Students' Perspectives on a School-based English Proficiency Test

Wenxia Zhang
Tsinghua University, China

Meihua Liu [*]
Tsinghua University, China

This paper reports on a study on students' perceptions of a school-based English proficiency test. Administering a 60-item survey to 526 students in three different study years at a university in Beijing, the study revealed that (1) most students perceived different parts of the test difficult, (2) the majority of the participants felt anxious during the test, (3) first-year students considered the proficient test the easiest, felt the least anxious, and performed the best in the test, (4) students who considered one part of the proficiency test more difficult tended to score lower in that part, perceive other parts more challenging and feel more anxious during the test, and (5) all the measured variables were significantly related to one another and had a negative relationship with students' achievement in the test. All these findings attest to the reliability and validity of the proficiency test. Based on these findings, some implications are discussed.

Key words: perception, proficiency test, university, study year

---

[*] **Corresponding author**

INTRODUCTION

Because foreign language learners usually have little access to the target language, they may not be so positive about and motivated to learn/use the target language as SL learners. Although increasing FL learners' exposure to and contact with the target language via certain programs can enhance their interest in and motivation to use/speak the language (Chlebek & Coltrinari, 1977; Park & Oxford, 1998), various tests are most often used as a main tool to motivate them to learn/use the target language, in addition to providing information for making decisions (Bachman, 1999a). The case is especially so in Mainland China: except that a high score in an English exam is compulsory for students to move into a higher grade, especially to be admitted into a good high school and later a good college or university, to pass the nationwide College English Test (CET) bands 4 and 6 targeting undergraduate and graduate non-English majors respectively is a must for university/graduate students to be granted the degree certificate. Considering that some students are really outstanding at English, the Bureau of Education set a law in 1996 that undergraduate students in eight key universities such as Peking University and Tsinghua University across the country do not need to pass the exam. Nevertheless, these students are often required to pass a self-designed school-based English proficiency test to obtain the certificate. The university where the present study took place belongs to this group.

While quite much research has been done to investigate learners' attitudes towards and/or perceptions of a SL/FL learning program or teaching method and so on (Pang & Liu, 2006; Park & Oxford, 1998), there is a dearth of research on test-takers' perspectives of a standardized SL/FL test, especially a school-based proficiency test. For this reason, the present research aimed to examine how students perceived a school-based English proficiency test and its relationship with their achievement in the test in a Chinese EFL context.

LITERATURE REVIEW

Concerning learners' attitudes towards and/or perspectives of a SL/FL teaching method, a learning program, and/or an activity and so on, much research can be found (Chlebek & Coltrinari, 1977; Li, 1998; Pang & Liu, 2006; Park & Oxford, 1998; Spratt, 1999). In order to investigate the adoption of the communicative approach in EFL classrooms in South Korea, Li (1998) administered a questionnaire to 18 South Korean secondary school EFL teachers studying at a Canadian university and conducted in-depth interviews with 10 of the participants. He found that students' resistance to class participation was a primary constraint in implementing CLT. Students were said to "have become accustomed to the traditional classroom structure, in which they sit motionless, takes notes while the teacher lectures, and speak only when they are spoken to" (1998, p. 691). Park and Oxford's (1998) study explored the possible outcomes of changing teachers' roles in a 5-week summer intensive program in a Korean university. All the participants of the program were required to speak only English and those who were found speaking Korean more than three times were asked to withdraw. Upon successful completion of program requirements, the participants would receive a certificate showing official recognition of their progress in English. Therefore, it was no wonder that the participants were actively involved in classroom activities and thought highly of the program. To explore students' perspectives on a summer school program, Pang and Liu (2006) administered a 15-item survey to 1331 students at six different grade levels at a Macao secondary school. The researchers found that (1) the majority of the participants held negative attitudes towards the school-based English program and were not motivated to learn spoken English, (2) students at higher grade levels tended to be significantly more motivated and held more positive attitudes towards the program, and (3) students' English-learning motivation and attitudes towards the program were significantly positively correlated with their English proficiency.

As to language tests, though much research can be reviewed, most aimed

at investigating the validity and reliability of various tests mainly from the aspects of construct, content, item difficulty, test performance, and/or washback from language teachers (Heaton, 2000; Henk & Holmes, 1988; Moss, 1995; Qi, 2004, 2005; Slomp & Fuite, 2005; Yang & Weir, 1998). In order to explore the validity of College English Test (CET) bands 4 & 6 in China, Yang and Weir (1998) conducted detailed analyses of the two tests in terms of construct validity, content validity, concurrent validity, predictive validity and face validity. They also analyzed the test structure, the scoring system, data-analyzing software, and test management and administration of the tests. In addition, they tested students from six universities in the country and conducted a large-scale survey to language teachers. Analyses of the data showed that CET bands 4 & 6 were highly reliable and valid, and that 92% of the teacher participants believed the two tests could truly test students' English proficiency. As a fairly complicated and comprehensive study, the report unfortunately did not present any clear discussion of the research method used. It also failed to take into consideration test-takers' perceptions towards the two tests. Likewise, though Qi's (2004, 2005) studies targeted the washback effect of the National Matriculation English Test (NMET) in China via interviews and questionnaires from 8 NMET constructors, 6 English inspectors, 388 teachers and 986 students, the reports somehow neglected the analyses and discussion of students' ideas of the washback effect of the NMET, which, however, might be upcoming.

Probably due to the fact that test-takers generally disperse soon after finishing a standardized test, it is difficult to ask enough of them for introspective ideas about the test. As a result, little can be reviewed about research on test-takers' perceptions of a standardized language test, especially a school-based proficiency test (Malabonga, et al., 2002). For different purposes, various kinds of language tests such as translation tests, objective and subjective tests have been designed. According to Heaton (2000, p. 5), a language test which "seeks to find out what candidates can do with language provides a focus for purposeful, everyday communication activities". As communicative competence is increasingly emphasized, this also becomes the aim of more and more

language tests. Whatever kind of test it is, however, its validity and reliability should be ensured to achieve its aim. For this reason, any test development is a complicated process of "creating and using a test, beginning with its initial conceptualization and design, and culminating in one or more archived tests and the results of their use" (Bachman, 1999b, p. 85). This implies that test-takers' perceptions of a test need to be incorporated into the development of a reliable and valid language test. For this reason, the present research sought to explore students' perspective of a self-designed school-based English proficiency test and its relationship with their achievement in the test, with the hope of providing some suggestions for test developers and curriculum designers at the University. Moreover, because language tests often drive students anxious (Covington & Omelich, 1987; Mackenzie, 1994; Young, 1991), the present study also aimed to examine how students felt during the test and its relationship with their achievement in the test. To achieve these aims, the following questions were formulated:

(1) How do students perceive the self-designed school-based English proficiency test?

(2) How are individual perception survey items related to students' achievement in the related part of the proficiency test?

(3) How do students feel during the school-based English proficiency test?

(4) What are the differences in their perceptions of and degrees of anxiety during the proficiency test among students in different study years?

(5) How are students' perception of the proficiency test, anxiety and achievement in the test related to one another?

RESEARCH METHODOLOGY

This paper reports on part of a study which investigated students' perceptions of a school-based English proficiency test and the strategies they employed to finish the test.

Context. Tsinghua University, the top 1 or 2 university in Mainland China, is famous for its outstanding students who are generally better than most other university students in different areas of learning, including English. At this university, English courses are compulsory for freshmen at the first term but become selective later on. Tsinghua English Proficiency Test I (TEPT1), developed by the Department of Foreign Language of the University and authorized by the Bureau of Education in 1999, has been existent for 10 years since 1996 and is correspondingly more difficult than the CET band 4. As an exit and proficiency test for non-English majors, the TEPT1 consists of two components: written (85 points) and oral (15 points) tests, the written component of which has three parts: listening comprehension (30 points), reading comprehension (40 points), and writing (15 points). The test is administered on the 8th Sunday of each 18-week long term and it is up to students to decide when to take the test during their 4-year university time.

Participants. This study involved 526 (411 male and 115 female) undergraduate non-English majors with an average age of 19.4, among whom, 157 were freshmen, 153 sophomores and 216 juniors.

Instrument. Along with a 83-item test-taking strategy survey, a 60-item self-designed survey was administered to students, which covered (a) a 10-item Perception of Listening Comprehension (PLC), (b) a 17-item Perception of Reading Comprehension (PRC), (c) a 6-item Perception of Writing Tasks (PWT), (d) 15-item Perception of the Oral Test (POT), (e) a 4-item Assessment of the TEPT1 2005 (ATEPT), (f) a 2-item Self-rated Performance in the TEPT1 2005 and (g) a 6-item Anxiety during the TEPT1 2005 (A-TEPT). The survey was designed on a 5-point Likert scale ranging from 'Strongly Disagree' to 'Strongly Agree' with values 1-5 assigned to the descriptors respectively.

An open-ended question. In addition to the survey, an open-ended question was designed to allow respondents to write down any other ideas about the TPET1 2005.

Background information. The background questionnaire was designed to

obtain information about participants' name, gender, and grade level.

Achievement in the TEPT1. All the participants' scores (total score and scores in different parts) were collected as their achievement in the test.

Procedure. The written part of the TEPT1 2005 was held for two hours on the 8th Sunday morning of the first term of the academic year 2005-2006. The oral test was held thereafter in the form of a 1-minute teacher-student conversation and a 5-minute student-student conversation. As soon as a student finished the oral test, s/he was asked to fill in the survey in about 20 minutes. Altogether, 547 questionnaires were collected, of which 526 were valid for statistical analyses.

Data analysis. The survey was analyzed in terms of mean, standard deviation, median, mode, and range to examine how students perceived different parts of the TEPT1 2005 and their degrees of anxiety during the test. Correlation analyses were then conducted to determine the relationship between the perception survey items and students' achievement in certain parts of the TEPT1 2005. In order to explore the differences in the measured variables among students in different study years, one-way ANOVA (Duncan's test) was conducted. Finally, correlation analyses were run to investigate the relationships between the measured variables and students' achievement in the TEPT1 2005. The open-ended question was analyzed and calculated according to different themes (Krippendorff, 1980), which was later integrated into the discussion of the survey data.

RESULTS AND DISCUSSION

Perceptions of the Proficiency Test

Perceptions of Different Parts of the Test

As previously described, the survey concerning perceptions of the proficiency

test covers fours parts: perception of the listening comprehension (PLC), perception of the reading comprehension (PRC), perception of the writing tasks (PWT) and perception of the oral test (POT). In order to explore the general tendency of students' perceptions of these parts, the mean, standard deviation, median, mode and range of each perception scale were computed. It holds true that the higher the score, the more difficult a respondent thought a part was. The results are presented in Table 1.

TABLE 1
Statistical Analyses of the PLC, PRC, PWT and POT (N = 526)

Measures	Mean	Standard deviation	Median	Mode	Range
PLC	28.32	5.79	28.50	30.00	10.00-48.00
PRC	53.72	7.81	54.00	52.00	23.00-84.00
PWT	16.59	3.67	17.00	18.00	6.00-30.00
POT	28.11	5.55	28.00	31.00	11.00-48.00

With a reliability score of .692 (Cronbach's alpha), the 10-item Perceptions of Listening Comprehension (PLC) had significant part-whole correlations with the total score, with the mean item-total correlation being .507. With a possible score range of 10.00 to 50.00, the actual score range for the 10-item PLC in the present study was 10.00 to 48.00 and the mean score for the 526 participants was 28.32 (SD = 5.79), as presented in Table 1. All slightly below the average score of 30.00, the mean, the median (28.50) and the mode (30.00) on the PLC clearly reveal that more than half of the participants did not think the listening test difficult. This was further confirmed by the statistical analyses of each PLC item (see items 1-6, 9-12, Appendix). Consistent with the responses to the open-ended question which indicated that some participants strongly believed the listening test was difficult in that it was difficult to answer questions after listening to the materials only once, nearly half of the respondents disagreed that "listening comprehension was easy", as indicated by the mean (3.26) and

standard deviation (1.10) of item 11. Nevertheless, more than half had enough time to finish the listening test (item 5 with a mean of 2.54). Although many respondents believed that little practice made the test difficult for them (item 10 with a mean of 3.27), more than half (strongly) rejected having difficulty comprehending the listening materials due to various reasons such as too many unfamiliar words (item 1 with a mean of 2.73), complicated grammar (item 2 with a mean of 2.37), fast speed (item 3 with a mean of 2.71), low voice quality (item 4 with a mean of 2.86 and item 12 with a mean of 2.85), and unfamiliarity with the topics (item 9 with a mean of 2.83).

Likewise, the 17-item PRC (a = .733) also had significant part-whole correlations with the total score, with the mean item-total correlation being .44. With a possible range of scores from 17.00 to 85.00, analyses of the PRC data revealed that the actual range for the participants in this study was 23.00 to 84.00 and that the mean was 53.72 (SD = 7.81). This, coupled with a median of 54.00 and a mode of 52.00 shown in Table 1, all above the average score of 51.00, implies that the majority of the participants thought the reading test rather difficult. This was further supported by the statistical analyses of each PRC item (see items 7-8, 13-27, Appendix). As indicated by the mean (3.59) and standard deviation (1.05) of item 19, most respondents vetoed the statement that "reading comprehension was easy". Although more than half (strongly) rejected having difficulty comprehending the reading materials due to too many complicated sentences (item 7 with a mean of 2.87) and/or unfamiliarity with the topics (item 18 with a mean of 2.95), nearly the same number of them attributed the difficulty of the reading test to little practice (item 20 with a mean of 3.25) and/or many strange words (item 8 with a mean of 3.06). Moreover, around half of the respondents could understand the main ideas but had difficulty deducing the writers' opinions/ intentions of the reading passages (item 14 with a mean of 3.08). All these were consistent with students' responses to the open-ended question, many of whom commented the reading passages were full of complicated sentences, unfamiliar words and too long to be finished in the set time.

In addition, as shown in the Appendix, most respondents endorsed item 15

"short-answer questions were more difficult than multiple-choice questions" (mean = 3.41). They thus felt diffident completing the sentences when doing reading comprehension (item 16 with a mean of 3.63) and claimed grammar as one of the causes (item 17 with a mean of 3.29). A similar trend was found with the cloze test and English-Chinese translation. The majority of the participants disfavored statements "the cloze test was easy" (item 21 with a mean of 3.29) and "the translation test was easy" (item 25 with a mean of 3.41) due to different reasons, the main one of which was unfamiliar words (item 23 with a mean of 3.10 and item 26 with a mean of 3.43).

As to the 6-item PWT (a = .591; the mean item-total correlation was .573), both the possible and actual score range was 6.00 to 30.00. As reported in Table 1, a mean of 16.59, a median of 17.00 and a mode of 18.00, all below the average score of 18.00, clearly suggest more than half correspondents considered the writing test easy. This conformed to the statistical analyses of each PWT item (see items 28-31, 34-35, Appendix). Although more than half of the correspondents disagreed that the writing tasks were easy (items 34 and 35 with a mean of 3.38 and 3.26 respectively), the majority believed writing task 1 was easy (item 29 with a mean of 2.39). They also vetoed statements like "writing task 1 was difficult because I had never practiced it before" (item 28 with a mean of 2.30), "writing task 2 was difficult because it required more words" (item 30 with a mean of 2.80) and "Writing task I was difficult because little information was given" (item 31 with a mean of 2.48).

Because the POT survey included 4 items indicative of respondents' perceptions of the test teachers, these items (37, 39, 40 and 43) were analyzed separately. Thus, the 11-item POT (a = .716; the mean item-total correlation was .514) had an actual score range of 11.00 to 48.00 in the present study (with a possible score range of 11.00- 55.00). As presented in Table 1, a mean of 28.11, a median of 28.00 and a mode of 31.00, all below the average score of 33.00, reveal that the majority of the participants considered the oral test easy, in conformity with students' responses to the open-ended question which even suggested that the oral test should have been more difficult. This was further supported by the statistical analyses of each POT item (see items

41-42, 44-51 and 54, Appendix). As presented in the Appendix, nearly half of the participants reported that "'teacher-student communication' was easy" (item 42 with a mean of 2.51) and "'student-student conversation' was easy because I had practiced it a lot before" (item 50 with a mean of 3.14). Though many participants reported to be unfamiliar with the topics (item 54 with a mean of 3.02), more than half (strongly) vetoed statements like "the card given to me during the oral test was hard to understand" (item 44 with a mean of 2.27), "I couldn't understand what I was required to do during the oral test" (item 46 with a mean of 2.42), "the part 'filling in the gap according to the card' was difficult" (item 48 with a mean of 2.74) and "the topics for the oral test were difficult" (item 49 with a mean of 2.74). They also disfavored statements about the negative influence of their partners on their own performance in the oral test (items 41, 45, 47 and 51 with a mean of 2.39, 2.13, 2.20 and 2.74 respectively).

Further, as reported in the Appendix, the majority of the correspondents (strongly) agreed that the teacher(s) spoke English clearly (item 37 with a mean of 4.16) and at a proper speed (item 43 with a mean of 3.75), were friendly (item 39 with a mean of 4.11), and gave clear instructions (item 40 with a mean of 3.88) during the oral test. Consequently, teacher factor(s) could hardly have negatively affected students' performance in the oral test.

Assessment of and Self-assessed Performance in TEPT1

In addition to expressing opinions about different components of the TEPT1 2005, participants were required to assess the overall difficulty of the written and oral tests of and their own performance in the TEPT1 2005. For this purpose, the mean, standard deviation, median, mode and range of the 4-item Assessment of the TEPT1 2005 (ATEPT1) and the 2-item Self-assessed Performance were computed. The higher the score, the more difficult the TEPT 2005 was or the worse a respondent had performed in the test. The results are reported in Table 2.

TABLE 2
Statistical Analyses of the ATEPT1 and Self-rated Performance (N = 526)

Measures	Mean	Standard deviation	Median	Mode	Range
ATEPT1	13.15	2.68	13.00	12.00	4.00-20.00
Self-rated performance	6.50	1.54	6.00	6.00	2.00-10.00

With a possible score range of 4.00 to 20.00, the actual score range for ATEPT1 (a = .518; the mean item-total correlation was .638) was the same, as presented in Table 2. In addition, a mean of 13.15, a median of 13.00 and a mode of 12.00 on the ATEPT1, all above the average score of 12.00, indicate that most participants considered the TEPT1 2005 difficult, consistent with the statistical analyses of each ATEPT item. As evidenced by the means of items 36 (3.01) and 55 (2.79) presented in the Appendix, nearly half of the respondents considered the oral test of the TEPT1 2005 difficult, contrary to the finding presented in the preceding section. And more believed the written test challenging. More than half of the participants disagreed that "the written test on the whole was easy" (item 33 with a mean of 3.50) and even more reported "the written test was difficult" (item 56 with a mean of 3.84). This was further supported by students' responses to the open-ended question. Many of them (in all study years) commented that it was difficult to answer the questions after listening to the dialogues or passages only once, that the reading passages were too long to be finished in the set time and that there was little time left to finish two writing tasks. Consequently, the TEPT1 2005, on the whole, seemed difficult for most participants, though many of the previous statistical analyses indicated that most of them rejected thinking in this way.

As reported in Table 2, a mean of 6.50, a median of 6.00 and a mode of 6.00 on the 2-item self-rated performance in the TEPT1 2005, all slightly above the average score of 6.00, clearly show that around half of the respondents thought they had failed to perform as well as they could have in the test. This was consistent with the statistical analyses of items 52 and 53

(with means of 3.31 and 3.21 respectively), as presented in the Appendix.

Correlations between Perception Survey Items and Students' Achievement in Related Parts of the TEPT1 2005

To explore the relationship between individual perception survey items and students' achievement in related parts of the TEPT1 2005, correlation analyses were conducted. The results are summarized in Table 3, which reveals that altogether 23 items had significant correlations with students' achievement in related parts of the TEPT1 2005.

TABLE 3
Correlations between Perception Survey Items and
Students' Achievement in Related Parts of the TEPT1 2005 (N = 526)

Items	A in listening	Items	A in reading	Items	A in writing	Items	A in oral test	Items	A in TEPT1
1	-.299**	7	-.132**	30	-.128**	41	-.150**	33	-.091*
2	-.256**	8	-.174**	34	-.111*	45	-.166**	36	-.106*
3	-.126**	22	-.139**			46	-.136**	56	-.141**
4	-.092*	23	-.238**			48	-.097*		
9	-.182**					49	-.125**		
10	-.252**					51	-.137**		
11	-.151**					39	.107*		

Notes: A → achievement;   * p ≤ .05; ** **p ≤ .01**

According to Table 3, 7 of 10 PLC items were significantly negatively correlated with students' achievement in the listening comprehension of the TEPT1 2005: item 1, "had difficulty understanding what I heard because there were too many unfamiliar words" (r = -.299); 2, "had difficulty understanding what I heard because of complicated grammar" (r = -.256); 3, "had difficulty understanding what I heard because the speaker(s) spoke too fast" (r = -.126); 4, "had difficulty understanding what I heard because I couldn't hear it clearly" (r = -.092), 9, "had difficulty understanding what I heard because I was not familiar with the topics" (r = -.182); 10, "listening comprehension was difficult because I had done little practice" (r = -.252);

and 11, "listening comprehension was easy" ($r = -.151$). As it stands, the first six items signified respondents' degree of comprehension of the listening materials; item 11 was implicative of overall assessment of the difficulty of the listening test. If a respondent scored higher on these items, the lower s/he scored in the listening test. Namely, the less a respondent understood the listening materials, and/or the more difficult s/he thought the listening test was, the lower s/he achieved in this part.

Among 17 PRC items, 4 significantly inversely correlated with students' achievement in the reading comprehension of the TEPT1 2005, as shown in Table 3: item 7, "had difficulty understanding the reading passages because there were too many complicated sentences" ($r = -.132$); 8, "had difficulty understanding the reading passages because there were too many strange words" ($r = -.174$); 22, "had difficulty doing the cloze test because I didn't understand the grammar" ($r = -.139$); and 23, "there were lots of unfamiliar words in the cloze test" ($r = -.238$). It is clear that items 7 and 8 revealed respondents' degree of comprehension of the reading materials and that items 22 and 23 reflected respondents' comprehension of the materials in the cloze test. The less a respondent comprehended the reading materials and/or the materials in the cloze test, the less s/he achieved in the reading test of the TEPT1 2005.

Likewise, 2 of 6 PWT items were significantly negatively related to students' achievement in the writing test, as revealed in Table 3: item 30, "writing task 2 was difficult because it required more words" ($r = -.128$) and 34 "the writing tasks were easy because I had a large vocabulary" ($r = -.111$). Alternatively, a respondent who perceived the writing test more challenging tended to perform worse in the test.

As noted from Table 3, among 15 POT items, 7 significantly correlated with students' achievement in the oral test: item 39, "the teacher(s) was/were nice and friendly during the oral test" ($r = .107$); 41, "my partner and I were at different spoken English proficiency levels, which affected my performance in the oral test" ($r = -.150$); 45, "couldn't understand my partner during the oral test" ($r = -.166$); 46, "couldn't understand what I was required to do during

the oral test" (r = -.136); 48, "the part 'filling in the gap according to the card' was difficult" (r = -.097); 49, "the topics for the oral test were difficult" (r = -.125); and 51 "my partner and I didn't cooperate well because we didn't know each other" (r = -.137). Except that item 39 implicative of the degree of friendliness of the test teacher(s) was positively related to students' achievement in the oral test, the other 6 items which reflected the degree of difficulty of the oral test were inversely correlated with students' achievement in the test. In other words, the more friendly a respondent perceived the test teacher(s) to be, the better s/he performed in the oral test; the more difficult a respondent thought the oral test, the worse s/he performed in it.

Table 3 also shows that 3 of 4 ATEPT1 items significantly negatively correlated with students' achievement in the TEPT1 2005: item 33, "the written test on the whole was easy" (r = -.091); 36, "the oral test on the whole was difficult" (r = -.106); and 56, "the written test was difficult" (r = -.141). Namely, the more difficult a respondent assessed the TEPT1 2005 to be, the lower s/he scored in the test.

In short, nearly half of the perception survey items significantly correlated with students' achievement in the related parts of the TEPT1 2005. Generally speaking, a student who perceived a certain part (e.g., listening comprehension and reading comprehension) of the proficiency test difficult tended to score lower in it. This implies the TEPT1 2005 was valid in that it tested students' proficiency in different areas of English as it had claimed.

Anxiety During TEPT1 2005

Related to the difficulty of the TEPT1 2005, a 6-item survey (A-TEPT1) (*a* = .763; the mean item-total correlation was .676) was developed to measure participants' degree of anxiety during the test (see items 32, 38 and 57-60, Appendix). In order to explore the general tendency of students' anxiety during the TEPT1 2005, the mean, standard deviation, median, mode and range of the A-TEPT1 were computed. The total score indicated a participant' degree of anxiety during the test; the higher the score, the more

anxious s/he felt. The results are summarized in Table 4.

TABLE 4
Statistical Analyses of the A-TEPT1 (N = 526)

Mean	Standard deviation	Median	Mode	Range
19.39	4.46	20.00	22.00	6.00-29.00

With a possible score range of 6.00 to 30.00, the actual score range for the 6-item A-TEPT1 in this study was 6.00 to 29.00 and the mean score for the 526 participants was 19.39 (SD = 4.46). As shown in Table 4, the A-TEPT1 mean, median (20.00) and mode (22.00) were all above the average score of 18.00, which clearly implies the majority of the participants felt anxious during the TEPT1 2005 due to various reasons. This was further supported by the statistical analysis of each A-TEPT1 item. As presented in the Appendix, except that about half of the participants reported they felt confident when doing writing during the test (item 32 with a mean of 2.72), the majority acknowledged they were anxious during the oral test (item 38 with a mean of 3.46) and became nervous when not knowing the answer (item 60 with a mean of 3.49) and/or not understanding what they heard (item 58 with a mean of 3.29) or read (item 59 with a mean of 3.27). Some respondents, when responding to the open-ended question, commented that they simply became nervous when facing the teacher(s) or speaking English in front of others.

Difference in Measured Variables Among Students in Different Study Years

In order to investigate differences in the measured variables among students in different study years, analysis of ANOVA was conducted, the results of which are reported in Table 5.

TABLE 5
ANOVA Results of the Measured Variables

Measures	F	P	F*	Study years (Mean) Year 1 = 157; Year 2 = 216; Year 3 = 153			Location of Sig. Difference (alpha = .05)
				1	2	3	
PLC	.359	.699	2.77	28.15	28.65	28.19	
PRC	3.55*	.030	2.77	55.07	53.36	52.92	1 & 3; 1 & 2
PWT	1.288	.277	2.77	16.26	16.52	16.87	
POT	.490	.613	2.77	27.89	27.95	28.41	
ATEPT1	.738	.479	2.77	13.36	13.08	13.04	
Self-rated performance	1.28	.278	2.77	6.63	6.35	6.52	
A-TEPT1	2.66*	.071	2.77	18.83	19.27	19.89	1 & 3
Written test score	2.45*	.087	2.77	60.32	59.03	58.14	1 & 3
Oral test score	10.87*	.000	2.77	12.26	11.48	11.72	1 & 3; 1 & 2
Total score	4.39*	.013	2.77	72.47	70.51	69.06	1 & 3

Notes: * $\rightarrow$ Significant differences are reached.

F* $\rightarrow$ Critical F value for Duncan's test at .05 level (Black, 1999).

As noted in Table 5, the majority of the participants in different study years considered the listening test easy (mean = 28.15, 28.65 and 28.19 for 1st-, 2nd- and 3rd-years respectively); the reading test challenging (mean = 55.07, 53.36 and 52.92 for 1st-, 2nd- and 3rd-years respectively); the writing test easy (mean = 16.26, 16.52 and 16.87 for 1st-, 2nd- and 3rd-years respectively); and the oral test easy (mean = 27.89, 27.95 and 28.41 for 1st-, 2nd- and 3rd-years respectively). As to the overall assessment of the TEPT1 2005, most participants in different study years (strongly) believed the TEPT1 2005 challenging (mean = 13.36, 13.08 and 13.04 for 1st-, 2nd- and 3rd-years respectively) and did not think they performed as well as they had expected in the test (mean = 6.63, 6.35 and 6.52 for 1st-, 2nd- and 3rd-years respectively). Probably for this reason, most of them in different study years became (highly) anxious during the TEPT1 2005 (mean = 18.83, 19.27 and 19.89 for freshmen, sophomores and juniors respectively).

A cross-comparison of the means for students in different study years on the measured variables suggests that freshmen thought the reading test the

most difficult but the listening, writing and oral tests the easiest, that sophomores considered the listening test the most challenging, and that juniors perceived the writing and oral tests the most challenging but the reading test the easiest. This is not surprising in that all freshmen were required to take English courses and still had the compassion for learning the language, while sophomores and juniors had to take no compulsory English courses at the University. Except that they might have not accumulated a large number of English words which might result in difficulty in reading comprehension, freshmen had had plenty of writing and listening practice to be admitted into a university and the memory had not yet faded. Coupled with the practice in University lessons, freshmen might find it easy to write in English effectively, understand listening materials and communicate with others orally. Nevertheless, this might become harder for sophomores and juniors, especially juniors who had not formally accessed English for quite a long time.

Surprisingly, a further review of the means presented in Table 5 shows that freshmen rated the TEPT1 2005 the most difficult while juniors believed it the easiest, contrary to the results found in previous analyses. This might be because it was the first time for freshmen to take the TEPT1 while students in other years had more experiences with the test and thus had more knowledge about it. Consequently, freshmen thought they had performed the worst while their second-year peers had done the best in the test. Even so, freshmen remained the least anxious while their third-year counterparts became the most nervous during the TEPT1 2005. This might be because first-year students had more chances to take the TEPT1 if they failed, whereas, juniors had fewer such chances before graduation.

Though the above results indicate participants in different study years differed from one another in terms of perceptions of different components of the TEPT1 2005, assessment of the TEPT1 2005, self-rated performance and anxiety during the test, differences were not always significant, as evidenced by F values shown in Table 5. The ANOVA results presented in Table 5 suggest that significant differences existed between the first- and third-year

students in anxiety during the TEPT1 2005 (A-TEPT1) and achievement in the written test of the TEPT1 2005 and the overall TEPT1 2005, while no significant differences were found between the first- and second-year or between the second- and third-year students in these aspects. Significant differences were also found between the first- and second-year students and the first- and third-year students but not between the second- and third-year students in perceptions of the reading test (PRC) and achievement in the oral test of the TEPT1 2005. Freshmen achieved the highest in the written test and the overall TEPT1 2005 and were the least anxious during the TEPT1 2005, whereas juniors achieved the lowest and felt the most anxious. This might lie in the fact that freshmen, fresh out of senior high school where they had received plenty of training for exams, knew better how to deal with different kinds of exams and that they still had enough opportunities to take the TEPT1 if they failed. Consequently, they felt more at ease during the TEPT1 2005 and performed better in it. By contrast, juniors did the worst in the TEPT1 2005 maybe because they had much less contact with English at the time and thus might have long forgotten how to tackle important exams. Also, because not many opportunities to take the TEPT1 before graduation waited for them, they unavoidably became more anxious during the test.

Correlations Among the Measured Variables

In addition to the statistical analyses of measured variables, correlation analyses were conducted to explore the relationships among these variables to better understand students' perception of the difficulty of the TEPT1 2005 and its impact on their achievement in the test. The results are reported in Table 6.

TABLE 6
Correlations among the Measured Variables (N = 526)

	PLC	PRC	PWT	POT	ATEPT1	Self-rated performance	A-TEPT1	Written test score	Oral test score
PRC	.444**	1							
PWT	.367**	.335**	1						
POT	.354**	.245**	.293**	1					
ATEPT1	.392**	.484**	.380**	.163**	1				
Self-rated performance	.148**	.130**	.041	.245**	.182**	1			
A-TEPT1	.256**	.349**	.372**	.259**	.369**	.168**	1		
Written test score	-.220**	-.136**	-.167**	-.086	-.153**	-.007	-.123**	1	
Oral test score	-.236**	-.115*	-.165**	-.170**	-.187**	-.061	-.191**	.411**	1
TEPT score	-.221**	-.124**	-.179**	-.088*	-.162**	-.012	-.129**	.987**	.498**

Notes: ** $p < .01$;   * $p < .05$

According to Table 6, all the measured variables were significantly positively correlated with one another. For example, respondents who scored higher on the PLC tended to score higher on the PRC, the POT, the PWT, the ATEPT, self-rated performance and the A-TEPT1. Alternatively, the more difficult students considered the listening comprehension to be, the more difficult they believed the reading comprehension, the writing tasks and the overall TEPT1 2005 to be, the more anxious they felt during the TEPT1 2005 and the worse they thought they had performed in the test as well.

Moreover, all these measured variables were significantly negatively correlated with students' achievement in the written and oral tests of the TEPT1 2005 and the overall TEPT1 2005. For instance, the more difficult respondents perceived the writing tasks to be, the worse they performed in the TEPT1 2005 as well as its written and oral parts.

CONCLUSIONS AND IMPLICATIONS

This study attempted to investigate students' perceptions of a school-based English proficiency test. It is found that these participants generally perceived the listening and writing parts of the TEPT1 2005 to be easy but the reading part and the overall TEPT1 2005 difficult. The majority of them also tended to feel anxious during the TEPT1 2005 and self-rated their performance worse than they could have done. All these partly attest that the TEPT1 2005 was valid and reliable in testing students' English proficiency. Concerning differences in the measured variables among students in different study years, first-year students were significantly less anxious than their third-year counterparts during the TEPT1 2005, achieved significantly higher in the written and oral tests and the overall TEPT1 2005, and meanwhile considered the reading part of the TEPT1 2005 the most challenging. As to correlations among the measured variables and students' achievement in the TEPT1 2005, respondents who perceived a certain part of the test difficult tended to score lower in that part, consider other parts challenging and felt anxious during the test. Most of these variables again had a significantly negative correlation with students' achievement in the written and oral tests of the TEPT1 2005 and the overall TEPT1 2005.

As described above, the TEPT1 2005 served as a proficiency and exit test, we thus had expected students in a higher study year should perceive the test easier and perform in it better. Contrary to our expectation, a reversed pattern was found in the present study. The reason might be that in the first year most of the courses students took were basic and general knowledge-based courses which required less time and efforts and thus they probably had more time to learn English. With the study load for content-based courses becoming heavier and heavier, students would spend less time and efforts in learning English, which might lead to lower grades as reflected in their test scores. There is a saying among students of Tsinghua University that "the higher your year of study, the lower your scores will be in school-based English tests, namely, TEPT1". This might also suggest that the teaching of English

at the University should be reconsidered. The overall syllabus and content of English teaching need to be re-investigated and re-designed so that students can really achieve more in English as their time progresses at the University.

A similar case might be found with other self-designed school-based or even nationwide English proficiency exams in China or even other Asian EFL contexts such as CET bands 4 and 6. It is often said that undergraduate non-English majors who are approaching graduation tend to have more difficulty passing CET band 4. This justifies continuous research on students' perspective on a proficiency and exit test to enhance its reliability and validity on one hand, and to shed light on adjusting and improving the teaching and learning of English on the other hand. It might be beneficial to specify in College English Syllabus that universities offer ESP courses throughout university years so that non-English majors can maintain continuous learning of English, which may result in not only continuous interest in the target language but a higher proficiency in the language (Liu, 2007). This, coupled with efforts made by curriculum- and test-designers and language instructors, could possibly make proficiency and exit tests really serve their purposes.

In addition, though as students at the top university in China, these participants still reported to feel anxious during the school-based proficiency test. Students from other universities in China might feel even more anxious during an important English exam. Thus, the issue of anxiety also deserves exploration. For example, causes for feeling anxious and strategies to tackle anxiety need to be researched to enhance students' performance in English exams.

THE AUTHOR

Dr. Wenxia Zhang is professor of English at the Department of Foreign Languages, Tsinghua University, China. Her major research interests include EFL teaching and learning, language testing and EFL writing.

Email: wxzhang@mail.tsinghua.edu.cn

Dr. Meihua Liu is lecturer of English at the Department of Foreign Languages, Tsinghua University, China. Her research interests mainly include second/foreign language teaching and learning, classroom research, and second language writing.
Email: ellenlmh@yahoo.com

REFERENCES

Bachman, L. E. (1999a). *Fundamental considerations in language testing*. Beijing: Shanghai Foreign Language Education Press.

Bachman, L. E. (1999b). *Language testing in practice*. Beijing: Shanghai Foreign Language Education Press.

Black, T. R. (1999). *Doing quantitative research in the social sciences*. London: Sage Publications.

Chlebek, A., & Coltrinari, H. (1977). Summer immersion programs abroad. *Canadian Modern Language Review, 33,* 348-53.

Covington, M. V., & Omelich, C. L. (1987). "I knew it cold before the exam": A test of the anxiety-blockage hypothesis. *Journal of Educational Psychology, 79*(4), 393-400.

Heaton, J. B. (2000). *Writing English language tests*. Beijing: Foreign Language Teaching and Research Press.

Henk, W. P., & Holmes, B. C. (1988). Effects of content-related attitude on the comprehension and retention of expository text. *Reading Psychology, 9,* 203-225.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills: Sage Publications.

Li, D. F. (1998). "It's always more difficult than you plan and imagine": Teachers' perceived difficulties in introducing the Communicative Approach in South Korea. *TESOL Quarterly, 32(4)*, 677-703.

Liu, M. (2007). Chinese students' motivation to learn English at the tertiary level. *Asian EFL Journal*, *9(1)*, 126-146.

Mackenzie, A. M. (1994). Examination preparation, anxiety and examination performance in a group of adult students. *International Journal of Lifelong*

Education, 13(5), 373-388.

Malabongna, V., Carpenter, H., & Kenyon, D. (2002). Computer assisted rating: Reliability, efficiency and perceptions on the COPT. Paper presented at *the 24ᵗʰ Language Testing Research Colloquium*, Hong Kong.

Moss, P. A. (1995). Themes and variations in validity theory. *Educational Measurement: Issues and Practices,* 14, 5-13.

Park, Y. Y., & Oxford, R. (1998). Changing roles for teachers in the English Village Course in Korea. *System, 26*(1), 107-113.

Pang, S. M., & Liu, M. (2006). Chinese students' perspectives on a school-based English program. *Indonesian Journal of English Language Teaching, 2*(1), 54-71.

Qi, L. (2004). Washback of NMET. *Foreign Language Teaching and Research, 36*(5), 357-363.

Qi, L. (2005). Stakeholders' conflicting aims undermine the washback function of a high-stakes test. *Language Testing, 22*(2), 142-173.

Slomp, D. H., & Fuite, J. (2005). Following phaedrus: alternate choices in surmounting the reliability/validity dilemma. *Assessing Writing, 9*(3), 190-207.

Spratt, M. (1999). How good are we at knowing what learners like? *System, 27*(2), 141-155.

Yang, H., & Weir, C. (1998). *Validation study of the national College English test.* Shanghai: Shanghai Foreign Language Education Press.

Young, D. (1991). Creating a low anxiety classroom environment: What does language anxiety research suggest? *The Modern Language Journal, 75*(4), 426-438.

Appendix: Survey on students' perceptions of the TEPT1 2005

Direction: This questionnaire is designed to seek your perceptions of TEPT1. Your answers will be kept confidential and <u>*WON'T*</u> be used to evaluate you or your teachers.

Name ＿＿＿＿ Sex ＿＿＿ Department ＿＿＿＿＿ Age ＿＿＿ Year of study ＿＿＿＿

	M	SD
Listening compression		
1. I had difficulty understanding what I heard because there were too many unfamiliar words.	2.73	1.16
2. I had difficulty understanding what I heard because of complicated grammar.	2.37	1.01
3. I had difficulty understanding what I heard because the speaker(s) spoke too fast.	2.71	1.13
4. I had difficulty understanding what I heard because I couldn't hear it clearly.	2.86	1.19
5. I had enough time to answer questions when doing listening comprehension.	2.54	1.29
6. I could understand what I heard but just didn't know the answers.	2.90	1.12
9. I had difficulty understanding what I heard because I was not familiar with the topics.	2.83	1.07
10. Listening comprehension was difficult because I had done little practice.	3.27	1.16
11. Listening comprehension was easy.	3.26	1.10
12. The voice quality of the listening test was very good.	2.85	1.25
Reading Comprehension		
7. I had difficulty understanding the reading passages because there were too many complicated sentences.	2.87	1.06
8. I had difficulty understanding the reading passages because there were too many strange words.	3.06	1.14

13. I understood every word of the reading passages but
    just didn't know what they talked about.   2.37   1.04
14. I had no difficulty understanding the main ideas of the
    reading passages but had difficulty deducing the authors'
    opinions or intentions.   3.08   1.06
15. Short-answer questions were more difficult than
    multiple-choice questions.   3.41   1.20
16. I was not confidant completing the sentences when
    doing reading comprehension.   3.63 1.02
17. Completing the sentences was difficult because correct
    grammar was required.   3.29   1.04
18. I had difficulty understanding the reading passages
    because I was not familiar with the topics.   2.95   1.03
19. Reading comprehension was easy.   3.59   1.05
20. Reading comprehension was difficult for me because
    I had practiced little before.   3.25   1.03
21. The cloze test was easy.   3.29   .93
22. I had difficulty doing the cloze test because I didn't
    understand the grammar.   2.97   .93
23. There were lots of unfamiliar words in the cloze test.   3.10   .98
24. Few clues could be found to help do the cloze test.   2.96   .97
25. The translation test was easy.   3.41   1.13
26. The translation test was difficult because it involved
    lots of strange words.   3.43   1.10
27. The translation test was difficult because it involved
    complicated grammar.   2.99   1.14

Writing Tasks

28. Writing task 1 was difficult because I had never
    practiced it before.   2.30   1.01
29. Writing task 1 was easy.   2.39   1.05
30. Writing task 2 was difficult because it required more words. 2.80   1.23

31. Writing task 1 was difficult because little information
    was given.   2.48 1.03
34. The writing tasks were easy because I had a large
    vocabulary.   3.38 1.01
35. The writing tasks were easy because I was good at writing.   3.26 1.06

Oral Test

41. My partner and I were at different spoken English
    proficiency levels, which affected my performance in the
    oral test.   2.39 .99
42. The part "teacher-student communication' was easy for me.   2.51 .83
44. The card given to me during the oral test was hard to
    understand.   2.27 1.02
45. I couldn't understand my partner during the oral test.   2.13 .89
46. I couldn't understand what I was required to do during
    the oral test.   2.42 1.04
47. My partner was too nervous during the oral test, which
    negatively affected me as well.   2.20 .92
48. The part "filling in the gap according to the card" was
    difficult.   2.74 1.02
49. The topics for the oral test were difficult.   2.67 1.05
50. The part "student-student conversation" was easy because
    I had practiced it a lot before.   3.14 1.02
51. My partner and I didn't cooperate well because we didn't
    know each other.   2.74 1.05
54. I was not familiar with the topics for the oral test.   3.02 1.05

Test Teacher(s)

37. The teacher(s) spoke English clearly during the oral test.   4.16 .83
39. The teacher(s) was/were nice and friendly during the
    oral test.   4.11 .78
40. The teacher(s) gave clear instructions during the oral test.   3.88 1.03

43. The teacher(s) spoke English at an appropriate speed.   3.75 .89

Assessment of TEPT1

33. The written test on the whole was easy.   3.50 1.11
36. The oral test on the whole was difficult.   3.01 1.08
55. The oral test on the whole was easy.   2.79 1.04
56. The written test was difficult.   3.84 .97

Self-assessed performance in TEPT1

52. I didn't perform as well as I could during the oral test.   3.31 .98
53. On the whole, I performed as well as I could during
    the written test.   3.21 1.02

Anxiety During TEPT1

32. I felt confident when doing writing during the test.   2.72 1.09
38. I was nervous during the oral test.   3.46 1.10
57. I became anxious before the test started.   3.16 1.14
58. I got nervous when I couldn't understand what I heard.   3.29 1.07
59. I got nervous when I couldn't understand what I read.   3.27 1.10
60. I became nervous when I didn't know what the answer was.   3.49 1.07