

**APTIS IN CHINA:
EXPLORING TEST-TAKER PERCEPTIONS OF
ITS TEST VALIDITY AND PRACTICALITY**

AR-G/2017/6

**Ying Zheng, University of Southampton
Yanyan Zhang, Wuhan University**

ABSTRACT

This project investigated potential Aptis test-taker perceptions of its validity and test practicality in China. Eighty-three students from Wuhan University China participated in the study and took the Aptis test. A survey design with follow-up interviews was adopted. A questionnaire was used to investigate test-taker perceptions on aspects of their test-taking experience, test items, skills measured, and some practicality issues. Semi-structured interviews were carried out to further probe participant perceptions of the test.

The study was designed to compare English skills measured in Aptis to skills measured by other English tests on the Chinese market. Gaining a better understanding of test-taker perceptions of this new test can provide valuable information, informing test development, as well as test preparation practices.

Authors

Dr. Ying Zheng is an Associate Professor in the Faculty of Humanities, University of Southampton. She worked as a Psychometrician and Director of Research in the Language Testing Division of Pearson, London before she joined Southampton. Her area of work covers language tests from China, Canada and the UK. She currently supervises MA/PhD students working in the fields of assessment literacy, large-scale testing and psychometrics.

Dr. Yanyan Zhang is a Professor in the English Department, Wuhan University, China. She specialises in World Englishes, Second Language Acquisition and Applied Linguistics. Dr. Zhang obtained her PhD in Linguistics and Applied Linguistics from the Hong Kong Baptist University in 2008. She was a visiting research scholar at the University of Southampton from 2013/2014.

CONTENTS

1. RESEARCH BACKGROUND	4
2. METHODS	5
2.1 Participants	5
2.2 Instruments	6
2.3 Procedures for data collection and analysis	6
2.4 Research questions	6
3. RESULTS	7
3.1 Descriptive statistics: Aptis performance	7
3.2 Comparability of computer-based and paper-based tests	10
3.3 Relationship between topics of the writing task and candidate experience	11
3.4 Test-taker perceptions of task difficulty	11
3.5 Contrast and comparison between Aptis and other tests	12
3.5.1 Aptis vs. TEM-4 & CET-4	12
3.5.2 Aptis vs. international tests	12
3.6 Suggestions to Aptis	13
4. CONCLUSIONS	14
4.1 Test-taker performance on Aptis	14
4.2 Aptis test-taker experiences	14
4.3 Test-taker perceptions of Aptis	15
REFERENCES	17
APPENDIX A: Questionnaire	19

Figures

Figure 1: Cumulative percentages of overall CEFR levels comparing English majors and non-English majors	9
Figure 2: Cumulative percentages of overall CEFR comparing females and males	10

Tables

Table 1: Participants' profile	5
Table 2: Descriptive statistics of test-taker performance on Aptis	7
Table 3: Score comparison by major	7
Table 4: Score comparison by gender	8
Table 5: CEFR levels (%)	8
Table 6: CEFR levels by major (%)	9
Table 7: CEFR levels by gender (%)	10

1. RESEARCH BACKGROUND

Aptis is a newly-launched international test developed by the British Council which aims to offer flexible, affordable, and reliable assessment of test-takers' English ability for practical purposes, such as recruitment and career development (O'Sullivan, 2012). There are 15 Aptis packages with different combinations of the four communicative skills, Listening, Speaking, Reading and Writing. Stakeholders are free to choose which skill(s) they want to test. Accordingly, the score report provides a score on a numerical scale (0–50) for the Grammar and Vocabulary section, and scores on a numerical scale (0–50) and a CEFR level (A1–C) for each communicative skill they have chosen to test (Aptis Brochure, 2012, p. 4). Compared with TOEFL and IELTS, Aptis is much less well known and less frequently used in China, with only four testing agencies currently available (Beijing, Shanghai, Guangzhou and Chongqing). Therefore, empirical studies on Aptis are relatively scarce.

The most popular national English tests in China are the College English Test Band 4 (CET-4) and the College English Test Band 6 (CET-6), with millions of test-takers using these tests annually (Jin, 2011; Zheng & Cheng, 2008). These two national tests are high-stakes tests usually associated with degree awarding and have often been used by employers for recruitment purposes (e.g. Jin & Cheng, 2013). However, neither CET-4 nor CET-6 seem to provide a clear reflection of test-takers' English ability for their potential jobs. Anecdotes and research have indicated that graduates with a certificate of CET-6 often fail to perform in jobs that require a proficient use of English in the workplace.

Under these circumstances, Aptis may offer an alternative for organisations and employers in China to select potential employees with the required level of English. However, to ascertain this, research needs to be undertaken to investigate stakeholder perceptions of its validity and practicality in the context of China. To the best of our knowledge, with the exception of Tang (2017) who investigated the usefulness of Aptis as a diagnostic tool in a tertiary context in China, studies exploring the validity and practicality of Aptis in China are not found in the literature.

As a new test, more evidence-based arguments need to be built to support a wider use of Aptis in China. Considering the current status of Aptis worldwide, it is important not only to gather more evidence on its construct validity, i.e. how it compares to other English tests of a similar nature, but also vitally important to obtain stakeholder perceptions, in particular, how Aptis compares with other national and international tests they are familiar with. Therefore, the aim of this study is to investigate the perceptions of potential Aptis test-takers on their test-taking experience, test constructs, test difficulty and test practicality.

The current research is based on O'Sullivan's (2011; 2012) model of test validation that was developed based on Weir (2005) and O'Sullivan and Weir (2011). This model consists of three basic elements: the test-taker, the test system and the scoring system. The current study focuses on the test-taker element by involving both the individual characteristics and cognitive resources. More specifically, we looked at the student test-taking experience, their knowledge of content and language abilities. By examining test-taker perceptions of the skills measured and their perception of the practicality of Aptis, taking into consideration the physical, psychological and experiential characteristics of the test-taker, the project relates the test-taker component to the test system and performance. The results of this study could help relevant stakeholders – including potential test-takers, test users and test developers – to understand Aptis in-depth and enrich their knowledge of the test in the context of China.

2. METHODS

In order to investigate stakeholder perceptions of Aptis, a survey design with follow-up interviews was adopted. A questionnaire (see Appendix A) was developed to examine test-taker perceptions in terms of their test-taking experience, test items, skills measured, and practicality of the test. For this purpose, students of different academic disciplines from Wuhan University were recruited to participate in the survey in September 2014. Since Chinese students are mostly unfamiliar with Aptis, a series of seminars was organised during September and October 2014 to introduce the test and provide help for them to take the test. In addition to the survey study, semi-structured interviews were conducted with nine randomly chosen students from the survey participants. The interview study aimed to explore participant perceptions of test constructs and the practicality element of Aptis in comparison to other high-stakes tests more widely used in China.

2.1 Participants

The participants of this study were all second-year students from Wuhan University, one of the leading universities in China. Since Aptis is a new test for Chinese university students, none of them had heard about or taken the test before the study. The researchers delivered a 45-minute long seminar on Aptis to second-year students and invited interested students to participate in the study in the first week of September 2014. The students were informed that they would be provided with training on Aptis to attend the test in October 2014. In return, the participating students were offered the following: 1) free Aptis training; 2) a free Aptis test (whole package, including all five components: Grammar and Vocabulary, Listening, Speaking, Reading, and Writing); 3) detailed test results; 4) evaluation of their test results; and 5) an incentive for their participation.

In all, 91 students were recruited, comprising 30 Science students, 31 Arts students, and 30 English-major students. They signed consent forms expressing their willingness to participate in the study and giving the researchers the right to use their data for research purposes. A few students did not turn up on the test day, resulting in a final number of 83 participants: 58 non-English major students and 25 English major students. These 83 participants consisted of 53 females and 30 males. All participants had been learning English for at least six years. Table 1 presents a profile of the participants.

Participants	Non-English Majors				English Majors		Total
Major	Science		Arts		English		
Gender	Male	Female	Male	Female	Male	Female	
Number	17	10	10	21	3	22	83
Average age	19.07		19.32		19.04		
Sum	58				25		83

Table 1: Participants' profile

2.2 Instruments

The instruments used for this project consist of:

1. the Aptis test: the British Council offered test codes for the students to take the test
2. a questionnaire on test-taker perceptions of Aptis (see Appendix A), in terms of their test-taking experience, test items, difficulty levels, skills measured, and practicality of the test (the questionnaire is based mainly on a five-point Likert scale and takes about 15 minutes to complete)
3. semi-structured interviews with nine randomly chosen candidates on their perceptions of test constructs and practicality in comparison to other high-stakes tests with which they are familiar.

2.3 Procedures for data collection and analysis

The data collection process took approximately four months from September to December 2014. As Chinese students were not familiar with Aptis, seminars were given to familiarise them with the test format by using the sample Aptis tests online. To remind the students of the testing procedure, a brief introduction was presented before the test started on the day of their test. After participants had completed the test, they were reminded to submit their answers online immediately.

The participants took the test in three groups: Science students, Arts students, and English major students. On average, the Grammar and Vocabulary component takes about 25 minutes to finish, the Listening component 50 minutes, the Reading component 30 minutes, the Writing component 50 minutes, and the Speaking component 12 minutes. In total, each participant took approximately 2 hours and 50 minutes to complete the Aptis test. When the participants finished their tests, a questionnaire was administered immediately to investigate their perceptions in terms of their test-taking experience, test items, skills measured, and practicality of the test.

After the test-takers had received their results, nine of them (three English major and six non-English major students) were invited to participate in a follow-up interview. Each interview was carried out individually with the interviewee sitting in front of a computer going over the Aptis test, item by item. Each interviewee answered questions on the basis of his or her reflections on the Aptis test and the test-taking experience. On average, each interview lasted approximately one hour. The interview data were later transcribed for in-depth analyses.

2.4 Research questions

The fundamental purpose of this study is to find out what test-takers think about the Aptis test and how this new test compares to other English tests in China with which they are familiar. The study addressed the following research questions.

1. What are the experiences of Aptis test-takers in terms of their perceived difficulty levels and opinions of taking a computer-based test of this nature?
2. What are test-taker perceptions of Aptis and what are the perceived differences between Aptis and other national or international tests in terms of test constructs measured and practicality?

3. RESULTS

3.1 Descriptive statistics: Aptis performance

The scores of the 83 participants were collected and analysed to gain a general picture of their performance on Aptis. The maximum score for each test component is 50. As Grammar & Vocabulary is considered to be a core skill in this test, the total maximum of the four communicative skill components is 200, as indicated in the test result reports. Table 2 shows descriptive statistics for the results. As can be seen, the Speaking average score is the lowest among the four skill scores. The reading average score is the highest, followed by Grammar & Vocabulary and Writing. Listening and Reading have higher standard deviations (SDs) than the other three test components, i.e., Grammar & Vocabulary, Speaking, and Writing.

	N	Range	Min	Max	Mean	SD
Grammar & Vocab	83	28	21	49	38.76	6.00
Listening	83	28	20	48	35.08	6.72
Reading	83	32	18	50	43.95	6.36
Speaking	83	33	10	43	27.02	6.07
Writing	83	27	21	48	38.41	6.03
Total	83	107	73	180	144.47	19.73

Table 2: Descriptive statistics of test-taker performance on Aptis

Table 3 shows the averages and SDs comparing English major and non-English major ($n = 25$ vs $n = 58$). Unsurprisingly, English majors outperformed non-English majors in all sub-scores, as well as in total scores. The former group also shows smaller standard deviations.

	Major	N	Mean	SD
Grammar & Vocab	English	25	42.0	5.0
	non-English	58	37.4	5.9
Listening	English	25	38.8	6.0
	non-English	58	33.5	6.4
Reading	English	25	46.3	4.5
	non-English	58	42.9	6.8
Speaking	English	25	30.8	4.2
	non-English	58	25.4	6.1
Writing	English	25	42.8	4.2
	non-English	58	36.5	5.8
Total	English	25	158.6	13.1
	non-English	58	138.4	19.0

Table 3: Score comparison by major

Table 4 shows the averages and SDs comparing female and male students ($n = 53$ vs $n = 30$). Female students outperformed their male counterparts in all reported scores. In terms of the spread of test performance, female students seem to cluster together more than their counterparts, except for Grammar & Vocabulary.

	Gender	N	Mean	SD
Grammar & Vocab	F	53	40.5	5.6
	M	30	35.7	5.5
Listening	F	53	37.0	6.1
	M	30	31.7	6.5
Reading	F	53	45.5	5.2
	M	30	41.3	7.3
Speaking	F	53	29.5	4.7
	M	30	22.7	5.9
Writing	F	53	40.2	5.3
	M	30	35.3	6.0
Total	F	53	152.1	15.6
	M	30	130.9	19.2

Table 4: Score comparison by gender

In addition to numerical score reporting, the Aptis score report also includes CEFR levels for each skill score (excluding Grammar & Vocabulary) and overall scores. Test-takers' CEFR level breakdown is summarised in Table 5. Overall, 59.52% of the cohort's level is at B2 as measured by Aptis, with variation in the skill levels. The highest percentages of the CEFR levels in each skill are highlighted in bold. To be specific, 41.67% of students were rated at B2 for Listening, 63.10% were at the C level for Reading, 50.00% were at B2 in Speaking, and 48.81% were at B1 for Writing.

	Listening	Reading	Speaking	Writing	Overall
A1		1.19	1.19		1.19
A2	1.19	1.19	4.76	1.19	1.19
B1	19.05	9.52	38.10	48.81	27.38
B2	41.67	25.00	50.00	35.71	59.52
C	38.10	63.10	5.95	14.29	10.71

Table 5: CEFR levels (%)

Table 6 and Figure 1 show a comparison of the CEFR levels of the English major group and the non-English major group. Two patterns can be observed from this comparison. Firstly, English major students' CEFR levels are, in general, one level higher than those from the non-English majors. Secondly, it seems that the test managed to differentiate among non-English majors' English proficiency levels better than those among the English majors. Non-English majors' proficiency covers four levels in Overall, Reading, and Speaking, instead of three levels among English majors. This result probably indicates the Aptis test is a better measuring instrument for general test-takers. English majors are rarely at A1 and A2 levels, except for 12% at A2 in Speaking. Of the non-English majors, 6.9% are at A1 and 50% are at A2 in Speaking. Neither group reaches the C level in speaking.

	Listening		Reading		Speaking		Writing		Overall	
	G1	G2	G1	G2	G1	G2	G1	G2	G1	G2
A1						6.90				
A2				1.72	12.00	50.00				1.72
B1	4.00	25.86	4.00	12.07	76.00	39.66	20.00	62.07	8.00	36.21
B2	32.00	46.55	16.00	29.31	12.00	3.45	40.00	34.48	60.00	60.34
C	64.00	27.59	80.00	56.90			40.00	3.45	32.00	1.72

Note: G1= English major group; G2 = non-English major group

Table 6: CEFR levels by major (%)

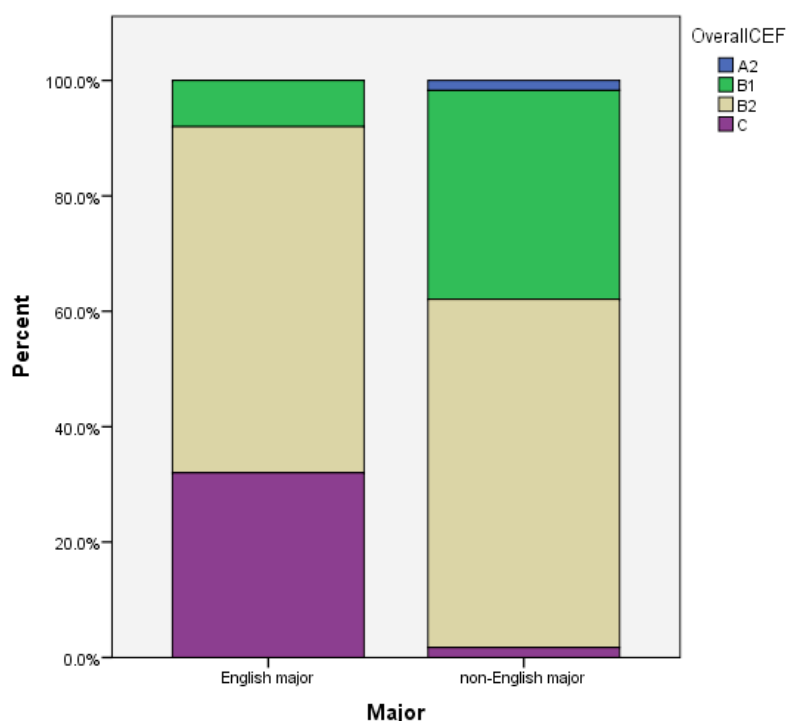


Figure 1: Cumulative percentages of overall CEFR levels comparing English majors and non-English majors

When comparing the test-takers' performance by gender, the picture looks very different. A comparison of the female and male groups is summarised in Table 7 and Figure 2. The male group covers four CEFR levels in Reading and Overall, while the female group covers three levels in those two scores. Most of the test-takers are one CEFR level lower in the male group than in the female group. There are no test-takers in the A1 levels, apart from 13.33% of the male group. Very few scores fall into the A2 levels. In contrast, no male test-takers achieved B2 and C levels in Speaking, while no female test-takers achieved the C level.

	Listening		Reading		Speaking		Writing		Overall	
	F	M	F	M	F	M	F	M	F	M
A1						13.33				
A2				3.33	26.42	60.00				3.33
B1	9.43	36.67	7.55	13.33	64.15	26.67	35.85	73.33	15.09	50.00
B2	41.51	43.33	16.98	40.00	9.43		45.28	20.00	69.81	43.33
C	49.06	20.00	75.47	43.33			18.87	6.67	15.09	3.33

Note: F= female group; M = male group

Table 7: CEFR levels by gender (%)

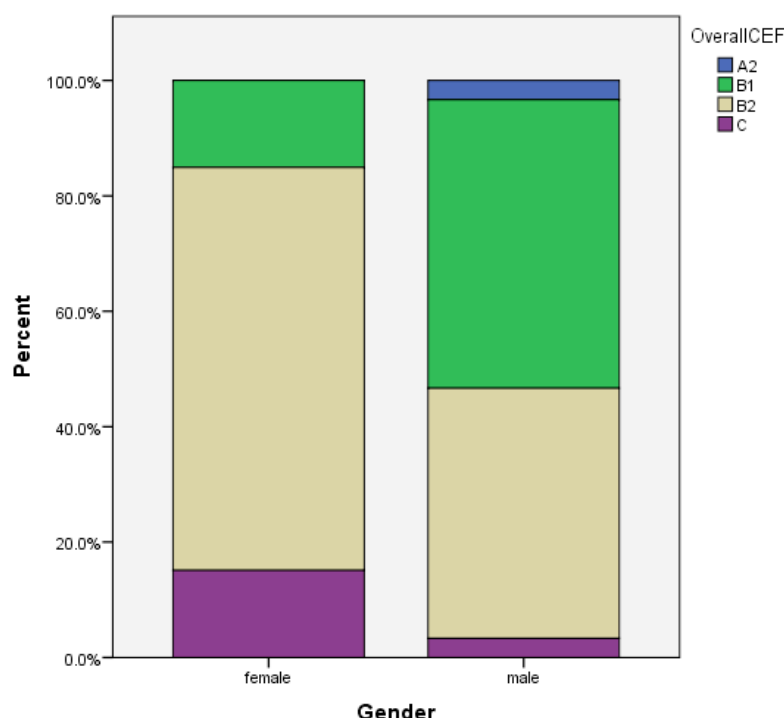


Figure 2: Cumulative percentages of overall CEFR comparing females and males

3.2 Comparability of computer-based and paper-based tests

Judging from the test-takers' responses to the open questions in the questionnaire, as well as from interview data of the nine interviewees, computer-based tests (CBT) have both advantages and disadvantages compared with paper-based tests (PBT). On the one hand, study participants stated that CBT embraces a series of merits. The test is "easy and convenient to operate, safe and confidential to save and upload candidates' answers, and paper-saving and environmentally friendly". Additionally, it enables candidates to "avoid nervousness and face-to-face embarrassment". Its "prompt feedback and appropriate item numbers and length" left them with a very good impression.

On the other hand, participants commented that they observed CBT has shortcomings as well. The test is “inconvenient to use computer to write and modify”, so candidates need time to accommodate to the online testing system and take a long time to be in a good state. It is also easy for candidates to “miss items or fail to record their responses” during the test. Moreover, it is likely that candidates find it “difficult to concentrate” when using a computer. CBT is especially inconvenient for those who are unfamiliar with computers and, thus, the test results are likely to “fail to represent candidates’ genuine English proficiency”, according to one participant.

Although test-takers in the study stated both advantages and disadvantages of CBT, they did not reach an agreement on whether or not CBT should be replaced by PBT. It is, therefore, difficult to tell whether CBT is preferable to PBT among this group of participants. Test-takers suggested that they would welcome an opportunity to voluntarily choose to take either CBT or the equivalent of PBT. One participant went further in expressing her opinion on the relationship between task types and testing modes and said that different task types may favour different testing modes:

What need to improve is that the computer-based test should be replaced by paper-based in terms of the writing task. Although computer-based test is open, fair and reasonable, it is hard to test students’ real writing competence since typing takes great efforts. (Interviewee LY)

3.3 Relationship between topics of the writing task and candidate experience

Test-takers appreciated that Aptis seems to give them a sense of “high authority” since it is developed by the British Council. They also felt the test items establish close links with daily life in that it appears to emphasise the application of English. As such, topics in each task are closely associated with British or European life. Amongst all tasks, the writing and speaking tasks are typically topic-oriented. Some test-takers claimed their unfamiliarity with topics in the writing task had a great impact on their performance. Some believed that test-takers were capable of finishing the writing task in a given context even if they were not familiar with the topic:

It is difficult for those who knew little about the topic like gardening to accomplish the writing task well. (Interviewee PZX)

3.4 Test-taker perceptions of task difficulty

Test-taker responses in questionnaires and interviews demonstrate that the difficulty level of Aptis varied depending on the test components. Specifically, they considered the speaking task to be the most difficult, but their perceptions of the task difficulty in the four other test components differed. One participant commented that if the difficulty scale is 1 to 5 from easy to difficult, Grammar and Listening are 2, Reading 1, Writing, 4 and Speaking is 5. (Interviewee LY). Interviewee HP commented that Listening and Grammar are easy, Reading is a bit difficult, and Writing is of medium difficulty. “I got a low score in the speaking task because I spoke too fast.” She also mentioned that, apart from the last item, there is no time provided for candidates to think about how to complete the tasks. Similar comments were expressed by several other participants. One of their main concerns is the lack of time given to prepare for the speaking tasks as can be seen from the quotes below.

Grammar and Vocabulary, Listening, Reading, and Writing sections are of moderate difficulties, but Speaking is difficult. (Interviewee PZX)

Listening is fairly easy, within my ability. Writing is not difficult either. On the whole, the test is not difficult. We only need time to get used to speaking. (Interviewee LRS)

The speaking task of Aptis is supposed to provide time for candidates to prepare. Some students regarded the 30 seconds as time for preparation, so no word is recorded. (Interviewee LY)

Speaking part doesn't give enough time for preparation. We can't use past fixed thinking mode to answer. It mainly tests your performance on the spot. (Interviewee QZH)

It is probably not convincing to argue that the difficulty level of the speaking task can be attributed to insufficient preparation time, but may be justified to conclude that there is a connection between preparation time in the speaking task and test-taker perceptions of task difficulty.

3.5 Contrast and comparison between Aptis and other tests

Participants indicate that Aptis is more difficult than CET-4 but easier than TEM-4, IELTS and TOEFL. Aptis shares several similarities with other tests but differs from them in a few facets. Test-takers mentioned that the other tests seem to be designed to test listening, speaking, reading, writing and other English competence integratively, while Aptis tests those skills separately. The final report can be regarded as a reference of a candidate's English competence as a whole. What is emphasised in test-taker responses are the differences between Aptis and other tests. In general, the Aptis test is more innovative and diverse. Differences also emerge in separate tasks.

3.5.1 Aptis vs. TEM-4 & CET-4

When comparing Aptis to local English tests in China, students thought that in the listening tasks, the speed of TEM-4's listening materials was faster and its texts longer. In addition, TEM-4 has dictation, unlike Aptis. As for reading, there are more item tasks in Aptis, but there are only reading and multiple-choice items in TEM-4. In terms of grammar and vocabulary, there is larger coverage, more knowledge points and more difficult content in TEM-4 than in Aptis. As for writing, there are more items in Aptis but candidates need to write more words in TEM-4. Candidates are required to write a text but in several paragraphs, which is appropriate to test text organisation and writing competence and is considered better than CET-4 and CET-6.

3.5.2 Aptis vs. international tests

When compared to other international English tests, students commented that Aptis tests basic English vocabulary and grammar in a single task while other international tests test more advanced vocabulary, and there is no single task for grammar since it is integrated with listening or reading tasks. International tests are aimed not only at the general public but people with an academic background. However, Aptis aims at non-English major students and focuses on practical writing. Compared to IELTS, participants commented that Aptis was more challenging. In IELTS, the examiner may ask a few easy lead-in questions in the speaking task, so examinees can gradually get used to the exam (*Interviewee QZH*). However Aptis offers no time to think and prepare. One participant stated that he is quite familiar with TOEFL and he made a detailed comparison between Aptis and TOEFL.

In Reading, the vocabulary range of TOEFL is around 10,000 to 13,000. Both two tests test vocabulary via synonyms but TOEFL is more difficult. In terms of speaking, TOEFL involves topics relevant to campus while Aptis covers everyday topics. TOEFL asks candidates to listen for a longer time and do a larger number of items. In TOEFL, preparation time was provided to all speaking items. There are six parts in TOEFL while four parts with several items in Aptis. TOEFL focuses on argumentation while Aptis focuses on narration. Both of them need good preparation in advance. In terms of writing, TOEFL provides topics for candidates to write while the other requires them to finish practical writing such as emails, letter. (Interviewee LY)

3.6 Suggestions to Aptis

Test-takers claimed that Aptis achieved “a large coverage” and “more various item tasks” in a general sense, which provided more room for candidates to demonstrate their English competence. The test is “original and comprehensive”, and is “useful to future life, work and communication”. Nevertheless, our participants still suggested a few areas for improvement.

In the vocabulary task, the vocabulary tested is easy but some options are ambiguous. Most of them are everyday words with less professional language involved. Participants claimed that, although words in the vocabulary task are easy, there are several words beyond their vocabulary range. They also pointed out that the test sometimes tests uncommon meanings of common vocabulary. One participant commented:

The grammar part is too easy and its coverage is too narrow. As for vocabulary, it is a bit difficult for those who just finished CET-4. It is better to be less difficult in vocabulary, more difficult in grammar, and have a broader coverage of knowledge points and item tasks. (Interviewee PZX)

They also thought there was no need for candidates to listen to the recordings for a second time in the listening task since it was very easy for them. They thought it would be better to lengthen the text of listening materials and add more items that can better test listening comprehension. One participant commented that the reading task does not test specific words in a text or the main idea of a text, and felt that it would be better to add more text reading to test reading competence in a comprehensive way.

The item design of the writing tasks is considered innovative, especially the last part that requires candidates to respond to an email. They think the design of the content and format is sensible. For improvement, an extension of the word count limit of the task would be needed.

Regarding the speaking tasks, the main complaint made by test-takers was that the speaking task had no preparation time. One participant suggested:

The test developers should consider posting the scoring criteria of Aptis on the Internet to help candidates prepare for the test. (Interviewee LY)

4. CONCLUSIONS

4.1 Test-taker performance on Aptis

The results of this study demonstrate that, on average, the speaking component of Aptis had the lowest score, while the reading component was the highest among all four skill sets of our participants. In terms of CEFR levels, the majority (59.52%) of participants achieved level B2 as measured by Aptis. More specifically, about 88.10% and 79.77% students achieved a level at and above B2 for reading and listening, respectively. In contrast, only 55.95% and 50% of students achieved a level at and above B2 for speaking and writing, respectively. As also reported in the questionnaire and interview, the participants stated that the speaking component was the most difficult, and reading the easiest among all the components, in line with their actual performance on the Aptis test. These findings suggest that our participants had a better performance in the receptive skills (reading and listening) than in the productive skills (speaking and writing) in Aptis. Such findings are in line with the performance of Chinese students in various communicative skills on other high-stakes national tests, such as CET-4 and CET-6 (e.g., Jin & Wu, 2010; Yang & Weir, 1998), and international tests, such as IELTS and TOFEL (cited from the official score reports).

In addition, it has been found that English majors outperformed non-English majors and the differences are significant. This result indicates that the Aptis test is probably a better instrument for measuring general test-takers' ability. English majors are rarely at A1 and A2 levels, except for 12% at A2 in speaking. Of the non-English majors, 6.9% are at A1 and 50% are at A2 in speaking. The English major students also reported that they feel the Aptis test is quite easy for them. The different performance on Aptis in terms of major can probably be attributed to the different English proficiency levels between the two groups of students. Although in the same university and at the same age, the two groups of students have undergone different study programs since they entered university. Compared with non-English major students, the English major students spend much more time learning and studying English language and culture. It would be rather unexpected if English majors were outperformed by non-English majors after two years of study in university. Nonetheless, neither the non-English major group nor the English major group reached the C level in speaking. The participants of the current study are only second-year university students in China, which may explain their relatively weak performance in the speaking component.

Further research is needed to investigate the performance of students with a more varied distribution of English proficiency levels in the Aptis test.

4.2 Aptis test-taker experiences

In contrast to many of the English tests in China, the Aptis test is computer-based. It was found that our test-takers had little experience in taking a computer-based English test. This finding is consistent with previous research. In a trial report on Aptis, O'Sullivan (2012) also found that less than 50% of the respondents had experience in taking an English test on a computer. As reported earlier, the test-takers in this study believe that CBT has both advantages and disadvantages compared with PBT.

On the one hand, CBTs embrace numerous merits such as being secure, environmentally friendly and allowing prompt score reports, thus confirming Noyes & Garland's (2008) contention that standardised CBTs provided quick and objective reports for candidates. On the other hand, CBT may influence candidates' genuine English proficiency, as shown in Hosseini, Jafre & Baghdarnia's (2014) study, in which some test-takers complained that their test score was not representative of their language proficiency because of their unfamiliarity with the CBT modes. Jin and Wu (2010) also found that computer familiarity exerts a significant influence on test-taker performance on the Internet-Based CET-4. However, Barkaoui (2014) found that the effect of keyboarding skills on test scores was weak and test scores are more strongly influenced by variables such as English language proficiency and task types.

Thus, there is no clear evidence demonstrating that CBT is less valid than PBT (Hosseini, Jafre & Baghdarnia, 2014). The test-takers in this study stated both advantages and disadvantages of CBT and suggested that candidates should be given an option to voluntarily choose the testing mode, if possible. O'Sullivan (2012) recommends that candidates should be monitored with regard to CBT experience and computer literacy experience.

It is obvious that amongst all tasks, the writing and speaking tasks are typically topic-oriented. Some test-takers claimed that their familiarity with topics in the writing task had a great impact on their performance. There are plenty of scholars who believe topic knowledge might influence test-taker perceptions and performance of writing tasks. McNamara (2001) doubted whether testing methods and test content should be linked with test-taker needs. Broadfoot (2005) questioned how insufficient topic knowledge and preparation in assessment practices might put a negative influence on test-taker learning and motivation. Test-takers believed it was better to diversify topics of the writing task, rather than merely confining them to British life. Powers and Fowles (1999) suggested high-stakes writing tests should use topics that enable test-takers to put what they know best into writing practice.

4.3 Test-taker perceptions of Aptis

In general, the test-taker —especially the English majors—did not perceive the Aptis test to be particularly difficult, except for the speaking component. Jin & Cheng (2013) found that test-taker perceptions of the test's value have significant effects on motivation and test anxiety, which has been confirmed as having a significant predictive power on CET-4 scores. As is clearly shown in the current study, the test-takers had never encountered Aptis before and did not associate the Aptis test with any high-stakes. In other words, for the test-taker, the Aptis test is not a high-stakes test, whereas CET-4, which the non-English major test-takers have to take, is a crucial test in their life. Their perception of Aptis may have influenced their performance on Aptis.

Test-taker responses seem to agree with the perception that the speaking task is difficult but diverge in their perceptions of the task difficulty in the other four tasks. This divergence raises the question of to what extent test-takers' perceptions of task difficulty match the actual difficulty of different tasks and the whole test. Chen (2006) found that there was a mismatch between test-taker perceptions of task difficulty and actual difficulty. It is recommended that the purpose of the test, types of test-takers, and other relevant factors should also be taken into consideration.

Test-takers reported that the main difficulty in the speaking task resulted from having too little preparation time. Researchers in previous studies have explored the effect of preparation time on tape-based tests. Participants in Brown's (1993) study mentioned a series of factors relevant to the difficulty of tape-based tests, among which was inadequate response times. Participants in Elder, Iwashita and McNamara's study (2002) reported that providing three minutes of 'planning' time was generally reckoned to make it easier to conduct the task than having no planning time regardless of proficiency. Brown's (1993) study further found that, in addition to preparation time difficulty, tape-based formats were also influenced by inadequate response time, unfamiliar vocabulary, speed of voices on the tape, lack of clarity in instructions, unclear prompts, too much input material to process and lack of familiarity with the task type. From the results of this study, it may not be convincing to argue that the difficulty level of the speaking tasks perceived by test-takers could be attributed to insufficient preparation time, but it is reasonable to state that there is a connection between preparation time in the speaking task and test-takers' perceptions of task difficulty.

In comparison with other national and international tests, the participants indicate that Aptis is more difficult than CET-4 but easier than TEM-4, IELTS, TOEFL and GRE. The Aptis test shares several similarities with other tests while differing from them in several facets. Test-takers mentioned that all of the tests seem to be designed to test listening, speaking, reading, writing and other English competence in a comprehensive way. Like IELTS, TOEFL, GRE and other international tests, the Aptis test is computer-based and tests comprehensive English competence via separate tasks. The participants recognised that Aptis could reflect their language ability.

Candidates also commented on their perceived differences between Aptis and other English tests. In comparison with other well-known tests in China, our participants considered Aptis to be more innovative, diverse and practical. As indicated clearly in the questionnaire and interviews, the Aptis test has been perceived to be more closely related to everyday life and communication, which is in line with the original purposes of Aptis. Some participants mentioned that the speaking and writing components are especially relevant to daily life. For example, one speaking task requires candidates to compare two pictures and discuss the similarities and differences between them. Such a task is seldom found in the national English tests in China according to the participants.

As is known, Aptis aims to offer a flexible, affordable and reliable assessment of English ability for practical purposes, such as recruitment and career development (O'Sullivan, 2012). In contrast, CET does not emphasise the practicality element. The major aim of CET, especially the spoken test, is to assess linguistic and communicative competence (Jin, 2000). Jin (2000) discussed the washback effects of the CET spoken English test (CET-SET) and investigated Chinese test-taker motivation for taking the CET-SET. Her study found that 79.6% of test-takers took the test in order to know their own English communicative competence, and only 17.9% students wanted to use the test for job-hunting purposes.

In conclusion, this study found that test-takers' perceived difficulty may be influenced by a few factors, such as the purpose of the test, types of test-takers and other relevant factors. Compared with other tests, our participants reported that the Aptis test is more practical and useful for the workplace and career development. These findings may prove valuable to its further development and marketing in China.

REFERENCES

- Aptis brochure. (2012). British Council 2012 / C447. Also available on www.britishcouncil.org/aptis
- Barkaoui, K. (2014). Examining the impact of L2 proficiency and keyboarding skills on scores on TOEFL-iBT writing tasks. *Language Testing*, 31(2), 241–259.
- Broadfoot, P. M. (2005). Dark alleys and blind bends: Testing the language of learning. *Language Testing*, 22, 123–141.
- Brown, A. (1993). The role of test-taker feedback in the test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing*, 10, 277–303.
- Chen, X. (2006). 陈水仙 (2006), 全国英语专业四级口试的任务难度与感知。广东外语外贸大学, (Test taker perception of task difficulty: A study on national English major band-4 speaking test, *Guangdong Foreign Language Institute*). Published in Chinese.
- Elder, C. Z., Iwashita, N. & McNamara, T. (2002). Estimating the Difficulty of Oral Proficiency Tasks: What Does the Test-taker Have to Offer? *Language Testing*, 19 (4), 347–368.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18, 333–349.
- Hosseini, M., Jafre, M. Z. A. & Baghdarnia, M. (2014). Comparability of Test Results of Computer Based Tests (CBT) and Paper and Pencil Tests (PPT) among English Language Learners in Iran. *Procedia – Social and Behavioral Sciences*. 98, 659–667.
- Jin, Y. (2000). 金艳, 2000, 大学英语四、六级考试口语考试对教学的反拨作用, 外语界, 第4期, 56–61页. (Washback effects of College English Test Band 4 & 6 speaking test. *Foreign Languages* 4, 56–61). Published in Chinese.
- Jin, Y. (2011). 金艳, 2011, 浅析标准化语言测试的教学评价功能, 外语测试与教学, 第1期, 26–34页. (Pedagogical functions of standardised language tests. *Foreign Language Testing and Teaching*, 1, 26–34). Published in Chinese.
- Jin, Y. & Cheng, L. (2013). 金艳、程李颖, 2013, 影响高风险考试效度的心理因素研究, 现代外语, 第1期, 62–69 (Affective factors that impact on the validity of high-stake tests. *Modern Foreign Languages*, 62–69). Published in Chinese.
- Jin, Y. & Wu, J. (2010). 金艳、吴江, 2010, 大学英语四级网考效度初探---影响考生评价和考试成绩的因素分析, 外语电化教学, 第132期, 3–10页. (An investigation of computer-based College English Test Band 4 – Factors influencing test takers' evaluation and performance. *Computer-Based Foreign Language Teaching*. 132, 3–10). Published in Chinese.
- Noyes, J. M. & Garland, K. J (2008). Computer- vs. paper-based tasks: Are they equivalent? *Ergonomics*. 51 (9), 1352–1375.
- O'Sullivan, B. (2011). Language testing. In J. Simpson (ed). *The Routledge Handbook of Applied Linguistics*. Oxford: Routledge.
- O'Sullivan, B. (2012). Aptis test development approach. Aptis Technical Report (ATR-1). British Council.
- O'Sullivan, B. (2012). Aptis formal trials feedback report. Aptis Technical Report (ATR-2). British Council.
- O'Sullivan, B. and Weir, C. J. (2011). Language testing = validation. In B. O'Sullivan (ed). *Language Testing: Theories and Practices*. Basingstoke: Palgrave Macmillan.

- Powers, D. E. & Fowles, M. E. (1999). Test-takers' judgments of essay prompts: Perceptions and performance. *Educational Assessment*, 6, 3–22.
- Tang, J. (2017). Assessment for learning: Using Aptis as a diagnostic tool in the EFL writing classroom. ARAGs Research Reports Online, AR-G/2017/4. Available from <https://www.britishcouncil.org/exam/aptis/research/publications/arags/assessment-learning-using-aptis>
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke: Palgrave Macmillan.
- Yang, H. & Weir, C.J. (1998). 杨惠中, Weir C, 1998, 大学英语四、六级考试效度研究。上海外语教育出版社. (Validity studies of College English Test Bands 4 & 6. *Shanghai Foreign Language Publishing*). Published in Chinese.
- Zheng Y. & Cheng L. (2008). College English Test (CET) in China. *Language Testing*, 25(3): 408–417.

APPENDIX A:

Questionnaire

Section I: Background Information

1. What is your gender? ☐ Male ☐ Female
2. What is your birth date? Month: Day: Year:
3. What's your major? _____
4. How long have you been studying English? ____ Years ____ Months
5. What are your scores of CET-4?
Total score _____ Grammar and vocabulary _____
Listening score _____ Reading score _____ Writing _____
6. Have you ever done an English language test on a computer before? ☐ Yes ☐ No

Section II: Aptis Test-taking Experience

In the following are some statements, please indicate whether you agree or disagree with the statements by choosing from numbers 1 to 5, corresponding to “strongly disagree” to “strongly agree”.

7. The look of the Aptis test was attractive and appealing. 1 2 3 4 5
8. The instructions were clear and easy to follow. 1 2 3 4 5
9. The Aptis test gave me an opportunity to show my true level of English. 1 2 3 4 5
10. On the whole, I had no difficulty in using the computer during the test. 1 2 3 4 5
More specifically, I had no difficulty in using the computer to take
 - the **reading** test. 1 2 3 4 5
 - the **listening** test. 1 2 3 4 5
 - the **writing** test. 1 2 3 4 5
 - the **speaking** test. 1 2 3 4 5
11. I would recommend this test to other people. 1 2 3 4 5

12. The main advantage of the Aptis test is

13. The main disadvantage of the Aptis test is

Section III: Perceptions of Aptis Test Items

14. I think the grammar part of Aptis is easy. 1 2 3 4 5
15. I think the grammar item type of Aptis is a good way to measure my English grammar. 1 2 3 4 5
16. I think the grammar tested in Aptis is useful for my future job. 1 2 3 4 5
17. I think the vocabulary part of Aptis is easy. 1 2 3 4 5
18. I think the vocabulary item types of Aptis are a good way to measure my English vocabulary. 1 2 3 4 5
- **Word usage** items are a good way to measure my English vocabulary. 1 2 3 4 5
 - **Word matching** items are a good way to measure my English vocabulary. 1 2 3 4 5
 - **Word combination** items are a good way to measure my English vocabulary. 1 2 3 4 5
 - **Word definition** items are a good way to measure my English vocabulary. 1 2 3 4 5
19. I think the vocabulary tested in Aptis is useful for work. 1 2 3 4 5
20. I think the reading part of Aptis is easy. 1 2 3 4 5
21. I think the reading item types of Aptis are a good way to measure my English reading ability. 1 2 3 4 5
- **Sentence comprehension** items are a good way to measure my English reading ability. 1 2 3 4 5
 - **Text cohesion** items are a good way to measure my English reading ability. 1 2 3 4 5
 - **Short text comprehension** items are a good way to measure my English reading ability. 1 2 3 4 5
 - **Long text comprehension** items are a good way to measure my English reading ability. 1 2 3 4 5
22. I think the reading items tested in Aptis are useful for my future job. 1 2 3 4 5
23. I think the listening part of Aptis is easy. 1 2 3 4 5
24. I think the listening item types of Aptis are a good way to measure my English listening ability. 1 2 3 4 5
- **Word and number recognition** items are a good way to measure my English listening ability. 1 2 3 4 5
 - **Literal meaning** items are a good way to measure my English listening ability. 1 2 3 4 5
 - **Inference meaning** comprehension items are a good way to measure my English listening ability. 1 2 3 4 5
25. I think the listening items tested in Aptis are useful for my future job. 1 2 3 4 5
26. I think the writing part of Aptis is easy. 1 2 3 4 5

27. I think the writing item types of Aptis are a good way to measure my English writing ability. 1 2 3 4 5
- **Word-level writing** items are a good way to measure my English writing ability. 1 2 3 4 5
 - **Short text writing** items are a good way to measure my English writing ability. 1 2 3 4 5
 - **Three written responses to written input** items are a good way to measure my English writing ability. 1 2 3 4 5
 - **Formal and informal text writing** items are a good way to measure my English writing ability. 1 2 3 4 5
28. I think the writing items tested in Aptis are useful for my future job. 1 2 3 4 5
29. I think the speaking part of Aptis is easy. 1 2 3 4 5
30. I think the speaking item types of Aptis are a good way to measure my English speaking ability. 1 2 3 4 5
- **Personal information** items are a good way to measure my English speaking ability. 1 2 3 4 5
 - **Description of picture and comparison with own situation** items are a good way to measure my English speaking ability. 1 2 3 4 5
 - **Describe, compare and speculate** items are a good way to measure my English speaking ability. 1 2 3 4 5
 - **Discuss personal experience or opinion in relation to an abstract topic** items are a good way to measure my English speaking ability. 1 2 3 4 5
31. I think the speaking items tested in Aptis are useful for my future job. 1 2 3 4 5

Section IV: Perceptions of Aptis Test Items in Comparison with CET-4

32. Compared with CET-4, Aptis is more difficult for me on the whole. 1 2 3 4 5
33. The grammar test of Aptis is more difficult than CET-4. 1 2 3 4 5
34. The vocabulary test of Aptis is more difficult than CET-4. 1 2 3 4 5
35. The reading test of Aptis is more difficult than CET-4. 1 2 3 4 5
36. The listening test of Aptis is more difficult than CET-4. 1 2 3 4 5
37. The writing test of Aptis is more difficult than CET-4. 1 2 3 4 5
38. The speaking test of Aptis is more difficult than CET-4. 1 2 3 4 5
39. Compared with CET-4, Aptis shows my true level of English better. 1 2 3 4 5
40. The grammar test of Aptis shows my true level of English grammar better than CET-4. 1 2 3 4 5
41. The vocabulary test of Aptis shows my true level of English vocabulary better than CET-4. 1 2 3 4 5
42. The reading test of Aptis shows my true level of English reading ability better than CET-4. 1 2 3 4 5
43. The listening test of Aptis shows my true level of English listening ability better than CET-4. 1 2 3 4 5

44. The writing test of Aptis shows my true level of English writing ability better than CET-4. 1 2 3 4 5
45. The speaking test of Aptis shows my true level of English speaking ability better than CET-4. 1 2 3 4 5
46. Compared with CET-4, I think Aptis is more useful for my future job. 1 2 3 4 5
47. The grammar test of Aptis is more useful for my future job than CET-4. 1 2 3 4 5
48. The vocabulary test of Aptis is more useful for my future job than CET-4. 1 2 3 4 5
49. The reading test of Aptis is more useful for my future job than CET-4. 1 2 3 4 5
50. The listening test of Aptis is more useful for my future job than CET-4. 1 2 3 4 5
51. The writing test of Aptis is more useful for my future job than CET-4. 1 2 3 4 5
52. The speaking test of Aptis is more useful for my future job than CET-4. 1 2 3 4 5

53. In your opinion, what are the major differences between CET-4 and Aptis tests?
Please be as specific as possible.

54. Have you ever taken an international test before, such as TOEFL, IELTS, etc.?
If yes, what tests have you taken?

55. In your opinion, what are the major differences between Aptis test and the international tests you have taken, such as TOEFL, IELTS? Please be as specific as possible.

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.

APTIS IN CHINA: EXPLORING TEST-TAKER PERCEPTIONS OF ITS TEST VALIDITY AND PRACTICALITY

AR-G/2017/6

Ying Zheng
Yanyan Zhang

**ARAGs RESEARCH REPORTS
ONLINE**

ISSN 2057-5203

© **British Council 2017**

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

