one. It's like, sounds okay to prepare because you know much more about the topics and stuff. [T-10]

In contrast, there were some trainees who felt that the generic nature of the IELTS speaking test questions actually made them less challenging and, for this reason, would allow them to better demonstrate their speaking ability compared with the Aptis tasks. For example, T-3 finds the IELTS questions easier precisely because they are 'really general' (Excerpt 124). Another test-taker (T-7) concurs with the view that the IELTS questions are easier to answer because they do not rely on test-takers having any specific background knowledge or experience; in Excerpt 125, T-7 describes his concern that the classroom scenarios in the Aptis tasks ('those situations') may not reflect the 'real life' experiences of all pre-service teachers.

> Excerpt 124: The IELTS one is really, really, really general. It doesn't have anything to do with teaching, per se…I would actually choose IELTS just because it's a bit easier. [T-3]

> Excerpt 125: And ah and another thing, they won't ask you, you know, about those situations, you might not have met in your real life. But the IELTS test, they always set up situations you will meet in your real life. [T-7]

Not all trainees, however, expressed a preference for one test over the other. T-2, for instance, while regarding the content-relevance of Aptis for Teachers as 'an advantage', at the same time felt that it is possible to do well on IELTS, regardless of the ideas expressed, if you 'structure your language well' (Excerpt 126).

> Excerpt 126: A big advantage, because we learn the theory because we know there's technical terms, so it's a lot more easier like when you're speaking out so there's, they will ask you a lot of different concepts of questions, you know this is a language test, but those questions really get you thinking of about in the teaching concept. And then, you would automatically focus less on your language…Whereas in IELTS, no matter even your opinion, it's like, what you gonna talk about? It's absolutely nothing but as long as you structure your language well, you can get higher score…So I wouldn't say which one I would take. [T-2]

In addition to the content, another consideration for trainees was test format – whether face-to-face, or automated and computer-delivered. A number of trainees described feeling less comfortable when taking an automated/computer-delivered speaking test, and preferred a format that could provide a context for interactive speaking. For example, T-6 prefers a live interlocutor because, unlike 'a machine', this person can help her to 'feel calm' in the test situation (Excerpt 127). Moreover, T-7 (Excerpt 128) points out that face-to-face interaction with a live interlocutor, such as in the IELTS speaking test, provides a more authentic context for test-taker performance ('in the classroom, you prefer talk to a real person'), and contrasts this with the 'uncomfortable' one-sidedness of speaking to a computer, i.e. where the 'real person' does not 'take over' in a natural turn-taking sequence.

> Excerpt 127: For me, when I think about I'm talking to a machine, I just don't like that one. I like talk to person. So when I was doing my IELTS, I feel more – I mean, I feel nervous, but if that ah, assessor can talk to me and then, makes me feel calm, then I will feel very comfortable talking more. But the machine, I just [laughter] what's that? [T-6]

> Excerpt 128: Cos IELTS you talk to a real person. That's very different. That is very, very different…For example, in the classroom, you prefer talk to a real person. Because when you stop, the real person will take over. Not like when you talk to the computer. You've got ten seconds left on the clock, and there's ten seconds silence. It makes it a bit uncomfortable and then in this situation, you try to make up things to talk. So that's the biggest difference. [T-7]

Likewise, despite valuing the 'education-related' content of Aptis for teachers, T-8 favoured IELTS for the interactivity afforded by its face-to-face delivery. As T-8 explains (Excerpt 129), she would still prefer the interactive possibilities of the IELTS test (albeit 'programmed'), because it is 'more of a conversation'.

> Excerpt 129: I think definitely the Aptis is definitely more, um, education-related. At least the responses need to be education-related also, I feel. But, if I'd been given the choice I'd still do the IELTS because it is more interactive. I understand it's still programmed. But having that other person sitting across from me makes it more interpersonal, as opposed to a computer. And there are always those follow-up questions. Um, I'm not sure if it's set in stone or if they're slightly modified based on my response, but I've always felt like it's been more of a conversation. Whereas here, the questions come up and, okay, can you tell me the answers to these, and okay! So it doesn't seem very interpersonal, conversation-like. [T-8]

Additionally, one trainee (T-3) preferred IELTS over Aptis because of the non-verbal dimension of face-to-face communication, which she felt gave her the opportunity to perform at her best (Excerpt 130).

> Excerpt 130: And I had a bit of experience with like computer-based, and for some reason, they tend to mark my speaking as lower, as opposed to a real person where you can speak and there's eye contact and you can use your, yeah, you can use gestures, and stuff. [T-3]

However, despite some reluctance to embrace a computer-delivered format, for some trainees, their preference ultimately depended upon on which test they could earn the higher score. For example, T-5, who had a clear preference for a face-to-face speaking test, also remarked that the 'better' test would be the one on which he earned 'better marks' (Excerpt 131).

> Excerpt 131: I will see how, like how many marks I got first. Like, if I got better marks in here, then, of course, this one is better. Otherwise IELTS is better. [T-5]

For educators, their comparisons of Aptis for Teachers Speaking with IELTS hinged upon the content-relevance of Aptis and the interactive dimension of the IELTS speaking test format. As already noted (Section 3.2.2), although educators observed that the Aptis tasks don't elicit classroom genres — including, notably, those associated with dialogic practices involving teacher-student spoken interactions — they also regarded the questions as relevant to key issues in contemporary pedagogical theory and practice for the Australian context. In contrast, IELTS was seen to lack any relevance to the domain, but to contain an interactional dimension which educators regarded as reflecting an important aspect of classroom speaking skills. With one exception (Ed-3, shown further below), educators felt that Aptis was the more appropriate speaking test for teachers, in terms of its content-relevance, at the same time that they regarded IELTS as the more appropriate measure of teachers' speaking skills due to the interactive dimension of the IELTS speaking tasks. This tension between content-relevance on the one hand, and interaction on the other, is encapsulated in comments from Ed-4 (Excerpt 132), who describes his 'ideal' speaking test as an amalgam of the content of Aptis (his choice for teachers 'in a heartbeat') and the 'conversation' of IELTS (in his own test-taking experience, allowing him to 'authentically demonstrate his ability').

> Excerpt 132: I think the one benefit [of IELTS] was the fact that, because it was a conversation, I felt confident that I could accurately demonstrate, or authentically demonstrate my ability… In terms of the content, definitely it was very kind of vague. It was kind of much more generic. So, where we're talking about for teachers, I would say, in a heartbeat, I would, I would choose this [Aptis] over an IELTS because – if you could somehow merge the two, a conversation, an organic dialogue happening and the content that is relevant in a classroom, as opposed to: When did you come to Australia? What's your favourite food? [Ed-4]

The exceptional view amongst educators' was that of Ed-3, who felt that Aptis is not an appropriate test for teachers in the Australian context. He argued that, since the test doesn't reflect 'a dialogical and creative approach' (Excerpt 133), it doesn't align well with the pedagogical values of the local context, and test preparation practices could therefore have a negative impact on teachers' preparedness for the classroom. In the case of IELTS, on the other hand, Ed-3 claimed (Excerpt 134) that this test represents a better fit to local values and standards ('it's directly relational to here'), and that IELTS test preparation has positive washback, helping candidates to develop language skills necessary for the classroom.

> Excerpt 133: I find it [Aptis] too deterministic in instructional approach, and not a dialogical and creative approach. And I think that's important. [Ed-3]

> Excerpt 134: I don't think it compares. I think the IELTS because it's, um, directly relational to here, and it's taken, it's factored in the [APST] standards…And is much more focused on developing skills. Rather than determining skills. [Ed-3]

4. SUMMARY AND DISCUSSION OF FINDINGS

This study aimed to examine the speaking processes and strategies engaged by the Aptis for Teachers Speaking tasks, and to explore the perceptions of two stakeholder groups — EAL trainee teachers and teacher educators — of the appropriateness of the tasks as measures of teachers' classroom-ready speaking skills in the Australian context. Using a stimulated recall method, 10 test-takers provided verbal reports while undertaking the speaking tasks. These reports were coded thematically to classify the reported strategic behaviours engaged by the tasks. To gather stakeholder perceptions of the tasks, interviews were conducted with 10 trainee teachers and 5 teacher educators. The interviews were analysed qualitatively to identify themes in the participant commentary. This section summarises the key findings for the research questions of the study, and discusses these findings in relation to the study aims.

RQ1 investigated the speaking processes and strategies engaged by the Aptis for Teachers Speaking tasks. Analysis of the verbal reports collected for the investigation revealed that test takers engaged in a range of processes and strategies, grouped for this study into five superordinate categories as follows: Task orientation, Communication, Cognition, Metacognition, and Affect. With one exception, all test-takers in the study reported using processes and strategies across all five superordinate categories (only one test-taker did not report engaging in any affective processes or strategies). As will be highlighted below, in using processes and strategies associated with task orientation and communication, test-takers drew on their teacher training and classroom experiences, which supports the relevance and appropriateness of the test tasks for assessing if graduating teachers are language-ready for professional practice.

Task orientation was largely concerned with framing the content of the response. Thus, in orienting themselves to the task, test-takers sought to relate the test questions to their own professional background. They also reported drawing on their training and authentic experiences from the classroom to generate and select suitable ideas. For Communication, test-takers reported using a range of communicative strategies when they encountered communication problems while undertaking the tasks. The various strategies were focused on dealing with different kinds of communication issues, including specific linguistic difficulties, length of response, and quality and quantity of ideas. The most frequently and widely reported communication strategy (reported by all 10 test-takers) was to relate the task to their own teaching experience or background knowledge when facing problems of 'what to say', or 'how to say it'. In using this strategy, all test-takers looked first to their professional training and experience to find relevant ideas and appropriate terminology to enable them to continue or complete their message. This reliance on accessing professional knowledge and real-world experience, as seen in test-takers' orientation and communicative strategies, is some indication of the authenticity of the tasks.

The process and strategies in the remaining superordinate categories were not directly related to test-takers' teaching experiences or knowledge, but nonetheless represented a range of strategies relevant to models of second language speaking proficiency (Bachman & Palmer, 1996; Bygate, 1987; Fulcher, 2003; Luoma, 2004). These processes and strategies can be summarised as follows. Reported cognitive processes and strategies were most frequently concerned with engaging with the picture stimuli (to understand what was depicted in the images), and taking notes during the preparation time provided in Part 4 of the test. Reported metacognitive strategies were related to planning and goal-setting, monitoring the time allowances, reflecting on communication problems, and evaluating the quality of language production and overall performance. Metacognitive strategies also included evaluating the content and format of the tasks – specifically, the accessibility of the stimulus photographs and, in relation to Part 4 of the test, the correspondence between the test instructions and the number of questions included in the task. Affective processes and strategies entailed reflecting on positive and negative affective responses to the test situation, and seeking to construe a

causal link between performance quality and task-related situational factors (such as lack of time) and negative affect associated with these factors.

RQ2, using semi-structured interviews, explored trainee and educator perceptions of the relevance and suitability of the speaking tasks as measures of the classroom readiness of teachers' oral skills. In the interview commentary, the following speaking skills were identified as necessary for meeting the oral communication demands of the classroom:

- flexible use of linguistic resources to enable teachers to accommodate their language to diverse receptive needs of individual learners; referred to by trainees as the practice of 'differentiation'; regarded by educators as a special area of challenge for EAL background trainees
- the ability to listen and give meaningful responses to student questions and contributions to classroom conversations and discussion; described as dialogic teaching skills
- vocabulary knowledge
- modelling language
- being skilled in classroom genres: classroom management; feedback; clear and concise instructions and explanation of curriculum content.

The interview commentary also revealed the following perspectives on the speaking tasks, which will be discussed below: (i) the tasks were viewed positively for the relevance of the task input to the teaching domain; (ii) at the same time, the tasks were seen to offer a partial reflection of the key oral skills needed for classroom; (iii) some oral skills were felt to be highly contingent on specific classroom contexts, and the extent to which these contingencies could, or should, be captured by the test was questioned; (iv) test fairness was seen as a concern, to the extent that some of the skills needed for effective classroom communication are deeply embedded in cultural knowledge and values.

Due to the domain-relevance of the topics, the test content was viewed favourably by trainees who, being comfortable talking about topics that were largely familiar to them, felt, therefore, that the tasks could allow them to demonstrate their best speaking performance. It was also suggested that test users, such as schools/employers, might find the score information more useful than that from other speaking assessments, due to the relevance of the tasks. Educators similarly favoured the domain-relevance of the content, and generally (although not unequivocally, as will discussed below) saw intrinsic value in engaging test-takers in reflective talk about pedagogical theory and practice relevant to the Australian context. The broadly positive attitudes of stakeholders towards the speaking tasks suggests potential for positive test impact in the local context. As argued by Murray, Riazi and Cross (2012), positive attitudes to a test may be associated, not only with better test performance, but also benefits for language development. Given the professional relevance of the content, positive washback might also be anticipated from the test via test preparation practices which could help trainees to develop their skills for the classroom.

Regarding the language of the expected task output, educators observed that the tasks elicited functional language associated with classroom discourse, including explaining, highlighted as necessary for the classroom, given the importance of clear and concise explanations of curriculum content. They also observed that the speaking tasks tapped into professional discourses for communication external to the classroom, such as with colleagues, or parents, for example. However, educators noted that the test tasks did not target some of the key oral skills for the classroom, including those associated with 'differentiation' (accommodative communication supported by flexibility in the use of linguistic resources), dialogic teaching (listening, responding meaningfully, facilitating discussion) and classroom genres such as giving feedback to students, or managing students' behaviour. Moreover, due to the monologic nature of the response format, interactive speaking skills, perceived to be inseparable from teacher oral proficiency for the classroom, are not tested. So, while the tasks were seen to tap into some of the oral communication demands faced by teachers, because they do not test classroom genres — and more fundamentally, communication skills for interaction between teachers and students — it appears the construct of classroom speaking is under-represented in the test.

Further construct implications are suggested by the diverse and contingent communication demands of teaching, given the range of classroom contexts that teachers may encounter. As noted by educators and trainees, the contexts of educational level (primary or secondary), disciplinary area, and the needs of individual learners for instance, are associated with particular communicative demands. Tailoring instructional language to suit younger learners in primary school (such as breaking down explanations, or allowing more time for students to respond to questions), using differentiated language to match the needs of students in linguistically diverse classrooms, and appropriate use of disciplinary-specific vocabulary in the context of specialist subjects (such as mathematics, science, or geography) were cited as examples. Arguably, any standardised test, even a specific purpose test, is unlikely to be able to capture such diverse and contingent communication demands as might be found in the classroom, as has been noted by Elder and Kim (2013). Indeed, for any language assessment intended to relate to a particular domain, ensuring that tasks are at once, on topics that are familiar and accessible to target test-takers, at the same time as reflecting a sufficient range of contingencies in the target language use context, is a perennial challenge. As already mentioned above, the design of the Aptis for Teachers Speaking tasks around topics intended to be familiar to test-takers was generally perceived to be a strength of the test. However, some trainees felt that the contextualised content of the test was too specific for them, arguing that knowledge of particular classroom scenarios or teaching specialties cannot be assumed, particularly as some test-takers may have had limited pre-service experience and/or exposure to local classrooms.

The content of the tasks, and its relationship to the local context, was also a concern for some educators. As outlined above, there was a common understanding of the key oral communication skills needed by teachers for effective classroom practice in the Australian context. However, the extent to which these skills may be embedded in local cultural values and whether, therefore, such skills should be assessed in a language test for EAL teachers, was a source of disagreement amongst two educators in the study. In the opinion of Ed-5 (who remarked that the questions in the Aptis tasks seem to be 'based on Western educational pedagogy'), it is inappropriate for the pedagogic values of a particular cultural context to be embedded in test tasks. On the other hand, according to Ed-3, a test for teachers should align with the professional values of a given context. Certainly, teacher education in Australia aims to prepare pre-service teachers for meeting the expectations of the local context, and supporting trainees to develop their knowledge of 'cultures of schooling' is considered a necessary part of this preparation (Cruickshank, Newell & Cole, 2003). However, if knowledge of local values and cultures were a prerequisite for success on a language test, test fairness clearly would be a concern. In the case of Aptis for Teachers Speaking, a risk of cultural bias may persist in the Speaking tasks for candidates with limited experience of local education systems, particularly those who have not completed their qualifications in Australia. Given the small sample of views represented in this study, this concern warrants a broader canvassing of stakeholder perceptions of the cultural-specificity of the Speaking tasks, and the appropriateness of this for the local context.

5. CONCLUSION

This study investigated the test-taker processes and strategies engaged by the Aptis for Teachers Speaking tasks, and EAL trainee teachers' and teacher educators' perceptions of the tasks, in order to explore their relevance to the oral communication demands of classrooms in the Australian context. The findings show that the tasks tap into classroom spoken genres only in a limited sense because of the non-interactive nature of the task format, and the audience (i.e., not students) prescribed by the test content. In other respects, however, the test content was felt to be highly relevant to the teaching domain because the input of the tasks (stimulus images and questions) is concerned with pedagogic scenarios that teachers would be likely to encounter. The tasks were also found to engage test-takers in processes of reflection on teaching practice, which trainees and educators regarded as valuable for enhancing the development of professional expertise in teachers.

Overall, while the findings suggest that the tasks under-represent the construct of classroom speaking, because they do not test interaction, at the same time, stakeholder confidence in the tasks was strong on the basis of their content-relevance to the teaching domain. The findings also provide insight into the oral communication demands of the classroom, including the skills that may be most challenging for EAL teachers.

5.1 Limitations

There are some limitations to this study, which should be acknowledged. First of all, due to the size and scope of the study, a small number of participants from the two stakeholder groups were involved. With a larger number of participants, a greater diversity of stakeholder views and a wider range of test-taker processes, and/or a different pattern of distribution in these processes, may have been found. With a larger data set across both participant groups, the findings may have been more generalisable to other similar contexts.

Secondly, there are some methodological limitations in how the test was administered. Although the test-takers were encouraged to perform as if they were taking 'a real test', the test was administered only for the purposes of the study and therefore, test-takers' motivation may not have matched that of an authentic testing situation. Also, efforts were made in the research design to reproduce the online test delivery experience; however, it was not possible to fully replicate this because it was necessary for test-takers to be able to pause the test (between Parts 2 and 3, and Parts 3 and 4) to produce the stimulated recalls. For the educators, similarly, although they were carefully briefed on how the test is normally administered, reviewing the test in hardcopy did not allow them the opportunity to fully appreciate the how the test would normally be experienced.

Finally, verbal protocol reports were collected as evidence of the test-taker processes and strategies engaged by the speaking tasks. However, there are limitations associated with verbal protocol methods, in particular, the risk of reactivity (Jourdenais, 2001; Wigglesworth 2005), or altering task performance as a result of the procedures. In the case of the present study, a stimulated recall method (rather than a concurrent reporting method) was used to minimise disruption to performance. However, some impact on task performance cannot be ruled out. For example, the very knowledge that they would produce a report retrospectively may have affected test-takers' thinking and performance. Similarly, as noted by Barkaoui (2016), awareness of an envisaged audience (Barkaoui, 2011; Cohen, 2011), could potentially influence the contents of a verbal report. A further limitation of verbal protocol reports is that they can't be expected to be fully representative of all the processes and strategies used, since they are known to be limited to what participants are able, or choose, to report. Thus, verbal reports may be biased towards processes that are more easily described, and they are likely to preclude the more automated processes that are engaged in while performing a task (Cohen, 2000). In the present study, although the delay between performance and reporting was minimal (since reports were immediately retrospective), test-takers' ability to remember what they were thinking and doing may also have been a factor in the accuracy and completeness of their reports.

5.2 Future research and recommendations

In view of the limitations mentioned above, the current data set could be explored further through a comparative analysis of test-takers' reported processes and strategies and their test performance. Such an analysis could be put towards validating the verbal report data by comparing the reported processes and strategies with behaviours observable in performance. Analysis of the performance data could also be used to investigate the effectiveness of the reported behaviours i.e., the relationship between strategy use and the quality of test-takers' performances. Future research could therefore compare reported strategy use with measurable features of the performance discourse (such as quality of ideas, accuracy, cohesion, and so on). Also, while the test performances in the present study were not scored, future studies could make use of score data to investigate the impact of strategy use on performance.

This study has presented evidence of the relevance of the Aptis for Teachers Speaking tasks to classroom communication by drawing on reports of test-taker cognition, and insights from domain insiders, as well as those preparing to enter the domain. Other forms of empirical evidence could be put forth to support the validity of inferences based on Aptis for Teachers Speaking scores, including analyses of test-taker performances (as mentioned), correlation studies of test results and real-world performance, comparative studies of test-taker and classroom spoken discourse, or corpus analyses of real-world tasks in the domain. Further, the findings of the present study would lead us to predict that test-taker discourse from the Aptis for Teachers Speaking tasks would not resemble the discourse of classroom genres, particularly those involving teacher-student interactions. However, the findings showed that the speaking tasks did tap into other valued forms of oral communication, notably, reflective talk on teaching practice, which stakeholders saw as important for contexts such as employment interviews, collaborating with colleagues, or communicating with parents. It is also possible that monologic tasks tap into abilities that underlie interactional competence, and it would be worthwhile for future studies to examine the extent to which the linguistic tools, as well as cognitive processes and strategies, that underlie test task performances are relevant to the more interactional activities in classrooms.

The study findings also suggest potential areas for revision to test design. Although test-takers' reactions to the speaking tasks were generally positive, at the same time, aspects of the input were confusing to some participants. In some instances, the stimulus photographs were difficult to interpret, and certain topics were felt to be too specific and potentially beyond the experience of candidates. For Part 4 of the test, there was a perception that the instructions did not match the questions, with resulting confusion over the number of questions test-takers were required to answer. We therefore recommend that the Aptis test providers review the selection of picture stimuli to minimise any problems of interpretation and accessibility, and make minor revisions to Part 4 to address any issues with the alignment of the task instructions and corresponding questions.

REFERENCES

Arkoudis, S., Hawthorne, L., Baik, C., Hawthorne, G., O'Loughlin, K., Leach, D. & Bexley, E. (2009). *The impact of English language proficiency and workplace readiness on the employment outcomes of tertiary international students*. Melbourne: Centre for the Study of Higher Education, The University of Melbourne.

Bachman, L. F., & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Barkaoui, K. (2011). Think-aloud protocols in research on essay rating: An empirical study of their veridicality and reactivity. *Language Testing, 28*(1), 51–75.

Barkaoui, K. (2015). Test takers' writing activities during the TOEFL iBT® Writing tasks: A stimulated recall study. *ETS Research Report Series*, 2015: 1–42. doi:10.1002/ets2.12050.

Barkaoui, K. (2016). Examining the cognitive processes engaged by Aptis Writing task 4 on paper and on the computer. AR-G/2016/1. *ARAGs Research Reports Online*. London: British Council.

British Council (n.d.). *English test for teachers*. https://www.britishcouncil.org/school-resources/aptis/teachers

Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.

Cohen, A. D. (2000). Exploring strategies in test-taking: fine-tuning verbal reports from respondents. In G. Ekbatani & H. Pierson (Eds.), *Learner directed assessment in ESL* (pp. 127-150). Mahwah, NJ: Lawrence Erlbaum.

Cohen, A. D. (2011). *Strategies in learning and using a second language: Research and practice* (2nd ed.). Harlow, UK: Longman/Pearson Education.

Cohen, A. D. (2012). Test-taking strategies and task design. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 262–277). New York: Routledge.

Cohen, A. D., & Olshtain, E. (1993). The production of speech acts by EFL learners. *TESOL Quarterly, 27*(1), 33–56.

Cruickshank, K., Newell, S., & Cole, S. (2003). Meeting English language needs in teacher education: A flexible support model for non-English speaking background students. *Asia Pacific Journal of Teacher Education, 31*(3): 239–247.

Elder, C., & Kim, S. H. O. (2013). Assessing Teachers' Language Proficiency. In A. J. Kunnan (Ed.*), The companion to language assessment*. I:4:27:454–470. Hoboken, NJ: Wiley-Blackwell. DOI: 10.1002/9781118411360.wbcla138.

Fulcher, G. (2003). *Testing second language speaking*. Essex, UK: Longman.

Gass, S. M. & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.

Green, A. (1998). *Verbal protocol analysis in language testing research: A handbook*. Cambridge, UK: Cambridge University Press.

Hall, G. (2010). International English Language Testing: A critical response. *ELT Journal*, 64(3), 321–28.

IEAC. (2013). Australia – Educating Globally: Advice from the International Education Advisory Council. Canberra: Australian Government.

Jourdenais, R. (2001). Cognition, instruction and protocol analysis. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 354–375). Cambridge, UK: Cambridge University Press.

Leow, R. P., & Morgan-Short, K. (2004). To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition 26* (1): 35–57.

Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.

Miles, M. B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks: Sage Publications.

Murray, J., Cross, J., & Cruickshank, K. (2014). Stakeholder perceptions of IELTS as a gateway to the professional workplace: the case of employers of overseas trained teachers. *IELTS Research Reports*, 1 1–78. IDP: IELTS Australia.

Murray, J., Riazi, M. & Cross, J. (2012). Test candidates' attitudes and their relationship to demographic and experiential variables: The case of overseas trained teachers in NSW, Australia. *Language Testing*, 29, 577–595.

O'Sullivan, B. (2015). Technical report: Aptis Test development approach. TR/2015/001. London: British Council. Available for download: www.britishcouncli.org/aptis

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing, 20*(1), 26–56.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13: 111–129.

Purpura, J. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test-takers: A structural equation modelling approach. *Language Testing, 15*(3), 333–379.

Purpura, J. (2014). Cognition and language assessment. In A. J. Kunnan (Ed.*), The companion to language assessment* (pp. 1452-1476). Hoboken, NJ: Wiley-Blackwell.
DOI: 10.1002/9781118411360.wbcla150.

Saldana, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). London: Sage Publications Ltd.

Sinclair, J., & Coulthard, M. (1992). Towards an analysis of discourse. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 1–34). London, England: Routledge.

Swain, M., Huang, L., Barkaoui, K, Brooks, L., & Lapkin, S. (2009). The Speaking section of the TOEFL iBT (SSTiBT): Test-takers' reported strategic behaviors. *TOEFL iBT ™ Research Report.* Princeton: ETS.

Victorian Institute of Teaching (VIT). (2009). Talking to learn: Dialogue in the classroom. *The Digest*, Issue 2.

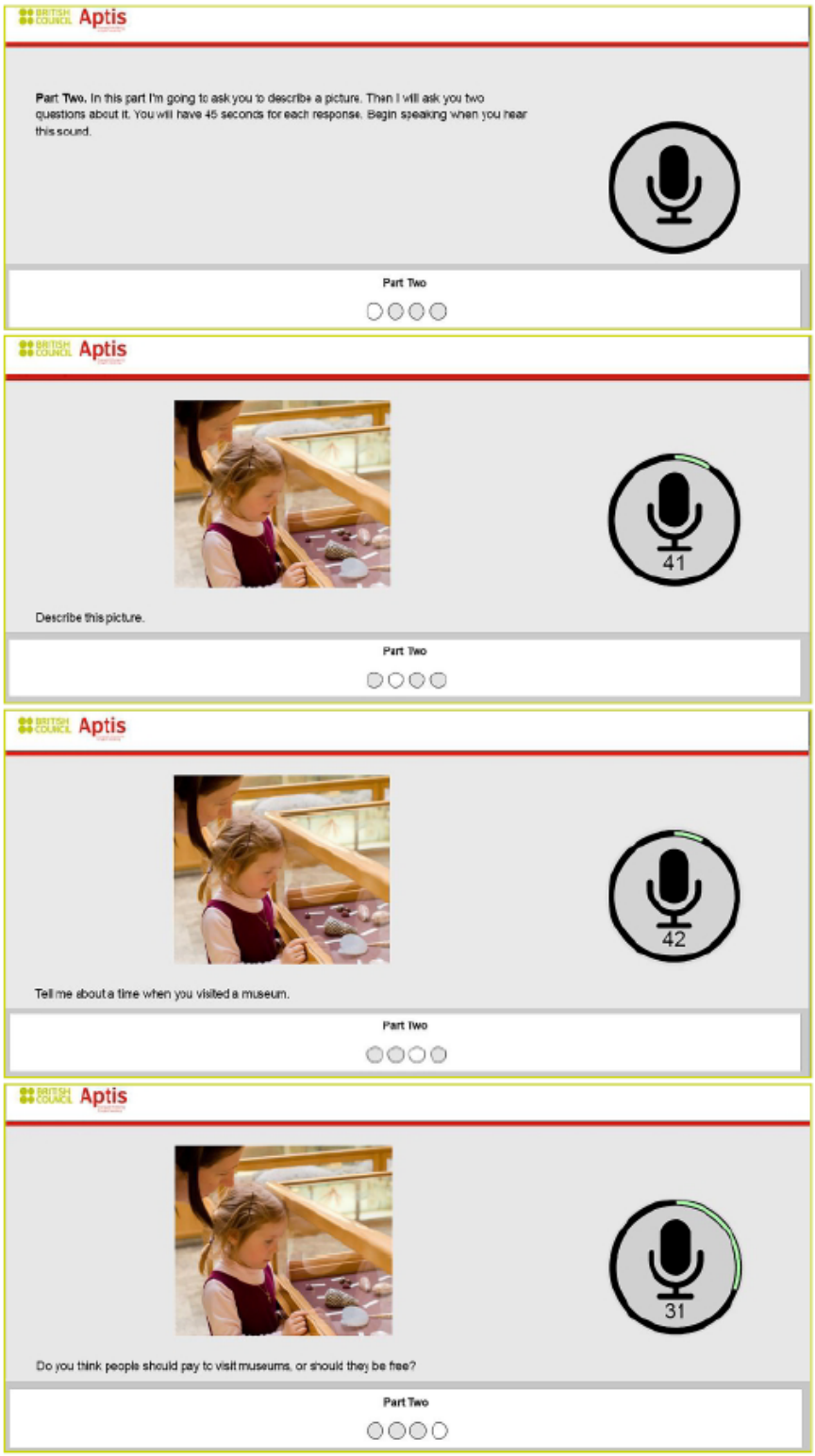
Victorian Institute of Teaching (VIT). (2015). http://www.vit.vic.edu.au Accessed 24 June, 2015.

Wigglesworth, G. (2005). Current approaches to researching second language learner processes. *Annual Review of Applied Linguistics 25*: 90–111.

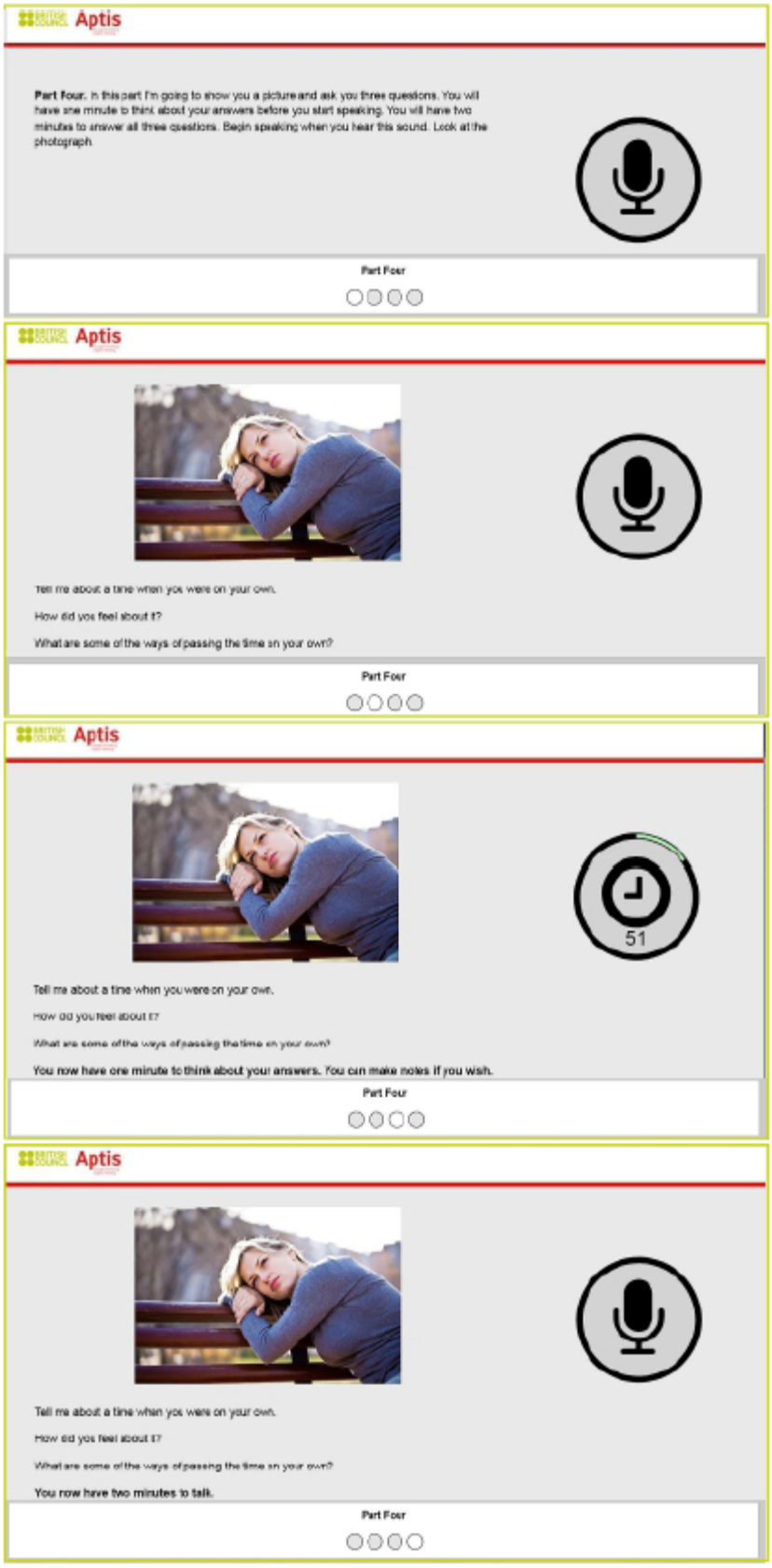
Yoshida-Morise, Y. (1998). The use of communication strategies in LPIs. In R. Young & W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 205–238). Amsterdam: John Benjamins.

APPENDIX A:
Sample speaking questions (screenshots)

Part 2



Part 4



British Council (2018). *Aptis Candidate Guide* https://www.britishcouncil.org/exam/aptis

APPENDIX B:
Background questionnaire (educators)

Name of participant:

1. Gender (please circle): Male / Female / Unspecified
2. Age (please select):

   □ 21 – 30

   □ 31 – 40

   □ 41 – 50

   □ 51 – 60

   □ 61 and over
3. What is your highest qualification?
4. What is your current workplace(s) and role(s)?
5. How many years of experience do you have as a classroom teacher?
6. What type of classroom teaching experience do you have?
   (e.g. primary, secondary, higher education, adult education)
7. In what discipline areas do you possess teaching specialisations?
8. Are you familiar with the English language requirements for teacher registration?
   (please circle): Yes/ No
   If yes, please provide details:
9. Are you familiar with the Aptis test? (please circle): Yes / No
   If yes, please provide details:
10. Have you worked as a teacher in a country other than Australia? (please circle): Yes / No

    If yes:
    Where?
    For how long?
    Type of experience? (e.g. primary, secondary, higher education, adult education)
    Main language(s) of instruction:

APPENDIX C:
Background questionnaire (trainees)

Name of participant:

1. Gender (please circle): Male / Female / Unspecified
2. Age (please select):

   □ 21 – 30

   □ 31 – 40

   □ 41 – 50

   □ 51 – 60

   □ 61 and over
3. Country of birth:
4. What is your first language?
5. In what language did you receive the majority of your pre-university education?
   If more than one, give the approximate number of years for each language:
6. Did you study English at school or university in your home country? (please circle):
   Yes / No
   If yes, for how long?
7. Have you spent time studying in an English-speaking country before coming to Australia? (please circle): Yes / No
   If yes, for how long?
8. How long have you been in Australia?
9. What is your previous degree before commencing your graduate teacher qualifications?
10. In what country did you obtain your previous degree?
11. In your previous degree, what discipline did you major in?
12. Have you worked as a teacher in a country other than Australia? (please circle):
    Yes / No

    If yes:
    Where?
    For how long?
    Type of experience? (e.g. primary, secondary, higher education, adult education)
    Main language(s) of instruction:

APPENDIX D: Verbal report protocol

Introduction

[Researcher provides an overview of the study and session, and describes the four parts of the test]

In this study, we are investigating what test-takers think and do as they complete an Aptis for Teachers Speaking test. To do this, I am going to ask you to complete each of the four parts of the test, and at the end of parts 2, 3 and 4, which relate to teaching, to talk about what was going through your mind as you completed the questions in each of those parts. We call this a 'verbal report' and we will repeat the process 3 times, at the end of each of the three relevant parts of the test.

To begin, I will provide you with some training. I will give you a short practice test, similar to one part of the actual test, with one sample question to try. You will do this practice question under timed conditions, and I will video you while you speak. When you have finished, you will watch the video and as you are watching, you will provide a practice verbal report. I will explain what we want you to do in the verbal report before you start, then I will ask you to start talking and I won't interrupt you or say anything unless I have a question about something you have said. When you have finished, we will discuss again what you need to try to do in the actual verbal report.

After this training, you will be asked to do the real test, and to provide verbal reports at the end of parts 2, 3 and 4 of the test. At the end of the last verbal report, I will ask you some further questions about your test performance, and your thoughts about the test in relation to the kind of speaking you do as a teacher. Do you have any questions about what I have just told you?

Participant training

[Researcher explains procedure]

Start the test when you are ready, and try to complete the question as if you were in a real test situation. [Test-taker completes practice test]

I am interested in learning what is going through your mind as you were responding to the test question, so I would like you now to provide a verbal report. By 'verbal report,' I mean that I want you to try to remember and to say out loud everything that came into your mind as you were completing the question. I want you to talk as much as possible about what you remember thinking as you were speaking in the test. It is important that you keep talking. If you are silent for any period of time, I will remind you to keep talking. Do you understand what I am asking you to do? Do you have any questions?

Now you will start to watch the video clip of you completing the practice test. As I said before, as you watch the video, I'd like you to say what you were thinking as you were completing the test question. I am interested in what was in your mind as early as when you began working on the question up until the time when you finished speaking. You can pause and rewind the video as much as you want. So, when you want to say something about what you were thinking, pause the video and talk. I want you to talk as much as possible. If I have a question about what you were thinking or what you have said, then I will pause and ask you to clarify that part of the video. Otherwise, I'll remain silent and leave you to think and talk. Please start the video when you are ready, and start talking about what you can remember thinking as you did the task.

Completion of test and verbal report

I am now going to ask you to complete the actual speaking test. You will do each of the four parts of the test under the same timed conditions that apply in the real test, but we will pause at the end of parts 2, 3 and 4 so that you can provide a verbal report.
[Researcher repeats the description of the four parts of the test]

Please give your best performance as if you were really in a test situation. I am going to video-record your performance during each part of the test. When you have finished the task, I will ask you to watch the video, to help you remember what you were thinking, while you provide a verbal report.
[Test-taker completes test parts 1 and 2]

I want you to try to remember and to say out loud everything that came into your mind as you were completing the questions in part 2 only. I want you to talk as much as possible about what you remember thinking as you were speaking in response to each question. Try to keep talking. Don't think about what you should have said or done, I don't want you to re-do the questions, just try to remember and say out loud what was going through your mind as you were speaking. I am interested in what was in your mind as early as when you began working on the first question up until the time when you finished speaking. If you are silent for any period of time, I will remind you to keep talking. Do you have any questions?

Now you will start to watch the video clip of you performing the task. As you watch the video, when you want to say something about what you were thinking, pause the video and talk. You can rewind the video if you want to. I want you to talk as much as possible. I will pause the video if I have any questions. Otherwise, I'll remain silent and leave you to think and talk. Please start the video when you are ready, and start talking about what you can remember thinking as you did the task.

[Repeat on completion of test Part 3]

[Repeat on completion of test Part 4]

APPENDIX E:
Interview protocol (educators)

Thank you for taking the time to participate in this study. The purpose of this study is to investigate the speaking demands faced by classroom teachers and to see how well the Aptis test reflects these demands. You have just reviewed the test, and now we are interested in hearing about how you think the test tasks compare to the types of actual speaking demands EAL teachers are expected to be able to handle when they begin teaching in the classroom.

- Based on your experience, what kinds of oral communication demands do teachers face in classrooms in Australia? [seek elaboration relevant to discipline and school level]
- To what extent does your course focus on preparing trainee teachers for the oral communication demands they face in classrooms in Australia?
    - What kinds of oral communication skills do you focus on developing in trainee teachers during your course?
    - How are these skills assessed within your course?
- In your experience are there any particular oral communication challenges faced by trainee teachers from EAL backgrounds?
    - How do these sorts of challenges impact your teaching?
    - Are there additional resources available for EAL trainees to support their oral skill development?
- Having reviewed the Aptis for Teachers Speaking test materials, how do you perceive the relevance of the test to the oral communication demands of classrooms in Australia?
    - Do you think the Aptis speaking tasks tap into the skills needed to teach in classrooms in Australia across diverse subjects and school levels?
    - Do you think the Aptis Speaking tasks tap into the particular challenges faced by EAL trainee teachers?
    - Do you think the kind of speaking in the Aptis test is related to the kind of oral communication you are focusing on teaching in your course? How is it similar/different?
- In terms of the English language test requirements for teacher registration, how does the Aptis Speaking test compare with the existing tests approved for this purpose?
  [prompt: IELTS, ISLPR, PEAT-specific purpose]

Thank you for your time.

APPENDIX F:
Interview protocol (trainees)

Thank you for taking the time to participate in this study. The purpose of this study is to investigate the speaking demands faced by classroom teachers and to see how well the Aptis test reflects these demands. You have just completed the test and verbal report procedures, and now we are interested in hearing about how you think the test tasks compare to the kind of speaking you actually do when you are teaching in the classroom.

- Based on your pre-service teaching experience, what kinds of speaking skills are needed for teaching in Australian classrooms?
- What are the main oral communication challenges for trainee teachers from EAL backgrounds?
- What kinds of oral communication skills do you focus on developing in your course?
    - What kinds of speaking skills does your course help you to develop?
    - How is this done? [additional prompts: part of the course, additional resources/support, self-select/referred by lecturer, optional/compulsory]
- How are oral communication skills assessed in your course?
- How well do you think your graduate course is preparing you for the speaking demands you face in the classroom?
- Do you think the kind of speaking you did for the Aptis test was related to the kind of oral communication you are focusing on in your course? How is it similar/different?
- Does the Aptis test tap into the kinds of speaking skills you will need in the classroom? [probe for discipline, level, EAL challenges, and diverse areas e.g. remote, indigenous, special learning needs]
- Are you planning to apply for registration when you graduate? Do you have the required IELTS score for board registration? If not, do you think you will get the required score?
- How does the Aptis Speaking test compare with IELTS in terms of difficulty and relevance to spoken language use in classrooms?
  [if aware of other approved tests i.e. ISLPR, PEAT, how does Aptis compare with these in terms of difficulty and relevance to classroom speaking demands]

Thank you for your time.

APPENDIX G: Verbal report coding categories with examples

Transcription conventions: x: incomprehensible syllable; comma: short pause

Category	Definition	Examples [Participant ID]
ORIENTATION TO TASK		
Generating ideas	Generating ideas for a response; includes drawing on professional knowledge, training, experience.	Just thinking about my time as a teacher, and what, the groupwork or the projects that we've done together as a class in groups and what were the good things that happened out of it. [T-8] I was thinking about um probably, I should just say something about, about the like the teaching method or the pedagogy that I have learned. [T-10]
Selecting ideas	Making choices from ideas generated for the response; considering and selecting options.	Because as a teacher you help students individually a lot. So there is a lot of moment going on. So I pick an recent one to start. [T-2] So a lot of things came to mind because I've done group projects with students, and ah, I was trying to pick and choose which one to talk about. [T-9] Yeah, so for the first question, do you remember one time that you're doing the lesson plan, um, yeah. So, I just talk about, ah, the time I remember, the first time. So, cos I didn't do well, that's why I still remember it. [T-6]
Reflecting on the question	Reflecting on the question/task; recalling the question; considering what the question requires of them; commenting on question difficulty.	Um, when I heard the question, I actually think about like, when, um, when exactly is that? Like, individually it could mean the first time, or it could mean in the recently [T-2] That's when I re-read the question and it said 'why is it good for young children to play sports?' [T-8] She say 'try to describe the picture', right? And then I'm actually not sure about the question because 'describe the picture' means to describe the things in, the objects in the picture. Or imagine what actually happened there. [T-6] I think the 'why' questions are always the most difficult questions, I guess. Because, there may be lots of reasons and you have to pick one. [T-4]
Linking questions and image (Part 4)	Establishing relationship of questions to picture stimuli.	I wouldn't have known what it was until I read the question. It was just a folder or something. It wasn't obvious what it was so I don't think the picture would've made a difference in this question. [T-1] I just first saw the picture, and then, ah heard the question, and then I tried to find out the relation between the picture and the question. And it took a little bit ah, of thinking time, because they were not closely related. [T-4] And then I realised, oh, I don't actually need to look at the picture, and then I just read the questions. It's a bit distracting. [T-10]
COMMUNICATION		
Avoiding	Using repetition, message substitution or abandonment to avoid linguistic difficulty	And there's one thing I wanna add but there's not a word in English I could describe it. [T-2] And then I start describing it, because I was too scared to think about um, what sentence I should use to compare. [T-2]
Approximating	Using lexical or grammatical substitution as a replacement for an unknown word/structure.	I can't think of the name of that sport, that's why I was, you know, using 'martial art'. What is it, judo? [T-7] Yeah, I was, um, when I said 'ethnic backgrounds', so I was sort of looking for different words for 'ethnic', and that's just what I was thinking. [T-3]
Slowing delivery to gain time	Slowing down speech rate or pausing while thinking of what to say or how to formulate the message; repeating the question to gain planning time; pausing to reflect on memories.	So I was just pausing and thinking of ways to link imagination into that. [T-3] So I repeat the questions again, so, ah, help me organise what I'm going to say. So I repeat the question again. [T-2] And the hesitation was also where I was thinking of an article that talks directly about this, and I was going 'what does it say?'. [T-9]

Category	Definition	Examples [Participant ID]
Expanding the message to fill time	Extending response (ideas, details, embellishments) to fill remaining time allowance for speaking.	And, ah, to fill up the time, you can see in the end I try to say something, you know, to make up the time, not quite relevant to a lot of things, you know. [T-7] And then, since I had time and I couldn't think of, ah, anything else in my part, then I focused on the picture, and just to fill the time, I said, 'oh, that teacher is doing that'. [T-6]
Expanding the message to improve response	Elaborating or adding to response (details, embellishments) to improve quality of the response.	I was trying to expand on my answer…just to make my answer a little more complex. [T-3] I realised I wasn't just talking about children, I was talking about students in general. Yeah, um, then I was like okay, when you're young it's easier to grasp things as opposed to when you're older. [T-8]
Word choice	Exercising careful (more generic) word choice with attention to accuracy or sensitivity in relation to interpretation of photograph.	Then I changed my mind and said, 'okay, don't tell them teachers, just tell them adults'. [T-4] So I say things like 'one person' and 'people' instead of 'children' and 'teacher'. [T-9]
Relating task to experience, knowledge	Drawing on background knowledge and professional experience (training and classroom) in formulating response.	I used a lot of words, um I just roll out of my mouth because we've been saying it during the placement. [T-2] I had no idea what to say. I was, yeah, sort of 'oh no', trying to sort of remember whether I learned this back in my degree or whether I heard it. [T-3]
Organising and structuring response	Sequencing and linking ideas while speaking; structuring (streamlining) response to complete within time allowance.	Cos like I sort of like structuring like what should I say. So after saying the details thing I can actually just go back and look at the structure and the colour. [T-2] I was trying to connect that with the first statement that I said, and I was trying to link it with the second picture. So the form of comparison. [T-3] I kind of combine the second question with the third one [T-6]
Paraphrasing to avoid repetition	Paraphrasing to avoid repetition of part of response already given.	I realised that I touched on that in my previous response, so I was a little bit unsure of, okay, do I just repeat it or do I rephrase it. [T-8]
Speeding delivery	Increasing the rate of speech to complete full response within time allowance.	I can say more about those, and you can also hear it when I talk, I talk faster here. [T-1]
Referring to notes (Part 4)	Referring to notes to organise response and to recall ideas for response.	Because I was starting to lose track of what I was saying, because I had dotted points on my notes, and just follow through that. [T-3] And then I looked down, I saw I write down 'inquiry' and 'website' and so that's what I'm gonna stick on with. [T-2]
COGNITION		
Engaging with stimulus images	Attending to stimulus images; establishing what is depicted in the photographs; interpreting the images; scanning images for details.	I'm like looking at the timer and I'm like, okay, I should say more, and I look back at the picture and see if I can pick out things to say. [T-1] The first part, 'describe this picture', I just quickly, ah, saw the picture and started to physically describe the things that I, ah, I mean I just focus, the number of people and tried to guess which one is the teacher. [T-4] I was trying to count like how many people there are there in the picture, but then at first I was gonna say seven, but then I saw obviously there's a teacher sitting there, so I just change to six students plus a teacher. [T-10]
Using imagery	Drawing on mental images of memories of classroom experiences as stimuli for language production, supporting generation and recall of ideas/details.	As I was talk I had like a video playing in my head of what happened last time. So I was trying to describe like what was in my head, yeah. [T-3] When I describing those, I actually can (see) this child being off task and talking about another party, xx party or something and they're going on, and it's a visual thing that I actually see, so those little video that's playing on my mind. [T-2]

Category	Definition	Examples [Participant ID]
Using L1	Thinking in L1 to generate and formulate ideas; taking notes in L1 (Part 4).	My mind was in between kind of like Mandarin and in between English. So I use my Mandarin kind of thinking to describe English. [T-2] When it comes to this question, I want to talk about like how important sunshine is to the human body, and that area is what I learn in Mandarin. So when I structuring the dot points, they are in Mandarin. [T-2]
Notetaking (Part 4)	Writing notes during preparation time to plan and organise response, and to use later to aid recall of ideas and organisation of response while speaking.	So basically I did 'one two three' for the three questions, and then I wrote on them. So the first question, I didn't write the full question, but I just wrote 'detailed lesson plan' and I started to write a time that I did…So it was just things that would remind me of what I should say. [T-1] When I was planning in the one minute, um, I just break it down because three questions will break it down into three parts, and each part I just write down maybe four keywords, and then yeah, just in a logical way so that later on I can skip things if I want. [T-6] I realised I had one minute, which isn't enough to write out whole sentences, so I was trying my best to write down keywords to guide my thinking while I'm doing the talking. [T-9]
METACOGNITION		
Evaluating performance	Evaluating quality of language production, response to question, or whole task performance.	I sounded quite vague, I realised, in that answer. [T-9] So I think that I did really badly in this question and also the question afterwards. [T-5]
Identifying communication issues	Identifying communication difficulty or problem with quality of language production.	I'm totally out of things to say. I had no idea what to say after that. [T-1] But for all of a sudden, how should I name that term? My mind just went blank. [T-5] I was trying to say there are students like of different races and I think, oh, probably it's not appropriate thing to say, so I just stop. [T-10]
Monitoring time	Checking the on-screen timer to monitor time allowance remaining.	Then, I was constantly checking the time, too, so, um I saw that I had a lot of time, so I started to talk about the things that I saw on the table. [T-4] And, um, in terms of advantages, I wanted to share more on the different aspects of engagement, and then I looked at the time going, 'just the word engagement should be sufficient' and I moved on. [T-9]
Planning	Making advance decisions about organisation and rhetorical structure of response.	I want to decide whether I gonna tell the story long or short. If it's long version, there's not enough time, and I'm probably gonna tell a different story. [T-2] I was trying to think of…whether I should summarise each of the picture first, or whether I should start comparing them. So I was like, hm, which technique should I go? [T-3] I try to answer the three questions one by one, but don't make them sound too separately, but when I'm answering the first question, I will start to answer the second question, and by connecting the language. [T-7]
Self-correcting linguistic error	Monitoring language production and self-correcting grammatical and lexical errors.	I wasn't quite sure whether it should be 'arts' or 'art'. Yeah, just like the preposition, I'm not sure if it's called preposition, I tried to remember to add it on. [T-10] I was try to um, how to say, double check whether I'm talking it correctly. Speaking to it correctly. So, sometimes I might say things again in a different way. [T-7]
Reflecting on orientation	Considering orientation to task in relation to envisioned assessment criteria; setting performance goals.	So if, like, I'm trying to thinking, should I use some fancy words to get me a higher score? [T-2] I was wondering, are you assessing my speaking performance, or do you really want me to describe what the picture is? [T-5] I cared about the words that I was saying. 'Cos I treated it like a real test, like displaying your range of vocabulary that you have I learnt is an important thing. [T-10]

Category	Definition	Examples [Participant ID]
Reflecting on picture stimulus	Making evaluative reflections on the photos; includes negative reactions to accessibility of content, and value as stimulus for ideas.	When I look at those two pictures, I don't find them very comparable. Like, they can't really be compared together. Because the first one looks like a classroom and the second one looks like a computer lab. So if you give me two classroom-like pictures, it will make more sense. [T-5] But, sometimes planning doesn't require a computer. You just take a pen and paper and some things came to your mind and you just take notes to a paper and so, that's why it wasn't the, ah, um, it limited my thinking. [T-4] When I first saw those pictures, sports isn't really an area of my interest…so my first thought wasn't very positive [T-8]
Reacting to test format	Reacting to test format; includes presentation of questions (Task 4) and alert ('beep') sounding the onset of speaking time.	I'm not so sure if it's only two questions or three questions. Cos I only see two question marks. [T-5] I think like the beep sound comes too soon after the question, so I, I probably will prefer it after like one or two seconds, then I hear the beep sound, then I can start. [T-10] I straight to looking at the questions to make sure I didn't miss anything because I thought maybe the questions would disappear after she finishes. So I didn't look at the picture at all, I kinda looking straight to the questions, writing them down. [T-1]
AFFECT		
Justifying performance	Appealing to contextual, situational or affective factors to explain quality of performance; includes test format and stimulus e.g. preparation time, presentation of questions, familiarity of topics (such as sport, imagination).	Maybe if I wasn't waiting for that bell or something, if it wasn't test conditions or something, I'd probably be able to speak more. [T-1] Then I got distracted, and so yeah, couldn't speak fluently. [T-2] So that's why I didn't have much to say, only because it isn't an area of my interest. [T-8] I didn't see the first part of the question, so I didn't write anything about the last time that I wrote a detailed lesson plan. I don't think I actually talked much about it. [T-10] If I have more time, like give me some more time, like give me some time to write down the answers, then I can probably elaborate it better. [T-5]
Reflecting on affect	Reflecting on mood or emotional state, especially as a response to test situation (stress, anxiety, confidence).	I hope this this won't be a real thing because it's really stressing me out. Yeah. By doing it in front of the computer. [T-5] I think this one I'm just not ready. Like, I haven't think properly to answer that, so I'm bit nervous…so when I see the timer, I just get nervous. Yeah. [T-6] Probably using words like stimulus, I was feeling quite good, because it sounds quite good. [T-10]

British Council Assessment Research Awards and Grants

If you're involved or work in research into assessment, then the British Council Assessment Research Awards and Grants might interest you.

These awards recognise achievement and innovation within the field of language assessment and form part of the British Council's extensive support of research activities across the world.

EXAMINING TEST-TAKER PROCESSES AND STRATEGIES AND STAKEHOLDER PERCEPTIONS OF RELEVANCE OF THE APTIS FOR TEACHERS SPEAKING TEST IN THE AUSTRALIAN CONTEXT

AR-G/2020/1

Sally O'Hagan
Kellie Frost
The University of Melbourne

ARAGs RESEARCH REPORTS ONLINE

ISSN 2057-5203

© British Council 2020
The British Council is the United Kingdom's international organisation for cultural relations and educational opportunities.