

**EVALUATING APTIS TESTS WITH TEACHER TRAINING
STUDENTS AND IN-SERVICE PRIMARY SCHOOL
TEACHERS IN DURBAN, SOUTH AFRICA**

AR-G/2015/002

**Julie Douglas, Talk English
Charlotte Mbali, University of KwaZulu-Natal**

ABSTRACT

The British Council's Aptis test was used with three groups of candidates in South Africa to investigate the potential value of introducing the test in that specific sociocultural context. The candidates were first-year and fourth-year Bachelor of Education (BEd) teacher trainees, and in-service primary school teachers in Durban, South Africa.

The 68 participants in the study took the test on computer. Their views on the experience were then gathered by means of a post-test questionnaire, focus groups and individual interviews. Their Aptis scores were analysed to determine the profiles of scores for each subgroup, and were correlated with other English assessments that they had taken during their studies.

Overall, the participants' attitudes were very positive, both towards the test and towards the mode of delivery – even candidates with little or no computer familiarity found the experience positive. However, a number commented on problems of lack of time. The comparison with other measures of proficiency indicated that the local tests are, to some extent, in line with Aptis, but there is less correlation than would be ideal. This suggests that some aspects of the local tests may merit further investigation.

At the same time, there are also grounds for arguing that certain features of the Aptis test may need to be reconsidered. Two aspects in particular emerged as potentially problematic. First, there were technical difficulties arising from the delivery by computer: low bandwidth and inconsistent internet connection made the process cumbersome at times; and lack of experience in using computers made it hard for some of the candidates to complete the tasks, which may have distorted, in some respects, the picture of their proficiency given by the test. Secondly, the test included reference to a number of British sociocultural constructs and terms rarely, or differently, used in South Africa, which may have confused some candidates.

We concluded that Aptis could be effectively used in South Africa provided that problems of bandwidth and culturally unfamiliar stimulus items are resolved.

Authors

Julie Douglas

Originally a high school English teacher in Nottinghamshire, Julie Douglas moved to South Africa in 1976. After completing a Master of Education cum laude in 2000, she lectured and worked on several education research projects. She became a senior researcher for three years with the Education Policy Unit (Natal) and the Centre for Education Policy Development. From 2005, Julie worked as an education consultant specialising in the evaluation of social, education and language projects. She has evaluated projects from most major English language non-government organisations (NGOs) in South Africa and several major South African school improvement projects. From 2006 to 2012 she monitored and evaluated the Toyota Teach Primary Schools Project in KwaZulu-Natal.

In 2005, after completing a CELTA, Julie started Talk English with Sarah Ralfe, and a group of volunteer language teachers offering free adults' and children's English classes, including medical English for doctors and nurses. Students include South African isiZulu speakers and refugees and immigrants from African countries, Eastern Europe and the East.

Julie currently consults in evaluation and language projects for several clients while completing a PhD through the University of KwaZulu-Natal, studying the oral English praxis of township and rural teachers in the Durban area, South Africa. Charlotte Mbali is her PhD supervisor.

Charlotte Mbali

Charlotte Mbali worked at the University of KwaZulu-Natal, South Africa for 16 years, running the Centre for Higher Education Studies. Now semi-retired, she supervises some doctoral candidates, including Julie Douglas. Her own PhD at the Institute of Education, University of London, investigated writing assessment of near-tertiary students and, during that time in the 1980s, she was one of the early members of the special interest group on Assessment (TEASIG) at IATEFL. Her career in TEFL teaching spans 40 years, culminating in short intensive contracts such as pre-sessionals for foreign students in UK.

Charlotte has also written graded readers, a teenage novel and BBC English radio programmes for Africa. She reviews for some journals and presents conference papers in collaboration with her students. Having resumed TEASIG membership in 2011 in order to follow European test development aligned to CEFR bands, she realized that the Aptis test would be useful to Julie's research into foundation phase English teaching in local Black schools here in KwaZulu-Natal.

Acknowledgements

The authors would like to thank Sarah Ralfe and Geoff Thompson for their help in preparing this report.

CONTENTS

1. INTRODUCTION	5
2. RELEVANT RESEARCH BACKGROUND	6
2.1. English in South Africa	6
2.2. Aptis relevance for South Africa	6
2.3. Theoretical background	7
2.3.1. The test system	7
2.3.2. The test-taker	8
3. METHODS	9
3.1. Overview	9
3.2. Piloting	9
3.3. The main test phase	10
3.3.1. Hotel candidates	10
3.3.2. Teacher trainees and teacher candidates	10
3.3.3. Technical issues	11
3.3.4. Data collection and analysis	11
4. FINDINGS	12
4.1. The questionnaire, focus group interviews and individual interviews	12
4.1.1. Pre-test tutorial and sound and recording test	12
4.1.2. The test process	12
4.1.3. Timing	14
4.1.4. Test content: Individual components of the test	15
4.1.5. Perceptions	16
4.1.6. Using the test in South Africa	17
4.1.7. Improving the Aptis experience	17
4.1.8. Sociocultural knowledge	18
4.2. Test results and comparison	19
4.2.1. Summary of test scores	19
4.2.2. Comparison of test scores with other measures of English language ability	20
5. DISCUSSION	23
5.1. Test-takers' views of Aptis	23
5.2. Correlations with other measures of language ability	23
5.3. Accessibility of Aptis	24
5.4. Appropriateness of Aptis	25
6. CONCLUSION	26
REFERENCES	27
Appendix 1: Aptis questionnaire for TT candidates	29
Appendix 2: Aptis focus group interview	32
Appendix 3: Guide questions for individual interview	33
Appendix 4: Comparison of Aptis results for the three groups: one-way ANOVA	34
LIST OF TABLES	
Table 1: Questionnaire responses for time	14
Table 2: Reactions to the five test components	15
Table 3: Summary of descriptive statistics for all groups	19
Table 4: Correlations with other tests at the TT institute	21

1. INTRODUCTION

English is well established and widely used throughout South Africa among speakers from many linguistic and ethnic backgrounds, despite it being a home language for less than 10% of the population (Stats SA, 2011). As the primary language of government, business and commerce, English is considered to be the language of prestige and power. It is the preferred medium of instruction in most South African schools and tertiary institutions. In such a context, reliable, accurate assessment of English proficiency is clearly of fundamental importance.

This motivated us to investigate the potential value of introducing the British Council's Aptis Test in South Africa. One major advantage of this test, in comparison with current measurement instruments in South Africa, is that the grading of proficiency is based on the Common European Framework of Reference for Languages (CEFR). Despite the clear benefits of calibrating local testing with an internationally recognised benchmarking system, the CEFR is neither widely known in South Africa nor applied in the domain of language testing. Therefore, our research aimed, among other goals, to raise awareness in South Africa about international language testing criteria.

Aptis claims to be a "major new English language testing system, based on the joint concepts of flexibility and accessibility" (O'Sullivan, 2012a, p 1). The 'flexibility' of Aptis makes it attractive in principle for South African use since, according to O'Sullivan (2012a, p 5), "The British Council will...localise (i.e. make appropriate to the particular context and domain of language uses) the test, thus ensuring it will meet the expectations and requirements of the client while maintaining its internal integrity (from a content and a measurement perspective)". At the same time, the quality of 'accessibility' raises potentially problematic issues, since this is dependent on delivery by computer. This project probes both the local applicability of the test content and the accessibility of the computer delivery.

Thus, the aim of researching the Aptis test in South Africa was to trial the developed test in a selected range of local contexts with EL2 speaking South Africans and to investigate questions about its suitability, validity, replicability and potential. This was done by organising and delivering the Aptis test itself, observing the implementation process, and gathering feedback from the participants who took the test. We felt that an in-depth focus on one particular country would complement the geographically broader survey of test-takers reported in O'Sullivan (2012b).

We also compared Aptis results as a predictor of proficiency with other direct and indirect tests of candidates' English language ability. This step had two aims: (1) to investigate how far Aptis seems to be in line with other local ways of assessing the English language skills of student-teachers and teachers in a context where English is not the L1 but a lingua franca and the preferred medium of education; and (2) to demonstrate to the language teaching and testing community in South Africa, the viability and potential value of aligning with global developments in the field, particularly the CEFR.

The specific research questions that guided our research were:

1. To what extent is the Aptis test seen by test-takers in South Africa as an acceptable way of measuring their language proficiency?
2. To what extent do test-takers' results on the Aptis test correlate with the results of other measures of their English language proficiency?
3. To what extent is Aptis "accessible" within the practical constraints of South Africa, a mix of urban centres and vast deep rural areas?
4. To what extent is the Aptis test "appropriate to the particular context and domain of language use" in South Africa?

For each of these questions, we planned to investigate in more depth the factors which seemed to affect the outcomes.

2. RESEARCH BACKGROUND

2.1. English in South Africa

The history of English in South Africa has been well documented (Bekker, 2008). For our purposes, it is enough to note that English in South Africa has been politicised from the outset and, therefore, compromised by power relations, with social and emotional connotations that depend on the position of either coloniser or colonised (Mufwene, 2002).

English is one of the 11 official languages in South Africa (SA languages.com), spoken as a first language in households by 9.6% of the population. It is the *lingua franca* (Seidlhofer, 2005, p 339) for international communication, commerce and the professions, and the preferred medium of instruction in higher and secondary education for most students. South Africa is not recognised as an English-speaking country for UK citizenship purposes and South Africans are required to achieve an English speaking and listening qualification at the B1 level of the CEFR (*The South African*, 2014).

After apartheid, despite official attempts to establish African home languages at the core of foundation phase in primary education, there has been growing pressure from township and rural parents (who have the democratic right to choose the medium of instruction for their child) for “straight to English” education. Also, formerly Afrikaans-medium universities have all added English-medium courses. This means there is more pressure on teachers whose home language is not English to improve their English for teaching purposes. Primary school teachers must teach children enough English in the first three years of schooling for them to understand, conceptualise, read, write and learn in it during the rest of their education.

Traditionally, some years ago, prospective teachers had to prove their oral competence in English through an interview or test (Spingies, 1991). This tradition was lost in the 1990s when South African teacher training colleges were amalgamated into the universities. The standard of South African education is rated very low in international comparative testing (see e.g. Howie et al., 2011; Reddy, 2006), and increasing numbers of educationists believe better training of foundation phase teachers is needed (Dale-Jones, 2013). We were swayed by this background to focus our research on the education sector, with South African primary school teachers and Bachelor of Education (BEd) students as our participants.

2.2 Aptis relevance for South Africa

We were initially attracted to Aptis for its potential to calibrate South African candidates and local tests with CEFR. International high stakes English tests and exams used in South Africa, such as IELTS or TOEFL, are costly, conducted externally by non-South African organisations, and often require a one-on-one examiner–candidate situation for the oral component, so in particular, a speaking test that is easier to administer is worth trying out.

Local institutions tend to construct their own language tests as assessment of their courses rather than for diagnostic purposes. General dissatisfaction with the Senior Certificate as the matriculation gate to higher education has led to the development of alternative assessments such as the National Benchmarking of 2009 (Cliff & Hanslo, 2009) and a suite of tests developed by the Nelson Mandela Metropolitan University. The CEFR for languages is still relatively unknown in South Africa and we could identify no local tests graded on CEFR levels.

The use of online and computer technology was also an advantage, as this seemed to promise a time- and cost-efficient way of language testing. Components could be tailored to specific needs and delivered without external supervision or invigilation. A quick turnaround of results was promised. The online delivery through SecureClient appeared quick, effective and robust. Importantly, given the South African context of large distances between developed centres, Aptis could be implemented immediately anywhere in the country with online access, urban or rural. Aptis promised an inexpensive way of benchmarking English skills for workplace, academic or personal use.

2.3 Theoretical background

O'Sullivan (2012a, p 1) points out that the theoretical background for the development of Aptis was provided by the socio-cognitive model set out in Weir (2005), with modifications proposed by O'Sullivan (2011) and supported by O'Sullivan and Weir (2011). The model of validation that was applied details three elements that determine performance. Prior to testing, the *test system* with its linguistic demands and performance parameters, and the *test takers*, with their individual characteristics, cognitive processes and language resources, combine to delimit performance. Post-testing, the *scoring system*, with its theoretical fit, accuracy and examiner decisions, measures candidates' performance. Since this particular research report on South African candidates covers mainly the first two of the elements – the test system and the test-taker – our review focuses on these in relation to Aptis and our project.

2.3.1 The test system

The test system depends on the supporting language model, since criteria for both model and language must match. If language is conceived as a pattern of grammar and vocabulary, then a multiple-choice question format on grammar and vocabulary can be said to be appropriate for the test. Morrow (1979) argued against such test formats, and for test tasks to be authentic, mirroring communication in the real world. Communicative language teaching theories (Canale & Swain, 1980) paid greater attention to features of discourse, and to the social context of communication, adding 'context' through verbal or picture cues.

However, a drawback of context-laden trigger material is its potential to be culturally or regionally alien and, thereby, disadvantage some candidates. There may be a dissonance between authentic trigger material appropriate to the task and a test useful for candidates around the world. This may impact in various ways on the design of the test.

Underpinning Aptis development was the principle that 'greater topic understanding tends to result in superior performance', from which the Aptis development team determined that topics would need to be known by candidates (O'Sullivan, 2011, p 8). However, cultural understandings in day-to-day life may differ greatly: a situation which is perfectly familiar in one culture may be odd or completely unknown in another. This is the case, for instance, with the concept of 'book clubs' (referred to in one version of the Aptis test that we used in our study) – the majority of South Africans have little or no idea what these are.

Another aspect which may be problematic is the selection of the language to be tested. Kirkpatrick (2010) describes emerging and dynamic forms of English as 'World Englishes' with their tendency to assimilate local meanings and commandeer words from other local languages. Given this situation, Canagarajah and Ben Said (2011, p 391) point out the anachronistic concept of a 'native speaker'. English as a Lingua Franca (ELF), according to Kirkpatrick and Deterding (2011, p 382) citing Jenkins (2007), is the most common use of English in the world for international communication and is characterised by its dynamic, rapidly changing nature. This seems to be at odds with the idea of an international test which is the same wherever it is administered. For those tests, such as IELTS and TOEFL, which assess a candidate's readiness to work or study in Anglophone countries (primarily the UK, Australia and the USA), there are arguments for focussing on the varieties used in those countries, that is, the Inner Circle (Kachru, 1985). On the other hand, these arguments do not apply with the same force to tests like Aptis which are intended to be used by companies and organisations

to benchmark the English language levels of their employees, potential employees, students or teachers – that is, for ‘local’ applications. In such cases, the desirable qualities of consistency and comparability of assessment across all administrations of the test may clash with the need to take into account the localised characteristics of ELF in particular areas.

This emerges most saliently in decisions on the vocabulary to be tested. In designing the Aptis test (O’Sullivan, 2012a, p 6), Vocabprofile software such as ‘Range’ that ranks words according to frequency in a learner corpus (Nation, 2012) was used to establish the items that should be tested. However, lexical software works by identifying broad patterns of occurrence in large amounts of data, which may render insignificant the smaller patterns from atypical populations, such as the group who speak colloquial South African English. Words may have different meanings, and meanings may be expressed in different ways. To take just one example, the word ‘yard’ is likely to be in an international learner corpus of the most frequent 2000 words, but South Africans (like Americans) understand ‘yard’ to mean a garden rather than an enclosed area paved with stone or concrete as understood in Britain.

A further potential complication for the test situation might arise when communication is via a machine and not a person, as with the computerised Aptis. This is clearly very cost-effective, but it runs the risk of disadvantaging candidates who are not familiar with computer-mediated communication. In addition, it depends on the availability of the required infrastructure – not only equipment but a reliable internet system and constant power supply.

The implications of these factors for Aptis will be discussed below.

2.3.2 The test-taker

The second element of the socio-cognitive framework, the test taker, refers to ‘Individual – physical, psychological, experiential’ and ‘Cognitive – processes, resources’. A considerable body of research into what personal resources constitute a ‘good language learner’ (Rubin, 1975; Skehan, 1989; Norton & Toohey, 2001) includes speculation about factors such as age, gender, whether introvert or extrovert, risk-taker or perfectionist, the degree of assimilation by the candidate of the culture of the target language, etc. However, qualities of good language learner are not necessarily transferred to the test-taking situation, which carries its own constraints. Mouti (2008) cites Weir (2005) on test-taker characteristics with potential to affect test performance which are categorised as physical, psychological and experiential. The physical can include short and long-term disability. The psychological includes personality qualities but also the emotional state of the candidate which can particularly influence motivation and concentration. Experiential includes education, examination and communication experience, and country of residence.

Of these characteristics, there are several which merit attention in regard to the administration of Aptis in the context of South Africa. The effect on motivation of certain practical constraints associated with computer-based administration of tests has been little researched, but can be argued to be a potentially important factor. The effect of computer familiarity on candidates’ performance has been extensively researched (e.g. Taylor et al., 1999; Weir et al., 2007), and is also a factor that is certainly relevant in the South African context. With regard to experiential factors and cognitive resources, if a test demands memory of an experience or scene that a candidate has never encountered, it can be queried whether what is being tested is language ability or experience. Another contentious point is whether cultural aspects of the target language are requisite ‘cognitive resources’ that can be regarded as within the remit of a testing instrument.

As with the test system factors noted above, these issues will be addressed below.

3. METHODS

3.1 Overview

Our original plan was to include a broader sample of participants: not only participants working in, or training for, the field of education but also workers and trainees in the hospitality industry, on the grounds that their jobs require them to have a degree of language proficiency, particularly in spoken English. However, practical difficulties meant that, although some participants from the latter group were included in the pilot stage, we were unable to gather sufficient relevant data for the group in the main phase of the study. Therefore, we amended our plans and focused only on the participants from the educational sector. The implications of the practical problems we encountered will be explored in Section 5.

The project consisted of two stages: piloting and main data collection. The piloting involved two kinds of candidates: participants from the hospitality industry and from education. The latter consisted of students on a Bachelor of Education program at a major teacher training institute in Durban preparing them to work as primary school teachers. The main data collection involved trainees from the institute, and a group of primary school teachers working in a township in Greater Durban, north of Durban.

3.2 Piloting

To prepare for the Aptis testing and gain some insights into the test process and experience, the two researchers took a demonstration Aptis test on their personal computers. Tests and other instruments were then piloted with five candidates at a hotel and 10 candidates at the teacher training (TT) institute. The aims of the piloting were to:

- ensure the test could be conducted smoothly in the two venues
- trial the test process in both venues and remedy emerging problems
- test the two instruments designed for participant responses: the questionnaire for teachers and hotel workers, and the focus group discussion schedule for teachers
- report and reflect on the questionnaire and discussion responses and adjust testing conditions and feedback instruments for the testing proper.

The five hotel workers took the Aptis test in listening, speaking, and grammar and vocabulary. At the TT institute, 10 volunteer fourth-year BEd students took all five components of the pilot Aptis test.

A questionnaire was administered on completion of the pilot test (see Appendix 1). It was designed to capture candidates' immediate post-Aptis impressions and feedback using yes/no or multiple-choice questions, with space for explanation of some answers. Candidates could tick more than one answer and explain their response or add another response not provided in the choices. The version of the questionnaire in Appendix 1 was used for the trainee teachers, with a slightly adapted version for the hotel candidates.

The questionnaire findings were probed further in a focus group discussion with five of the TT students (see Appendix 2 for the stimulus questions).

Questionnaire and focus group responses showed a strong positive attitude to the test and appreciation for the opportunity to experience it. Some felt their English skills were improved, suggesting a positive washback role for the test.

The 10 TT participants were the only volunteers from a group of 120 and it was feasible to assume that they were keen to do the test and not apprehensive or nervous about being subjected to a computer-generated assessment (their questionnaire responses confirmed this assumption). However, in the subsequent focus group discussion with five of the group, they pointed out that it might not be true of other students in the cohort, since most were unlikely to have encountered computer-generated assessment and most did not have strong computer skills. Also, many were understood to have poor English language skills and correspondingly low confidence in their use of English. From this feedback, we decided to ensure sufficient support staff were present at all test sessions specifically to assist those candidates with limited computer skills.

The emerging responses proved valuable, and, since it worked well, we decided not to change the instrument or process. We were then able to incorporate the pilot questionnaires and focus group responses from the TT participants into the full data set for analysis.

3.3 The main test phase

The main test phase commenced in mid-August 2013, and continued until mid-November 2013. Due to the nature of Aptis testing and the constraints in both the education and hospitality sectors from candidates' limited availability, testing had to be conducted in several phases and sessions. Each group was given a brief pre-test tutorial to prepare them for the test. After they took the test, they were asked to complete the questionnaire, and some were then invited to participate in focus group discussions. Because large numbers of candidates wanted to take part in the focus groups, we decided subsequently to conduct several individual interviews to validate data and elicit some personal responses (see Appendix 3 for the guide questions).

3.3.1 Hotel candidates

The attempt to test hotel candidates failed for various reasons, managerial and technical, and so insufficient data were collected for inclusion in the final results. The managerial factors related to the fact that hotel workers were under time pressures from shifts. There would need to be strong institutional support for their English at their initial training stage; and marketing of Aptis to this sector would need to take the scheduling problems into consideration.

See Section 3.3.3 below for an initial account of the technical factors.

3.3.2 Teacher trainees and teacher candidates

After the pilot testing, lecturers at the teacher training institute identified 30 first and second-year BEd candidates, as well as a further 20 fourth-year in-service BEd candidates to take part in the study. Many fourth-year students were teaching in remote areas of KwaZulu-Natal and could not be contacted earlier than the day before the tests had been scheduled, which meant some students did not show for the first two test sessions. However, all but one first-year student came to the final session which was still within the five-day window period when Aptis tests were available to registered candidates. The study also included nine isiZulu-speaking primary school teachers.

Although we had asked five candidates to take part in the focus group interview for each group, larger numbers arrived wanting to participate and we allowed them to do so. In order to elicit more fine-grained information, we conducted individual interviews by telephone with a small number of TT students and teachers.

The final total number of completed questionnaires was 68 (as noted above, this includes the data from the fourth-year students who participated in the piloting). Although all these participants took the test before completing the questionnaire, technical problems meant that we did not receive full test results for 11 of them. The comparisons with other 'local' measurements of proficiency reported in Section 4.2.2, therefore, involve different numbers of participants depending on the sections of the Aptis tests for which we had results.

In all, 35 of the students and all nine of the teachers attended focus group discussions. Subsequently, individual telephone follow-up interviews were conducted with eight participants.

3.3.3 Technical issues

With all groups of participants, a number of technical issues arose in various administrations of the test. They included very slow speeds of downloading at times and general instability of the internet connections. In some cases, speaking test responses were not correctly uploaded, and some results were not returned to us in time to include them in the statistical tests. The implications of these problems will be further discussed in Section 5 below.

3.3.4 Data collection and analysis

To summarise, data for this report were collected through:

- the researchers' own trial of Aptis
- the piloting of test and instruments
- the main testing phase
- candidates' performance in other local English tests and exams administered at the teacher training institute.

The phases of analysis comprised:

- analysis of questionnaire responses, for each group and then collectively
- analysis of focus group discussions
- analysis of individual interviews, for each individual and then collectively
- analysis of Aptis scores for each group and then collectively
- correlation of Aptis scores with other direct and indirect assessments and tests of English language ability prior to Aptis testing.

4. FINDINGS

Findings from the questionnaire, focus groups and interviews are reported in Section 4.1 below. Aptis scores for the research populations and comparisons and correlations of Aptis with other indicators of English performance are reported in Section 4.2. The main findings are summarised and discussed in Section 5.

4.1 The questionnaire, focus group interviews and individual interviews

The post-test questionnaire (Appendix 1) was completed by 68 participants: 59 TT students and nine primary teachers. The focus group discussions were conducted with:

- five candidates from the pilot group of 10 fourth-year BEd students
- 10 from the subsequent 20 fourth-year BEd students
- 20 from the 30 first-year students
- nine primary school teachers.

Individual interviews were conducted by telephone with eight participants: three TT students and five teachers.

In the following, results from the questionnaire are reported, with comments from the focus groups and the interviews introduced where appropriate to complement the questionnaire data.

4.1.1 Pre-test tutorial and sound and recording test

In regards to the pre-test tutorial, 66 out of 68 respondents reported it as helpful, and 67 candidates found the sound and recording pre-test helpful. Among the reasons given for this positive reaction were:

- aiding their computer use / typing (10 respondents)
- raising their concentration (5)
- making them aware of what the test would be like (4).

It was clear from our observation that the tutorial and sound test helped candidates learn how to interact with the computer, sometimes for the first time. We noticed anxiety diminish as they became intrigued and engaged in the test proper.

4.1.2 The test process

The younger first-year BEd students were mostly studying straight from school and had more computer experience than the older fourth-year group. The primary school teachers had the least computing experience: in fact, we discovered that three had never used a computer. Nevertheless, the great majority of the candidates (62) reported that they found that it was easy / fairly easy / okay using a computer for the test. Only seven found it a little difficult; and no candidate found it very difficult or far too difficult. Similarly, nearly all (64 of 68) indicated that they felt comfortable using the computer.

The following four questions explored the participants' attitudes towards the test as a whole. General reasons for finding it easy to do the test included the following responses.

- Instructions were clear (64 responses).
- Questions were clear (54).
- We were given sufficient time to answer (48).
- The computer worked well (43).
- It was easy to use the computers (35).

The few negative responses included the following reasons.

- It was too long (15 responses).
- The computer didn't work well (4).

It is worth highlighting the fact that the majority of the responses to these items are related to general features of the test, such as the length, rather than specifically to working on the computer.

The aspects of being tested by a computer that the participants found difficult included:

- counting words in my answers (23)
- answering quickly (14)
- reading quickly (11)
- dragging words or sentences (9)
- the time counter (6)
- computer not working (6)
- not being able to hear (5)
- the drop down menu (4).

Reactions in the focus groups were also generally positive towards the test in general and the use of computers. Several stated that they enjoyed working on the computer. Comments included:

It was easier because no-one was listening to you

It's handy...very handy...because you can erase when you make a mistake and then you go forward

To me it was like playing a computer game...it wasn't that serious and you didn't have just to answer...you had to think about what you had to say before you could talk...and maybe it took some time to finish it.

There were just a few negative comments, some of which concerned practical problems due to internet speed:

Sometimes the computer was slow and we waited

We had to wait for the computers to reload.

Specific aspects that the participants liked included:

Picking the correct word and dropping it into the box

I liked the choosing where I was supposed to choose the answers

Just to be tested in all skills...I liked that

As English teachers so we would get an understanding how to assess the whole language

I liked the passage when you have to select the topic...it was tricky.

Parts or questions that they found hard included:

The grammar was hard – there were new words

The reading was a little bit hard...putting the labels

I think rearranging those sentences to make a paragraph wasn't easy.

Generally, the focus group candidates agreed that they had benefited from doing the test. All those interviewed said they felt more confident as a result.

The researchers observed that the candidates were extremely pleased with themselves after completing the test, even euphoric as they left the test room. The older candidates in particular were excited and inspired by the novelty of being tested in several components of English on a computer rather than through a written test.

Indeed, the test had a knock-on impact that might not have been predicted: in subsequent interviews, some teachers described how they had changed their teaching as a result of the test. In particular, they stressed that they had realised the importance of encouraging their students to use the language to communicate. One representative comment by a teacher gives a flavour of this reaction:

Aptis changed me a lot...now I try to change the way I teach...I use pictures and ask them to speak more freely in most lessons.

4.1.3 Timing

Despite the overall positive reaction to the test, one particular issue emerged as slightly problematic for this group of candidates: the time allowed (this was also highlighted by respondents in the survey reported in O'Sullivan 2012b). Although 48 candidates said that they had sufficient time overall and only four said that time was insufficient, many said they found the test too long and could not complete some sections. In all, participants indicated in 79 responses that they could not answer one or other component quickly enough. This issue of length, time and speed cropped up throughout the questionnaire responses. These are summarised in Table 1.

Table 1: Questionnaire responses for time

Comment	General	Listen	Read	Speak	Write	Gram/ vocab	Totals
Too long	23						23
Insufficient time	4		15			7	26
Questions came too quickly		4		7		6	17
Couldn't answer quickly enough		13	14	20	38	12	97
No time to check		6	8	11		6	31
TOTALS	27	23	37	38	38	31	194

As Table 1 shows, the highest number of time-related problems were reported for Writing, Speaking and Reading, while the specific problem of not being able to answer quickly enough was reported as affecting Speaking and Writing above all. We noticed that several candidates were timed out during the writing test, and five of seven interviewed in one focus group indicated that they were timed out in the writing, but in no other component.

4.1.4 Test content: Individual components of the test

The questionnaire asked the participants for their reactions to the five individual components of the test. Table 2 gives an overview of the responses.

Table 2: Reactions to the five test components

	Negative responses			Positive responses
	Most difficult part	Time problems	Other problems	Not difficult
Listening	7	23	7	26
Reading	8	37	11	19
Speaking	11	38	12	17
Writing	24	38	12	11
Vocabulary and Grammar	18	31	11	23

The **Listening** component was generally felt to be one of the easier parts: only seven candidates rated it as the most difficult component. Twenty-six candidates indicated that they did not find it difficult; and in fact, many scored better for listening than for the other components. As shown in Table 2, however, 23 participants had problems with the time allowed. In addition, three said that they found it difficult to use the computer for listening.

Four candidates indicated that they did not understand the speakers. We followed this up in the focus groups, asking if the variety of English voices and accents caused problems. One teacher found the accents challenging but most understood well:

There were a lot of different ways of speaking English...it wasn't difficult...those people were not speaking fast...they were speaking normally.

One student took far longer than the others and told us in the focus group that he had a hearing problem. We had not known this before the test. It indicates the advisability of either screening students beforehand for this disability or ensuring sufficient volume through the headphones for them to hear the listening component.

The **Reading** component was perceived as fairly easy, with only eight candidates rating it as the most difficult test component. Nineteen students did not find reading difficult. According to one:

I found the reading was much easier than the speaking...because we were seeing the words clearly...and the English was just the English you would speak...not a complex one.

Despite this evaluation of the component as relatively easy, 37 participants reported time problems, as shown in Table 2; and 11 more had difficulty in understanding the questions. In the focus group, one teacher explained:

From reading I think the other problem was understanding...you had to read and write the passage...at the end of the passage you realised...it's not the thing that you're supposed to write.

Although we noticed that students struggled to select the headings for paragraphs, interview responses were mixed on this aspect: some reported difficulty while others found it straightforward.

The **Speaking** component gave mixed results. It was not rated as particularly difficult – only 11 participants marked it as the most difficult part of the test, and 17 assessed it as not difficult. However, it caused some problems, with many candidates not doing well in the assessment (see Table 3). A number felt that the time for speaking was short – as many as 38 responses indicated problems with time (Table 2); and the layout of the computer room, with computers in close proximity, made it more difficult for some:

The set up was a bit cramped so it was hard...it was short for speaking.

On the question of speaking to the computer, the reactions were equally mixed. Twelve found it hard to use the computer for this purpose, and some teachers were shy about speaking. On the other hand, when we raised the issue in the focus groups, most candidates were amused and there was laughter when they were asked how it felt to speak to a computer. Though some were also bemused by the experience, most were positive. Comments included:

I liked using the computer...speaking to the computer...it was a new experience

It felt like you were speaking to yourself

It was just like speaking to a person that I don't see and don't know.

The **Writing** component was rated as the most difficult part of the test by the highest number of participants (24 of the 68) and as not difficult by only 11 participants. This component also shared with Speaking the highest number of responses (38) indicating problems of time. In addition, six participants said that they did not understand the questions, and six found the computer difficult to use for writing. One interviewee explained:

And the other thing with the writing was we had to type rather than write...rather than clicking as with the others and then we were thinking...

It is worth noting, however, that the mean score for Writing (Table 3) was not markedly lower than those for Listening and Reading, and was higher than for Speaking.

As with Speaking, reactions to the **Grammar and Vocabulary** component were somewhat mixed. It received the second highest number of questionnaire responses (23) for not difficult; and all groups achieved their highest mean scores on this component (Table 3). At the same time, it also received the second highest number of responses (18) for the most difficult part of the test. Thirty-one participants reported difficulties with time, and 11 did not understand some of the questions. This mixture was neatly encapsulated by one interviewee who said:

We were not good at it...but it was easy.

4.1.5 Perceptions

To gauge the affective impact of Aptis, we asked candidates how the test made them feel. Most were positive: a total of 143 responses reported that doing the test made them feel proud, confident, comfortable, etc.

The few negative perceptions included:

- anxious (5 responses)
- panicky (4)
- scared (3)
- stressed (2).

One candidate rated it as the worst experience ever and two as a horrible experience. However:

- 29 responses indicated that it was a good experience
- 22 indicated that it was a great experience
- 22 participants marked that it was the best (test) experience they had ever had.

A total of 71 selected responses described the experience in positive terms, indicating that the candidates felt that it was fun and pleasant to do. And a total of 99 responses rated the Aptis test as well designed, user friendly, etc.

In focus groups, when participants were asked if they enjoyed the test, the response was unanimously positive. All said that they were glad they had done the test and that it gave them confidence in their English.

As a whole, the participants' perceptions of the test were thus overwhelmingly positive, in many ways echoing the responses of test-takers reported in O'Sullivan (2012b).

4.1.6 Using the test in South Africa

Regarding the candidates' views on using the test in South African education, the picture that emerged was again extremely positive. Not only were there 57 responses indicating that the test should be used for teachers or trainee teachers of English, but 46 participants felt that it should be used for all South African teachers (recall that English is the preferred medium of instruction in most schools).

In the focus group interviews, candidates responded that the test should be used for teachers and learners. Comments relating to use with teachers included:

It should be used with teachers because some can't speak English well

Yes I do think so because in most cases even though English (speaking) especially the teachers...even though we are teachers we can't speak English

It would help if you were tested before you qualified as teachers...even at the very beginning and then again towards the very end.

Three questionnaire respondents suggested that Aptis should also be used with learners. One interviewee made a perceptive comment on this possibility:

It will benefit the learners because they learn best with games and computers and they like to watch television.

However, another interviewee reflected the situation regarding the availability of technological resources in South Africa more accurately:

It can be used with high school students like grade 8 to grade 12...they would do it but not with computers.

4.1.7 Improving the Aptis experience

Twenty-six questionnaire candidates made positive comments about the test when asked how it could be improved. The small number of critical suggestions that were made fell into three main groups.

First, the issue of time and timing came up. Four participants felt more time should be allocated; in the focus groups, two participants mentioned the possibility of allowing more flexibility in timing:

You could make it so that you can choose your own speed

In a way you could push the pace yourself with questions like the multiple choice ones.

Secondly, the practical problems of the equipment were highlighted by two questionnaire respondents; one focus group member complained:

Sometimes the computer was slow and we waited.

The third area was one that we ourselves were concerned about on the basis of our experience of the test: the role of sociocultural knowledge. Two respondents suggested that the questions should relate more to their own lives; and one interviewee used the example of travelling:

I'm not into travelling...not all people know what is happening inside and outside the country.

4.1.8 Sociocultural knowledge

The issue of the sociocultural knowledge required emerged as problematic for some participants in relation to various parts of the test. For Writing, for instance, comments like the following were made in the focus groups:

they ask you to write an email...we didn't even know the format for an email

Writing...took me a lot longer because the context they were using of...they were talking about the gardening...I didn't have the knowledge...it took me some time and I thought why was I doing this

Some of the things...they were not as usual in our everyday lives.

As researchers, we experienced the trial test ourselves, and we identified some stimulus items where we thought South African English second language speakers might encounter problems. Examples are given below of the prompts and materials we observed that were specific to English society and culture.

- Candidates shown pictures of five-story townhouses or apartments were asked to comment whether they were suitable for families living with children. Most South Africans live in rural areas where buildings are almost all single story with one or two rooms and so they would not be familiar with the type of living associated with apartments.
- One student was shown a picture of the Colosseum in Rome and asked if he would like to visit the building. He said that he would not visit because the building was broken down and would be uncomfortable. This highlighted three issues. First, South Africans use the term *visit* when meaning *stay at* or *with*. Secondly, the Colosseum is not a familiar tourist destination for the vast majority of South Africans. Thirdly, being a tourist is also a concept foreign to many South Africans. Travel tends to be undertaken when necessary and then mostly to see family.
- We noted some unusual responses about various topics. For example in South Africa a *health club* is normally termed a *gym* and is a feature of affluent urban society. One candidate when asked what activities she would like held in a *health club* replied *counselling sessions as many people in the area were dying*. She was referring to health related to disease, notably AIDS, where the purpose of a health club would be to promote community wellness in the face of the ever-present AIDS threat.
- Similarly a *book club* in South Africa tends to be mainly white women who meet every month to share and talk about books purchased for the group. Though increasingly found across racial and language groups, they tend to be a feature of affluence.
- Few South Africans, except those in more affluent urban areas, visit the cinema (locally termed *bioscope*, particularly in Afrikaans-speaking or rural areas), or watch films other than on the television.
- Some candidates understood *gardening* as subsistence farming, rather than as a pursuit for leisure.

4.2 Test results and comparison

In this section, we first present the test scores achieved on the different components of Aptis by each group and comment on the patterns that emerge. We then compare the Aptis scores with the candidates' scores on other measures of their language proficiency. It should be noted that, because of the technical problems mentioned in Section 3.3.3 above, Aptis results for some candidates for some components were not available. Therefore, the numbers of participants involved in the different calculations below vary.

4.2.1 Summary of test scores

Table 3 shows a summary of the mean, range and variance of scores for each Aptis component for each group tested, and the corresponding collective figures for all groups.

Table 3: Summary of descriptive statistics for all groups

Aptis test	BEd 4 th yr in-service	BEd 1 st yr full-time	Primary teachers	Mean all groups	Range all groups	SD all groups
Reading	No. 23 Mean 33.04 Range 32.00 S.D. 8.09	No. 29 Mean 30.55 Range 34.00 S.D. 8.77	No. 9 Mean 34.44 Range 34.00 S.D. 11.82	32.68	33.33	9.70
Speaking	No. 17 Mean 33.35 Range 15.00 S.D. 4.11	No. 25 Mean 28.88 Range 22.00 S.D. 6.53	No. 8 Mean 23.38 Range 20.00 S.D. 6.55	28.54	20.33	5.95
Listening	No. 23 Mean 33.04 Range 24.00 S.D. 6.44	No. 29 Mean 31.93 Range 22.00 S.D. 5.64	No. 9 Mean 26.67 Range 24.00 S.D. 7.87	32.17	23.33	6.72
Writing	No. 23 Mean 30.04 Range 31.00 S.D. 7.80	No. 29 Mean 31.90 Range 19.00 S.D. 5.25	No. 9 Mean 31.56 Range 19.00 S.D. 6.52	31.17	23.00	6.61
Grammar/ vocabulary	No. 23 Mean 36.26 Range 32.00 S.D. 7.45	No. 29 Mean 34.24 Range 27.00 S.D. 6.59	No. 9 Mean 35.89 Range 22.00 S.D. 7.24	35.46	27.00	7.10
Mean all components	33.15	31.50	30.39	32.00	25.40	7.33

Table 3 allows comparisons between the different educational groups within the whole set. One might, all things being equal, have expected to see an increase in the mean Aptis scores from first-year to fourth-year students to practising teachers; but in fact, the primary teachers score lowest overall, mainly because of low scores on Speaking and Listening. Our own prediction was that the first-year BEd students would score higher than the other two groups. As teacher trainers, we were aware of the poor spoken English of the older teachers (both fourth-year part-time BEd students and primary teachers) compared to the younger first-year full-time students, since the latter, although also EL2 speakers, had come straight from schools where English had been the medium of instruction.

In fact, as Table 3 shows, it was the fourth-year BEd students who scored highest overall, although the differences here are not significant (see ANOVA results in Appendix 4). Our tentative explanation is that this group were volunteers and, therefore, likely to be the more confident and possibly more proficient students in their group, whereas the first-year students were selected purposively. A further possible factor is suggested by the fact that, while in Reading, Writing and Vocabulary/Grammar, the three groups achieved closely comparable scores, there were significant differences on Speaking [$F(2, 47) = 8.29, p = 0.001$] and, to a lesser extent, on Listening [$F(2, 58) = 3.41, p = 0.04$], with the fourth-year students scoring higher than the other two groups. This group of students had been given a newly designed spoken English course specifically to address their poor oral skills. We believe that this explains, to some extent, the large difference between their listening and speaking scores and those of the primary teachers who had not been given a dedicated course in spoken English (although see discussion in Section 4.2.2 below).

On the basis of our experience of working with these and similar groups, we expected to see better results for the skills of Reading, Listening and Writing than for Speaking; and this was confirmed through the collective means and, on the whole, in the means for each group within the set (though interestingly, the fourth-year students achieved their lowest mean score on the Writing). Grammar and vocabulary was also expected to show a higher score, since the teaching of English in South African schools and teacher training establishments has focused more on writing and correct grammar than on spoken English.

The Aptis scores thus generally corresponded to our predictions, indicating that the test seems to be offering an accurate picture of the participants' language proficiency. The scores also confirmed our suspicions that speaking English is a challenge for the EL2 candidates and EL2 speakers in South Africa generally when measured against the other English skills that were tested. This has serious implications for the use of English as a medium of instruction in schools, since a good deal of teaching and learning is carried out orally, and pupils need to be given a secure basis of language proficiency in the primary school in order to cope successfully. However, this issue, though serious for those involved in education in South Africa, is outside the scope of the present study.

4.2.2 Comparison of test scores with other measures of English language ability

The Aptis results were compared with results that the candidates had received in other English modules and tests. This was to ascertain the concurrent validity of Aptis with other predictive tests for entering tertiary education. We also plan to disseminate the results amongst education practitioners in South Africa in order to persuade them of the viability and value of aligning their assessments of language proficiency with global developments in the field.

The comparison involved the candidates studying at the teacher training (TT) institute. We also intended to compare the primary teachers' Aptis results with their English school-leaving and teaching qualification grades. However, the nine teachers qualified at different times with different qualifications, some with no mark or assessment for English.

For the TT candidates, we used various English language results they received throughout their studies at the institute. The fourth-year students completed a number of English modules: specifically English in their first and second years (referred to as E100 and E200 below) and a conversational English course in their fourth year (referred to as LCE400 below). The other students all completed an English language support course in their first year (ELS100) and an independent English placement test administered by a major South African university (referred to below as EPT). In addition to the modules and tests at the teacher training institute, all candidates took English in their matriculation (matric) qualification (the South African exam taken at the end of 12 years of schooling). Their matric English as an Additional Language (EAL) results were also correlated with the Aptis results.

Table 4 shows the correlations between Aptis and other 'local' tests of English, indicating those that are statistically significant.

Table 4: Correlations with other tests at the TT institute

No. of participants	Test	Aptis L	Aptis R	Aptis S	Aptis W	Aptis GV
39	Matric EAL	-.466*	-.303	-.054	-.402*	-.486**
23 (18 for S)	E100	.431*	.380	.668**	.151	.410
21 (17 for S)	E200	.466*	.325	.649**	.157	.543*
22	LCE400	.600**	.496*	.455	.546**	.671**
27	ELS100	.374	.251	.401	.210	.393*
18	EPT	.750**	.399	.250	.557*	.672**

** Correlation is significant at the 0.01 level (two-tailed).

* Correlation is significant at the 0.05 level (two-tailed).

The first point that might, at first sight, appear surprising is that the matriculation results all correlate negatively with the Aptis results (though not all these correlations are significant). However, not too much can be read into this finding, since the English matriculation test has very different aims from Aptis: it focuses on literature, essay writing, comprehension and summaries. In addition, the EAL results are from different cohort years, so the bases for comparison are not consistent. Nevertheless, the fact that the two tests appear to give such different results raises intriguing questions for possible future research.

Leaving aside the matriculation scores and looking down the columns in Table 4, two of the Aptis components – Grammar/Vocabulary and Listening – show more consistently significant correlations with the local assessments than the three others. It is not surprising that all the local tests, apart from E100, show a significant correlation with the Aptis Grammar and Vocabulary component. Grammar was specifically tested in the EPT test; and the students' grammar and vocabulary proficiency would also have been taken into account in most assessments in all of their courses. For example, most scoring rubrics used in the Languages Department of the TT institute have a descriptor relating to use of grammar. This correlation might be helpful in reassuring traditionalists in the South Africa TEFL community that what many regard as the core knowledge required by language learners is adequately accounted for in Aptis. It is perhaps more surprising that the other Aptis component which correlates significantly with all but one of the local tests is Listening. However, it is worth noting that the EPT test includes a separate test of listening, and that LCE400 focuses on oral skills.

Aptis Reading, on the other hand, shows significant correlations with only one of the local tests – LCE400 – which is a course focusing on speaking skills. This is an issue which would merit further investigation. It may be that the Aptis test probes aspects of reading skills which are, for whatever reason, given less emphasis in the students' courses. It may also be that academic reading skills, which are presumably of importance for the TT students, are less well covered in the Aptis test. More encouragingly, it is noticeable that E100 (first-year English) and E200 (second-year English) show significant correlations with the Speaking scores, and slightly lower correlations with the Listening scores, of the Aptis test. This perhaps reflects the fact that both courses had an oral component which was taken into account when calculating the students' final mark.

Looking now across the rows in Table 4, it can be observed that one institute course, ELS100, which is intended to cover all the language skills, shows a significant correlation, at the 0.05 level, with only one Aptis component, Grammar/Vocabulary. Given that the validity of Aptis is better established than that of the assessment of a single-institution course such as ELS100, one possible implication is that the course, or the way it is tested, needs to be re-visited. If problems with the course do emerge on investigation, it would again help support the case for introducing Aptis in the context of the institute.

Another notable 'gap' can be seen with LCE400, Conversational English. This shows significant correlations with all components of the Aptis test other than Speaking. This is unexpected as LCE400 is specifically a course to improve English speaking skills and much of the final mark is based on the quality of students' speaking in class during prepared and unprepared orals. Furthermore, the fourth-year candidates who took LCE400 achieved the highest mean score in Aptis Speaking. The non-significant correlation with Aptis Speaking suggests again that it would be useful to scrutinise the institute course to check if the assessment is appropriate. However, the fact that LCE400 correlates with all the other components raises the possibility that a crucial factor was the requirement in the Aptis test to interact with a computer, rather than engage in face-to-face communication as in the course. Some of the group may have coped with the novelty more or less successfully than others and may have performed better or worse in this mode than face-to-face. Moreover, the topics about which they were asked to speak may have been culturally unfamiliar (see Section 4.1.8 above), leading to untypical performances. Those factors would raise some doubts over the construct validity of the Aptis Speaking test.

Overall, the picture that emerges from Table 4 is not entirely reassuring, in that one might expect the patterns of correlation to be more consistent than they are: as many as 12 out of 25 of the possible matches between Aptis and local tests (excluding the matriculation results) do not show statistically significant correlations. There are clearly factors which partly explain the differences – for example, some of the local tests are achievement tests, tied to particular courses of study, rather than proficiency tests (although the EPT test is designed to be a placement test for higher education). However, all the local tests are intended to offer a valid and reliable picture of the English language proficiency of the students – particularly important since they are future teachers of English. Therefore, the observed degree of mismatch with a robustly validated test such as Aptis is a cause for concern.

5. DISCUSSION

We can now return to our research questions and summarise the findings and insights that have emerged from the project.

5.1 Test-takers' views of Aptis

Research Question 1 asked:

To what extent is the Aptis test seen by test-takers in South Africa as an acceptable way of measuring their language proficiency?

Before the research, we expected that candidates would have poor computing skills and that this might affect their reactions adversely. But in fact, although some did indeed have little or no experience of using computers, the overall attitude of the candidates was strongly positive. We were impressed by their extremely enthusiastic reception of Aptis, especially the older ones who we had expected to be somewhat resistant to the innovative –but, for them, quite alien – way of testing. This positive reaction was seen specifically when they were asked how they felt about the test and being tested on a computer, but also emphatically throughout the responses in the questionnaire, focus groups and interviews. Older teachers were also more triumphant and excited by the Aptis experience: even those with poor computer skills enjoyed the sense of achievement as they learnt to use the computer routines during the test. All of this bodes well for any future use of the test with EL2 speaking teachers in South Africa.

One aspect of the test that many participants commented on less positively was the problem of time. On the one hand, this is presumably a deliberate part of the design, since the test has to assess a wide range of proficiency: candidates complete as much of the test as they can in the time allowed, with those whose proficiency is lower not being expected to complete all the tasks. However, the way the test is administered also played a part – this will be discussed in Section 5.3.1 below.

5.2 Correlations with other measures of language ability

Research Question 2 asked:

To what extent do test-takers' results on the Aptis test correlate with the result of other measures of their English language proficiency?

As a whole, the comparison indicates that the local tests are to some extent in line with Aptis, but that there is less correlation than would be ideal. This means that there are aspects of local tests which merit further investigation. Given the mismatches between the results, and taking into account the fact that the design and trialling of the Aptis test is clearly much more stringent and theoretically-grounded than local tests can be, a strong case can be made to bodies concerned with language testing in South Africa that it is worth seriously considering the introduction of Aptis in the English language education domain.

It is worth noting, however, that our findings also suggest that some aspects of the Aptis test, especially relating to the Speaking component and to the possible impact of culturally inappropriate items, may benefit from further research.

5.3 Accessibility of Aptis

Research Question 3 asked:

To what extent is the Aptis test “accessible” within the practical constraints of South Africa, a mix of urban centres and vast deep rural areas?

The survey results and the test scores indicate that Aptis has both good face and construct validity, at least in the eyes of our group of participants. However, two areas of concern emerged which are connected with the South African context, but which may well be relevant to other geographical regions: the impact of using computers, and the inclusion of socially or culturally specific items. The evidence suggests that these may have had a small but potentially important effect on the candidates' scores. This section and the next will address each of these issues in turn.

There are two aspects of the impact of computers that need to be considered. The first involves the technological infrastructure of South Africa. As noted in Section 3.3.3 above, we had a number of technical difficulties. These forced us to abandon part of our proposed study; but they also affected the main study. Downloading the tests was slow at times because of bandwidth problems. This may well have affected the candidates' concentration and also undermined their confidence in the professionalism of the test administration (although the latter point was not raised explicitly by any of the participants). Ideally, all the required tests would be loaded onto the computers in advance, but this was not possible because of limited access to the computers. In addition, the computer system at the teacher training institute was sometimes unstable and, at times, a candidate was disconnected from the test computer and could not be reconnected. The invigilator then had to move the candidate to another computer, inevitably disrupting the candidate's concentration. Uploading the Speaking component after a candidate completed the test also took time, which may have been the reason why five of the Speaking tests were not received by British Council. A number of Speaking tests were returned as zero, meaning they had not been uploaded, and some test results were not returned at all.

The crowded and cramped computer room for the first few sessions also required students to speak loudly – some tests were inaudible because of background noise, and some Speaking tests had to be repeated.

Although these problems are in some respects specific to this project, they do serve as a warning signal: the feasibility of using the test depends greatly on the availability of adequate computer and internet facilities. In the South African context, such availability cannot be taken for granted; and the accessibility of the test is, therefore, diminished. This does apply equally, of course, to any computer-based test; but, unlike, say, a test such as computer-based TOEFL which is only offered at dedicated sites, an important characteristic of the Aptis test is that it can be used in any organisation at any time. It may be advisable, therefore, to highlight the technical requirements in the publicity materials.

The second aspect relates to the test-takers. Almost all candidates needed help to manipulate the drop-down lists and the drag and drop sentences, although the majority quickly mastered the techniques involved. The younger full-time first-year students completed their test considerably more quickly and took between 30 minutes and one hour *less* than the older part-time fourth-years and the primary teachers. Many of the older candidates did not have good computer or typing skills and did not have access to computers in their day-to-day lives. In itself, this was not a problem: it mainly meant that starting the tests and moving between components took longer for some candidates; and we were pleasantly surprised at how quickly the participants picked up the skills and techniques needed.

However, the unfamiliarity with computers did affect how well the candidates were able to cope with the demands of the test, particularly for Speaking and Writing. Marks for Speaking were among the lowest for many participants, particularly the primary teachers who were the group with the least computer familiarity. Candidates found it hard to respond quickly and said they needed more planning time. It is uncertain how far the requirement to listen and speak to a computer was a factor in the low scores, since it did not appear to affect all candidates, and the focus groups expressed positive feelings about the method. But it cannot be entirely discounted that the strangeness of talking to a machine for the first time played some part in some candidates' hesitancy.

The effect on Writing appears to have been more direct and more serious. Poor computer skills, especially typing ability, inevitably slowed down some candidates to the extent that they were not able to complete certain test components, particularly writing, in the given time. One older student was observed typing her response to the Writing test, letter-by-letter, using only one thumb. There have been numerous studies comparing computer-based (CBT) and paper-and-pencil (PPT) tests, but relatively few (e.g. Weir *et al.* 2007) have focused on tests of writing, and none of the latter that we know of have included candidates with no, or very little, computer familiarity. Again, it is worth recalling that Aptis is not aimed at an academic clientele (where current students can usually be assumed to have a reasonable degree of computer familiarity), but at a more general market (including, for example, the hospitality sector), where at least some potential test-takers may not have regular – or any – access to computers. It would seem advisable to carry out further research comparing CBT and PPT modes with the kind of test-takers that participated in our study.

5.4 Appropriateness of Aptis

Research Question 4 asked:

To what extent is the Aptis test “appropriate to the particular context and domain of language use” in South Africa?

Although the issue of culturally unfamiliar items in the test did not emerge as a major theme in comments, there were a number of scattered references in the responses to the strangeness of some of the Aptis content for the South African context; and our own analysis picked up several items that we felt could prove problematic (see Section 4.1.8 above).

Though APTIS is designed as a global test, our examples show that it may not sufficiently take into account the local cultural underpinnings of language. The few instances of mismatching or confusing sociocultural constructs and terms that we identified warned us that there could be more. In addition, South African English has its own syntax and vocabulary, with some cultural groups (Afrikaans speakers, and Indians) using simplified grammar, a feature of *lingua franca* (Seidlhofer). The question is whether the British English varieties used in Aptis are the most appropriate for candidates who aspire to a globally understood English, in which case it can be argued that varieties of English that fit the *lingua franca* definition (but could be from another part of the world) should be utilised by Aptis. Once more, the specific nature and marketing focus of Aptis, designed to assess not just a globally-valid standard but also language use in local contexts, means that this issue is of particular importance for the test.

In our view, further analysis of cultural bias, and language and terms alien to South African English would be essential, so that local Aptis candidates are not disadvantaged by the test giving a misleading picture of their language proficiency. As Djwandono (2006, p 81) stresses: “In the domain of language testing, the issue of cultural differences must be treated seriously if the test is to treat all test-takers fairly”. Kramsch (2011, p 310) also argues that: “Tests are skewed in favour of test-takers who are from the same culture as the test designers”. To counter such miscues, test designers should be strongly aware of the potential for parochial British culture to be misunderstood by candidates from other countries; and they should aim to avoid culturally specific items. Thus, in view of the examples we noted, we believe that attempts to avoid cultural misunderstandings still need strengthening.

A more radical option, but one that is in line with current theorising of the status of different varieties of English around the world (e.g. Jenkins 2007, Kirkpatrick 2010), would be to develop localised versions of Aptis, possibly using a franchising system with trained local test developers and markers. This would be relevant not just to South Africa, but to other Outer Circle areas such as India and Singapore. These versions would be designed to parallel the Inner Circle (Kachru 1985) version as closely as possible, to ensure broad comparability of assessment and conformity with the can-do statements of the CEFR, but with adjustments to make them linguistically appropriate to the sociocultural context in which they would be used. Clients could then decide whether the local or 'standard' version was more appropriate for their needs.

6. CONCLUSION

The enthusiastic response of most candidates to the test format shows the excellent potential for the use of Aptis in South Africa. We ourselves experienced Aptis as a test vehicle that was unthreatening, user-friendly and simple for candidates to manipulate once they had been introduced to the computer techniques required. The study has provided us with important empirical insights into testing theory and practices which we are already disseminating amongst the ELT community in South Africa. The aim is to contribute to the slowly-growing awareness in that community and the society at large of the value of using a globally recognised measure of proficiency couched in terms of what the candidates can do in the language.

One welcome by-product of the project was that candidates claimed in questionnaire and interview responses to have benefited from awareness-raising of general language issues, particularly the separation of language skills into different components of the test. The research introduced a group of South African teachers to a new paradigm, particularly those who were school-based and not studying. They better understood the need to consider different aspects of language, and they experienced a new computer-based way of assessing them in terms of 'can do' criteria. As shown in the questionnaire and interview responses, they achieved some understanding of their own language skills and potential to control and improve them. Some used computers for the first time and we observed a newly developed self-esteem and sense of achievement.

There are certainly technical considerations to take into account. Bandwidth issues should be investigated and clearly stated to potential South African Aptis clients since many regions of South Africa are less technologically developed than most of Europe. In addition, the present study also brought out the important question of sociocultural mismatches in test items and prompts. In order for Aptis to be effective in South Africa, the tests would require checks for absence of bias.

As a further step along this path, we raised the possibility of developing a localised version of the test. We realise that this would have a large-scale knock-on effect, as Aptis would need to similarly cater for other regions; but it would contribute to making the Aptis test a trail-blazer in responding to our current understanding of the role of English as a *lingua franca*.

REFERENCES

- Bekker, I. (2008). *The history of the language acquisition question in South Africa and language attitudes at UNISA*. Department of Linguistics, University of South Africa: Pretoria.
- Canagarajah, S. & Ben Said, S. (2011). Linguistic imperialism. In J. Simpson (ed.), *The Routledge Handbook of Applied Linguistics*, pp. 388–400. London: Routledge.
- Canale, M. & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1: 1-47.
- Cliff, A. & Hanslo, M. (2009). The design and use of 'alternate' assessments of academic literacy as selection mechanisms in higher education. *Southern African Linguistics and Applied Language Studies* 27(3): 265-276.
- Clifford, M. & Kerfoot, C. (1992). English: Language of hope or broken dream? In B. Hutton (ed.), *Adult basic education in South Africa*, pp. 152–217. Cape Town: Oxford University Press.
- Council of Europe. (2011). *CEFR language levels*. <http://www.dw.de/common-european-framework-of-reference-for-languages/a-6529726>, accessed August 2011.
- CRELLA. (undated). Centre for Research in English Language Learning and Assessment's socio-cognitive framework for language testing. http://www.beds.ac.uk/data/assets/pdf_file/0003/222825/socio-cog.pdf, accessed July 2014.
- Dale-Jones, B. (2013). Tongue-tied on language policy. *Mail & Guardian online*, 22 March 2013. <http://mg.co.za/article/2013-03-22-tongue-tied-on-language-policy>, accessed May 2014.
- Djiwandono, P. I. (2006). Cultural bias in language testing. *TEFLIN Journal* 17(1): 81-89.
- Howie, S., van Staden, S., Tshela, M., Dowse, C. & Zimmerman, L. (2011). *PIRLS 2011: South African children's reading literacy achievement* [online]. Pretoria: University of Pretoria. http://www.up.ac.za/media/shared/Legacy/sitefiles/file/publications/2013/pirls_2011_report_12_dec.pdf, accessed 23 January 2015.
- Jenkins, J. (2007). *English as a lingua franca: attitude and identity*. Oxford: Oxford University Press.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk & H. Widdowson (eds.), *English in the world: teaching and learning the language and literatures*, pp. 11–30. Cambridge: Cambridge University Press.
- Kirkpatrick, A. (2010). *The Routledge handbook of World Englishes*, New York: Routledge.
- Kirkpatrick, A. & Deterding, D. (2011). World Englishes. In J. Simpson (ed.), *The Routledge Handbook of Applied Linguistics*, pp. 373–387. London: Routledge.
- Kramsch, C. (2011). Language and culture. In J. Simpson (ed.), *The Routledge Handbook of Applied Linguistics*, pp. 305–317. London, Routledge:
- Morrow, K. (1979). Communicative language testing: revolution or evolution? In C. Brumfit & K. Johnson (eds.), *The communicative approach to language teaching*, pp. 143–157. Oxford: Oxford University Press.
- Mouti, A. (2008). Measuring fairness in terms of test-taker characteristics, <http://www.ealta.eu.org/conference/2008/docs/posters/mouti.pdf>, accessed May 2014
- Mufwene, S. (2002). Colonisation, globalisation, and the future of languages in the twenty-first century. *International Journal on Multicultural Societies* 4(2): 1-47.
- Nation, P. (2012). The BNC/COCA word family lists. http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf, accessed July 2014
- Norton, B. & Toohey, K. (2001). Changing perspectives on good language learners. *TESOL Quarterly* 35(2): 307-322.
- O'Sullivan, B. (2011). Language testing. In J. Simpson (ed.) *The Routledge Handbook of Applied Linguistics*, pp. 259–273. London: Routledge.

- O'Sullivan, B. (2012a). *Aptis test development approach (ATR-1)*, <http://www.britishcouncil.org/sites/britishcouncil.uk2/files/aptis-test-dev-approach-report.pdf>, accessed June 2013
- O'Sullivan, B. (2012b). *Aptis formal trials feedback report (ATR-2)*, <http://www.britishcouncil.org/sites/britishcouncil.uk2/files/aptis-formal-trial-report.pdf>, accessed June 2013.
- O'Sullivan, B. & Weir, C. (2011). Language testing and validation. In B. O'Sullivan (ed.) *Language testing: Theory and practice*, pp. 13–32. Oxford: Palgrave.
- Reddy, V. (2006). *Mathematics and science achievement at South African schools in TIMSS 2003*. Cape Town: HSRC Press.
- Rubin, J. (1975). What the 'Good Language Learner' can teach us. *TESOL Quarterly* 9(1): 41-48.
- SA Languages.com. (undated). <http://www.salanguages.com/>, accessed June 2014.
- Seidlhofer, B. (2005). Key concepts: English as a lingua franca. *ELT Journal* 59(4): 339-341.
- Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics* 18: 268-286.
- Spingies, C. (1991). *The "Big E": The English (first language) endorsement on the Teacher's Bilingualism Certificate: an investigation into the background to and origin of this requirement, an evaluation of past and current examining practices and standards and criteria, a needs assessment, and suggestions and recommendations for sound practice*. Rhodes University (eprints.ru.ac.za/4225).
- Stats SA. (2011). http://www.statssa.gov.za/Census2011/Products/Census_2011_Census_in_brief.pdf, accessed June 2014.
- The South African. (2014). <http://www.thesouthafrican.com/business/are-you-eligible-to-apply-for-ilr-or-british-citizenship-this-month.htm>, accessed June 2014.
- Taylor, C., Kirsch, I. Eignor, D. & Jamieson, J. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning* 49(2): 219-274.
- Weir, C. J. (2005). *Language testing and validation*. Basingstoke: Palgrave Macmillan.
- Weir, C. J., O'Sullivan, B. Yan, J. & Bax, S. (2007). Does the computer make a difference? The reaction of candidates to a computer-based versus a traditional hand-written form of the IELTS Writing component: effects and impact. *IELTS Research Reports Volume 7*. Canberra: IELTS Australia and London: British Council.

Appendix 1:

Aptis questionnaire for TT candidates

Pre-test

1. Was the tutorial helpful? Yes No
2. Why?
.....
3. Were the pre-tests for sound and recording helpful? Yes No
4. Why?
.....

Test Process

1. Using the computer for the test was:
 Easy Fairly easy OK A little difficult Very difficult Far too difficult
2. I was comfortable to be tested using the computer: Yes No
3. It was easy because (you may tick more than one):
 Instructions were clear The computer worked well It was easy to use the computers
 Questions were clear I could hear well We were given sufficient time to answer
 I could answer well I could answer most I could do the whole test
 Other:
4. It was difficult because (*you may tick more than one*):
 I didn't understand instructions The computer didn't work well
 I don't know computers I didn't understand questions I couldn't hear well
 I could only answer a few questions I didn't know many answers
 I didn't have enough time It was too long
 Other:
5. What aspects were difficult? (*you may tick more than one*)
 Using the computer Computer not working Drop-down menu
 Finding the place to click Dragging words or sentences The time counter
 Too little time Being able to hear Speaking well Multiple choice questions
 Questions too difficult Grammar questions Reading quickly
 Counting words in my answers Answering quickly
 Other:
6. What was the most difficult part of the test process? (*you may tick more than one*)
 Using the computer Understanding instructions Answering the questions
 Too long Not enough time
 Other:

Content

7. What were the most difficult parts of the test (*you may tick more than one*)

Vocabulary and Grammar Reading Speaking Listening Writing
Other:

8. Explain what was difficult for each part (*you may tick more than one*):

Vocabulary and Grammar:

Too little time Questions too difficult Didn't understand questions
Questions came too quickly Couldn't answer quickly enough
Couldn't check answers Too difficult to use the computer for this Not difficult
Other:

Reading:

Too little time to read Questions too difficult Didn't understand questions
Questions came too quickly Couldn't answer quickly enough
Couldn't check answers Too difficult to use the computer for this Not difficult
Other:

Speaking:

Questions too difficult Didn't understand questions Questions came too quickly
Hard to speak to the computer Couldn't answer quickly enough
Couldn't check answers Too difficult to use the computer for this Not difficult
Other:

Listening:

Didn't understand questions Questions too difficult Didn't understand speakers
Questions came too quickly Couldn't answer quickly enough
Couldn't check answers Too difficult to use the computer for this Not difficult
Other:

Writing:

Couldn't write quickly enough Questions came too quickly
Couldn't answer quickly enough Couldn't check answers Questions too difficult
Didn't understand questions Too difficult to use the computer for this Not difficult
Other:

Perception

9. This test made me feel (*you may tick more than one*):

Proud confident pleased comfortable competent clever stressed
panicky scared anxious stupid unhappy angry
Other:

10. I thought the test was (*you may tick more than one*):

Excellent Very good Well-designed Clever Fun Nice to do Easy
A good experience Fair A great experience The best test experience I've had
User-friendly Pleasant Not good Badly designed Silly Nerve-wracking Difficult
A horrible experience Unfair The worst test experience I've had Not user-friendly
Other:

Judgement

11. In my opinion:

- This test should be used for teachers of English
- This test should be used for those training to teach English
- This test should be used for all South African teachers
- This test should not be used for anyone

12. Please make any other comments to help improve this test experience:

.....
.....

Thank you very much for your help

Appendix 2: Aptis focus group interview

Now you've done the APTIS test, I want to ask you generally about the test experience etc.

1. Was the test a good thing for you all?
2. How?
3. Why not?
4. We noted that best marks seemed to be in listening. Why do you think that was?
5. Most of you didn't do so well in the speaking section as in the others?
Why do you think this happened?
6. How did it feel speaking to the computer?
7. What did you think of the other components? Let's go through them... Reading, Writing and Grammar and vocab.
8. You all seemed to cope with being tested on the computer. What did you think of it?
9. Was there anything that you found really difficult?
10. Was there anything that didn't make sense or seemed strange to you?
11. What did you think about the speed of the test?
12. What was the thing you liked most about the test?
13. Which section or question did you like most?
14. Do you recall any parts or questions you found hard? (go round group)
15. How did you benefit from doing this test?
16. To what extent do you think the test can be used in education?
17. What could be done to make the experience better?
18. Can anyone say any more to help us with this research?

Appendix 3:

Guide questions for individual interview

Use this as a guide but allow and encourage the interviewee to lead if they are willing.

1. Matric grade? (for non-TT teachers)
2. Other rests and exams for English and grades?
3. Do you have any question about the test and your certificate?
4. What were the three things you remembered most? Questions? Pictures? What you did?
5. What was the most important thing you remembered?
6. Did anything confuse you? Questions? Pictures?
7. Has anything changed in your teaching?
8. In your own life?
9. Has it changed the way you feel about English?
10. How is it showing?

Appendix 4: Comparison of Aptis results for the three groups: one-way ANOVA

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
AptisL	Between Groups	270.099	2	135.050	3.407	.040
	Within Groups	2298.819	58	39.635		
	Total	2568.918	60			
AptisR	Between Groups	139.387	2	69.693	.859	.429
	Within Groups	4708.351	58	81.178		
	Total	4847.738	60			
AptisS	Between Groups	562.083	2	281.041	8.285	.001
	Within Groups	1594.397	47	33.923		
	Total	2156.480	49			
AptisW	Between Groups	45.804	2	22.902	.542	.584
	Within Groups	2449.868	58	42.239		
	Total	2495.672	60			
AptisGV	Between Groups	56.677	2	28.339	.575	.566
	Within Groups	2856.634	58	49.252		
	Total	2913.311	60			
AptisSum	Between Groups	2316.445	2	1158.222	1.743	.186
	Within Groups	31227.175	47	664.408		
	Total	33543.620	49			

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.

EVALUATING APTIS TESTS WITH TEACHER TRAINING STUDENTS AND IN-SERVICES PRIMARY SCHOOL TEACHERS IN DURBAN, SOUTH AFRICA

AR/2015/002

Julie Douglas and Charlotte Mbali

**ARAGs RESEARCH REPORTS
ONLINE**

ISSN 2057-5203

© British Council 2015

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