



How do test takers prepare for computerized speaking tests? A comparative study of the PTE Academic and the CET-SET

Jason Fan, University of Melbourne
Yan Jin, Shanghai Jiao Tong University
Xiaoxiao Kong, University of Melbourne
Xiaoyi Zhang, Fudan University

Language Testing Research Centre, University of Melbourne

Final report - May 2021

Table of Contents

Abstract.....	3
1. Introduction.....	4
2. Literature review.....	5
2.1 Preparing for language tests: Test takers’ preparation practices and strategies	5
2.2 The speaking subtest of the PTE-A and the CET-SET4.....	7
2.2.1 The PTE-A speaking subtest	7
2.2.2 The CET-SET4.....	8
3. Methodology.....	9
3.1 Participants.....	10
3.1.1 interview and focus group participants	10
3.1.2 Questionnaire participants	11
3.2 Instruments	14
3.2.1 Interview protocols.....	14
3.2.2 Questionnaires	14
3.3 Data collection.....	15
3.3.1 Interview and focus group data collection.....	15
3.3.2 Questionnaire data collection.....	15
3.4 Data analysis	16
3.4.1 Analysis of interview and focus group data.....	16
3.4.2 Analysis of questionnaire data.....	16
4. Results.....	17
4.1 Findings from interviews and focus groups (PTE-A)	17
4.1.1 Participants’ background.....	19
4.1.2 Participants’ preparation activities and strategies	20
4.1.3 Participants’ perceptions of the PTE-A speaking subtest	28
4.2 Findings from questionnaire survey (PTE-A)	33
4.2.1 Participants’ background.....	33
4.2.2 Test preparation activities and strategies.....	34

4.2.3 Participants' perceptions of the PTE-A speaking sub-test.....	38
4.3 Findings from interviews and focus groups (CET-SET4).....	40
4.3.1 Participants' background.....	42
4.3.2 Test preparation activities and strategies.....	43
4.3.3 Participants' perceptions of the CET-SET4	48
4.4 Findings from questionnaire survey (CET-SET4).....	50
4.4.1 Participants' background.....	50
4.4.2 Test preparation practices and strategies.....	51
4.4.3 Participants' perceptions of the CET-SET4	55
5. Summary and conclusions.....	57
5.1 Test takers' preparation practices and strategies on the PTE-A speaking subtest and the CET-SET4.....	58
5.2 Test takers' preparation practices and strategies on independent and integrated speaking tasks.....	62
5.3 Concluding remarks.....	64
References	65
Appendices	68
Appendix 1. Interviewee and Focus Group Background Questionnaire.....	68
Appendix 2 Interview Protocol (PTE Academic).....	69
Appendix 3. Interview Protocol (CET-SET4).....	71
Appendix 4 Survey questionnaire (PTE Academic).....	73
Appendix 5 Survey questionnaire (CET-SET4).....	79

Abstract

Despite the abundance of research on test preparation, few studies have focused on how test takers prepare for speaking tests, let alone a fully automated one. This study investigated and compared test takers' preparation practices and strategies on two computerised English speaking tests, that is, the speaking subtest of the Pearson Test of English – Academic (PTE-A) and the College English Test – Spoken English Test Band 4 (CET-SET4). Though both speaking tests are delivered via computer, they are different in terms of their task formats, speaking constructs, scoring methods, and test stakes. Two research questions were investigated in this study: 1) What are test takers' preparation practices and strategies on the speaking subtest of the PTE-A and the CET-SET4? 2) Do test takers engage in different preparation practices or adopt different preparation strategies on the independent and integrated tasks? To investigate the two research questions, this study adopted a sequential mixed-methods design, with a qualitative phase and a follow-up quantitative phase. During the qualitative phase of this study, we conducted interviews and focus groups with both the PTE-A (n = 18) and CET-SET4 test takers (n = 16). The qualitative data were coded and analysed thematically. The findings that emerged from the qualitative phase were then used to develop two questionnaire surveys, focusing on test takers' preparation practices on the PTE-A speaking subtest (n = 193) and the CET-SET4 (n = 257) respectively. Our findings indicate that test takers engaged in highly strategic test preparation practices when preparing for the PTE-A speaking subtest, including intensive practice of the test items in the item bank, memorisation of 'canned' responses or templates, and a strong emphasis on fluency which was perceived as the key to success on the PTE-A speaking subtest. Though most of these practices could be considered as *construct-irrelevant* which do not contribute to test takers' meaningful improvement on the target speaking constructs, they were perceived as highly effective in 'gaming' the machine and boosting their test scores on the speaking test. Given the perceived low-stakes of the CET-SET4, test takers did not report prolonged or intensive test preparation practices. Though construct-irrelevant test preparation also emerged in the data, test takers' preparation activities on the CET-SET4 seemed to be more aligned with the improvement of the target speaking constructs. Test takers indicated more positive perceptions of the integrated tasks in the speaking tests largely due to their perceived authenticity. Not surprisingly, preparations on these tasks were characterised by activities involving the language skills that are critical to success on an integrated speaking task (e.g., listening and speaking). This study further highlights the complexity of test preparation, suggesting that it could be significantly influenced by the language learning and testing context as well as the target test, in terms of, for example, the task formats, test stakes and scoring method. In addition, the findings also shed light on the validity of both the PTE-A speaking subtest and the CET-SET4, with implications for future revisions or refinement of both speaking tests.

1. Introduction

Test preparation is a recurring topic in test washback and impact studies, and has been examined from multiple perspectives, including test takers' preparation practices (Green, 2006), test preparation courses (Gan, 2009), the effect of test preparation on test performance (Hong & Peng, 2008), and teachers' perceptions and practices (Alderson & Hamp-Lyons, 1996; Lai & Waltman, 2008). Most test preparation studies have focused on high-stakes language tests because test takers and teachers are more likely to engage in special test preparation practices with a view to enhancing test takers' performance. Previous test preparation studies have targeted some international high-stakes language tests such as the Test of English as a Foreign Language (TOEFL, e.g., Alderson & Hamp-Lyons, 1996; Yu et al., 2017), the International English Language Test System (IELTS, e.g., Elder & O'Loughlin, 2003; Gan, 2009; Green, 2006), and the Pearson Test of English (PTE) Academic (PTE-A, e.g., Knoch, Huisman, Elder, Kong, & McKenna, 2020; Ma, 2017).

Despite the abundance of research on test preparation, few studies have focused on how test takers prepare for speaking tests (Fan & Yan, 2020; Yu et al., 2017), let alone a fully automated one. Farnsworth (2013) is the only study that we identified targeting an automated speaking test, that is, the Versant English Test (VET), developed by Knowledge Technologies Pearson, Inc. Nonetheless, the focus of this study was the effect of short-term coaching on test takers' performance. Commenting on the rise of computer-assisted language assessment, Chapelle (2008, p. 127) argued that 'test takers have needed to reorient their test preparation practices to help them prepare for new test items'. As such, it is essential to examine test takers' practices and strategies more broadly – not only targeted preparation practices such as short-term coaching – when they prepare for high-stakes computerised speaking tests. In addition to the paucity of research on test preparation for automated speaking tests, no research, to the best of our knowledge, has compared test takers' preparation practices and strategies on two high-stakes computerised speaking tests which differ in purpose, format, scoring methods, and indeed, the speaking constructs that are measured.

Another area which is worth exploring is how test takers prepare for integrated speaking tasks. Integrated test formats are increasingly used in international academic English tests because they are perceived as more effective than independent tasks in reflecting authentic language use in academic settings. It has also been argued that compared with independent tasks, integrated tasks are less sensitive to short-term coaching (e.g., Butler, Eignor, Jones, McNamara, & Suomi, 2000). Nonetheless, there is to date very limited empirical evidence regarding the coachability of integrated speaking tasks (Yu et al., 2017).

This study aims to examine and compare test takers' preparation practices and strategies on two computerised speaking tests: the speaking test of the Pearson Test of English Academic (PTE-A) and the College English Test – Spoken English Test Band 4 (CET-SET4). It focuses on the Chinese context where an unprecedented number of students are learning English at different levels and

taking various English language tests, both local (e.g., the CET) and international (e.g., the PTE-A). China has an exam-driven culture where high-stakes language tests play a prominent role in society (e.g., Yu & Jin, 2016), and where test preparation is extremely prevalent and has indeed developed into a huge and lucrative industry in recent years (Yu, et al., 2017). In addition, research has demonstrated that speaking is a language skill which proves to be a challenge for many Chinese students in terms of English language learning and test performance. (e.g., Jin & Zhang, 2016).

Though both the speaking subtest of the PTE-A and the CET-SET4 are delivered via computer, the two tests differ significantly with respect to purpose, format, scoring method, the speaking constructs that are assessed, and how test scores are used. The speaking subtest of the PTE-A is fully automated, and all its tasks can be considered integrated. The CET-SET4, on the other hand, is not fully automated; indeed, except the first task (i.e. *read aloud*), all other tasks in the CET-SET4 are rated by human raters, and most of its tasks are independent. More details about the two speaking tests are provided in the Literature Review section. Specifically, this study investigated the following two research questions:

- 1) What are test takers' preparation practices and strategies on the speaking subtest of the PTE-A and the CET-SET4?
- 2) Do test takers engage in different preparation practices or adopt different preparation strategies on the independent and integrated tasks when preparing for the speaking subtest of the PTE-A and the CET-SET4?

2. Literature review

2.1 Preparing for language tests: Test takers' preparation practices and strategies

There has been a dearth of research in the test preparation literature which investigates test takers' self-access test preparation practices (Alderson, 2004; Knoch, et al., 2020). Of the studies which focus on students' test preparation, the vast majority looked at all four skills (i.e. listening, reading, speaking, and writing), with very few studies focusing exclusively on the speaking subtest (Yu et al., 2017). Furthermore, the comparison between test takers' test preparation practices on integrated and independent speaking tasks, as well as automated and non-automated tasks, remains to be investigated.

Three studies to date have explored test takers' specific self-preparation practices for speaking subtests. Though none has discussed the differences in test preparation between integrated and independent tasks or between automated and non-automated tasks, their findings could still give insights into the various practices and strategies employed by test takers as well as how self-preparation practices could potentially differ depending on task characteristics. In her large-scale study, Liu (2014) sought to understand the relationship between test preparation practices and test performance on the TOEFL iBT through surveying 14,593 test takers in China. She found that

general English learning strategies, such as reading English books and watching movies in English, positively predicted total test scores on the TOEFL iBT as well as scores on the related subtests. In relation to the speaking subtest, test takers utilised both general test practices such as practising spoken English with a native English speaker, and test specific strategies such as practising speaking on topics similar to that of the speaking tasks in TOEFL iBT. Through multiple regression analyses, it was found that most test preparation practices employed by test takers in relation to the speaking subtest positively predicted speaking scores. Furthermore, some test preparation practices aiming to improve speaking scores also resulted in score increases in listening and reading, reflecting the integrated nature of the TOEFL iBT test tasks.

As integrated and automated test formats become increasingly commonly used, they have gradually become a point of interest for studying test preparation practices in recent years. To explore test preparation practices for the speaking subtest of the TOEFL iBT, which is both integrated and automated, Yu et al. (2017) elicited over 1,500 Chinese test takers' test preparation experiences through questionnaires and interviews and analysed these along with their TOEFL iBT scores. In addition to specific test preparation practices, Yu et al. (2017) also examined test takers' perceptions of the usefulness of the different practices. It was found that test takers perceived memorisation activities (e.g., memorising model essays) to be the least useful, while 'taking notes while listening or reading' and 'reading aloud' were considered the most useful (Yu et al., p. 14). When test takers' scores on the TOEFL iBT speaking test were taken into account, however, only one test preparation activity, namely, using TOEFL's official practice test, TOEFL Practice Online (TPO), was able to predict test takers' speaking test scores. Apart from test preparation strategies used, it was also found that the vast majority of test takers (75.4%) spent more time preparing integrated tasks than independent tasks (Yu et al., 2017, p. 13). These findings can be received as initial evidence suggesting the impact of integrated and automated task types on test takers' preparation practices.

The literature thus far has provided little comparison between test takers' preparation practices for integrated and independent tasks. While there is also almost no comparison between test preparation practices for automated and non-automated tasks, Knoch et al.'s (2020) study provided some further insights into how they might differ. They interviewed 60 recent repeat test takers of the fully automated PTE-A and analysed their test preparation practices thematically along with their score gains in the four language skills. Among the four skills, speaking stood out as it was the language skill in which the largest score increases were identified; furthermore, it was also in speaking that the most construct-irrelevant test preparation practices were found. These test preparation practices included making deliberate changes in fluency and pronunciation, as many test takers reported being informed by high-scoring friends or in coaching lessons to speed up their speech to gain higher scores (Knoch et al., 2020, p. 561). Additionally, a number of test takers stated that practising particular speaking task types (i.e. *describe image*) helped increase scores in speaking dramatically. Furthermore, in an attempt to understand how machine scoring works, many test takers also practiced through speaking into

voice recognition systems, such as Google. Taken together, it is evident that test takers' preparation practices for speaking will likely differ depending on the nature of the subtest, whether it be integrated or independent, automated or non-automated. The specific differences between test takers' preparation practices for different types of speaking subtests remain to be uncovered.

2.2 The speaking subtest of the PTE-A and the CET-SET4

2.2.1 The PTE-A speaking subtest

Launched globally in 2009, the PTE-A is used to measure test takers' ability to use English in an academic environment. Its results are used to make decisions about a student's readiness for academic study in an English-medium environment. The test is fully automated and administered on computers in Pearson VUE's test centres. The score report that a test taker receives includes an overall score, scores on each of the four language skills (i.e. listening, reading, speaking and writing), and scores on the enabling skills, including grammar, oral fluency, pronunciation, spelling, vocabulary, and written discourse. In the speaking subtest, test takers are asked to speak directly into the microphone. The total speaking time for each candidate is about 13 to 16 minutes.

As shown in Table 1 below, the speaking subtest of the PTE-A consists of five tasks: *read aloud*, *repeat sentence*, *describe image*, *retell lecture*, and *answer short question*. All five tasks are all designed in an integrated manner, requiring a synthetic operation of listening, reading, and speaking skills to produce speech content. For example, *repeat sentence* requires a test taker to listen to an English sentence and repeat it; *describe image* requires a test taker to read an image first before describing it orally in English. Test takers' spoken responses are machine scored. Two types of automated scoring systems are employed by the speaking subtest of the PTE-A: correct/incorrect system for the *answer short question* task, and partial credit system for the other tasks (Pearson, 2019). In terms of correct/incorrect scoring, a score of 1 point is assigned to a correct response, and a score of 0 is assigned to an incorrect response. In the partial credit system, the responses are scored as correct, partially correct or incorrect, depending on test takers' performances. The machine-generated raw scores on each task are processed via complex calculations to generate the composite score for the speaking subtest (Pearson, 2019).

Table 1. Test format of the PTE-A speaking subtest (adapted from Pearson, 2011, p. 2)

Task	Description	Preparation time	Speaking time	Skills assessed
1. Personal introduction*	Make a short self-introduction	25 secs	30 secs	Speaking
2. Read aloud	Read aloud a written text (each text up to 60 words, 6-7 items)	30-40 secs, depending on	Varies by item,	Reading & speaking

			the length of the text.	depending on the length of text.	
3.	Repeat sentence	Repeat a sentence after listening to its recording (each recording from 3 to 9 secs, 10-12 items)	None	15 secs	Listening & speaking
4.	Describe image	Describe an image (6-7 items)	25 secs	40 secs	Reading & speaking
5.	Re-tell lecture	Re-tell a lecture in one's own words after listening to or watching it (each lecture up to 90 secs, 3-4 items)	10 secs	40 secs	Listening & speaking
6.	Answer short question	Answer a short question after listening to its recording (each question from 3 to 9 secs, 10-12 items)	None	10 secs	Listening & speaking

2.2.2 The CET-SET4

The College English Test (CET), first administered in 1987, is a standardized test battery designed to measure the English proficiency of non-English major undergraduates in China mainland (Zheng & Cheng, 2008). The battery is comprised of the CET written test at two levels, that is, CET Band 4 and 6 (CET-4 and CET-6), and the CET Spoken English Test (CET-SET). Students usually take the CET-4 after they have completed the foundational stage of College English (CE) learning, typically at the end of the first year of their degree program, and the CET-6 after they have completed the more advanced stage of CE learning, typically at the end of the second year. The CET-SET was introduced as an optional component to the CET test battery in 1999, which was available only to those who had passed the CET written test with distinction.

The test format that the CET-SET adopted from 1999 to 2014 was the IELTS-style face-to-face interview, with one interlocutor, one examiner, and three students in a test session. To meet the increasing call for assessment of spoken English proficiency, the face-to-face CET-SET was replaced by a computer-based test in 2015 (Jin, 2019). Different from its predecessor, the computer-based CET-SET is available to all non-English undergraduates, and is administered at two levels: CET-SET4 and CET-SET6. Since the introduction of the computer-based CET-SET, it has been attracting about an increasing number of test takers). The test population reached one million before the outbreak of the Covid-19 pandemic in early 2020 (Fan & Frost, under review). In this study, we focused on the CET-SET4, as it has a larger test population as compared with the CET-SET6. As indicated in Table 2, four tasks are included in the CET-SET4, including *read aloud*, *questions and answers*, *individual presentation*, and *pair work*. Of the four tasks, two can be

considered as integrated as they entail the use of more than one language skill for task completion, including *read aloud* and *questions and answers*; the other two tasks in the test, that is, *individual presentation* and *pair work*, are independent tasks as they do not entail language skill integration.

Table 2. Test format of the CET-SET4

Task	Description	Preparation time	Speaking time	Skills assessed
1. <i>Read aloud</i>	Read aloud a text of about 120 words.	45 secs	1 min	Reading & speaking
2. <i>Questions and answers</i>	Answer two questions related to the short text in the <i>read aloud</i> task	None.	40 secs	Listening & speaking
3. <i>Individual presentation</i>	Present on a given topic or picture	45 secs	1 min	Speaking
4. <i>Pair work</i>	Discuss a given topic in pairs	1 min	3 mins for the pair	Speaking

In terms of scoring, the *reading aloud* task is automatically scored. The machine is trained to evaluate a test taker's performances on a five-point task-based scale. The other three tasks are all scored by human raters on a five-point analytic rating scale (National College English Testing Committee, 2016). Trained raters assign scores on the three dimensions of a) accuracy and range, b) size and discourse management and c) flexibility and appropriateness to evaluate a test taker's overall performance on the three tasks of *questions and answers*, *individual presentation*, and *pair work*. The total score for the CET-SET4 is 20 points (15 points from human rating and 5 points from automated scoring). Test scores are transformed into band levels of A, B, C, and D for score reporting. Passing certificates are awarded to test takers with a band level of C or above.

It should be noted, however, that the CET-SET4 has much lower stakes compared with the PTE-A, whose result is typically used to make high-stakes decisions such as admissions into English-medium universities or immigration (Wang, Choi, Schmidgall, & Bachman, 2012). The CET4 written test is high-stakes because various high-stakes decisions can be made of its results. For example, some universities in China mainland require students to pass the CET4 written test to be conferred the bachelor's degree. More controversially, applicants need to have a CET4 certificate when applying for permanent residential permits in first-tier cities such as Shanghai and Beijing (Jin, 2014). The CET-SET4, however, is not used for these purposes, though university students are generally encouraged to take the test.

3. Methodology

To respond to the two research questions, this study adopted a sequential mixed-methods design (e.g., Moeller, Creswell, & Saville, 2016), including a qualitative and quantitative phase (see Figure 1). During the first phase of this study (i.e. the qualitative phase), we conducted one-on-

one interview and focus-groups with the PTE-A and CET-SET4 test takers respectively. The purpose of Phase 1 was to explore test takers' preparation practices and strategies on the two speaking tests in focus. During the second phase (i.e. the quantitative phase), two survey questionnaires were developed based on the findings that emerged from the first phase of the study. The questionnaire surveys were aiming to investigate whether the findings derived from Phase 1 held for a much larger sample of the PTE-A and CET-SET4 test takers. In what follows, we provide details of the research methodology of this study.

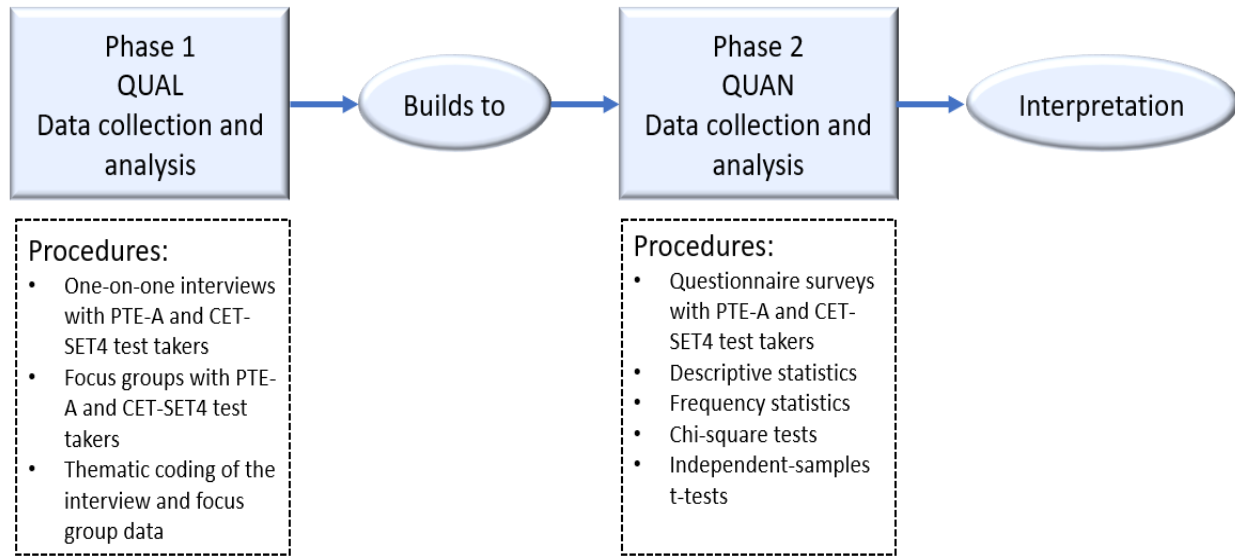


Figure 1. The procedural diagram of the mixed-methods design in this study

3.1 Participants

3.1.1 interview and focus group participants

During Phase 1 of this study (see Figure 1), the participants were 18 PTE-A test takers (coded as Participant PA to PR in this study), with 12 participating in one-on-one interviews and six in two focus groups (three in each group; see Table 3). The participants included 5 males and 18 females, their ages ranging from 16 to 32 years old ($M = 24.44$, $SD = 3.59$). When it comes to their educational background, half of them ($n = 9$) obtained or were studying the bachelor's degree at the time of the interview; seven obtained or were studying a master's degree; one was studying her doctoral degree, and another one was studying in a senior secondary school. The PTE-A participants were from a range of academic or occupational background, including, for example, law, finance, civil engineering, and education. All participants reported that they had studied English for over 10 years ($M = 15.89$, $SD = 3.63$). Twelve participants reported that they took the PTE-A once; the other six participants indicated that they had taken the PTE multiple times. Two participants, for example, reported that they had taken the PTE-A six times each.

In addition, 16 CET-SET test takers participated in the first phase of this study (coded as Participant CA to CP in this study), with 10 participating in the one-on-one interviews and six in

two focus groups (three in each group; see Table 4). The participants included five males and 11 females, their ages ranging from 18 to 19 ($M = 18.38$, $SD = 0.50$). Since only university undergraduates in China are eligible to take the CET-SET4, all CET-SET4 test taker participants were university students with half from national key universities ($n = 8$) and the other half from non-key universities¹ ($n = 8$). Their self-reported years of English learning ranged from six to 15 ($M = 10.38$, $SD = 2.28$), most likely because the participants came from different parts in China where students start learning English from quite different ages. The CET-SET4 participants were studying a range of majors at the time of investigation, including, for example, environment science, IT, and public administration.

3.1.2 Questionnaire participants

During Phase 2 of this study (see Figure 1), 193 PTE-A test takers completed the questionnaire, with 80 males (41.5%) and 113 females (58.5%). Most of the participants were within the age ranges of 19 to 24 ($n = 70$, 36.6%) and 25 to 29 ($n = 73$, 37.8%), with 37 over 30 years old (19.2%). Over half of the participants were studying at university at the time of data collection ($n = 120$, 62.2%), among whom 66 (34.2%) were studying their bachelor's degree, 51 (26.4%) their master's degree, and three (1.6%) their doctoral degree. The participants reported a wide range of years of experience in learning English (4-30 years, $M = 10.98$, $SD = 5.23$). Of the 193 participants, 71 (36.8%) reported that they had the experience of taking the IELTS, whereas very few of them reported the experience of taking the TOEFL before ($n = 6$, 3.1%). Sixty-five (33.7%) reported that they had the experience of taking the CET4, and 45 (23.3%) had the experience of taking the CET6.

Two hundred and fifty-seven CET-SET4 test takers completed the questionnaire, with 145 males (56.4%) and 112 (43.6%) females. The participants were within the age ranges of either 16-18 years old ($n = 178$, 69.3%) or 19-24 years old ($n = 79$, 30.7%). Most of the participants were at the first year of their university study at the time of data collection ($n = 242$, 94.2%). About half of the participants were from national or regional key universities ($n = 132$, 51.4%), and the rest were from non-key universities ($n = 125$, 48.6%). The CET-SET4 participants reported a similar length of English learning experience to the PTE-A participants ($M = 10.02$, $SD = 2.39$). Very few participants, however, reported that they had the experience of taking international language tests such as the IELTS or TOEFL; neither did they have experience in taking other English language tests in China, such as the CET4 and CET6. These findings are not entirely surprising, given that most of them were studying their first year of university.

¹ In China mainland, universities can be broadly classified into national key and non-key universities, with the former typically being allocated more educational resources; students at national key universities on the whole demonstrate a higher English proficiency than those at their non-key counterparts.

Table 3. Background of the interview and focus group participants (PTE-A)

Participant	Interview/FG	Gender	Age	English learning (yrs)	Education	Occupation/Major
PA	Interview	Male	25	12	BA	Law
PB	Interview	Female	32	20	PhD	PhD student (Ecohydrology)
PC	Interview	Female	27	20	BA	Finance
PD	Interview	Female	21	15	BA	Undergraduate (English)
PE	Interview	Male	25	20	BA	Teacher (Earlier education)
PF	Interview	Female	22	15	BA	Undergraduate (English)
PG	Interview	Female	30	15	MA	Accounting
PH	Interview	Female	27	20	BA	Civil Engineering
PI	Interview	Female	16	10	Secondary school	N/A
PJ	Interview	Female	27	18	BA	Accounting
PK	Interview	Female	23	16	MA	MA student (Media communications)
PL	Interview	Female	23	16	MA	MA student (Education)
PM	FG1	Male	26	10	MA	MA student (Information technology)
PN	FG1	Male	24	14	MA	MA student (Information technology)
PO	FG1	Female	23	16	MA	MA student (Information technology)
PP	FG2	Female	22	10	MA	MA student (Marketing communications)
PQ	FG2	Female	25	19	MA	MA student (Food science)
PR	FG2	Male	22	20	MA	MA student (Engineering)

Notes. FG = focus group

Table 4. Background of the interview and focus group participants (CET-SET4)

Participant	Interview/FG	Gender	Age	University	English learning (yrs)	Major
CA	Interview	Female	18	Non-key	11	Environment science
CB	Interview	Female	18	Non-key	7	Food science and engineering
CC	Interview	Female	18	Non-key	6	Medial science
CD	Interview	Male	19	Non-key	8	Machinery
CE	Interview	Male	19	Key	10	Environment science
CF	Interview	Female	18	Key	12	Environment science
CG	Interview	Female	18	Non-key	10	Machinery
CH	Interview	Male	18	Key	10	Public administration
CI	Interview	Male	18	Key	10	Public Administration
CJ	Interview	Male	18	Non-key	11	IT
CK	FG1	Female	19	Non-key	12	IT
CL	FG1	Female	18	Non-key	15	Environment science
CM	FG1	Female	18	Key	10	Environment science
CN	FG2	Female	19	Key	14	Economics
CO	FG2	Female	19	Key	10	Environment science
CP	FG2	Female	19	Key	10	Environment science

Notes. FG = focus group

3.2 Instruments

3.2.1 Interview protocols

During the first phase of this study, two interview protocols were developed to explore how test takers prepared for the speaking subtest of the PTE-A and the CET-SET4 (see Appendices 2 and 3). Before the interview, each participant was asked to complete a background questionnaire (see Appendix 1). The two protocols adopted a similar structure, encompassing three sections:

- 1) Background and test-taking motivation
- 2) Test preparation practices and strategies
- 3) Perceptions of the speaking test

In the first section, we asked the participants to self-assess their English language ability and to report their motivations/purposes in taking the PTE-A and the CET-SET4. In the next section, we asked them to report the activities and strategies in which they engaged when preparing for the two speaking tests. The last section of the interview focused on their perceptions of the difficulty level and face validity of the two speaking tests in focus.

For those who prepared the test on their own, we were interested in examining the learning resources that they used when preparing for the two speaking tests; in addition, we also asked them to report the activities and strategies that they engaged with when preparing for each task in the two speaking tests, as well as their perceived effectiveness of these activities and strategies in improving their test scores and English speaking ability. For those attending preparation classes, we asked questions relating to their motivations and expectations in attending a test preparation class, the materials that were used in test preparation lessons, the typical learning activities in class, and their perceived effectiveness of the test preparation class for improving test scores and English speaking ability. At the end of the interview, each participant was given the opportunity to comment on the strengths and weaknesses of the two speaking tests in focus as well as recommendations, if any, for the future improvement. The focus group discussions were guided by similar questions.

3.2.2 Questionnaires

During the second phase of this study, two questionnaires were developed based on the findings that emerged from the first phase of this study (see Figure 1). The two questionnaires are attached to Appendices 4 and 5, focusing on the PTE-A speaking subtest and the CET-SET4 respectively. The two questionnaires adopted a similar design, consisting of three sections:

- 1) Participants' background
- 2) Test preparation activities and strategies
- 3) Participants' perceptions of the speaking test

The first section of the PTE-A questionnaire includes a few items tapping into the participants' background, including their gender, age, occupation, academic background (if the participant is studying at the university at the time of data collection), the length of English learning, and their self-assessed English language ability. The next section centres on the participants' test

preparation activities and strategies, including rank ordering the time that they spent on each task in the speaking test, reporting the frequency that they engaged in a list of test preparation activities, as well as assessing the effectiveness of each activity for improving their test scores on the speaking subtest of the PTE-A. If a test taker reported that they had attended a test preparation class, they were asked to provide more details about the test preparation class, including the materials that were used, teachers' teaching style, and the perceived effectiveness of such class. The last section of the questionnaire taps into the participants' perceptions of the speaking subtest of the PTE-A, including their perceived difficulty level of the overall speaking test as well as each task in the test. In this section, the participants were also asked to indicate their attitudes towards several statements about automated and non-automated speaking tests.

As mentioned above, the CET-SET questionnaire adopted a similar design. The first section of the questionnaire adopted almost the same design to the PTE-A questionnaire, except that the participants were asked to provide their academic background and year of study at the university because all of them were university undergraduates. The second section of the questionnaire taps into the preparation activities and strategies in which the participants engaged when preparing for the CET-SET4, as well as their perceived effectiveness of each activity for improving their scores on the CET-SET4. The last section of the questionnaire asked the participants to report their perceptions of the CET-SET4, in terms of both the difficulty level and the face validity of the CET-SET4. For both the PTE-A and the CET-SET4 questionnaires, the preparation activities and strategies as well as the items on test taker's perceptions and attitudes were generated based on the findings that emerged from the qualitative phase of this study (see Figure 1).

3.3 Data collection

3.3.1 Interview and focus group data collection

The PTE-A participants were recruited with the assistance of our colleagues at Pearson Education for the interviews and focus groups. The CET-SET4 participants were recruited through our professional networks in English language teaching and testing in China. All interviews and focus groups were conducted on Zoom, a video-conferencing platform. Before the interviews or focus groups, we sent the interview protocol to the participants. All interviews and focus groups were conducted in Chinese, the participants' first language; however, they were allowed to use English as desired. With the participants' consent, the interview and focus groups sessions were audio-recorded and subsequently transcribed verbatim for analysis. Each interview lasted from half an hour to one hour; each focus group session lasted about one hour. At the time of data collection, all interview and focus group participants were either planning to take the two speaking tests in focus very soon or had finished them in the previous month. All of them confirmed that they could clearly remember how they prepared for the two tests in focus.

3.3.2 Questionnaire data collection

As indicated previously, the development of the questionnaires was informed by the findings that emerged from Phase 1 of this study. Due to the difficulty in recruiting PTE-A test takers, the

PTE questionnaire was not piloted. However, efforts were made to ensure that each item or statement in the questionnaire was clear to respondents. For example, after the questionnaire was developed, it was subjected to a rigorous review process undertaken by the project team members. It was also sent to ten colleagues working in the field of language assessment and/or English language education in China for comments, based on which the questionnaire was revised accordingly. The CET-SET questionnaire went through a similar iterative process of review and revisions. After that, it was piloted on 76 CET-SET4 test takers. The reliability (Cronbach's alpha) for the whole questionnaire was 0.92, suggesting very satisfactory internal consistency (Field, 2013).

Both questionnaires were set up in sojump.com, an online questionnaire distribution platform. The link to the PTE questionnaire was distributed to the PTE test takers with the assistance of Pearson Education, the owner of the speaking subtest of the PTE-A; the link to the CET questionnaire was distributed to the CET-SET4 test takers with the assistance of our networks in College English teaching and testing in China.

3.4 Data analysis

3.4.1 Analysis of interview and focus group data

The transcribed interview and focus group data were imported in NVivo 12, a qualitative data analysis program (QSR, 2012). Two researchers coded the data iteratively, using both closed and open coding methods to analyse the participants' responses to the questions in the interview protocols (Richards, 2014). The closed coding method was employed because all interviews and focus groups were semi-structured, and thus the guided questions provided the structure for the coding scheme; the open coding method was employed because other recurring themes were identified in the participants' responses which might shed light on their preparation activities and strategies but which were not covered by the interview protocol. The two researchers coded part of the data independently, including four interviews and one focus group for both the PTE-A and the CET-SET4 data, and then compared and discussed their coding results. Specifically, they discussed the coding scheme that they developed through this initial coding process. Based on their discussions, the coding scheme was revised. Inter-coder reliability was calculated (Cohen's kappa = 0.82), suggesting satisfactory inter-coder consistency. One of the researchers then finished coding the remaining data with this finalised coding scheme (see the 'Results' section).

3.4.2 Analysis of questionnaire data

Before we analysed the questionnaire data, a data cleaning procedure was implemented (Field, 2013). We removed some responses from the data set which clearly jeopardised its validity, such as those with the same option on every Likert-scale item. Following data cleaning, the data set was imported into SPSS 21.0 for analysis (IBM, 2012). Essentially, we used descriptive (mean and standard deviation) and frequency statistics (number and percentage) to explore the patterns in the participants' responses. For the items on a five-point Likert-scale, we combined the negatively

and positively oriented categories to explore the participants' overall orientations or attitudes. For example, when we investigated the participants' frequencies of engaging in test preparation activities and strategies, we combined the two categories of 'Never' and 'Seldom' into a broader category of 'Less frequent', and 'Often' and 'Always' into 'More frequent'. A non-parametric Chi-square test was then performed to examine whether there was a significant difference in the participants' frequencies in engaging in these activities and strategies. We also divided the PTE-A participants into two groups, that is, those attending test preparation classes and those engaging in solo preparation and employed independent-samples t-tests to compare the test preparation practices of the two groups of participants. A Bonferroni procedure was adopted to adjust the critical value to avoid Type 1 error (Field, 2013).

4. Results

In what follows, we present the findings from the two phases of this study. We present the findings from the qualitative phase of this study first before reporting the results from the quantitative phase (see Figure 1). The findings from both phases are then interpreted and integrated to address the two research questions of this project, that is, a) What are test takers' preparation practices and strategies on the speaking subtest of the PTE-A and the CET-SET4? b) Do test takers engage in different preparation practices or adopt different preparation strategies on the independent and integrated tasks when preparing for the speaking subtest of the PTE-A and the CET-SET4?

4.1 Findings from interviews and focus groups (PTE-A)

Table 5 below presents the categories and subcategories that were identified through the thematic coding the interview and focus group data about the speaking subtest of the PTE-A, as well as the topics that were included under each subcategory. As indicated in this table, three broad categories were extracted from the data: a) participants' background; b) test preparation activities and strategies; and c) participants' perceptions of the speaking subtest of the PTE-A. Next, we elaborate on each of the three categories.

Table 5. Categories and subcategories extracted from the data (PTE-A)

Category	Subcategory	Topic
1. Background	A. Purposes of taking the PTE-A	1-A-a. Study overseas 1-A-b. Immigration
	B. Selection of the PTE-A over other tests	1-B-a. Difficulty level 1-B-b. Fairness and efficiency in scoring
	C. English learning context	1-C-a. Self-assessed English ability 1-C-b. Opportunities to practice English speaking
2. Test preparation activities and strategies	A. Types of preparation	2-A-a. Solo preparation 2-A-b. Preparation school/class
	B. Preparing for each task	2-B-a. Read aloud 2-B-b. Repeat sentence 2-B-c. Describe image 2-B-d. Re-tell lectures 2-B-f. Answer short question
	C. Time management	2-C-a. Tasks with most time commitment 2-C-b. Tasks with least time commitment
3. Perceptions of the speaking subtest of the PTE-A	A. Perceived difficulty	3-A-a. Perceived difficulty of each task 3-A-b. Factors that contributed to task difficulty
	B. Perceived validity	3-B-a. Overall perceptions 3-B-b. Comparing computerised with live speaking tests
	C. Likes and dislikes	3-C-a. Perceived strengths of the test 3-C-b. Perceived weaknesses of the test

4.1.1 Participants' background

The participants reported that they took the PTE-A for either studying at an English-medium university overseas where the PTE-A was recognised as proof of English language proficiency, or meeting the English language requirements of the immigration departments of the countries to which they were planning to emigrate. In addition, the participants indicated several reasons why they selected the PTE-A over other English language proficiency tests in the market, among which the two most important ones were a) the PTE-A was perceived as overall less challenging as compared with some other English language tests; and b) the automated scoring practices of the PTE-A were perceived as fairer and more efficient. Excerpt 1 below illustrates this finding:

Excerpt 1

Researcher: Why did you select the PTE-A?

PA: Because I felt it was easier than some other tests.

Researcher: Could you elaborate a bit?

PA: The PTE-A is not as developed in China as some other tests such as the IELTS. It is not as interactive as the IELTS because our performance is scored by machines, not human raters.

Researcher: Does that mean machine scoring was an important factor when you selected the PTE-A?

PA: Yes, that's correct.

(PA, Interview 1)

As indicated by this excerpt, this participant (PA) perceived the PTE-A as overall less challenging as compared with some other language tests such as the IELTS, primarily because the test design was not as interactive as some other tests which were rated by human raters. This perception, in consequence, prompted him to select the PTE-A over similar language tests in the market. Another important reason is that the automated scoring method that the PTE-A adopted was perceived to be fairer and more efficient, as compared with human raters. In other words, the automated scoring method seems to have a significant impact on test takers' decision to select the PTE-A.

In this section, the participants were asked to self-assess their English language proficiency, in terms of the four skills of listening, reading, writing, and speaking. Over half of the participants indicated that they were stronger in reading but weaker in speaking. The reason was mostly attributed to their English learning context where reading tended to be given more emphasis than speaking, and where English teaching and learning were dominated by ensuring students' success on various English language tests, with the improvement of their communicative English language ability often side-tracked. Most participants indicated that they didn't have many opportunities to practice English speaking in their everyday life, except when they were preparing for the speaking subtest of the PTE-A.

4.1.2 Participants' preparation activities and strategies

4.1.2.1 Attending preparation schools or classes

About half of the participants reported that they attended test preparation schools or classes when preparing for the PTE-A, whereas the other half prepared on their own (i.e. solo preparation). The preparation classes were in either one-on-one or group format, with the latter varying significantly in terms of their sizes, ranging from a few to over a hundred students. Some of these classes were streamed online, whereas others were delivered face to face. Those who attended the test preparation classes indicated that they were hopeful of getting more familiar with the test with the assistance of the teachers, and of acquiring test-taking skills so that they could boost their scores on the PTE-A, as indicated in Excerpt 2 below.

Excerpt 2

Researcher: Why did you choose to attend the test preparation class?

PM: At that time, I wanted to get the test preparation materials that the preparation school had prepared. I had already taken the PTE-A for once or twice before attending the preparation class; however, I felt that I was ill-equipped with the test-taking skills. Before taking the PTE-A, I was focusing on the IELTS, which differed from the PTE-A in significant ways. Therefore, I thought the preparation school could help me get well prepared for the PTE-A.

Researcher: What did you expect to learn from the test preparation class?

PM: I was very keen to learn some test taking skills and strategies.

(PM, Focus Group 1)

According to the participants, most preparation schools developed their own teaching materials, including authentic test items from previous PTE-A administrations, the skills and strategies to tackle each type of task in the PTE-A, and some canned or template responses. The participants also reported that almost all test preparation classes were geared towards boosting a student's score on each task in the test. The coaching of template use was often the focus of such preparation classes. A typical test preparation class would start with the teacher explaining the templates for a task in the speaking test and demonstrating how to use them with a few sample tasks. Next, students would be asked to practice applying these templates to authentic test items. After that, teachers would provide feedback to students on their use of the templates. Overall, those who attended the test preparation classes spoke positively of them, believing that they were effective in helping them improve their test scores on the speaking subtest of the PTE-A; however, when asked whether these classes also improved their English speaking ability, most participants commented that they were much more geared towards improving test scores than improving their English speaking ability. In the following excerpt, PP, a focus group participant who attended a small-size preparation class with three students, reported the teaching style.

Excerpt 3

Researcher: How did your speaking teacher teach?

PP: Usually, he would show you how to use the templates. Basically, he delivered the teaching based on the task type in the PTE-A speaking test.

Researcher: Does that mean he demonstrated how to use these templates for different tasks?

PP: That's correct. Every student had plenty of opportunities in class to practice the templates, and we practiced using the templates for each task in the PTE-A speaking test. The teacher would give us feedback on whether we used the templates correctly or effectively. We usually used authentic test items for our practice.

Researcher: How did you like it?

PP: I feel it was highly effective.

(PP, Focus Group 2)

As illustrated in this excerpt, the preparation class was characterized by the following procedures: a) the teacher demonstrated the templates that they had developed for each task in the test; b) students practiced applying the templates to authentic tests; and c) the teacher provided feedback to students on how to improve the use of the templates. Based on the teaching style, it can be quite safely concluded that the preparation class focused much more on improving students' test scores than their spoken English ability.

4.1.1.2 Solo preparation

Those who prepared on their own reported that they referred to a wealth of test preparation resources that were available on the Internet or applications on their mobile devices, including the bank of authentic items (known as *ji-jing*) or simulated items. Specifically, the participants mentioned the following websites which were very popular among PTE test takers:

- Alpaca Education (*yangtuo*, <https://www.ytaxx.com/>)
- BlackTech (*heikeji*, <https://ikkedu.com/>)
- APEUni (*xingji*, <https://www.apeuni.com/>)
- FireflyEnlightenment (*yinghuochong*, <https://www.fireflyau.com/>)

Some participants mentioned that as part of their solo preparation, they watched the PTE coaching videos on YouTube (<https://www.youtube.com/>) or Bilibili (<https://www.bilibili.com/>), a popular video sharing website in China mainland. Zhihu (<https://www.zhihu.com/>), a popular question-and-answer website in China, was also mentioned where previous PTE-A test takers shared their test taking experience and prospective test takers could ask questions about the PTE-A. Some other participants also mentioned that they visited the Pearson official website (<https://pearsonpte.com/>) to search for related materials or practice test papers as part of their solo preparation.

It was common for a test taker to use multiple websites or mobile applications when preparing for the PTE-A speaking subtest because according to the participants, different websites or mobile

applications seemed to have their own unique strengths. For example, PL, an interview participant, recalled that ‘I referred to FireflyEnlightenment (*yinghuochong*) for the items that are likely to appear in the upcoming PTE-A test because they have a good reputation in predicting test items; however, I used the APEUni (*xingji*) on my mobile phone because it can rate my performance on each task in the PTE-A speaking subtest and provide me with detailed feedback.’ The Pearson official website was only sporadically mentioned in the interviews and focus group discussions. One participant (PA), for instance, mentioned that he did the practice tests on the Pearson website because he believed that it was the ‘most authoritative source of information’.

Despite a wide array of online resources which the participants reported that they would draw on, their primary motivations of engaging with these resources were a) familiarising themselves with the task formats in the PTE-A speaking subtest; b) getting more practice with authentic or simulated test items; and c) developing their test taking skills and strategies; in particular, their use of the templates, which they believed was conducive to boosting their scores on the PTE-A speaking subtest. When asked about their perceived effectiveness of the online resources that they drew on in their solo preparation, most participants indicated that they were very or somewhat useful. For example, PR, a focus group participant, indicated that these resources were ‘immensely helpful’ to her and she was ‘brimming with confidence after practicing the authentic items online for a week.’ Another interview participant, PC, however, commented frankly that ‘they were mainly oriented towards improving one’s test scores, with little effect on spoken English ability.’

4.1.2.3 Preparing for each task in the PTE-A speaking subtest

As mentioned previously (see Section 2.2.1), five tasks are used in the PTE-A speaking subtest, including *read aloud*, *repeat sentence*, *describe image*, *re-tell lecture*, and *answer short questions*. In what follows, we present the findings about how the participants prepared for each task in the PTE-A speaking subtest.

Reading aloud

Intensive reading aloud practice was repeatedly mentioned in the interviews and focus groups discussions. PM, for example, mentioned that when he was preparing for this task, he read aloud the texts from the item bank for about 20 minutes every day, and he believed that ‘practice makes perfect; more practice will make it easier for you to control your read aloud performance, and hence you wouldn’t feel very nervous when taking the test.’

The participants also mentioned that when they prepared for the *read aloud* task type, many of them used applications on mobile phones or tablets which could rate their read aloud performance and provide immediate and detailed feedback on aspects such as fluency, accuracy, and stress. Such feedback was considered very valuable because it helped the participants identify their strengths and weaknesses in their read aloud performance; as a result, they were made better aware of their weaknesses, which could be addressed through future remedial actions. The excerpt below from PR illustrates this point.

Excerpt 4

Researcher: How did you prepare for *read aloud*?

PR: I practiced about 10 texts every day, and I used an app on my mobile phone to rate my performance.

Researcher: What was the rating like? Was it useful?

PP: It was quite useful because it provided a total score as well as scores on several aspects of my reading aloud performance, such as pronunciation, fluency, and accuracy. If I didn't do well in an aspect, I would pay more attention to it next time.

(PR, Focus Group 2)

As this excerpt shows, by providing diagnostic feedback on PR's reading aloud performance, the mobile application helped to raise his awareness of the weaknesses in his reading aloud performance. As a result, he would try to take some actions to address them when attempting this task in the future.

Several participants also mentioned that to get a high score on this task, fluency was extremely important because their performance was rated by machines, which prioritized fluency over other aspects of the performance. PB, for instance, commented that it was essential not to pause when attempting this task, and 'if there is an unfamiliar word, just skip it because it is vital to make sure that your speech is very fluent, without any pauses.'

Repeat sentence

When it came to the second task, *repeat sentence*, the participants again reiterated the importance of intensive practice. PK mentioned that her teacher asked them to 'allocate time specifically to practice the items in this task; one should practice at least 100 to 200 sentences a day'. PL reported that she selected 200 'high-frequency' sentences and read them carefully every day. By 'high-frequency' she meant the sentences which were predicted to have a high likelihood of appearing in future PTE-A speaking subtests.

Like the first task, *read aloud*, fluency was mentioned as the key to success on this task. PK recalled that her teacher advised them to 'repeat a sentence even if you fail to understand it – you can pretend that you have understood, but it is important to sound fluent. If you can understand a few words in the sentence, just repeat these words confidently.' Echoing what PK said about her teacher's advice, PC, another participant, recalled the advice from her teacher on how to perform well on this task (see Excerpt 5).

Excerpt 5

Researcher: In addition to intensive practice, did your teacher give you any advice on *repeat sentence*?

PC: Yes. Our teacher told us that if we found it difficult to remember the whole sentence, we could just focus on the first or second half of the sentence. He said if we could focus on pronunciation and fluency at the sacrifice of content which wouldn't affect our scores very significantly anyway.

Researcher: Have you tried whether his advice worked?

PC: Surely it worked. I don't have a good working memory to remember the whole sentence. So, I found his advice very useful.

(PC, Interview)

It was also frequently mentioned that listening ability was very crucial to this task, which is not entirely surprisingly because test takers are supposed to understand the sentences before repeating them. Several participants mentioned that they were trying to strengthen their listening ability when preparing for this task. The excerpt below explains how PD prepared for this task.

Excerpt 6

Researcher: How did you prepare for *repeat sentence*?

PD: I spent quite a lot of time doing intensive listening training to reinforce my listening ability.

Researcher: What do you mean by 'intensive listening'? Could you explain it?

PD: By intensive listening I mean I would find some texts from the listening sections of other English tests such as the IELTS. Then I listened to these texts sentence by sentence. After listening to each sentence, I tried to write it down.

(PD, Interview)

PG, another interview participant, reported quite similar preparation practices. She argued that strong listening skills were critical to ensuring success on this task. She reported that her strategy was to integrate the preparation of this task into that for the listening subtest of the PTE-A. Different from PD who wrote down the sentences that she had heard, PG paraphrased the texts after hearing them. She noticed that 'the practice was quite effective; once your listening ability becomes stronger, it is easier to complete the task of *repeat sentence*.'

[Describe image](#)

A recurring theme that emerged from the data about how the participants prepared for *describe image* is the use of template responses or *muban* in Chinese. The participants almost unanimously reported that effective use of templates was vital to the success on this task. PR mentioned in the focus group discussions that 'when I saw some images at the beginning, I felt at a loss about what

to say, particularly given the limited preparation time. Later, I learned templates through visiting some PTE-A test preparation websites. I must acknowledge that they indeed made my life much easier.' PK remarked that the importance of templates couldn't be overemphasised for this task. She commented that 'if one can memorise the templates, he or she shouldn't have many troubles with this task. The templates can be applied to different types of graphs, even including some maps.' Given the effectiveness of templates, many participants reported that their preparation for this task mainly involved the memorization of templates and the intensive practice of applying the templates to authentic test items.

When asked whether they were concerned about being penalised for using templates, the participants argued that since their performance would be rated by machines, it was unlikely that they would be penalised for using templates; however, human raters might easily detect their use of templates and therefore penalised them. Excerpt 7 below explains this finding.

Excerpt 7

Researcher: How did you prepare for *describe image*?

PF: Templates are important for this task.

Researcher: What do you mean by 'template'?

PF: By 'template' I mean some response patterns which can be used to describe most input images, usually with a beginning, body and conclusion.

Researcher: Can you give me one or two examples?

PF: For example, one can always start with 'this picture describes the information about...'. Here you need to add the title or caption of the image; after that, you can talk about what is on the horizontal and vertical axis and make some comparisons. Well, we were asked to memorise different sentence patterns for different types of input graph such as a bar or line chart.

Researcher: Were you worried about being penalised?

PF: No. Because PTE-A is scored by machines, you wouldn't be penalised for using the templates because machines don't care; however, it is likely that you are penalised in the case of the IELTS, which is rated by human raters. They can discern that you are using templates.

(PF, Interview)

In addition to the use of templates, it was again mentioned that 'fluency' was very important when it comes to describing images. As PO, one of the focus group participants, reiterated, 'fluency is very important. You should just keep talking and don't pause, and then you should be okay.' The reason why fluency was important was again attributed to the machine scoring practice of the PTE-A, as PO commented, 'because the machine prioritises fluency over other criteria; keep talking, even if what you said is probably not correct'.

[Re-tell lecture](#)

Regarding *re-tell lecture*, the participants mentioned that listening ability was crucial; therefore, some of them spent quite a lot of time reinforcing their listening skills as part of the preparation for this task. As PF commented, 'I feel this task is challenging because first of all, I have to understand what the lecture is about. Maybe understanding the lecture is only the first step, I have to summarise it.' This finding can also be evidenced by the excerpt below:

Excerpt 8

Researcher: How did you prepare for *re-tell lecture*?

PD: Our teacher asked us to focus on listening. I listened to the BBC news every day to strengthen my listening ability.

Researcher: Did you also try to retell what you had heard?

PD: Yes, our teacher asked us to summarise the main ideas after we listened to the news. Usually, I would write a summary of the main ideas; then I would speak about it. I feel it was quite helpful.

(PD, Interview)

Templates were also mentioned as important to the success on this task. The participants reported that they usually needed to identify the topic sentences and key words in the lecture; then, they would put these details that they had identified into the templates that they had memorised for this task. As PR commented, 'the machine will be looking for the key words when rating my performance; if the key words in the rating system match those that I have identified, it is likely that I will get a high score on this task.' When commenting on the effectiveness of templates for this task, PB said, 'I think templates are necessary for this task because I would be at a loss about what to say and what not to say within the time limit; the template helps me put together the key words in the lecture'. This finding is further elaborated by PR in the excerpt below.

Excerpt 9

Researcher: How did you prepare for *re-tell lecture*?

PR: I memorised the templates. When I listened to the lecture, I would try to take notes of some key words and key points in the lecture. For example, I was listening to a professor giving a lecture on a certain topic, for example, 'smoking'. After listening to the lecture, I would be able to take notes of about a dozen key words. Then I would put four words into the first part of my template, and another four words into the second and third part of the template. Then when it was my time to speak, I would say 'in this lecture, the professor talks about...smoking...bad for our health'. This is the first part. In the second part, I would say 'he speaks more about the harms of smoking, for men, women and children.' Just like this, as you see, the key words...

Researcher: Okay. Talk around the key words.

PR: Yes, the key words. They are so important because the machine can't evaluate your logic; it can, however, detect the key words in the lecture.

(PR, Focus Group 2)

Answer short question

Several participants mentioned that since this task did not contribute much to the speaking and overall scores, they didn't spend much time preparing it. For instance, PG remarked that 'the items in this task have very little weighting; therefore, it doesn't make much sense to spend a lot of time on preparing this task. For me, I just tried my best to get the items that I could understand correct; I wouldn't worry too much about the items that I failed to understand.'

For those who prepared, listening ability was mentioned as important because to answer a short question, one had to understand it in the first place. PA used an example 'what does ice turn into when it melts' to illustrate the importance of listening ability: 'if you don't get 'ice' and 'melt', it would be quite impossible to get this question correct, though the sentence itself is admittedly very short.' Another recurring theme that emerged from the data is the intensive practice of authentic items. PM, for instance, remarked that he practiced over 1,000 items of *answer short question* (see Excerpt 10 below):

Excerpt 10

Researcher: How did you prepare for *answer short questions*?

PM: I practiced all the authentic items of this task in the item bank.

Researcher: How many?

PM: I guess over 1,000.

Researcher: You did all of them?

PM: Actually, I didn't do all of them; I read these questions whenever I had time. They are typically short and not complex. Therefore, it was not too bad.

Researcher: Why did you have to finish all the items in the item bank?

PM: Because the items in the test are basically from the item bank, though I know that the item bank is updated sometimes.

(PM, Focus Group 1)

As can be indicated in this excerpt, PM's preparation for this task was primarily characterised by intensive practice and familiarisation of authentic items in the item bank, motivated by the belief that the items that he would encounter would be drawn from the item bank.

4.1.2.4 Time management

When asked to reflect on the amount of time that they spent on preparing for the five tasks in the PTE-A speaking subtest, several participants mentioned that they spent the most time on *read aloud*, *repeat sentence*, *re-tell lecture*, or *describe image*; no participants, however, mentioned *answer short question*, which is not surprising given the finding that we reported in the previous section about this task, that is, it was considered as having little weighting and thus very limited impact on a test taker's speaking score and total score. It is worth noting that the task with the most time commitment was often attributed to its perceived difficulty level. PQ, for example, commented that he spent the most time on *read aloud* because 'after reading the PTE-A Official Guide, I felt

that this task was very challenging.’ When asked what was challenging about *read aloud*, he remarked that it was mainly related to the performance criteria of the PTE-A scoring algorithm (see Excerpt 11 below).

Excerpt 11

Researcher: On which task did you spend the most time when preparing the PTE-A speaking subtest?

PQ: I think it was *read aloud*. After reading the PTE-A Official Guide, I feel this task was really challenging.

Researcher: Why? Can you elaborate a bit?

PQ: It was mainly because the machine lays emphasis on some elements in *read aloud* performance to which I rarely pay any attention.

Researcher: Could you give me a few examples? Was it due to some difficult vocabulary which you didn’t know how to read?

PQ: No. The Official Guide mentions that elements such as linking, plosive sounds and strong and weak forms in pronunciation are important in determining my score on this task. In other words, our performance is evaluated in a very detailed way. Even though I know how to read every word in the text, my score may be quite low.

(PQ, Focus Group 2)

PP commented in a focus group discussion that he spent the most time on *read aloud* simply because there were templates for other tasks in the PTE-A speaking subtest: ‘when I was preparing for the other tasks, I have templates to memorise; for *read aloud*, however, there are no templates. The only way to prepare for this task is to read aloud as many texts as you can.’

When it comes to *repeat sentence*, PM mentioned that he spent the most time on this task because ‘I felt that it was quite difficult; therefore, I spent a lot time practicing so that I can overcome it.’ With regards to both *repeat sentence* and *re-tell lecture*, the participants mentioned that they spent a sizable portion of their time strengthening their listening ability which was perceived as the key to the success on these two tasks. The participants also mentioned that the practice of using templates took a lot of their time when they were preparing for *repeat sentence* and *describe image*.

When asked on which task they spent the least time in their preparation for the PTE-A speaking subtest, most participants mentioned *answer short question*. As mentioned previously, they felt that this task did not contribute much to their speaking and overall scores; therefore, they believed that it had ‘little value for time’, in PM’s words.

4.1.3 Participants’ perceptions of the PTE-A speaking subtest

4.1.3.1 Perceived difficulty of the PTE-A speaking subtest

Overall, the participants remarked that the PTE-A speaking subtest was reasonably difficult or relatively easier than some other language tests on the market, such as the IELTS. Having said that, there was an important caveat to this observation, that is, it was not difficult only after one

knew the tips or strategies to impress the machine which would evaluate their performance. PK, for instance, remarked that 'testwiseness is the key to success on the PTE-A speaking subtest; it is not difficult at all if you know the strategies to deal with each type of task in the test.' In a similar vein, PR commented that 'the PTE-A speaking subtest is a 'paper tiger' because once you know how to deal with it, it becomes very easy.' Another participant PD was comparing the PTE-A speaking subtest with the IELTS speaking test, concluding that 'it was not difficult, at least easier than the IELTS speaking test.'

Speaking of the first task, *read aloud*, most participants indicated that it was not difficult because 'in China, we have been taught to do a lot of reading aloud exercise since we started learning English; our English teachers also explain to us how effective it is for improving our English speaking.' PG commented that 'it is not so difficult because you don't need to think too much about it; you just read whatever that is given to you.' Interestingly, a few participants had the opposite observation, arguing that this task was quite difficult because 'I don't know how to impress the machine; fluency and accuracy don't seem to be enough' (PN). For the second task, *repeat sentence*, most participants reported that it was a challenging one because a) a strong listening ability is essential because one can't possibly repeat a sentence without understanding it; b) it is difficult to remember some sentences because they are quite long and complex, even after one has correctly parsed and understood them; and c) repeating sentences in exactly the same words as one has heard is considered very demanding. For instance, PF commented that 'they are read only once. Some sentences are quite long. Sometimes before you even understand what a sentence is about, it is gone and forever gone!' PG also perceived this task as very difficult. She explained that 'it is so difficult to repeat the exact words that you hear; I sometimes rephrase the sentences, but it is not what I am supposed to do.'

With regards to the third task, *describe image*, most participants indicated that it was not difficult because they all had memorized the templates for this task. PM commented that 'I memorised the templates, and I kept talking; it was a smooth process.' In a similar vein, PJ said that 'it is important to get familiar with two to three templates which could be applied to describe almost any kind of input images.' The next task, *re-tell lecture*, was perceived as difficult by most participants, who mentioned the following five factors that contributed to the difficulty of this task. First, to understand the lecture, one needs a strong listening ability. As PF commented, 'first and foremost, you need to understand what the lecture is about. Listening comprehension is the first step.' Second, it is challenging to remember the key details in the lecture. PQ remarked that 'the most difficult aspect about this task is to remember all the key details that are mentioned in the lecture'. Next, the way in which the lecture is presented (e.g., speaker accent) could also contribute to its difficulty. This is indicated in PP's comment on this task: 'the lectures are usually delivered in different accents, and this made it even more difficult to understand them.' The fourth factor is the background knowledge in the lecture. As PI remarked, 'if I know about the topic of the lecture, then it becomes so much easier; if the topic is unfamiliar to me, I feel it is very difficult to understand.' Finally, one needs the ability to organise the key details that they have

heard and deliver a speech in a fluent and coherent manner. Despite these difficulties, the use of templates was repeatedly mentioned which the participants believed helped them improve their performance on this task.

Regarding the last task, *answer short question*, the participants mentioned that it was quite difficult; however, given that it had only little weighting, they wouldn't worry too much about their performance on this task.

4.1.3.2 Perceived validity of the PTE-A speaking subtest

When asked whether the PTE-A speaking subtest could effectively measure their spoken English ability, the participants' views seemed rather polarised. Some participants commented positively on the overall effectiveness of the PTE-A in assessing one's spoken English ability, based on their test preparation and test-taking experience. In the excerpt below, for example, PE commented that the PTE-A speaking subtest could 'to a large extent reflect our spoken ability'. Having said that, this participant also remarked that the speaking constructs represented in the PTE-A speaking subtest were still quite different from 'how we use English in real life.'

Excerpt 12

Researcher: Do you think the PTE-A speaking subtest can effectively measure your spoken English ability?

PE: I think it could reflect 70% of my spoken English ability?

Researcher: Why? Can you explain a bit?

PE: I feel that it could to a large extent reflect our spoken ability; however, it is still quite different from how we use English in the real life.

(PE, Interview)

Another participant PO also commented positively on the overall effectiveness of the PTE-A speaking subtest in that 'it represents a more comprehensive assessment of one's spoken English ability. This observation, according to PO, was largely attributed to the integrated nature of the tasks in the PTE-A speaking subtest, which makes the tasks more authentic than independent speaking tasks (see the excerpt below).

Excerpt 12

Researcher: Do you think the PTE-A speaking subtest can effectively measure your spoken English ability?

PO: Yes, I think it can.

Researcher: Why do you think so?

PO: You know in some language tests like the IELTS, language skills like listening and speaking are assessed separately. However, in the PTE-A speaking subtest, they are assessed together, in an integrated manner. For example, in the *re-tell lecture* task, you need to listen to a lecture and then re-tell it.

Researcher: Do you think the task integrating multiple language skills is more authentic?

PO: Yes, absolutely.

(PO, Focus Group 1)

PJ approached this question based on her experience in attempting the PTE-A multiple times: 'I think the answer is yes because I noticed if I worked harder in preparing for the test, I would usually get a better score. This is quite different from when I was taking the IELTS.'

Some other participants held quite the opposite view, arguing that the PTE-A speaking subtest largely failed to reflect their spoken English ability. The reason was primarily attributed to their observation that the test was sensitive to the coaching of test taking skills and strategies, or more specifically, the use of templates, as illustrated by the excerpt below.

Excerpt 13

Researcher: Do you think the PTE-A speaking subtest can effectively measure your spoken English ability?

PK: I don't think so. It is a test which is very sensitive to testwiseness.

Researcher: Do you mean it can't reflect your spoken English ability.

PK: I think it is difficult to claim that one is proficient in spoken English if he or she gets a high score on the PTE-A speaking subtest. For example, I got a high score, but I still felt it difficult to communicate with others in real life.

Researcher: Okay.

PK: This test is perfect for Chinese students because we are so well trained in test taking in our education.

(PK, Interview)

PF commented that ‘some tasks in the PTE-A speaking subtest appear quite mechanical, such as *repeat sentence*.’ Therefore, ‘it is just a language test, which can’t reflect my real spoken English ability.’

Most participants indicated their preference to the computer-based spoken English test like the PTE-A speaking subtest because of the high level of anxiety caused by the presence of an examiner and/or interlocutor. As PI remarked, ‘sitting face to face with an examiner makes me feel nervous. Machines don’t have feelings; therefore, I feel much better sitting in front of the computer.’ The second reason for this preference is the perceived efficiency and fairness characterized by machine scoring. PE mentioned that ‘it is very rare that a test taker could get nine on the IELTS speaking test because I think the examiner was deliberately lowering the score; this would never happen when the machine rates your performance.’ A few participants mentioned that they would like to take the face-to-face speaking test ‘on the condition that test scores are not used for making important decisions’ (PF). As PF remarked, ‘face-to-face tests are different from machines which do not have logical ability and could only detect the keywords in my speech.’

When asked whether the rating method (i.e. human vs machine rating) would have any impact on their test preparation and test taking process, most participants answered positively. As PP responded in the focus group discussion, ‘surely I would take a different approach. For example, when I was taking the test, I gave priority to fluency. I knew that some of my sentences were not grammatical, but if I paused, that would affect my fluency score. I am sure I would do something different if I knew my performance was to be scored by a human rater.’ Some other participants commented that they would be careful about the use of templates in their responses if they knew that their performance was to be rated by a human rater, because ‘human raters might penalise the use of templates’ (PO); and accordingly, they wouldn’t spend much time memorising the templates and practice how to use them. Overall, the participants expressed their preference to machine scoring because it was believed to be fairer than human scoring.

4.1.3.3 Likes and dislikes of the PTE speaking subtest

The participants expressed likes about the PTE-A speaking subtest in terms of a) the machine scoring method which was perceived to be fairer and also enabled the speedy communication of test scores; b) the score report which was more detailed than some other English tests on the market. PP indicated that she received the score report on the evening of the day when she took the test: ‘it was very quick and delivered online; you don’t need to wait for ages for the score report to be delivered by a courier.’ Therefore, she concluded that ‘compared with some other English tests, PTE-A was very friendly to test takers.’ The machine scoring method, as mentioned in previous sections, was perceived to be fairer as compared with human raters whose rating results tended to be affected by personal traits or working contexts.

Only a few participants mentioned their dislikes about the PTE-A speaking subtest in terms of a) the quick response that is required of some tasks. As PF remarked, ‘what I dislike about the PTE-

A speaking subtest is that for some tasks, we are required to respond very swiftly. For example, for *repeat sentence*, I am required to respond within three seconds; otherwise, my response would not be recorded.’ In a similar vein, PB echoed that ‘in our daily communication, it is inevitable that we pause for three seconds or even five seconds, but you can’t pause for that long in the speaking test because you are required to respond within three or five seconds.’ b) the assessment of working memory. It was mentioned that an excellent working memory was crucial to the success on the PTE-A speaking subtest. PC indicated that ‘one aspect that I don’t like about the PTE-A speaking subtest lies in that it is not friendly to those with an unsatisfactory working memory.’

4.2 Findings from questionnaire survey (PTE-A)

As mentioned previously, the questionnaire survey was conducted with the view to exploring whether the findings that emerged in the qualitative phase of this project hold for a larger sample (see Figure 1). In this section, we report the findings based on the questionnaire data on the PTE-A speaking subtest. The findings are presented in accordance with the structure of the questionnaire, in terms of a) participant’s background, b) test preparation activities and strategies; and c) participants’ perceptions of the PTE-A speaking subtest.

4.2.1 Participants’ background

The participants self-reported English proficiency level indicates that they were reasonably proficient² (Overall: $M = 3.15$, $SD = 0.90$; Listening: $M = 3.21$, $SD = 1.06$; Reading: $M = 3.23$, $SD = 0.99$; Writing: $M = 2.99$, $SD = 0.95$; Speaking: $M = 3.01$, $SD = 0.99$). Their self-assessment results also seemed to indicate that they were stronger in their receptive skills (i.e. listening and reading) than productive skills (i.e. writing and speaking). Regarding opportunities to practice English speaking in their daily life, 67 participants (34.7%) reported they had no or few opportunities, and a similar number of participants ($n = 71$, 36.8%) indicated that they had relatively more opportunities to practice English speaking.

About half of the participants ($n = 97$, 50.3%) had taken the PTE-A for at least once within two months prior to our data collection; the other half ($n = 96$, 49.7%) were planning to take the PTE-A in the near future. Of the 88 participants (44.6%) who reported the number of attempts in taking the PTE-A, most of them made one ($n = 37$, 43.0%), two ($n = 17$, 19.8%) or three ($n = 10$, 11.6%) attempts. Three participants, however, reported that they took the PTE-A eight times. About half of the participants ($n = 90$, 46.6%) reported that they started preparing for the PTE-A over two months before the test; 34 (17.6%) started their preparation one month before the test, and another 34 (17.6%) started their preparation two weeks before the test. The remaining 35 participants (18.1%) started their preparation one week before the test.

When it comes to the purpose of or motivation in taking the PTE-A, about half of the participants ($n = 94$, 48.7%) took it because they were planning to seek further studies in English-medium

² The items are on a five-point Likert scale (see Appendix 3 for the questionnaire).

universities overseas; 59 (30.6%) reported that they took the PET-A for immigration purposes, and the remaining 40 (20.7%) took it for other purposes, such as work requirement, applying for work visa or assessing English proficiency.

4.2.2 Test preparation activities and strategies

4.2.2.1 Test preparation schools or classes

Of the 193 participants, 75 (38.9%) reported that they attended test preparation classes, with 55 (73.3%) attending online classes and 20 (26.7%) attending traditional face-to-face classes. As far as class size is concerned, 33 (44.0%) attended one-on-one tutoring class, 32 (42.7%) attended the classes with less than 20 students, and the remaining 10 (13.3%) attended the classes with over 20 students. Table 6 below presents a summary of findings relating to test preparation class. As shown in this table, most participants attended preparation classes mainly for two purposes: a) boosting their test scores on the PTE-A ($M = 4.35$, $SD = 0.80$), and b) getting more familiar with the PTE-A ($M = 4.16$, $SD = 1.04$). Also, participants reported that during the preparation class, their teachers mainly focused on introducing test-taking skills ($M = 4.24$, $SD = 0.93$), and believed that the class was important to enhance their performance on the PTE-A speaking subtest ($M = 4.15$, $SD = 0.83$).

Table 6. Participants' perceptions of test preparation class (PTE-A, n = 75)

Activity	Mean	SD	Agree n(%)	Neutral n(%)	Disagree n(%)
1) I attended the class to get more familiar with the PTE-A.	4.16	1.04	60(80.0)	8(10.7)	7(9.3)
2) I attended the class to more quickly enhance my test scores.	4.35	0.80	67(89.3)	6(8.0)	2(2.7)
3) I attended the class to know about the items in future tests.	3.73	1.12	48(64.0)	14(18.7)	13(17.3)
4) The teachers mainly used the authentic PTE-A test items.	4.05	0.96	57(76.0)	14(18.7)	5(5.3)
5) The teachers mainly introduced test-taking skills.	4.24	0.93	63(84.0)	7(9.3)	5(6.7)
6) The teachers provided feedback on our performance.	4.31	0.80	66(88.0)	7(9.3)	2(2.7)
7) I had many opportunities to practice spoken English in class.	3.89	0.97	52(69.3)	15(20.0)	8(10.7)
8) The class was important to enhance my test performance.	4.15	0.83	60(80.0)	12(16.0)	3(4.0)
9) The class was important to improve my spoken English.	3.84	0.94	48(64.0)	22(29.3)	5(6.7)

Notes. These items are on a five-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree).

4.2.2.2 Test preparation activities and strategies

Table 7 shows the descriptive statistics of the 14 test preparation activities and strategies included in the questionnaire. Note that these questionnaire items were generated based on the findings

that emerged from the qualitative part of this study (see Figure 1). As can be indicated in Table 7, the five activities in which the participants most frequently engaged are:

- *practice simulated or authentic items* (M = 3.49, SD = 1.25)
- *get familiar with the rating criteria for the PTE-A speaking subtest* (M = 3.44, SD = 1.20)
- *use mobile apps/computer software to rate read aloud performance* (M = 3.41, SD = 1.30)
- *learn test taking experience on mobile apps and websites* (M = 3.41, SD = 1.21).
- *learn and practice templates for describe image* (M = 3.31, SD = 1.29)

Meanwhile, the activities in which the participants least frequently engaged include:

- *find a partner to practice English* (M = 2.81, SD = 1.20)
- *practice English listening (e.g., news broadcasts, lectures, stories)* (M = 3.07, SD = 1.20)
- *read aloud English texts* (M = 3.07, SD = 1.23)
- *listen and take notes* (M = 3.08, SD = 1.22)
- *dictate English sentences* (M = 3.08, SD = 1.22).

We combined the responses to the two categories of ‘always’ and ‘often’ into the ‘more frequent’ category, and those to the categories of ‘rarely’ and ‘never’ into the ‘less frequent’ category. The frequency statistics for each activity are shown in Table 7. Next, we performed a non-parametric Chi-square test on each activity to explore whether the frequency of the participants selecting ‘more frequent’ and ‘less frequent’ differed significantly. As indicated in Table 7, the following test preparation activities have a significant Chi-square test result, with more participants reporting ‘more frequent’ than ‘less frequent’:

- *use mobile apps/computer software to rate read aloud performance*
- *listen and repeat English sentences*
- *learn and practice templates for describe image*
- *learn and practice templates for re-tell lecture*
- *practice how to obtain key details from English lectures*
- *learn test taking experience on mobile apps and websites*
- *practice simulated or authentic items*
- *get familiar with the rating criteria for the PTE-A speaking subtest.*

Only one activity, that is, *find a partner to practice spoken English*, had a significant larger number of participants selecting ‘less frequent’ than ‘more frequent’.

To compare whether attending test preparation class affected the participants’ test preparation practices and strategies, we performed independent-samples t-tests. Since there are 14 items in this section, we applied the Bonferroni correction procedure (Field, 2009) to adjust the critical value ($0.05/14 = 0.004$) to avoid Type 1 error. The results on the following six items were significant ($p < 0.004$), with those attending test preparation classes engaging in these activities and strategies more frequently than the solo group:

- *learn and practice templates for describe image* (Cohen’s $d = 0.73$)
- *learn and practice templates for re-tell lecture* (Cohen’s $d = 0.70$)
- *learn test taking experience on mobile apps and websites* (Cohen’s $d = 0.61$)

- *practice simulated or authentic items* (Cohen's $d = 0.64$)
- *get familiar with the rating criteria for the PTE-A speaking subtest* (Cohen's $d = 0.57$)
- *read the PTE-A test syllabus or official guide* (Cohen's $d = 0.47$)

In this analysis, Cohen's d was used as the effect size, which was interpreted based on Cohen (1988):

- $d < 0.2$ (no effect size)
- $0.2 \leq d < 0.5$ (small effect size)
- $0.5 \leq d < 0.8$ (medium effect size)
- $d \geq 0.8$ (large effect size)

The effect sizes in this analysis are in the medium range (i.e. 0.5-0.8), suggesting that these differences have reasonably significant implications in practice (Cohen, 1988).

Table 7. Test preparation activities reported by the questionnaire respondents (PTE-A, $n = 193$)

Activity	Mean	SD	More frequent n(%)	Less frequent n(%)
1) Read aloud English texts	3.07	1.23	72(37.3)	66(34.2)
2) Use mobile apps/computer software to rate read aloud performance*	3.41	1.30	104(53.9)	50(25.9)
3) Listen and repeat English sentences*	3.28	1.23	95(49.2)	53(27.5)
4) Dictate English sentences	3.17	1.31	86(44.6)	65(33.7)
5) Learn and practice templates for <i>describe image</i> *	3.31	1.29	92(47.7)	54(28.0)
6) Practice English listening (e.g., news broadcasts, lectures, stories)	3.07	1.20	73(37.8)	62(32.1)
7) Listen and take notes	3.08	1.22	73(37.8)	59(30.6)
8) Learn and practice templates for <i>re-tell lecture</i> *	3.27	1.29	88(45.6)	52(26.9)
9) Practice how to obtain key details from English lectures*	3.22	1.24	87(45.1)	54(28.0)
10) Learn test taking experience on mobile apps and websites*	3.41	1.21	99(51.3)	42(21.8)
11) Practice simulated or authentic items*	3.49	1.25	100(51.8)	42(21.8)
12) Find a partner to practice spoken English*	2.81	1.20	56(29.0)	79(40.9)
13) Get familiar with the rating criteria for the PTE-A speaking subtest*	3.44	1.20	101(52.3)	42(21.8)
14) Read the PTE-A test syllabus or official guide	2.96	1.19	63(32.6)	70(36.3)

Notes. *Non-parametric Chi-square test is significant ($p < 0.05$). These items are on a five-point Likert scale (1 = Never, 2 = Seldom, 3 = Sometimes, 4 = Often, 5 = Always).

In addition, we also asked the participants to indicate their perceived effectiveness of each activity in enhancing their performance on the PTE-A speaking subtest. We combined the responses to

the two categories of ‘very effective’ and ‘somewhat effective’ into the ‘more effective’ category, and those to the categories of ‘ineffective at all’ and ‘somewhat ineffective’ into the ‘less effective’ category. The results are summarised in Table 8. As indicated in this table, the participants reported that the following activities were perceived as the most effective:

- *practice simulated or authentic items* (M = 3.74, SD = 1.17)
- *get familiar with the rating criteria for the PTE-A speaking subtest* (M = 3.67, SD = 1.17)
- *learn test taking experience on mobile apps and websites* (M = 3.66, SD = 1.16)
- *practice how to obtain key details from English lectures* (M = 3.63, SD = 1.14).

The least effective activities as perceived by the participants include:

- *find a partner to practice spoken English* (M = 3.38, SD = 1.15)
- *read the PTE-A test syllabus or official guide* (M = 3.45, SD = 1.16)
- *listen and take notes* (M = 3.47, SD = 1.17)
- *practice English listening (e.g., news broadcasts, lectures, stories)* (M = 3.48, SD = 1.16)

Table 8. Perceived usefulness of test preparation activities reported by the questionnaire respondents (PTE-A, n = 193)

Activity	Mean	SD	More effective n(%)	Less effective n(%)
1) Read aloud English texts	3.52	1.09	96(49.7)	28(14.5)
2) Use mobile apps/computer software to rate read aloud performance	3.59	1.13	109(56.5)	31(16.1)
3) Listen and repeat English sentences	3.60	1.14	108(56.0)	32(16.6)
4) Dictate English sentences	3.62	1.17	105(54.4)	29(15.0)
5) Learn and practice templates for describe image	3.61	1.19	105(54.4)	31(16.1)
6) Practice English listening (e.g., news broadcasts, lectures, stories)	3.48	1.16	96(49.7)	35(18.1)
7) Listen and take notes	3.47	1.17	98(50.8)	36(18.7)
8) Learn and practice templates for re-tell lecture	3.62	1.19	106(54.9)	29(15.0)
9) Practice how to obtain key details from English lectures	3.63	1.14	112(58.0)	33(17.1)
10) Learn test taking experience on mobile apps and websites	3.66	1.16	110(57.0)	28(14.5)
11) Practice simulated or authentic items	3.74	1.17	119(61.7)	30(15.5)
12) Find a partner to practice spoken English	3.38	1.15	89(46.1)	40(20.7)
13) Get familiar with the rating criteria for the PTE-A speaking subtest	3.67	1.17	113(58.5)	30(15.5)
14) Read the PTE-A test syllabus or official guide	3.45	1.16	91(47.2)	36(18.7)

Notes. These items are on a five-point Likert scale (1 = Not useful at all, 2 = Not useful, 3 = Neutral, 4 = Useful, 5 = Very useful).

4.2.2.3 Learning materials

In the qualitative part of this study, the participants reported that they mainly referred to online learning resources when preparing for the PTE-A speaking subtest (see Section 4.1.1.2). Based on the findings from the qualitative part, we asked the participants to indicate the frequency of their use of eight online resources. As indicated in Table 9 below, they used all eight online learning resources quite frequently, with APEUni (*xinji*), Alpaca Education (*yangtuo*), and Bilibili the most frequently used.

Table 9. Use of online learning resources (PTE-A, n = 193)

Online learning resource	Mean	SD	More frequent n(%)	Less frequent n(%)
1) Alpaca Education (<i>yangtuo</i>)	3.62	1.39	106(54.9)	44(22.8)
2) BlackTech (<i>heikeji</i>)	3.32	1.42	88(45.6)	57(29.5)
3) FireflyEnlightenment (<i>yinghuochong</i>)	3.12	1.49	80(41.5)	71(36.8)
4) APEUni (<i>xinji</i>)	3.64	1.49	113(58.5)	48(24.9)
5) Zhihu	3.48	1.37	97(50.3)	49(25.4)
6) Youtube	3.35	1.33	84(43.5)	58(30.1)
7) Bilibili	3.60	1.40	108(56.0)	49(25.4)
8) Pearson official website	3.34	1.38	92(47.7)	56(29.0)

Notes. These items are on a five-point Likert scale (1 = Never, 2 = Seldom, 3 = Sometimes, 4 = Often, 5 = Always).

4.2.3 Participants' perceptions of the PTE-A speaking sub-test

The participants were asked to indicate their perceptions of the overall difficulty of the PTE-A speaking subtest and the difficulty of each task in the test. Table 10 below summarises the findings. As indicated in this table, most participants indicated that overall, the PTE-A speaking subtest was reasonably difficult (M = 3.36, SD = 1.10); *repeat sentence* was regarded as the most difficult task (M = 3.67, SD = 1.08), which was followed by *re-tell lecture* (M = 3.33, SD = 1.13), *read aloud* (M = 3.28, SD = 1.13) and *describe image* (M = 3.21, SD = 1.07). The least difficult task was *answer short question* (M = 3.20, SD = 1.09).

Table 10. Participants' perceptions of the difficulty of the PTE-A speaking subtest (n = 193)

Task	Mean	SD	Easy n(%)	Neutral n(%)	Difficult n(%)
1) Overall difficulty	3.36	1.10	35(18.1)	66(34.2)	92(47.7)
2) Read aloud	3.28	1.13	47(24.4)	62(32.1)	84(43.5)
3) Repeat sentence	3.67	1.08	27(14.0)	49(25.4)	117(60.6)
4) Describe image	3.21	1.07	41(21.2)	83(43.0)	41(21.2)
5) Re-tell lecture	3.33	1.13	39(20.2)	68(35.2)	86(44.6)
6) Answer short question	3.20	1.09	49(25.4)	70(36.3)	74(38.3)

Notes. These items are on a five-point Likert scale (1 = Very easy, 2 = Easy, 3 = Neutral, 4 = Difficult, 5 = Very difficult).

Findings relating to the participants' perceptions of automated scoring are presented in Table 11 below. This table indicates that overall the participants seemed to prefer automated to human

scoring ($M = 3.72$, $SD = 1.12$), and believed that automated scoring better reflected their spoken English proficiency ($M = 3.55$, $SD = 1.13$). In addition, the automated scoring practice that the PTE-A speaking subtest adopts seemed to have a significant impact on their test preparation practices ($M = 4.01$, $SD = 1.05$) and test-taking strategies ($M = 4.02$, $SD = 1.05$).

Table 11. Participants' perceptions of automated scoring (n = 193)

Statement	Mean	SD	Agree n(%)	Neutral n(%)	Disagree n(%)
1) Automated scoring is fairer compared with human scoring.	3.72	1.12	88(61.1)	35(24.3)	21(14.6)
2) Automated scoring better reflects my spoken English proficiency.	3.55	1.13	77(53.5)	43(29.9)	24(16.7)
3) If PTE-A was rated by human raters, I would adopt different preparation strategies.	4.01	1.05	102(70.8)	30(20.8)	12(8.3)
4) If PTE-A was rated by human raters, I would adopt different test-taking strategies.	4.02	1.05	105(72.9)	27(18.8)	12(8.3)

Notes. These items are on a five-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree).

Table 12 below summarises the findings concerning the participants' perceptions of the face validity of the PTE-A speaking subtest. Not surprisingly, most participants reported that the PTE-A speaking subtest score was important to them ($M = 3.89$, $SD = 1.12$). Also, the participants seemed to agree that the PTE-A speaking subtest was generally well-designed ($M = 3.54$, $SD = 1.08$) and could effectively reflect their spoken English proficiency ($M = 3.37$, $SD = 1.13$).

Table 12. Participants' perceptions of the face validity of the PTE-A speaking subtest (n = 193)

Statement	Mean	SD	Agree n(%)	Neutral n(%)	Disagree n(%)
1) The PTE-A speaking subtest is well-designed.	3.54	1.08	105(54.4)	60(30.1)	28(14.5)
2) The PTE-A speaking subtest score is important to me.	3.89	1.12	124(60.2)	50(25.9)	19(9.8)
3) The PTE-A speaking subtest well reflects my spoken English proficiency.	3.37	1.13	89(46.1)	66(34.2)	38(19.7)
4) I prefer live speaking test to computer-based speaking test.	3.12	1.20	68(35.2)	68(35.2)	57(29.5)
5) I feel more relaxed in live speaking test.	3.04	1.25	65(33.7)	63(32.6)	65(33.7)
6) Getting a high score on the PTE-A speaking subtest makes me more confident in English speaking.	3.55	1.14	103(53.4)	61(31.6)	29(15.0)

Notes. These items are on a five-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree).

4.3 Findings from interviews and focus groups (CET-SET4)

Table 13 below presents the categories and subcategories that were identified through coding the interview and focus group data about the CET-SET4, as well as the topics that were included under each subcategory. As indicated in this table, similar to what we found about the PTE-A speaking subtest, three broad categories were extracted from the data: a) participants' background; b) test preparation activities and strategies; and c) participants' perceptions of the CET-SET4. Next, we will elaborate on each of the three categories.

Table 13. Categories and subcategories extracted from the data (CET-SET4)

Category	Subcategory	Topic
1. Participants' background	A. Motivations in taking the CET-SET4	1-A-a. Gaining experience in spoken English tests 1-A-b. Promoting English learning
	B. English learning context	1-C-a. Self-assessed English ability 1-C-b. Opportunities to practice English speaking
2. Test preparation activities and strategies	A. Preparation materials	2-A-a. Online learning materials 2-A-b. Test preparation books
	B. Preparing for each task	2-B-a. Read aloud 2-B-b. Questions and answers 2-B-c. Individual presentation 2-B-d. Pair work
	C. Time management	2-C-a. Tasks with most time commitment 2-C-b. Tasks with least time commitment
3. Participants' perceptions of the CET-SET4	A. Perceived difficulty of the CET-SET4	3-A-a. Perceived difficulty of each task 3-A-b. Factors that contributed to task difficulty
	B. Perceived validity of the CET-SET4	3-B-a. Overall perceptions 3-B-b. Comparing computerised with live speaking tests
	C. Likes and dislikes about the CET-SET4	3-C-a. Perceived strengths of the test 3-C-b. Perceived weaknesses of the test

4.3.1 Participants' background

All participants mentioned that taking the CET-SET4 was not required by their institutions. Most of them took the test with a view to assessing their spoken English ability or gaining some experience in attending spoken English tests. As CN mentioned in the focus group discussion, 'I knew that it was not mandatory, but I still wanted to have a go to assess my spoken English proficiency.' This view was echoed by CI who commented that 'anyway, I didn't lose anything in attempting it.' In addition, some participants also mentioned that taking the CET-SET4 provided some impetus to their English learning. As CD commented of his motivation in taking the CET-SET4, 'I wanted to give myself some pressure in bettering my English through taking the CET-SET4.' The participants also mentioned that passing the CET-SET4 would give them some advantage when competing in the job market after they graduated. CG, for instance, remarked that 'if you seek employment with a company in the future, I believe they would not only look at your ability in English reading and writing, but also your proficiency in English speaking. A CET-SET4 certificate would be useful.'

When asked to self-assess their English ability, most participants indicated that they were not proficient in speaking. According to the participants, this was mainly ascribed to the test-oriented education which was prevalent in China's education context and the lack of opportunities to practice English speaking. Excerpt 14 below illustrates this finding.

Excerpt 14

Researcher: Why do you think you are weak in English speaking?

CE: In our daily life, we don't have many opportunities to speak English. Our teachers speak Chinese most of the time in the English class.

Researcher: Okay, I see.

CE: Another reason is that English teaching is more focused on students' scores on English tests than on improving how to use English in communicative contexts. Therefore, I don't feel very confident in speaking English.

(CE, Interview)

For most participants, they had opportunities to speak English only in their English classes, whereas the opportunities to speak English outside the classroom were scant. CK mentioned in the focus group discussion that 'our English teachers often asked us to speak English in the class, such as giving a presentation or participating in a group discussion. Some of these activities were spontaneous; some of them were prepared. Our teachers would give us some feedback after we finished speaking.' However, some participants also mentioned that even in their English class, they didn't have ample opportunities to speak English. CD, for instance, remarked that 'we don't have much time to speak English. Sometimes our English teachers organize speaking activities. However, there are so many students in the class that each of us had very few opportunities to speak.' As mentioned earlier, the participants reported scant opportunities to practice English speaking beyond the classroom. One opportunity to practice spoken English, according to some

participants, was the English Corner at their universities. Another way was to find a language partner, but it was ‘quite rare for two Chinese students to speak English, and I myself feel it a bit weird’ (CF).

4.3.2 Test preparation activities and strategies

All participants reported that they prepared the CET-SET4 on their own (i.e. solo preparation); no one attended test preparation schools or classes. Indeed, according to the participants, there were presumably no test preparation schools and classes in the market focusing on the CET-SET4. The primary reason was because the scores on the CET-SET4 were not used for high-stakes purposes. In other words, the CET-SET4 was not an important test for students. As CN commented in a focus group discussion, ‘I took the test almost without any preparation (*luokao* in Chinese) because it is not an important test; if the test result is used for important purposes, I would have spent more time preparing for it.’ Having said that, most participants still reported that they engaged in some test preparation activities to improve their performance on the CET-SET4. In what follows, we report the findings from the data in terms of the preparation materials that they used as well as the preparation activities in which they engaged to prepare for each task in the CET-SET4.

4.3.2.1 Preparation materials

The participants reported that their test preparation materials were mainly online learning resources either on mobile applications or English learning websites. One application, iSpeak (or *aitingshuo* in Chinese), developed by Unipus (<https://www.unipus.cn/iSpeak-mobile-i.html>), affiliated to Foreign Language Teaching and Research Press (FLTRP), was most frequently mentioned by the participants. According to the participants, this application was designed specifically for the CET-SET4, and it was very popular among university students who were preparing for the CET-SET4. The excerpt below explains its popularity.

Excerpt 15

Researcher: What learning materials did you use when preparing for the CET-SET4?

CC: I used iSpeak. Our English teacher recommended this app to us.

Researcher: Did you feel it was useful.

CC: Absolutely. It was designed for the CET-SET4. You can practice the test items on this app which can provide some helpful feedback on your performance. For example, it can highlight the areas for improvement in your *read aloud* performance.

(CC, Interview)

Another participant, CA, echoed what CC reported in Excerpt 15, ‘you can go through the whole test on this application; it can give you a score and some useful feedback for future improvement.’ Recall that in the CET-SET4, there is a pair work task where the test takers were paired up to communicate with each other on a given topic. According to the participants, the application

could pair one online user with another so that they could both practice this task. An important function of iSpeak lies in that it contains many simulated CET-SET4 tests for users to practice. CG, for example, reported that she did a simulated CET-SET4 test every day for a week prior to taking the test, and after that, she felt much more confident about the CET-SET4.

In addition to iSpeak, some participants also mentioned that they would consult some websites about the CET-SET4 where some previous test takers shared their test taking and/or preparation experience, such as zhihu (www.zhihu.com). It was also mentioned that some English learning websites were quite useful for preparing the CET-SET4. A test preparation textbook, called Spark English (*xinghuo yingyu* in Chinese), was also mentioned. For example, CE recalled that he used the Spark English textbook when preparing for the CET-SET4 because 'it contains some authentic test items from previous administrations, and it also teaches one how to deal with each task in the CET-SET4, such as using certain sentence structures or templates to express viewpoints for the *individual presentation* task'. CD used the website of Spark English (<https://www.sparke.cn/>) together with the textbook when he was preparing for the CET-SET4.

4.3.2.2 Preparing for each task in the CET-SET4

As mentioned previously (see Section 2.2.1), four tasks are used in the CET-SET4, including *read aloud*, *questions and answers*, *individual presentation* and *pair work*. In what follows, we present the findings relating to how the participants prepared for each task in the CET-SET4.

Read aloud

Several participants pointed out that reading aloud was an English learning activity that was familiar to Chinese students since their English teachers asked them to practice reading aloud back when they were studying in junior high school. Also, in many areas in China, reading aloud is a section in the spoken English test of *Gaokao*, the National Matriculation English Test (NMET). CM mentioned that 'I did a lot of reading aloud when I was studying at my junior and senior high schools; after I started university, however, I haven't done as much as I did before.' Shadowing was also a practice method which was frequently mentioned in our interviews and focus group discussions. The excerpt below illustrates this finding.

Excerpt 16

Researcher: How did you prepare for read aloud in the CET-SET4?

CN: Our English teacher recommended a method called shadowing which I found was quite effective. I did it quite often before the test.

Researcher: Can you explain?

CN: For example, I would listen to the audio files of the *Economist* magazine and mimic the way that the announcers read words and sentences. This helps to train my pronunciation and intonation.

(CN, Interview)

In a similar vein, CP mentioned in the focus group discussion that ‘when I want to practice reading aloud, I would watch a movie or an episode of the TV series that I like. While watching, I would listen carefully and mimic the way that the roles in these programs speak English.’ In addition to movies and TV series, TED talks and English-speaking documentaries were also mentioned as input materials.

The participants mentioned that a lot of practice was vital to the success on this task. Several participants recalled that they read aloud the texts in their College English textbooks; some other participants mentioned that they did the reading aloud exercise in iSpeak, the CET-SET test preparation application mentioned previously.

To succeed on this task, the participants mentioned a few strategies. First, according to CK, a focus group participant, it was important to ‘read the text and memorise some important details which you might need to answer the questions in the following task.’ She argued that ‘since we can’t see the text on the screen again when we answer questions in the following task, you can’t just read the text without remembering some key details.’ The other strategy was time management. CB mentioned that while preparing for this task, she noticed at the beginning that ‘I was reading too fast and finished it well before the signal; I felt at a loss about what to do in the remaining time. Therefore, I made conscious efforts to control my speed.’ This view was echoed by CH, who reported that ‘I would assess the length of the text first and plan my reading speed accordingly.’ The last strategy was reading the text closely during preparation time. By so doing, it was much easier for one to complete this task successfully because they already knew what the text was about.

Questions and answers

Most participants mentioned that they didn’t spend much time on this task in their preparations because it was related to the previous task. They argued that if one could understand the text in the *read aloud* section, he or she shouldn’t perceive this task as very challenging. Therefore, the participants reported that they would prepare for this task together with the previous task, that is *read aloud*. CD, for instance, commented that ‘this task is not very challenging. Understanding the text in the previous task is very important. Usually, the input texts in the *read aloud* section are fairly easy.’ Having said that, he admitted that this task ‘taps into one’s memory; if you can’t remember the input text that you have read aloud in the previous task, of course it is unlikely that you will respond to the questions correctly.’

Another reason why the participants didn’t spend much time on *questions and answers* in their preparations is that like *read aloud*, they were quite familiar with this task as it was included in the spoken English test of the *Gaokao* in some areas of China. Therefore, they made strenuous efforts to prepare for this task when they were studying in their senior high school, and they didn’t feel it was necessary to intensively prepare for this task again.

Individual presentation

When it comes to *individual presentation*, two topics were identified in the data: a) planning the presentation; and b) use of templates. CI, for instance, reported that he would make a simple mental draft which included the points that he planned to cover in his presentation. He mentioned that he usually organises his presentation by using expressions such as ‘first’, ‘second’, ‘third’, and ‘finally’. Similarly, CB indicated that she would build a ‘mental map’ which illustrated the points that she planned to deliver through the presentation. In addition to planning the points that one intended to cover, the participants also believed that it was vital to plan their time strategically if they had several points to elaborate in the presentation. The excerpt below illustrates how CP planned her time while preparing for this task.

Excerpt 17

Researcher: How did you prepare for individual presentation in the CET-SET4?

CP: I remember we only have one minute to respond to this task. Therefore, it is important to plan the time strategically in your response.

Researcher: Could you explain what you mean by ‘plan the time strategically’?

CP: Well, after you have seen the topic, you should think about how many points you want to cover in your presentation, and then plan your time accordingly. For example, if you plan to mention four points, you can’t spend a lot of time on the first point and end up having no time to cover the other three points.

(CP, Focus Group)

Talking about her preparations for this task, CP also reported that she visited some English learning websites where previous test takers shared their experience and some teachers taught some guidelines on how to respond to the task of *individual presentation*. She said that she would look at the topics in the IELTS speaking test and study the experience and guidelines on dealing with topic presentation. As she commented, ‘I would listen to their analysis. In this task, being logical and coherent is very important. Therefore, when I was responding to this task, I would make conscious efforts to ensure that my responses appear logical and coherent.’

The participants also mentioned the use of templates. CE, for instance, reported that he prepared a three-part template for this task: ‘in the first part, I should describe what the problem is; in the second part, I should analyse the reasons that might have caused this problem; finally, I should provide my own suggestions to resolve this problem.’ He mentioned that his English teacher taught him this template, which could be used effectively to deal with almost any task of such a nature. Having said that, CE commented that he was not keen to memorise and practice the templates because they represented ‘mechanical learning’, and because ‘I am already a university student, I don’t like rote memorisation. I have refrained from using the template, though I don’t resist the idea of memorising some useful vocabulary and expressions.’

In addition to the two topics, the importance of logical reasoning and critical thinking ability was also mentioned. CD reported that when he was preparing for this task, he was training his logical

reasoning ability through extensive reading. He said that ‘I would identify the problems when I was engaging with a text; then, I would try to express my viewpoints through oral presentation. I would deliberately create opportunities for me to practice presenting on a topic spontaneously.’

Pair work

Several participants mentioned that they found a conversation partner, either in real life or a virtual one, while preparing for this task. As CK mentioned in the focus group discussion, ‘I had a classmate who practiced this task with me. During our practice, we would simulate what happens in the test, that is, we had one minute to prepare, and then we started our conversation which last for three minutes.’ More participants, however, mentioned that they tried to find conversation partners online, mostly on iSpeak, which could pair online users. Despite the convenience of having a conversation partner online, CC commented that it was not easy for the system to pair users successfully: ‘I made quite a few attempts but was successfully paired with another person for only once.’ Also using iSpeak, CG reported a similar experience: ‘I was not successfully paired with another user, though I tried multiple times. In the end, I had to invite a classmate to talk to me online; however, the experience was different because we knew each other well. In the test, we are paired with complete strangers.’

Some participants also mentioned a few strategies which would help them improve their performance on this task. In the excerpt below, CO commented that it was vital to get to know about the conversation partner that the system had assigned to you at the beginning of the test.

Excerpt 18

Researcher: Do you have anything to share about *pair work* in the CET-SET4?

CO: I would say it is very advisable to get to know about your conversation partner prior to the test. You know, when we were testing the equipment at the beginning, your partner was already assigned to you. You could have a quick chat with him or her during this time.

Researcher: What do you chat about?

CO: In my case, we make a rough plan for our conversation. For example, we would basically agree on the turn-taking between us. I would raise a question first, and he would respond to my question. By so doing, the communication was much smoother.

(CO, Focus Group)

CP expressed strong agreement with CO’s comments: ‘I totally agree. It is very important to get familiar with your conversation partner at the beginning. For example, you can talk about each other’s background while testing the equipment. This would make the subsequent communications much easier.’

4.3.3 Participants' perceptions of the CET-SET4

4.3.3.1 Perceived difficulty of the CET-SET4

Overall, the participants perceived the CET-SET4 as not very challenging or relatively easy. CC, for instance, commented that 'it is pitched at the English proficiency level of first-year university students.' CG mentioned that 'it is quite like the spoken English test in the *Gaokao*. We are very familiar with the test format, except *pair work*.' While voicing similar sentiments about the difficulty level of the CET-SET4, CF argued that 'it doesn't make much sense to spend a lot of time preparing for the test because it is not very difficult.'

The participants commented that *read aloud* was probably the easiest among the four tasks in the CET-SET4. The input text was usually not difficult; only very occasionally did the test takers encounter unfamiliar vocabulary in the text. An important strategy to perform well on this task was to make the best use of the preparation time to read and understand the input text. Similarly, the participants believed that *questions and answers* was not very difficult to complete. However, as mentioned in the previous section, some participants argued that it was vital to make conscious efforts to remember some key details in the input text while working on the previous task; it was also necessary to 'predict what questions might be asked in the following task so that one can plan in advance.' Only a couple of participants commented that this part was reasonably challenging because, according to CE, 'you need to multitask; that is, while reading aloud, you need to think simultaneously about the *questions and answers* task because they are related.'

Individual presentation was considered the most challenging of the four tasks. The participants mentioned that the topic was the most important factor that affected the difficulty of this task. CA remarked that 'it really depends on the topic that was given to you. The topic that I got this time was about exchange students. However, I have never been an exchange student, and hence I didn't have much to say. If I got a topic with which I was familiar, my performance would be very different.' CE also indicated that 'luck is so important. If you are lucky enough to get an easier topic, it will make a huge difference.' In addition to topic, some participants also mentioned that it was a task that tapped into one's comprehensive English language ability. As CI remarked, 'it is quite a complicated process for a Chinese student. You need to think about the topic and search everywhere in your brain for the words and expressions to express your viewpoints. It is a difficult task for those who are not very proficient in English.'

The participants mentioned that it was their first time to encounter *pair work* in a computer-mediated spoken English test. The perceived difficulty of this task was closely associated with the English proficiency level of the conversation partner that was randomly assigned by the system. CJ, for instance, remarked that 'it was unfair because I felt that my performance was significantly affected by my partner whose English proficiency was very limited. When we were having the conversation, he could barely formulate a complete English sentence. As a result, the interaction was a failure.' CM reported a similar experience: 'my partner had a very limited English proficiency level. I asked him a question, and he stumbled and had a lot of pauses. I didn't know what to do. It took him a long time to finish what he wanted to express. He also spoke some

Chinese in our communication which was frustrating to me. Anyway, we ended up not talking much during the three minutes.’ Another problem that the participants mentioned was that it was quite difficult to talk to a stranger on a given topic. Therefore, as mentioned in the previous section, it was necessary to get familiar with the conversation partner prior to this task.

4.3.3.2 Perceived validity of the CET-SET4

Overall, the participants perceived the CET-SET4 as an effective or reasonably effective indicator of their spoken English proficiency, though one or two participants also mentioned that the test was too short for them to adequately demonstrate their English proficiency. The participants also perceived the CET-SET4 as low-stakes because the test results were not used to make any important decisions. CF, for instance, commented that ‘I didn’t make a lot of preparations for this test because I only wanted to experience what a computer-based spoken English test was like. I didn’t intend to get a high score, anyway, because the score would not be included in the CET4 written test. It’s not an important test.’

Most participants reported that the test format of the CET-SET4 was similar to the spoken English test of the *Gaokao*, except the *pair work* task. This also partly explains why they didn’t feel that the CET-SET4 was very difficult. Some participants pointed out that the CET-SET4 motivated them to study English, particularly spoken English. CA mentioned that because of the CET-SET4, she would pay more attention to spoken English, whose importance, in her opinion, was not highlighted in English teaching and learning in China. In the excerpt below, CN argued why Chinese students needed a test like the CET-SET4, despite its perceived low stakes.

Excerpt 19

Researcher: Do you think Chinese students need the CET-SET4?

CN: Absolutely. Many students don’t feel motivated to study English after they start their university. A test like the CET-SET4 is a good reminder that it is time to study English.

Researcher: Does that mean that a test can motivate students to learn English, even though the test is probably not very important?

CN: That’s correct. Tests play an important role in motivating us to learn English.

(CN, Focus Group)

When asked about their views on computer-based and face-to-face spoken English tests, about half of the participants indicated that they would prefer computer-based tests, mainly due to the high level of anxiety that was associated with interacting with a human examiner or interlocutor. The other half of the participants, however, expressed their preference to the face-to-face format because ‘it is more like authentic communication’ (CP), and because ‘it can better reflect my communicative language ability’ (CO).

4.3.3.3 Likes and dislikes about the CET-SET4

Most participants liked the CET-SET4 because of *pair work* which was not typically included in other computer-based spoken English tests such as the *Gaokao*. CA indicated that she liked the CET-SET4 because of 'the *pair work* task which reflects communications in real life.' In a similar vein, CM argued that her favourite task in the CET-SET4 was *pair work* because 'it is about interaction and communication which I have never seen in other computer-based speaking tests.' CD argued that 'we use language for communication; the most fascinating part of language lies in communication. I think it's a great idea to include *pair work* in the test.' In addition to *pair work*, the participants also mentioned the multiple task formats in the CET-SET4 represented a comprehensive assessment of spoken English proficiency.

When it comes to dislikes about the CET-SET4, interestingly, *pair work* was again most frequently mentioned. Based on their test taking experience, the participants pointed out that pairing was a challenge for test design. They argued that test takers at similar proficiency levels should be paired together as opposed to the current practice of random pairing of test takers. It was also suggested that some warm-up time should be built in the task of *pair work* for the pair of conversation partners to get familiar with each other before starting their conversation. A few participants also mentioned that as quite a number of test takers took the test in the same room and they sat very close to each other, their performance was somewhat affected by the noise around them.

4.4 Findings from questionnaire survey (CET-SET4)

Similar to the PTE-A speaking subtest, the CET-SET4 questionnaire aimed to explore whether the findings that emerged from the qualitative part of this study held for a much larger sample of the CET-SET4 test takers (see Figure 1). The findings in this section are presented in accordance with the structure of the questionnaire, in terms of a) participant's background, b) test preparation practices and strategies; and c) participants' perceptions of the CET-SET4.

4.4.1 Participants' background

Overall, the participants seemed to be moderately proficient in English language proficiency, as indicated by their self-assessment results ($M = 3.09$, $SD = 0.67$)³. They seemed to be stronger in reading ($M = 3.23$, $SD = 0.76$) and writing ($M = 3.03$, $SD = 0.72$) than in listening ($M = 2.73$, $SD = 0.92$) and speaking ($M = 2.75$, $SD = 0.88$). As far as their English learning environment was concerned, 43 (16.7%) reported that they had few opportunities to speak English in class, whereas 89 (34.6%) indicated that they had relatively more opportunities. Most of the participants seemed to have very limited opportunities to practice English speaking beyond the classroom. Of the 257 participants, 167 (65.0%) reported that they had few or no opportunities; only 17 (6.6%) reported that they had some opportunities to practice English outside the English class.

³ The self-assessment items are on a five-point Likert scale (1 – Very Poor; 2 – Poor; 3 – Fair; 4 – Good; 5 – Excellent).

Most participants started preparing for the CET-SET4 either one week (n = 154, 59.9%) or half a month prior to the test (n = 49, 19.1%). Table 14 below summarises why or what motivated the participants to take the CET-SET4. As indicated in this table, the two most important reasons for participants to take the CET-SET4 were a) assessing their spoken English proficiency (n = 185, 72.0%); and b) gaining the experience of taking the CET-SET4 (n = 185, 72.0%).

Table 14. Reasons for or motivations in taking the CET-SET4 (n = 257)

Reasons	N	%
Assess my spoken English proficiency	185	72.0
Gain the experience of taking the CET-SET4	185	72.0
Get another certificate for future employment	105	40.9
Follow the advice of my English teacher	73	28.4
Follow the crowd	59	23.0
Institutional requirement	52	20.2

4.4.2 Test preparation practices and strategies

Table 15 shows the descriptive statistics of the 18 items in the questionnaire about test preparation activities and strategies. Note that these items were generated based on the findings that emerged from the qualitative part of this study (see Figure 1). As indicated in Table 15, the preparation activities and strategies in which the participants engaged most frequently are:

- *memorise useful words and expressions* (M = 3.49, SD = 0.96)
- *memorise useful sentence patterns* (M = 3.47, SD = 0.99)
- *prepare templates for individual presentation* (M = 3.18, SD = 1.01)
- *strengthen English listening* (M = 3.12, SD = 0.88)
- *learn test taking experience on some websites* (M = 3.10, SD = 1.04)

Meanwhile, the activities in which the participants least frequently engaged include:

- *attend English corner activities* (M = 2.04, SD = 0.91)
- *practice pair work through finding partners on mobile apps or the internet* (M = 2.18, SD = 1.00)
- *find pair work partners in real life* (M = 2.27, SD = 0.94)
- *talk to myself on certain topics* (M = 2.35, SD = 0.97)
- *paraphrase an English text after listening* (M = 2.49, SD = 0.91)

Table 15. Test preparation activities and strategies reported by the questionnaire respondents (CET-SET4, n = 257)

Activity	Mean	SD	More frequent n(%)	Less frequent n(%)
1) Read aloud English texts	3.01	0.86	68(26.5)	63(24.5)
2) Shadow*	2.88	0.89	57(22.2)	85(33.1)
3) Listen to English sentences and repeat	3.00	0.95	79(30.7)	76(29.6)
4) Strengthen English listening*	3.12	0.88	87(33.9)	60(23.3)
5) Listen and take notes	2.87	0.99	69(26.8)	94(36.6)
6) Record performance on read aloud and self-evaluate*	2.49	0.96	30(11.7)	131(51.0)
7) Practice pair work through finding partners on mobile apps or the internet*	2.18	1.00	22(8.6)	164(63.8)
8) Find pair work partners in real life*	2.27	0.94	20(7.8)	156(60.7)
9) Attend English corner activities*	2.04	0.91	12(4.7)	182(70.8)
10) Talk to myself on certain topics*	2.35	0.97	26(10.1)	143(55.6)
11) Paraphrase an English text after listening*	2.49	0.91	134(52.1)	32(12.5)
12) Memorise useful words and expressions*	3.49	0.96	140(54.5)	41(16.0)
13) Memorise useful sentence patterns*	3.47	0.99	140(54.5)	41(16.0)
14) Prepare templates for <i>individual presentation</i> *	3.18	1.01	97(37.7)	61(23.7)
15) Prepare some possible topics	2.96	1.04	81(31.5)	86(33.5)
16) Learn test taking experience on websites*	3.10	1.04	95(37.0)	69(26.8)
17) Do simulated CET-SET4 items	2.88	1.11	81(31.5)	96(37.4)
18) Get familiar with the rating criteria	2.91	1.05	79(30.7)	88(34.2)

Notes. *Non-parametric Chi-square test is significant ($p < 0.05$). These items are on a five-point Likert scale (1 = Never, 2 = Seldom, 3 = Sometimes, 4 = Often, 5 = Always).

We combined the responses to the two categories of ‘always’ and ‘often’ into the ‘more frequent’ category, and those to the categories of ‘rarely’ and ‘never’ into the ‘less frequent’ category. The frequency statistics for each activity are shown in Table 15. Next, we performed a non-parametric Chi-square test on each activity to explore whether the frequency of the participants selecting ‘more frequent’ and ‘less frequent’ differed significantly. As indicated in Table 15, the following test preparation activities/strategies have a significant Chi-square test result, with significantly more participants reporting ‘more frequent’ than ‘less frequent’:

- *strengthen English listening*

- *paraphrase an English text after listening*
- *memorise useful words and expressions*
- *memorise useful sentence patterns*
- *prepare templates for individual presentation*
- *learn test taking experience on websites*

Meanwhile, the following activities had a significant larger number of participants selecting ‘less frequent’ than ‘more frequent’:

- *shadow*
- *record performance on read aloud and self-evaluate*
- *practice pair work through finding partners on mobile apps or the internet*
- *find pair work partners in real life*
- *attend English corner activities*
- *talk to myself on certain topics*

For the activities in which the participants engaged most frequently, we also examined their perceived usefulness for improving their test performance on the CET-SET4. Table 16 below summarises the results. As the statistics in this table suggest, the participants perceived these activities as reasonably useful (e.g., all mean values are over 3.5 on a five-point Likert scale).

Table 16. Perceived usefulness of test preparation activities reported by the questionnaire respondents (CET-SET4, n = 257)

Activity	Mean	SD	More useful n(%)	Less useful n(%)
1) Read aloud English texts	3.63	0.79	153(59.5)	15(5.8)
3) Listen to English sentences and repeat	3.55	0.83	144(56.0)	23(8.9)
4) Strengthen English listening	3.69	0.80	159(61.9)	13(5.1)
12) Memorise useful words and expressions	3.77	0.85	166(64.6)	16(6.2)
13) Memorise useful sentence patterns	3.79	0.82	168(65.4)	11(4.3)
14) Prepare templates for individual presentation	3.74	0.84	158(61.5)	15(5.8)
16) Learn test taking experience on websites	3.66	0.83	151(58.8)	18(7.0)

Notes. These items are on a five-point Likert scale (1 = Not useful at all, 2 = Not useful, 3 = Neutral, 4 = Useful, 5 = Very useful).

Tables 17 below show the descriptive statistics of the 12 test taking strategies, which were generated based on the findings that emerged from the qualitative part of this study (see Figure 1). As indicated in this table, the test taking strategies in which the participants engaged most frequently are:

- *read the text beforehand* (M = 3.74, SD = 0.98)
- *memorise key points while reading the text beforehand* (M = 3.62, SD = 0.94)
- *memorise key points while reading aloud the text* (M = 3.59, SD = 0.95)
- *control the speed and timing while reading aloud* (M = 3.57, SD = 0.93)

Meanwhile, the testing taking strategies in which the participants engaged least frequently include:

- *communicate with the partner beforehand* (M = 2.56, SD = 1.21)
- *plan the structure of conversation with the partner beforehand* (M = 2.60, SD = 1.17)
- *predict how the partner is going to respond* (M = 3.01, SD = 1.09)
- *use templates that have been prepared* (M = 3.03, SD = 1.00)

Table 17. Test taking strategies reported by the questionnaire respondents (CET-SET4, n = 257)

Strategy	Mean	SD	More frequent n(%)	Less frequent n(%)
<i>Read aloud and questions and answers</i>				
1) Read the text beforehand	3.74	0.98	161(62.6)	28(10.9)
2) Memorise key points while reading the text beforehand	3.62	0.94	148(57.6)	30(11.7)
3) Memorise key points while reading aloud the text	3.59	0.95	143(55.6)	33(12.8)
4) Control the speed and timing while reading aloud	3.57	0.93	140(54.5)	30(11.7)
<i>Individual presentation</i>				
5) Plan the points to be covered in the presentation beforehand	3.39	0.93	124(48.2)	42(16.3)
6) Plan the structure of the presentation beforehand	3.28	1.00	113(44.0)	53(20.6)
7) Use templates that have been prepared	3.03	1.00	84(32.7)	75(29.2)
8) Use words and expressions that have been prepared	3.16	0.99	99(38.5)	66(25.7)
<i>Pair work</i>				
9) Communicate with the partner beforehand	2.56	1.21	63(24.5)	125(48.6)
10) Plan the structure of conversation with the partner beforehand	2.60	1.17	59(23.0)	120(46.7)
11) Predict how the partner is going to respond	3.01	1.09	97(37.7)	76(29.6)
12) Take initiative in pair work	3.11	1.11	101(39.3)	71(27.6)

Notes. *Non-parametric Chi-square test are significant ($p < 0.05$). These items are on a five-point Likert scale (1 = Never, 2 = Seldom, 3 = Sometimes, 4 = Often, 5 = Always).

In Table 18 below, we present the findings on the participants' perceived effectiveness of the test taking strategies in which the participants tended to engage (mean values over 3.00 on a five-point Likert scale).

Table 18. Perceived usefulness of test taking strategies in improving test performance (CET-SET4, n = 257)

Strategy	Mean	SD	More useful n(%)	Less useful n(%)
<i>Read aloud and questions and answers</i>				
1) Read the text prior to the task	3.80	0.87	167(65)	16(6.2)
2) Remember some key points while reading the text prior to the task	3.84	0.83	169(65.8)	10(3.9)
3) Remember some key points while reading aloud the text	3.80	0.84	165(64.2)	13(5.1)
4) Control the speed and timing while reading aloud	3.76	0.86	159(61.9)	16(6.2)
<i>Individual presentation</i>				
5) Plan the points to be covered in the presentation beforehand	3.82	0.84	170(66.1)	12(4.7)
6) Plan the structure of the presentation beforehand	3.78	0.89	164(63.8)	16(6.2)
7) Use templates that have been prepared	3.67	0.90	151(58.8)	21(8.2)
8) Use words and expressions that have been memorised	3.73	0.88	161(62.6)	17(6.6)
<i>Pair work</i>				
11) Predict how the partner is going to respond	3.66	0.91	156(60.7)	21(8.2)
12) Take initiative in pair work	3.74	0.93	165(64.2)	22(8.6)

Notes. These items are on a five-point Likert scale (1 = Not useful at all, 2 = Not useful, 3 = Neutral, 4 = Useful, 5 = Very useful).

As indicated in this table, the participants perceived all the test taking strategies included in this table as useful for improving their performance on the CET-SET4 (all mean values are over 3.50 on a five-point Likert scale).

4.4.3 Participants' perceptions of the CET-SET4

The participants reported that the CET-SET was overall reasonably difficult ($M = 3.39$, $SD = 0.76$). They perceived *individual presentation* as the most difficult task ($M = 3.61$, $SD = 0.74$), followed by *pair work* ($M = 3.56$, $SD = 0.89$) and *questions and answers* ($M = 3.44$, $SD = 0.81$). *Read aloud* was perceived as the least difficult task in the CET-SET4 ($M = 2.84$, $SD = 0.91$).

Table 19. Participants' perceptions of the difficulty of the CET-SET4 (n = 257)

Task	Mean	SD	Easy n(%)	Neutral n(%)	Difficult n(%)
1) Overall difficulty	3.39	0.76	24(9.3)	124(48.2)	109(42.4)
2) Read aloud	2.84	0.91	90(35.0)	117(45.5)	50(19.5)
3) Questions and answers	3.44	0.81	25(9.7)	111(43.2)	121(47.1)
4) Individual presentation	3.61	0.74	14(5.4)	96(37.4)	147(57.2)
5) Pair work	3.56	0.89	23(8.9)	101(39.3)	133(51.8)

Notes. These items are on a five-point Likert scale (1 = Very easy, 2 = Easy, 3 = Neutral, 4 = Difficult, 5 = Very difficult).

Table 20 below summarises the findings relating to the participants' perceptions of the face validity of the CET-SET4. In terms of test design, about half of the participants agreed that the CET-SET 4 was well-designed (n = 136, 52.9%), and that they could demonstrate their spoken English proficiency in the CET-SET4 (n = 123, 47.9%). Regarding *individual presentation*, most participants agreed that the topic that was assigned to them was an important factor in affecting their performance on this task (n = 182, 72.0%). They also tended to agree that the English proficiency of the conversation partner that the system assigned to them had a significant impact on their performance (n = 151, 58.8%). The CET-SET4 seemed to have some positive washback effects on the participants' learning of spoken English. For example, 160 participants (62.3%) reported that they paid more attention to spoken English because of the CET-SET4; 164 participants (63.8%) indicated that the CET-SET4 motivated them to study English, and 167 participants (65.0%) reported that a good performance on the CET-SET4 made them feel more confident in spoken English communication. Finally, about half of the participants indicated that they preferred a computer-based spoken English to a live one (n = 135, 52.5%), that they felt more relaxed when taking a computer-based spoken English test (n = 144, 56.0%), and that computer-based tests could effectively measure their spoken English proficiency (n = 120, 46.7%).

Table 20. Participants' perceptions of the face validity of the CET-SET4 (n = 257)

Statement	Mean	SD	Agree n(%)	Neutral n(%)	Disagree n(%)
1) CET-SET4 was well-designed.	3.60	0.82	136(52.9)	107(41.6)	14(5.4)
2) I can demonstrate my spoken English proficiency in CET-SET4.	3.42	0.94	123(47.9)	93(36.2)	41(16.0)
3) In <i>individual presentation</i> , the topic might affect my performance.	3.90	0.88	185(72.0)	56(21.8)	16(6.2)
4) In <i>pair work</i> , my partner's proficiency level might affect my performance.	3.64	0.95	151(58.8)	77(30.0)	29(11.3)
5) I am familiar with the CET-SET4 rating criteria.	3.34	0.93	113(44.0)	106(41.2)	38(14.8)
6) CET-SET4 scores are important to me.	3.50	0.91	128(49.8)	102(39.7)	27(10.5)

7) I paid more attention to spoken English learning because of CET-SET4.	3.74	0.87	160(62.3)	81(31.5)	16(6.2)
8) CET-SET4 motivated me to study English.	3.80	0.91	164(63.8)	75(29.2)	18(7.0)
9) I have improved my spoken English proficiency through preparing the CET-SET4.	3.58	0.96	140(54.5)	89(34.6)	28(10.9)
10) A good score on the CET-SET4 makes me more confident in spoken communications.	3.76	0.91	167(65.0)	70(27.2)	20(7.8)
11) I prefer computer-based spoken English test to a live one.	3.57	0.99	135(52.5)	92(35.8)	30(11.7)
12) I feel more relaxed when taking a computer-based spoken English test.	3.63	0.99	144(56.0)	87(33.9)	26(10.1)
13) Computer-based spoken English tests can effectively measure our spoken English proficiency.	3.46	0.93	120(46.7)	107(41.6)	30(11.7)

Notes. These items are on a five-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree).

5. Summary and conclusions

In this study, we investigated test takers' preparation practices and strategies on two computer-mediated English speaking tests, that is, the PTE-A speaking subtest and the CET-SET4, which differ in quite significant ways in their design, tasks, speaking constructs, scoring methods, and test stakes. As mentioned previously, the PTE-A speaking subtest is an English proficiency test, designed to assess test takers' ability to use English in an academic environment. It is a high-stakes test because the test scores are used to make decisions about students' readiness to study in an English-medium university or for the purpose of immigration, in particular the skilled migration pathway (Knoch, Huisman, Elder, Kong, & McKenna, 2020; Wang et al., 2012). The test is fully automated which means test takers' performance is rated by an automatic scoring system developed by Pearson, the owner of the PTE-A speaking subtest. All five tasks in the PTE-A speaking subtest could be considered as integrated, though the level of integration varies. In comparison, the CET-SET4 is a spoken English test targeting non-English major undergraduates at universities in China mainland. The test is low-stakes in the sense that it is not mandatory for students, and its results are not used to make important decisions with a considerable impact on students, at least at the time when the data were collected for this study. Apart from *read aloud* which is automatically scored, students' performance on the other tasks in the CET-SET4 is rated by human raters. Compared with the PTE-A speaking subtest, the tasks in the CET-SET4 are much less integrated. Except the first two tasks, that is, *read aloud* and *questions and answers*, which entail a certain extent of skill integration, the other two tasks could be considered as independent.

In their study on the relationship between repeat test takers' test preparation practices and their score gains on the PTE-A, Knoch et al. (2020) found that among the four skills, test takers made the largest number of meaningful improvements from one test session to the next on the speaking subtest. Through in-depth interviews, they found that test takers engaged in strategic *construct-irrelevant* activities which resulted in the improvements in their scores on the PTE-A speaking subtest, including deliberate changes in fluency and pronunciation or changes in their voice quality. The findings brought them to the speculation that 'speaking test may be a particular focus for test takers in terms of gaming' (p. 565). Since their study did not focus exclusively on the speaking test, they called for future research looking more closely into the test takers' preparation activities when preparing for the speaking test. They also argued for the importance of comparing test takers' preparation activities on speaking tests that are automatically scored and rater-mediated. By comparing test takers' preparation practices and strategies on the PTE-A speaking subtest and the CET-SET4, this study therefore fills an important gap in the current literature on test preparation. Furthermore, according to Wang et al. (2012), very limited research has focused on the impact and consequences of the PTE-A, considered as an important inference in the argument-based validation framework (Chapelle, Enright, & Jamieson, 2008), despite the high stakes that are typically attached to it. By exploring how test takers prepare for the PTE-A speaking subtest, this study contributes significant evidence to the development of a coherent and comprehensive validity narrative for the PTE-A speaking subtest.

In what follows, we discuss the findings of this study in accordance with the two research questions:

- 1) What are test takers' preparation practices and strategies on the speaking subtest of the PTE-A and the CET-SET4?
- 2) Do test takers engage in different preparation practices or adopt different preparation strategies on the independent and integrated tasks when preparing for the speaking subtest of the PTE-A and the CET-SET4?

5.1 Test takers' preparation practices and strategies on the PTE-A speaking subtest and the CET-SET4

Messick (1982, p. 70) defined test preparation broadly as 'any intervention procedure specifically undertaken to improve test scores, whether by improving the skills measured by the test or by improving the skills for taking the test, or both.' He argued that test preparation activities could be classified into three types. Type 1 preparation improves test scores through improving the aspects of the constructs that the test assesses, or the *construct-relevant* aspects. This type of test preparation is considered as positive washback because test takers improve their ability on the target constructs through test preparation activities. In this study, such activities could include, for example, 'find a partner to practice spoken English' or 'listen to an English lecture and orally summarise it'. Type 2 preparation improves test scores through reducing *construct-irrelevant*

inferences in the test. For example, test takers' lack of familiarity with the task formats in a computerised speaking test might result in high levels of anxiety, which prevent them from performing to the best of their speaking ability. Therefore, it is essential for test takers to engage in some preparation activities which help them get familiar with the speaking test. Type 2 preparation is generally considered as conducive to test validity because it reduces *construct-irrelevant* variance in test score interpretations. Type 3 preparation improves test scores by enhancing test takers' test taking skills or strategies which are irrelevant to the constructs that the test assesses (also known as *testwiseness*). Because these activities or strategies are *construct-irrelevant*, they are considered as a threat to test validity. In other words, test takers might have improved their test scores without any improvement in the target constructs. A typical example of Type 3 preparation is the memorisation of 'canned' responses to a writing or speaking task.

In their study of test takers' preparation practices on the PTE-A, Knoch et al. (2020) identified all three types of test preparation. Similar to numerous previous studies (Dong, Fan, & Xu, 2021; Lai & Waltman, 2008; Xie & Andrews, 2013; Yu et al., 2017), their findings reiterate the complex nature of test preparation as they found that 'test takers often engaged in various processes at the same time and some of these activities were overlapping' (p. 567). The findings of our study, in relation to both the PTE-A speaking subtest and the CET-SET4, provide further evidence pointing to the complexity of test preparation. By comparing test takers' preparation practices on two different computerised speaking tests, we are also able to demonstrate test preparation could be significantly influenced by the language learning and testing context and the target test, in terms of, for example, the task format, test stakes or scoring method. In what follows, we discuss test takers' preparation activities and strategies on the two speaking tests and explore the factors which might influence or mediate their preparation practices.

Compared with the CET-SET4, participants of the PTE-A seemed to engage in more prolonged and intensive test preparation with a wider range of preparation activities. In this study, about half of the PTE-A participants attended test preparation classes whereas all CET-SET4 participants reported that they prepared for the test on their own. Indeed, the participants mentioned that there were no test preparation classes for the CET-SET4 in the market. This finding could be interpreted in relation to the different stakes of the two speaking tests. The PTE-A speaking subtest, as mentioned previously, has much higher stakes as compared with the CET-SET4 which is not required for students and most of them took the test with a view to assessing their spoken English proficiency.

For the PTE-A participants who attended the test preparation classes, their primary purpose was to improve their test scores through developing effective test taking skills and strategies, or testwiseness (Type 3 preparation), though some of them also mentioned that they wanted to get more familiar with the PTE-A speaking subtest (Type 2 preparation). The teaching content and activities at the preparation classes clearly indicate that they were oriented to improving students' test scores through intensive practice and coaching of testwiseness (Type 3 preparation). For those who prepared the test on their own (i.e. solo preparation), they reported that they drew on a wide

range of online resources in their test preparation. Similar to those who attended the preparation classes, however, they mainly engaged in intensive practice of authentic or simulated test items, and also developed their test testwiseness through either watching videos on the internet or learning from the experience shared by previous successful test takers (Type 3 preparation).

Similar to the findings from Knoch et al. (2020) about test takers' preparation for the PTE-A speaking subtest, this study uncovered that test takers made highly strategic decisions or choices in their preparation practices. This could be exemplified by some strategies that they deployed with a view to enhancing their test performance. For example, several participants mentioned that *fluency* was the key to success on the PTE-A speaking subtest. Therefore, they would follow the advice from either the teachers in the test preparation classes or the experience that previous successful test takers shared on websites. As PB reported about how she responded to the *read aloud* task in the PTE-A speaking subtest, 'if there is an unfamiliar word, just skip it because it is vital to make sure that your speech is very fluent, without any pauses.' In addition, the use of 'templates' also featured prominently in our data. Most participants mentioned that their preparation for some tasks in the PTE-A speaking subtest, in particular, *describe image* and *re-tell lecture*, centred on the memorisation of templates and how to use them effectively in their responses. Indeed, the effective use of templates was considered as highly significant to the success on the PTE-A speaking subtest. These activities are clearly Type 3 preparation as they are construct-irrelevant, thus posing a threat to the validity and score interpretations of the PTE-A speaking subtest.

Several reasons might explain the prevalent Type 3 preparation activities in which the participants engaged while preparing for the PTE-A speaking subtest. First, the PTE-A is a high-stakes test, and thus not surprisingly, test takers would engage in strategic practices which they believe would improve their performance. Another possible reason is that the data in this study seemed to capture test takers' more overt test preparation practices, given that most participants started preparing for the PTE-A within two months prior to the test and that they tended to engage in more overt test preparation practices as the test was approaching (Cheng, Watanabe, & Curtis, 2004). The most important reason, based on the data that we collected, is the automatic scoring method that the PTE-A speaking subtest uses to score test takers' performance. The participants in this study mentioned several times that their strategic practices were aiming to 'game' the machine. For example, PO argued that 'the machine prioritises fluency over other criteria'. Indeed, fluency was perceived as carrying so much weight in the algorithm that communication of meaning could be sacrificed for the sake of fluency, as PO continued to argue that 'keep talking, even if what you said is probably not correct.'

When it comes to the use of templates, automatic scoring again plays a significant role in shaping test takers' preparation practices. The participants reported that they would refrain from using templates if they knew that their performance would be rated by human raters because they could detect the templates and therefore penalise them for using templates in their responses. However, since they knew that their performance would be scored by machines, their perceptions shifted.

This was explicitly evidenced by what PF reported, ‘because PTE-A is cored by machines, you wouldn’t be penalised for using the templates because machines don’t care; however, it is likely that you are penalised in the case of IELTS, which is rated by human raters. They can discern that you are using templates.’ Arguably, these Type 3 activities pose a threat to the validity of the PTE-A speaking subtest and might explain the low correlation identified by a recent study between the scores on the PTE-A and IELTS speaking subtest ($r = 0.44$; Elliot & Blackhurst, 2021).

Compared with the PTE-A speaking subtest, the CET-SET4 participants didn’t seem to engage in prolonged, intensive preparation activities as most test takers started preparing the CET-SET4 within a week prior to the test. The reason was ascribed to the perceived low stakes that are associated with the CET-SET4. Having said that, the CET-SET4 participants reported that they also drew on online resources, in particular iSpeak, which focuses specifically on the CET-SET4. Also, the participants reported a narrower range of preparation activities and strategies, as compared with the PTE-A speaking subtest. Nonetheless, a considerable proportion of the test preparation activities in which the CET-SET4 participants engaged seemed to centre around Type 1 (e.g., ‘watch English movies, Ted talks, documentaries’ or ‘read aloud texts in the College English textbooks’) or Type 2 preparation (e.g., ‘get familiar with the tasks through practicing authentic test items’). Admittedly, the CET-SET4 participants also reported some strategic moves in their test preparation or test taking process (Type 3 preparation). For example, CB reported that she was making conscious efforts to control her speed when responding to the *read aloud* task and carefully managing her time.

Some participants also mentioned the practice and use of ‘canned’ responses or templates when responding to the *individual presentation* task. However, these practices did not feature prominently in our data, particularly in comparison with the PTE-A speaking subtest. It should also be noted that some participants expressed unwillingness about using templates because such practice was perceived as inconsistent with their identity as a university student. As CE argued, ‘I am a university student; I don’t like rote memorisation. I have refrained from using the template, though I don’t resist the idea of memorising some useful vocabulary and expressions.’ It was also repeatedly mentioned in the data that except the *pair work* task, the other tasks in the CET-SET4 are similar to the speaking test of the *Gaokao*, with which most participants were already very familiar. This again explains the lack of duration or intensity of their preparation practices.

Most PTE-A participants clearly indicated that they preferred computerised speaking tests and in particular, the automatic scoring method, which was indeed one of the most important reasons why they had selected the PTE-A in the first place. The primary reason underlying their preference was the perceived fairness of the automatic scoring method, as human raters were believed to behave erratically sometimes. Interestingly, this was not the case with the CET-SET4 participants, many of whom expressed their preference to the traditional IELTS-style live speaking tests because they believed that live speaking tests were more authentic and better reflected communicative language use in real life settings. The finding may be again interpreted in relation to the different nature of the two tests. As mentioned previously, the PTE-A is a high-

stakes test whose results are very important to test takers. In fact, many test takers took the PTE-A multiple times, as indicated in this study and also in Knoch et al. (2020), because they failed to meet the score requirements from overseas universities or immigration authorities. Therefore, it is not surprising that fairness in scoring is a primary consideration for test takers. By contrast, the CET-SET4 is a curriculum-based test with low stakes attached to it. As such, students seemed to prefer more authentic tasks which mirror communicative language use in real life. This is also evidenced by the participants' positive perceptions of the *pair work* task in the CET-SET4, though they also pointed out its problems which will be discussed in the next section. The finding resonates with Fan (2014) who concluded that in low-stakes contexts, test takers demonstrate a clear preference for the speaking tasks which align well with communicative language use in real life.

5.2 Test takers' preparation practices and strategies on independent and integrated speaking tasks

As mentioned previously, all tasks in the PTE-A speaking subtest can be considered as integrated, as their completion requires the use and integration of multiple language skills, albeit the degree of integration varies across tasks (see Section 2.2.1). Previous literature on the use of integrated tasks in language assessment indicates that test takers usually demonstrate positive perceptions and strong endorsement of these tasks because they are believed to be more authentic and better reflect language use in target language use (TLU) domains (Cumming, 2013; Plakans, 2013). The finding of this study generally resonates with previous literature on integrated assessment, indicating that the integrated task design of the PTE-A speaking subtest won strong endorsement from the test takers. Not surprisingly, the participants tended to focus on the target language constructs when preparing for integrated tasks. For example, 'repeat sentence' is a task which integrates listening and speaking ability. The participants reported that listening ability was important to the success on this task. As a result, many of them engaged in preparation activities which helped to strengthen their listening ability. Indeed, as one interview participant PG reported, her strategy was to prepare the PTE-A listening subtest together with this task.

When it comes to *re-tell lecture*, the participants again highlighted the importance of listening ability to the successful completion of this task. Arguably, compared with *repeat sentence*, this task entails a stronger integration of listening and speaking, as test takers need to listen to a short lecture, as opposed to a single sentence (in the case of *repeat sentence*). Indeed, understanding the input text was considered as a challenge for some test takers who are presumably not very strong in listening ability. This is clearly illustrated in the comment made by PF, 'I feel this task is challenging because first of all, I have to understand what the lecture is about.' Having said that, the automatic scoring method seems to play a significant role in shaping the participants' preparation practices and test taking strategies, an argument which we have mentioned repeatedly in this report. As far as 're-tell lecture' is concerned, the participants reported that a useful strategy that they employed was to identify the topic sentence and key words in the lecture which were then organised into a template. As PR, a focus group participant, later explained, 'the

machine can't evaluate your logic; it can, however, detect the key words in the lecture.' On both tasks (i.e. *repeat sentence* and *re-tell lecture*), and indeed on other tasks in the PTE-A speaking subtest, an essential criterion that the participants kept reiterating was fluency because it was believed that one would be penalised for disfluency or pauses in their speech. This finding, as noted previously, dovetails with Knoch et al. (2020), which clearly showcases how the automatic scoring method shapes test preparation and test taking practices.

Read aloud is a task that is common to the two speaking tests in focus, that is, the PTE-A speaking subtest and the CET-SET4. As demonstrated in our data, both groups of participants seemed to engage in intensive read aloud practice when preparing for this task; in addition, both group of participants also utilised mobile applications which could provide detailed feedback on their read aloud performance. However, two differences should be noted here. One is the PTE-A test takers' emphasis on fluency, which, as discussed previously, is related to the automatic scoring method that the PTE-A speaking subtest employs. Interestingly, though the read aloud task in the CET-SET4 is also automatically scored (see Section 2.2.2), most CET-SET4 participants did not seem to be aware of that and assumed that it was scored by human raters, like the other tasks in the test. This might explain the different findings. Another reason might be associated with the test stakes. Compared with the CET-SET4, the PTE-A speaking subtest has much higher stakes. Therefore, it is not surprising that the test takers try every means that they deem as effective to game the machine to succeed on the test. Another difference is that compared with the PTE-A participants, the CET-SET4 participants seemed to engage in more Type 1 test preparation (i.e. *construct-relevant* preparation activities) such as 'shadowing' (in the case of participant CN), or 'mimic the ways of speaking in English movies or TV series' (in the case of participant CP). This difference might be ascribed to the nature of the CET-SET4, that is, a curriculum-based test that is aligned with the College English teaching syllabus, as well as to its perceived lower stakes. The other integrated task in the CET-SET4, *questions and answers*, is related to the previous task, that is, *read aloud*. As suggested by our data, the participants believed that to succeed on this task, it was crucial to understand the input text well in the previous task; it was also essential to 'multitask', in the word of CE, which means one needs to predict the questions in the following task while performing on the *read aloud* task.

Regarding the two independent tasks in the CET-SET4, that is, *independent presentation* and *pair work*, the participants' preparation practices and strategies centre around both Type 1 (i.e. *construct-relevant* preparation) and Type 3 test preparation (i.e. *construct-irrelevant* preparation). For example, the participants reported that they would plan the presentation or build a 'mental map' (Type 1 preparation); meanwhile, some participants also mentioned memorisation and use of templates (Type 3 preparation). Regarding the *pair work* task, they would look for conversation partners to practice (Type 1 preparation); some of them also mentioned working collaboratively and strategically with their assigned partners on developing an interlocutor frame at the time

when they were allowed to test the facilities⁴ (Type 3 preparation). Topic familiarity seems to be an important variable which affects test takers' performance on the *individual presentation* task. When it comes to *pair work*, the participants mentioned that lack of familiarity with the assigned partner as well as their English proficiency level were important factors affecting their performance.

5.3 Concluding remarks

By exploring test takers' preparation practices and strategies on two different computerised speaking tests, this study reiterates the complex nature of test preparation which interacts with various factors related to a) the context in which a test is developed and used, in particular the policies concerning its use; b) the target test, including the speaking constructs that are assessed, task format, mode of delivery, test stakes and scoring method. In view of the findings that emerged from this study, we concur with Knoch et al. (2020) that it is necessary to expand the current test washback models to include: a) the broader policies concerning test use; b) new or innovative features which have been incorporated in language test design such as automated scoring method; and c) how test takers' perceptions of a test impact the agency that they exercise in test preparation, which, in turn, shapes their preparation practices and potentially test results.

The findings of this study have implications for future revisions and/or refinement of the two speaking tests. For the PTE-A speaking subtest, it is essential to continue to improve the scoring algorithm. As suggested by this study, test takers engage in some highly strategic test preparation practices such as putting an emphasis on fluency in their responses, even at the sacrifice of meaningful communication, with a view to 'gaming' the machine and to boosting their test scores. In addition, the memorisation of 'canned' responses or templates featured prominently in test takers' preparation activities because they believed that scoring mechanism would not penalise such practices. As several participants in this study reported who had attempted the PTE-A speaking subtest multiple times, these strategies were highly effective, yielding a significant improvement in their test scores. However, as mentioned previously in this report, these testwiseness strategies are *construct-irrelevant* (Messick, 1982; 1989), hence posing a threat to the validity of the PTE-A speaking subtest. In addition to improving its scoring algorithm, we also advocate that Pearson, the provider of the PTE-A speaking subtest, make the scoring algorithms transparent to relevant stakeholders, most notably the PTE-A test takers (see also Xi, 2010). By so doing, test takers are better informed of the criteria and mechanisms against which their performance will be evaluated and with which they may align their test preparation practices.

⁴ We later contacted the national CET committee, the developer of the CET-SET4, who confirmed that test takers are not supposed to talk to each other at the beginning of the CET-SET4. Our study, however, reveals that test administration procedures in different areas seemed to vary, with test takers in some areas given the opportunity to interact with their assigned partners at the beginning of the test, whereas those in other areas not given such an opportunity. As our study suggests, whether test takers are given such an opportunity to interact with their partners seems to have a quite significant effect on their performance.

Furthermore, we also recommend that Pearson continues to improve the design of the PTE-A speaking subtest. As suggested by this study, the integrated task design won strong endorsement from the participants, thanks to its perceived authenticity. However, some tasks in the PTE-A speaking subtest, such as *repeat sentence* and *answer short question*, remain quite constrained, failing to mirror communicative language use in TLU domains. As uncovered by this study, test takers tend to engage in intensive practice when preparing for these tasks. However, given the apparent lack of alignment between these tasks and communicative language use in academic settings, it is doubtful that such intensive practice would yield meaningful improvements in their academic English ability. As such, Pearson may consider replacing these tasks with more communicative tasks in future test revisions.

The findings of this study have the following implications for the CET-SET4. First, it is necessary for the national CET committee, the developer of the CET-SET4, to reconsider the role of the CET-SET4 in university students' College English learning. As this study suggests, university students do not take the CET-SET4 seriously because they cannot see much value in taking the test or because the test is only optional to them. The CET committee may consider working collaboratively with universities in China mainland to formulate local policies concerning the use of the CET-SET4. Second, we recommend the national CET committee improve the design of the CET-SET4. Several participants in this study indicate that the test is very similar to the speaking test of the *Gaokao*, except for the *pair work* task. The CET committee may consider adding one or two innovative task formats which seem novel to first- or second-year university students in China. For the *pair work* task, it is recommendable to give students the opportunity to engage in a short warm-up activity before the task. Given the enormous scale of the CET-SET4, it is also worth exploring the possibility of converting it into a fully automated spoken English test. Finally, as a teaching test which is supposed to be closely aligned with College English teaching and learning, it is important to keep monitoring the impact of the CET-SET4 on the teaching and learning of English speaking in China's universities.

References

- Alderson, J. C. (2004). The shape of things to come: Will it be the normal distribution? In M. Milanovic & C. J. Weir (Eds.), *European language testing in a global context: Proceedings of the ALTE Barcelona Conference July 2001* (pp. 1–26). Cambridge: Cambridge University Press.
- Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York and London: Routledge, Taylor & Francis Group.
- Cheng, L., Watanabe, Y., & Curtis, A. (Eds.). (2004). *Washback in language testing*. London: Lawrence Erlbaum.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Erlbaum Associates.
- Cumming, A. (2013). Assessing integrated skills. In A. J. Kunnan (Ed.), *The companion to language assessment* (Vol. 1, pp. 216–229). New York: John Wiley & Sons, Inc.

- Dong, M., Fan, J., & Xu, J. (2021). Differential washback effects of a high-stakes test on students' English learning process: evidence from a large-scale stratified survey in China. *Asia Pacific Journal of Education*, 1-18. Doi: <https://doi.org/10.1080/02188791.2021.1918057>
- Elliot, M., & Blackhurst, A. (2021). Investigating the relationship between Pearson PTE scores and IELTS scores. Accessed on May 5, 2021 at https://takeielts.britishcouncil.org/sites/default/files/investigating_the_relationship_between_pearson_pte_scores_and_ielts_bands_-_january_2021.pdf
- Fan, J. (2014). Chinese test takers' attitudes towards the Versant English Test: A mixed-methods approach. *Language Testing in Asia*, 4(6), 1-17.
- Fan, J., & Frost, K. (under review). At the intersection of language testing policy, practice and research: An interview with Yan Jin. *Language Assessment Quarterly*.
- Fan, J., & Yan, X. (2020). Assessing speaking proficiency: A narrative review of speaking assessment research within the argument-based validation framework. *Frontiers in psychology*, 11, 1-14.
- Field, A. (2009). Discover statistics using SPSS. In. London: Sage.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*: Sage.
- IBM. (2012). *IBM SPSS statistics version 21*. Boston, Massachusetts.
- Jin, Y. (2014). The limits of language tests and language testing: challenges and opportunities facing the College English Test. In D. Coniam (Ed.), *English Language Education and Assessment* (pp. 155-169). Singapore: Springer.
- Jin, Y. (2019). Testing tertiary-level English language learners: The College English Test in China, in Su, L I-W, Weir, C. J., & Wu, J. R. W. (Eds) *English Language Proficiency Testing in Asia: A New Paradigm Bridging Global and Local Contexts*. London/New York: Routledge, 101-130.
- Knoch, U., Huisman, A., Elder, C., Kong, X., & McKenna, A. (2020). Drawing on repeat test takers to study test preparation practices and their links to score gains. *Language Testing*, 37(4), 550-572.
- Lai, E. R., & Waltman, K. (2008). Test preparation: Examining teacher perceptions and practices. *Educational Measurement: Issues and Practice*, 27(2), 28-45.
- Liu, O. L. (2014). Investigating the relationship between test preparation and TOEFL iBT® performance. *ETS Research Report Series*, ETS RR-14-15.
- Messick, S. (1982). Issues of effectiveness and equity in the coaching controversy: Implications for educational and testing practice. *Educational Psychologist*, 17(2), 67-91.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (Third ed., pp. 13-103). McMillan: American Council on Education.
- National College English Testing Committee. (2016). *Test Syllabus for the College English Test* (2016 Version). Shanghai: Shanghai Jiao Tong University Press.
- Plakans, L. (2013). Assessment of integrated skills. In C. A. Chapelle (Ed.), *The Encyclopedia of applied linguistics*. New York: Blackwell Publishing Ltd.
- QSR. (2012). *NVivo qualitative data analysis software*. Melbourne: QSR International Pty Ltd.
- Richards, L. (2014). *Handling qualitative data: A practical guide*. London: Sage.
- Wang, H., Choi, I., Schmidgall, J., & Bachman, L. F. (2012). Review of Pearson Test of English Academic: Building an assessment use argument. *Language Testing*, 29(4), 603-619.

- Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27(3), 291-300.
- Xie, Q., & Andrews, S. (2013). Do test design and uses influence test preparation? Testing a model of washback with Structural Equation Modeling. *Language Testing*, 30(1), 49-70.
- Yu, G., He, L., Rea-Dickins, P., Kiely, R., Lu, Y., Zhang, J., Zhang, Y., Xu, S., & Fang, L. (2017). Preparing for the speaking tasks of the TOEFL iBT® test: An investigation of the journeys of Chinese test takers. *ETS Research Report Series*, 2017(1), 1-59.

Appendices

Appendix 1. Interviewee and Focus Group Background Questionnaire

Interviewee number: _____

Please select one of the options or fill the blanks below:

1. Gender: Male / Female
2. Age: _____
3. Number of years learning English: _____
4. Occupation/ Major: _____
5. Educational level: _____
6. Test prep institution for the PTE Academic (if attended): _____
7. Test prep institution for the CET-SET (if attended): _____
8. Times of taking the PTE Academic: _____
9. Times of taking the CET-SET: _____
10. Contact information (mobile/email): _____

Appendix 2 Interview Protocol (PTE Academic)

I. Background and motivation

1. What is your experience in English language learning and PTE Academic?

- Do you think you are a successful English language learner?
- Which is your strongest language skill(s) – reading, listening, writing or speaking?
- Have you taken the PTE Academic speaking test before? If so, could you please share with us your experience in taking the test (e.g., score, test difficulty, readiness)?

2. Why are you taking the PTE Academic test?

II. Perceptions of test/task difficulty

3. Can you make some comments on the difficulty level of the PTE Academic speaking test and its five tasks?

- Overall test
- Read aloud
- Repeat sentence
- Describe image
- Retell lecture
- Answer short question

III. Test preparation practices and strategies (Preparation lessons & solo preparation)

4. When did you start to intensively prepare for the PTE Academic speaking test?

5. Are you attending any test preparation lessons for the PTE Academic speaking test? (If yes, answer Q6 and Q7; if no, move on to Q8 and Q9.)

6. (Test preparation lessons) How are you prepared for the PTE Academic speaking test in the test preparation lessons?

- Why did you decide to go to the test preparation school?
- What do you expect to learn from the preparation lessons and why?
- What learning resources (e.g., test syllabus, official guide, online courses, mock tests) do you use in the test preparation lessons? Any other resources you use on your own outside the lessons?
- What typical class activities do you do in the test preparation lessons?
- What opportunities do you have to speak in English in the test preparation lessons? Any other opportunities outside the lessons?
- What test-taking strategies are you recommended for taking the PTE Academic speaking test?

7. (Test preparation lessons) How useful do you find are the speaking test preparation lessons to improve:

- your chance of getting a higher score in the PTE Academic speaking test?
- your English speaking ability?

8. (Solo preparation) How do you prepare for the PTE Academic speaking test on your own?

- What learning resources (e.g., test syllabus, official guide, online courses, mock tests) do you use for test preparation?
- What opportunities do you have to speak in English?
- What test-taking strategies do you employ for taking the PTE Academic speaking test?

9. (Sole preparation) How useful do you find your test preparation to improve:

- you chance of getting a higher score in the PTE Academic speaking test?
- your English speaking ability?

10. What preparation activities do you engage for the following tasks in the PTE Academic speaking test?

- Read aloud
- Repeat sentence
- Describe image
- Retell lecture
- Answer short question

11. How useful do you find are these preparation activities to improve:

- 1) you chance of getting a higher score in individual tasks?
- 2) your English speaking ability?

12. What is the percentage of time, on average, that you spend on each of the five tasks in the PTE Academic speaking test? On which task(s) do you spend the most/least amount of time?

13. Are you familiar with the rating rubrics used in the PTE Academic speaking test? When preparing for the test, do you pay attention to the specific aspects of your speaking performance that are addressed in the rating rubrics (e.g., vocabulary, content)?

14. Are you aware that your performance on the PTE Academic speaking test is automatically rated by machines? Would that affect how you prepare for the test? If so, in what ways?

IV. Others

15. Based on your own test preparation experiences, what do you like or not like about the PTE Academic speaking test? Do you have any recommendations to improve its task design?

16. Is there anything you would like to add or clarify?

Appendix 3. Interview Protocol (CET-SET4)

I. Test motivation

- 1. What is your experience in English language learning and CET-SET?**
 - Do you think you are a successful English language learner?
 - Which is your strongest language skill(s) – reading, listening, writing or speaking?
 - Have you taken the CET-SET before? If so, could you please share with us your experience in taking the test (e.g., grades, test difficulty, readiness)?
- 2. Why are you taking the CET-SET?**

II. Test/task difficulty

- 3. Can you make some comments on the difficulty level of the CET-SET and its four tasks?**
 - Overall test
 - Read aloud
 - Questions and answers
 - Individual presentation
 - Pair work

III. Test preparation (Preparation lessons & sole preparation)

- 4. When did you start to intensively prepare for the CET-SET?**
- 5. Does your English teacher give lessons on test preparation for the CET-SET? (If yes, move on to Q6 and Q7; if no, move on to Q8 and Q9.)**
- 6. (Test preparation lessons) How does your English teacher prepare you for the CET-SET?**
 - What learning resources (e.g., test syllabus, textbooks, mock tests) does your teacher use in the test preparation lessons? Any other resources you use on your own outside the lessons?
 - What typical class activities do you do in the test preparation lessons?
 - What opportunities do you have to speak in English in the test preparation lessons? Any other opportunities outside the lessons?
 - What test-taking strategies are you recommended by your teacher for taking the CET-SET?
- 7. (Test preparation lessons) How useful do you find are the speaking test preparation lessons to improve:**
 - your chance of getting a higher score in the CET-SET?
 - your English speaking ability?
- 8. (Sole preparation) How do you prepare for the CET-SET on your own?**
 - What learning resources (e.g., test syllabus, textbooks, mock tests) do you use for test preparation?
 - What opportunities do you have to speak in English?
 - What test-taking strategies do you employ to take the CET-SET?
- 9. (Sole preparation) How useful do you find is your test preparation to improve:**
 - your chance of getting a higher score in the CET-SET?
 - your English speaking ability?

VI. Test preparation at task level

- 10. What preparation activities do you engage for the following tasks in the CET-SET?**

- Read aloud
- Questions and answers
- Individual presentation
- Pair work

11. How useful do you find are these preparation activities to improve:

- 1) your chance of getting a higher score in individual tasks?
- 2) your English speaking ability demonstrated in different tasks?

12. What is the percentage of time, on average, that you spend on each of the four tasks in the CET-SET? On which task(s) do you spend the most/least amount of time?

V. Familiarity with CET-SET rating rubrics

13. Are you familiar with the CET-SET rating rubrics? When preparing for the test, do you pay attention to the specific aspects of your speaking performance that are addressed in the rating rubrics (e.g., range and accuracy of language use)?

VI. Others

14. Based on your own test preparation experience, what do you like or not like about the CET-SET? Do you have any recommendations to improve its task design?

15. Is there anything you would like to add or clarify?

Appendix 4 Survey questionnaire (PTE Academic)

This project aims to investigate how test takers prepare for the speaking subtest of the Pearson Test of English Academic. Please answer the questions honestly according to your own context and experience by selecting your response(s) or filling in the blanks. Please note that this is NOT a test; there is no correct or wrong answer. Your answer will be kept strictly confidential and used ONLY for this project. Thank you for your contribution to this project.

SECTION 1

1.1 Gender:

- (a) Male
- (b) Female

1.2 Age range:

- (a) ≤ 15
- (b) 16-18
- (c) 19-24
- (d) 25-30
- (e) ≥ 30

1.3 What is your occupation?

- (a) High school student
- (b) University student (including undergraduates and postgraduates)
- (c) Other (please specify _____)

1.3.1 If you are a high school student, which grade are you in?

- (a) Grade 1
- (b) Grade 2
- (c) Grade 3

1.3.2 If you are a university student, which degree are you working on?

- (a) BA or associate degree
- (b) MA
- (c) PhD

1.4 How many years have you been learning English? _____ (years)

1.5 English language tests you have taken:

Tests	Have you taken it?	Total score	Year
IELTS			
TOEFL iBT			
CET4 written test			
CET6 written test			
CET-SET4			
CET-SET6			
TEM4			
TEM6			

1.6 Please self-assess your English proficiency:

	Excellent	Good	Average	Below average	Poor
Overall					
Listening					
Reading					
Writing					
Speaking					

1.7 How often do you practice spoken English?

- (a) Always
- (b) Often
- (c) Sometimes
- (d) Seldom
- (e) Never

SECTION 2

2.1 Have you taken the PTE?

- (a) Yes (Please go to 2.1.1)
- (b) No

2.1.1 If Yes, how many times have you taken the PTE? _____

2.1.2 When did you take the PTE most recently? _____

2.1.3 Do you plan to take the PTE in the next six months? Yes/No

2.2 When did you start preparing for the most recent PTE that you took?

- (a) a week before the test

- (b) two weeks before the test
- (c) a month before the test
- (d) more than a month

2.3 What is your main reason for taking the PTE-A?

- (a) Study bachelor’s degree overseas
- (b) Study MA degree overseas
- (c) Study PhD degree overseas
- (d) Immigration’
- (e) Other (please specify _____)

2.4 Please rank-order the time that you spent on the tasks in the PTE-A speaking subtest in your test preparation (from the most to the least):

- (a) Read aloud
- (b) Repeat sentence
- (c) Describe image
- (d) Re-tell lecture
- (e) Answer question

2.5 How often do you have the following activities to prepare the PTE-A speaking subtest? How useful are the following activities for improving your performance on the PTE-A speaking subtest?

Activities	How often?					How useful?				
	Several times a day	Once a day	Weekly	Monthly	Never	Absolutely essential	Very useful	Somewhat useful	Slightly useful	Not useful at all
Read aloud English texts	5	4	3	2	1	5	4	3	2	1
Use mobile apps/computer software to rate read aloud performance	5	4	3	2	1	5	4	3	2	1
Listen and repeat English sentences	5	4	3	2	1	5	4	3	2	1
Dictate English sentences	5	4	3	2	1	5	4	3	2	1
Learn and practice templates for <i>describe image</i>	5	4	3	2	1	5	4	3	2	1
Practice English listening (e.g., news broadcasts, lectures, stories)	5	4	3	2	1	5	4	3	2	1

Listen and take notes	5	4	3	2	1	5	4	3	2	1
Learn and practice templates for <i>re-tell lecture</i>	5	4	3	2	1	5	4	3	2	1
Practice how to obtain key details from English lectures	5	4	3	2	1	5	4	3	2	1
Learn test taking experience on mobile apps and websites	5	4	3	2	1	5	4	3	2	1
Practice simulated or authentic items	5	4	3	2	1	5	4	3	2	1
Find a partner to practice spoken English	5	4	3	2	1	5	4	3	2	1
Get familiar with the rating criteria for the PTE-A speaking subtest*	5	4	3	2	1	5	4	3	2	1
Read the PTE-A test syllabus or official guide	5	4	3	2	1	5	4	3	2	1

2.6 How often do you use the following online resources when preparing for the PTE-A speaking subtest?

Online resources	Always	Often	Sometimes	Rarely	Never
Alpaca Education (<i>yangtuo</i>)	5	4	3	2	1
BlackTech (<i>heikeji</i>)	5	4	3	2	1
FireflyEnlightment (<i>yinghuocong</i>)	5	4	3	2	1
APEUni (<i>xinji</i>)	5	4	3	2	1
Zhihu	5	4	3	2	1
Youtube	5	4	3	2	1
Bilibili	5	4	3	2	1
Pearson official website	5	4	3	2	1

2.7 Did you attend test preparation classes?

- (a) Yes (please go to 2.7.1)
- (b) No

2.7.1 Where did the preparation class take place?

- (a) Online
- (b) Offline

2.7.2 How many students were there in your preparation class?

(a) One on one

(b) < 20

(c) > 20 (incl.)

2.7.3 Do you agree with the following statements about the test preparation class that you attended?

Statements	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I attended the class to get more familiar with the PTE-A.	5	4	3	2	1
I attended the class to more quickly enhance my test scores.	5	4	3	2	1
I attended the class to know about the items in future tests.	5	4	3	2	1
The teachers mainly used the authentic PTE-A test items.	5	4	3	2	1
The teachers mainly introduced test-taking skills.	5	4	3	2	1
The teachers provided feedback on our performance.	5	4	3	2	1
I had many opportunities to practice spoken English in class.	5	4	3	2	1
The class was important to enhance my test performance.	5	4	3	2	1
The class was important to improve my spoken English.	5	4	3	2	1

Section 3

3.1 How do you evaluate the difficulty level of the PTE-A speaking subtest?

Task	Very difficult	Difficult	Somewhat difficult	Easy	Very easy
Overall	5	4	3	2	1
Read aloud	5	4	3	2	1
Repeat sentence	5	4	3	2	1
Describe image	5	4	3	2	1
Re-tell lectures	5	4	3	2	1
Answer short questions	5	4	3	2	1

3.2 Do you agree with the following statements on automated scoring?

Statements	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Automated scoring is fairer compared with human scoring.	5	4	3	2	1
Automated scoring better reflects my spoken English proficiency.	5	4	3	2	1
If PTE-A was rated by human raters, I would adopt different preparation strategies.	5	4	3	2	1
If PTE-A was rated by human raters, I would adopt different test-taking strategies.	5	4	3	2	1

3.3 Do you agree with the following statements on the PTE-A speaking subtest?

Statements	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The PTE-A speaking subtest is well-designed.	5	4	3	2	1
The PTE-A speaking subtest score is important to me.	5	4	3	2	1
The PTE-A speaking subtest well reflects my spoken English proficiency.	5	4	3	2	1
I prefer live speaking test to computer-based speaking test.	5	4	3	2	1
I feel more relaxed in live speaking test.	5	4	3	2	1
Getting a high score on the PTE-A speaking subtest makes me more confident in English speaking.	5	4	3	2	1

3.4 Is there anything you would like to add or clarify about getting prepared for the PTE-A speaking test?

Appendix 5 Survey questionnaire (CET-SET4)

This project aims to investigate how test takers prepare for College English Test – Spoken English Test Band 4 (CET-SET4). Please answer the questions honestly according to your own context and experience by selecting your response(s) or filling in the blanks. Please note that this is NOT a test; there is no correct or wrong answer. Your answer will be kept strictly confidential and used ONLY for this project. Thank you for your contribution to this project.

SECTION 1

1.1 Gender:

- (a) Male
- (b) Female

1.2 Age range

- (a) ≤ 15
- (b) 16-18
- (c) 19-24
- (d) 25-30
- (e) ≥ 30

1.3 Which year of study are you in?

- (a) 1st
- (b) 2nd
- (c) 3rd
- (d) 4th
- (e) Other (please specify _____)

1.4 Your major:

- (a) Humanities and social sciences
- (b) Science and engineering
- (c) Business
- (d) Medical science
- (e) Others (please specify): _____

1.5 How long have you been learning English? _____ years

1.6 English language tests you have taken:

Tests	Have you taken it?	Total score	Year
IELTS			
TOEFL iBT			
CET4			
CET6			
CET-Spoken English Test (SET)			

1.7 Please self-assess your English proficiency:

	Excellent	Good	Average	Below average	Poor
Overall					
Listening					
Reading					
Writing					
Speaking					

1.8 How often do you practice spoken English in class?

- (a) Always
- (b) Often
- (c) Sometimes
- (d) Seldom
- (e) Never

1.9 How often do you practice spoken English out of class?

- (a) Always
- (b) Often
- (c) Sometimes
- (d) Seldom
- (e) Never

SECTION 2

2.1 When did you start preparing for the most recent PTE that you took?

- (a) a week before the test
- (b) two weeks before the test
- (c) a month before the test
- (d) more than a month

2.2 Why did you take CET-SET? (you can select multiple options)

- (a) Required by my university
- (b) Assess my spoken English proficiency
- (c) Follow the advice of my English teacher
- (d) Experience the CET-SET
- (e) Follow the crowd
- (f) Get the certificate for future use
- (g) For other purpose, please specify _____

2.3 Please rank-order the time that you spent on the tasks in the CET-SET4 in your test preparation (from the most to the least):

- (a) Read aloud
- (b) Answer questions
- (c) Individual presentation
- (d) Pair work

2.4 How difficult do you think of the tasks in the CET-SET4?

Task	Very difficult	Difficult	Somewhat difficult	Easy	Very easy
Overall	5	4	3	2	1
Read aloud	5	4	3	2	1
Repeat sentence	5	4	3	2	1
Describe image	5	4	3	2	1
Retell lectures	5	4	3	2	1
Answer short questions	5	4	3	2	1

2.5 How often do you have the following activities to prepare the CET-SET4? How useful are the following activities improving your performance on the CET-SET4?

Activities	How often?					How useful?				
	Several times a day	Once a day	Weekly	Monthly	Never	Absolutely essential	Very useful	Somewhat useful	Slightly useful	Not useful at all
Read aloud English texts	5	4	3	2	1	5	4	3	2	1
Shadow	5	4	3	2	1	5	4	3	2	1
Listen to English sentences and repeat	5	4	3	2	1	5	4	3	2	1

Strengthen English listening	5	4	3	2	1	5	4	3	2	1
Listen and take notes	5	4	3	2	1	5	4	3	2	1
Record performance on read aloud and self-evaluate	5	4	3	2	1	5	4	3	2	1
Practice pair work through finding partners on mobile apps or the internet	5	4	3	2	1	5	4	3	2	1
Find pair work partners in real life	5	4	3	2	1	5	4	3	2	1
Attend English corner activities	5	4	3	2	1	5	4	3	2	1
Talk to myself on certain topics	5	4	3	2	1	5	4	3	2	1
Paraphrase an English text after listening	5	4	3	2	1	5	4	3	2	1
Memorise useful words and expressions	5	4	3	2	1	5	4	3	2	1
Memorise useful sentence patterns	5	4	3	2	1	5	4	3	2	1
Prepare templates for <i>individual presentation</i>	5	4	3	2	1	5	4	3	2	1
Prepare some possible topics	5	4	3	2	1	5	4	3	2	1
Learn test taking experience on websites	5	4	3	2	1	5	4	3	2	1
Do simulated CET-SET4 items	5	4	3	2	1	5	4	3	2	1
Get familiar with the rating criteria	5	4	3	2	1	5	4	3	2	1

2.6 Please list up to 5 books, computer software or other materials that you most frequently use in your preparation for the CET-SET4?

1. _____
2. _____
3. _____
4. _____
5. _____

2.7 Please list up to 5 websites that you most frequently visit in order to prepare for the CET-SET4 and state the main reasons for using them (from the most to the least frequent).

	Website	Main reasons
--	---------	--------------

1		
2		
3		
4		
5		

SECTION 3

3.1 How often did you engage in the following strategies when taking the CET-SET4? How useful were these strategies for improving your performance on the CET-SET4?

Activities	How often?					How useful?				
	Very often	Once	Sometimes	Rarely	Never	Absolutely essential	Very useful	Somewhat useful	Slightly useful	Not useful at all
<i>Read aloud and questions and answers</i>										
Read the text beforehand	5	4	3	2	1	5	4	3	2	1
Memorise key points while reading the text beforehand	5	4	3	2	1	5	4	3	2	1
Memorise key points while reading aloud the text	5	4	3	2	1	5	4	3	2	1
Control the speed and timing while reading aloud	5	4	3	2	1	5	4	3	2	1
<i>Individual presentation</i>										
Plan the points to be covered in the presentation beforehand	5	4	3	2	1	5	4	3	2	1
Plan the structure of the presentation beforehand	5	4	3	2	1	5	4	3	2	1
Use templates that have been prepared	5	4	3	2	1	5	4	3	2	1
Use words and expressions that have been prepared	5	4	3	2	1	5	4	3	2	1
<i>Pair work</i>										
Communicate with the partner beforehand	5	4	3	2	1	5	4	3	2	1

Plan the structure of conversation with the partner beforehand	5	4	3	2	1	5	4	3	2	1
Predict how the partner is going to respond	5	4	3	2	1	5	4	3	2	1
Take initiative in pair work	5	4	3	2	1	5	4	3	2	1

3.2 Do you agree with the following statements about the CET-SET4?

Statements	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
CET-SET4 was well-designed.	5	4	3	2	1
I can demonstrate my spoken English proficiency in CET-SET4.	5	4	3	2	1
In <i>individual presentation</i> , the topic might affect my performance.	5	4	3	2	1
In <i>pair work</i> , my partner's proficiency level might affect my performance.	5	4	3	2	1
I am familiar with the CET-SET4 rating criteria.	5	4	3	2	1
CET-SET4 scores are important to me.	5	4	3	2	1
I paid more attention to spoken English learning because of CET-SET4.	5	4	3	2	1
CET-SET4 motivated me to study English.	5	4	3	2	1
I have improved my spoken English proficiency through preparing the CET-SET4.	5	4	3	2	1
A good score on the CET-SET4 makes me more confident in spoken communications.	5	4	3	2	1
I prefer computer-based spoken English test to a live one.	5	4	3	2	1
I feel more relaxed when taking a computer-based spoken English test.	5	4	3	2	1
Computer-based spoken English tests can effectively measure our spoken English proficiency.	5	4	3	2	1

3.3 Is there anything you would like to add or clarify about getting prepared for the CET-SET4?