ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

Examining test fairness from the perspective of administration—A case of VETS

Weilie Lu

Guangzhou College of Technology and Business, Guangzhou, China
891285101@qq.com

Abstract: *This study examines test fairness from the perspective of test administration, taking the Vocational English Test System (VETS) as an example. After analyzing questionnaire completed by 2046 VEST takers and 40 test takers' interviews, it was found that VETS was perceived to be well administered. To be specific, test takers were satisfied with the administration in their testing sites, that they strongly believed in the the same condition across different test sites, and that they strongly believed in absence of cheating during test administration. However, some problems raised by a small number of test takers are also worth noticing, including unsteady equipment, seat arrangements in test rooms across different test sites, and the potential cheating in the speaking section of VETS. On the whole, the current research showed that examining test fairness from the perspective of administration is a necessity. No matter how good the quality of test items are, without good administration, the fairness of the test is undermined.*

Keywords: *Test Fairness, Administration, VETS*

1. Introduction

Fairness is arguably the most critical in test evaluation [9]. It is considered to be a top priority in virtually all context [8]. Test fairness investigation can be conducted from different perspectives. In Kunnan's (2004) Test Fairness Framework (TFF), five fairness qualities were listed as targets of investigation, including validity, absence of bias, access, administration and social consequences. Each fairness quality has its own focuses, which provide researchers with clear directions in test fairness examinations. In this research we focus on test administration.

Administration is the fourth fairness quality in Kunnan's (2004) TFF, which is further divided into physical conditions and uniformity or consistency. The former refers to appropriate conditions for test administration, such as optimum light and temperature levels and facilities which are considered relevant for administering tests while the latter refers to uniformity in test administration exactly as required so that there is uniformity and consistency across test sites and in equivalent forms, and that test manuals or instructions specify such requirements. Kunnan (2004) further explained that uniformity referred to length, materials and any other conditions (for example, planning time or the absence of planning time for oral and written responses) so that test takers (except those receiving accommodations due to disability) receive the test under the same conditions. In addition, test security is also relevant to this quality because the uniformity of a test is dependent on it being administered in secure conditions.

To ensure successful test administration, test holders or administrators will make great efforts to help everything go smoothly, however, whether the test is well administered or not, test takers have the right to judge, considering their full and active participation in the whole process of test-taking. Thus, in this research, we are going to explore test takers' perception on the administration of Vocational English Test System (VETS), which is administered to vocational students in different provinces in China.

2. Related literature

2.1 Vocational English Test System (VETS)

Vocational English Test System (VETS), consisting of three levels: elementary level, intermediate level, and advanced level, is developed and administered by Beijing Waiyan Online Digital Technology Co., Ltd (Beiwai Online afterwards), with students from vocational schools or colleges as its targeted test takers. VETS measures language users' English communication abilities in specific posts in the

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

workplace. It is outcome-oriented, task-driven and scene-based, reflecting the employers' need for English communication skills in different professional posts. VETS covers four language skills of listening, speaking, reading and writing and the tasks are selected from the following five work fields: 1) administrative support, 2) technical operation and product development, 3) customer service, 4) business promotion and 5) global trade. In each level, there are altogether 100 marks, and those who get 60 or more marks pass the test, and therefore will be awarded a certificate. At the same time, the learning outcomes are recognized by the National Credit Bank for Vocational Education. The test is administered via computer and test takers from different provinces across the country take the test on the same day. Considering the features of VETS, two aspects have to be guaranteed to ensure the fairness of the test: first, it should be guaranteed that the test-related equipment functions well on the day of test administration. If there are some problems with equipment like computer or headphones, then those test takers will be at an unfavorable position. Second, it should be guaranteed that uniformity of administration is achieved across different test sites, otherwise unfairness occurs due to different ways of administration.

2.2 Previous studies on administration

In any test, test administration is the most public and visible aspect [6]. Empirical studies generally agree on the importance of equality in standardized administration to ensure test fairness [1,10], which makes it a necessity to conduct research from this perspective. In literature, a number of researchers found potential problems with test administration, thus gives test developers directions as to how to improve test practice.

In Jang's (2002) study, it was found that test takers were concerned about construct-irrelevant factors such as test wiseness, guessing problems or cheating issues[7]. Some test takers complained that cheating downgraded the degree of test fairness. In Fox & Cheng's (2007) study, test takers were not allowed to use their dictionary during the test administration, which was not consistent with what students did in their actual writing task, and which might, according to some test takers' accounts, create an unfair disadvantage for them, so many of the L2 students considered such a rule to be unfair[5]. Other issues of fairness in test administration included different atmosphere (in some schools the atmosphere was relaxed and supportive while in others there was tension and stress) and approaches (some schools took a casual approach to breaks and exchange of information while others did not) in different schools on the day of the test, indicating lack of uniformity across different testing sites. During test administration, it is important for raters/examiners in the speaking section to be consistent. In Cheng & DeLuca's (2011) study, during the administration of interactional performance test like speaking test in live interview format, the kind of comments or gesture from the rater/examiner could influence how candidates felt (like feeling anxious or calm). It was also found that the nature of computer-based testing of large-scale tests was considered by test takers to be one type of systematic bias, which limited opportunity to revisit a test item. In such a case, computer-based testing was seen to bias test takers whose typical test taking strategy was to revisit difficult questions at the end of the test. What's more, computer-based tests were considered disadvantageous by test takers with low computer proficiency, or by those who had not prepared for the test using a computer [3]. Test takers also raised random biases concerns in test administration like distracting environmental conditions, poor quality of test format (e.g., low volume on tape recorder), and inconsistent protocols, which was consistent with Fan & Ji's (2014) findings. Fan & Ji (2014) took the Fudan English Test (FET), a university-based English proficiency test as their research target. The purpose was to investigate test candidates' attitudes to the test and at the same time explored the relationship between test candidates' attitudes, test taker characteristics and test performance[4]. Using questionnaire and interview as investigation tools, the researchers found that test candidates were most positive in their attitude to test administration and least positive in their attitudes to the mode of the computer-based speaking test. However, such rather negative attitudes to the speaking component were found not to be about the mode of test delivery itself, but due to their lack of familiarity with the test format and the noisy test environment, owing to which candidates were not able to perform to their best. According to the participants' perception in Song's (2014) study, many policies, guidelines and practices were used to ensure standardization, warrant test security and combat cheating[11]. In addition, various technologies were used to ensure test standardization. However, test takers still believed that irregularities and dishonesty might occur. Some test takers complained about the tight proximity of seats and distractive testing conditions, some mentioned the irresponsibility of the proctor like falling asleep or being unable to prevent test takers peeking at another test takers' test answers.

In the above studies, most focused on investigating uniformity or consistency [3, 4,5,7,11], especially concerned about test security like cheating, and some others focused on investigating physical conditions

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

[4,11], like whether there's noise or not in the test site/room. These studies show us that, during test administration, a lot of aspects (e.g. invigilation, environment, seat arrangement, procedure standardization) might pose threats to the validity and fairness of the test. If the test is not administered well, test takers' performance will be adversely affected. Thus, comparing test takers based on test scores is not meaningful at all.

3. Research design

3.1 Participants

The current research includes two kinds of participants: questionnaire participants and interview participants. Altogether 2046 test takers participated in the answering the questionnaire. They were students from 66 vocational colleges/schools in 17 provinces, ranging from 16 to 22 in age, and the ratio of male to female is 1 to 5. Altogether 40 VEST takers participated in the one-to-one or one-to-two interviews, they were selected from 4 vocational colleges/schools in 3 provinces, ranging from 19-22 in age, five were male and thirty-four female. At the time when test takers answered the questionnaire or attended the interview, they had just finished taking VETS.

3.2 Methods

In the questionnaire for a previous large-scale test investigation, there were four items that were pertinent to the current study (Table 1).

Table 1: Questionnaire Items

Number	Items	Examinee judgement
1	VETS was well organized and administered in my testing site.	1 2 3 4
2	I believe that students in different testing sites/rooms take the test under the same condition.	1 2 3 4
3	During VETS administration, there was no cheating.	1 2 3 4
4	I can adapt myself well to completing tasks on computer.	1 2 3 4

Note: '1' indicates strongly disagree, '4' indicates strongly agree

In addition to the questionnaires, the current research invited 39 test takers to participate in the one-to-one or one-to-two interviews, aiming to deeper our understanding of VETS administration through test takers' lens. Before the formal interviews began, the current researcher informed the participants of the aim of the research. We also told them that the whole process of interview would be recorded but only used in the research. If they found some questions difficult to answer or unwilling to answer them, they could just choose not to. After that, the formal interview began: first, test takers generally told us their feelings about the administration of VETS on the day they took the test. Then, the researcher asked them each of the four questions in Table 1 one by one. In this way, the result of the questionnaire and that of the interview can be compared and collaborated, furthermore, the researcher can dig out the reasons behind which test takers made their judgments. Before each interview ended, the researcher would ask the interviewees whether they had anything else to add up in case any views should be missed.

3.3 Analysis

To analyze the questionnaire, we used SPSS 19.0 to conduct reliability analysis and descriptive analysis.

To analyze the interview data, we first transcribed the collected data with the help of a recording pen SR302 produced by IFLY TEK. Forty interviewees' responses to the questions produced 11,000 Chinese characters, against which we conducted coding. When presenting results, we translated the relevant parts into English. We adopted the thematic analysis method proposed by Chen (2000), and followed the procedure of four steps: reading→classifying→labeling→ categorizing [2].That is, we go through the following steps: First step: reading. We got ourselves familiar with the original interview data by reading it three times. Second step: classifying. We grouped interviewees' transcripts according to specific questions, and classified their responses into positive category and negative category. Third step: coding/labeling: We carefully examined test takers' explanation for their judgments by underlined them

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

and gave them a name (label). Fourth step: categorizing. We read the coded reasons or explanation and classified them into different categories. To help the work of coding easier and more reliable, the current researcher first read the data several times and began coding. After that, I created a framework for coding access. The other coder also used this framework to code the data by following the above four steps. After we both completed the whole process, we compared the results of our coding, and then held discussion together until agreement was achieved.

4. Results

Administration is the fourth fairness quality in Kunnan's (2004) TFF [9]. It attaches great attention to what really happens while all test takers are taking the test. If the test is not well administered, there might be some occurrences to affect test taker performance, leading to the existence of construct irrelevant factors. And whether the test is well administered or not, test takers, as the test participants, the closest witness of the test and the test site, have the natural advantage in providing information on this aspect. Table 2 presents the result of test takers' perception on VETS administration.

Table 2: Test takers' perception on administration

Administration	Percentage of (dis)agreement		Mean	SD
	disagreement	agreement		
1 VETS was well organized and administered in my testing site.	4.4	95.6	3.72	.513
2 I believe that students in different testing sites/rooms take the test under the same condition.	2.2	97.8	3.67	.587
3 During VETS administration, there was no cheating.	3.8	96.2	3.67	.565
4 I can adapt myself well to completing tasks on computer.	3.6	96.4	3.60	.585

Note: n=2,046

Test takers' perception on administration is highly positive, reflected in the high means for all the question items (see the fourth column 'Mean') and the large percentage of questionnaire respondents showing their agreement with the statements. According to test takers' experience and perception, VETS was well organized and administered on the testing day (mean=3.72). They held strong belief that students in different testing sites/rooms take the test under the same condition (mean = 3.67) and that there was no cheating during VETS administration (mean = 3.67). These test takers thought that they could adapt themselves well to completing tasks via computer. These results from the questionnaire analysis suggest that during test administration test takers can perform to their best and that no test takers were unfairly advantaged (e.g. by cheating). Next, we move on to the interview analysis for deeper understanding.

During the interview, when asked the question 'Are you satisfied with the test administration in your testing site?', all the interviewees responded with positive answers. Then they would further provide their reasons for making that judgment. Or, although some interviewees answered 'yes', they still raised some problems with test administration or gave some advice on how to better administer the test. Of all the 40 interviewees, seven just indicated satisfaction without any further explanations. And the other 33 interviewees, further provided us with their comments on VETS administration. These comments could be classified into two categories: reasons for satisfaction with administration and potential problems with administration, which are summarized in Table 3.

Table 3: Test takers' comments on VETS administration

Category	Details	Frequency
Reasons for satisfaction with administration	Helpful teachers/workers	18
	Strict	17
	Good discipline	15
	Clear instructions on procedure	14
	Good equipment	4
	Quiet	3
Problems with administration	Unsteady equipment	6
	Unclear instructions on procedure	1
	Unfavorable temperature	1

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

From Table 3, we know that there were six reasons for interviewees' satisfaction judgment.The most frequently mentioned reason is that on the testing day, there were helpful teachers or workers. According to interviewees, workers included student volunteers, members from students unions or some technicians. Teachers or workers could provide help to test takers in two aspects. They could give guidance to test takers as to how and where to find their seats or when the computer could not work normally, instant help could be obtained. According to the experience of interviewee 8, in the testing site where she took the test, the teachers would guide the students to the examination room following certain standards. The following interviewee's experience clearly demonstrated the necessity and importance of helpful teachers/workers:

'At that time, there was a problem with a candidate's computer, and the teacher quickly helped him to find a well-functioned computer. What a relief!'

Interviewee 11

The key word of the second most frequently mentioned reason is 'strict'. By saying 'strict', test takers means that there were some rigorous procedures before they entered the test room, like putting unrelated materials to a designated place, checking candidate identification with face recognition device, and probing test takers' body with a metal detector. Everyone took the test seriously. And after they were seated and began doing the test tasks, invigilators strictly followed test administration rules to prevent accidents from happening. When responding to the interview question, some interviewees said that they were satisfied with administration, added by a brief explanation: 'because it is strictly administered' (Interviewee 05), while some other interviewees will give more detailed descriptions to show how they conceptualized 'strict', like the following interviewee:

'When you go upstairs to that floor, there were two rooms: one is for checking test takers' identity and the other is the test room. After test takers have been carefully checked, they were allowed to entered the test room, where there were three invigilators. '

Interviewee 21

The third most frequently mentioned reason for satisfying administration is good discipline, which means that everyone does his own things during the testing period, and no private communication among test takers, and no cheating would occur. Other reasons mentioned included clear instruction on procedure, good equipment and no noise in the testing site/room.

Besides reasons for test takers' satisfaction with administration, Table 3 also presents some problems and the one that is most frequently mentioned is unsteady equipment. It is understandable that test takers attach great attention to equipment like computer because it will decide the result of their performance. When commenting on the equipment in her testing site, interviewee 11 said:

'The equipment is unstable. During the simulation test, there were some cases where the equipment failed to work well, but there were still such cases when it comes to the real test administration. What's more, not only one computer had such a problem. The equipment is really unstable.'

Interviewee 11

Besides, one interviewee mentioned unclear instructions on procedure as a problem. In her opinion, the process arranged by the school was a bit messy and caused inconvenience and trouble to test takers (interviewee 4). The last problem is unfavorable temperature. One interviewee complained that it was too hot in the test room (interviewee 21), which might affect him to perform to his best.

When asked the question 'Do you think that students in different testing sites/rooms take the test under the same condition?', all interviewees responded to this question, 7 of them (17.5%) provided negative answers while a large portion of interviewees---33 of them (82.5%) provided positive answers, which can be clearly seen from Figure 1 below.

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

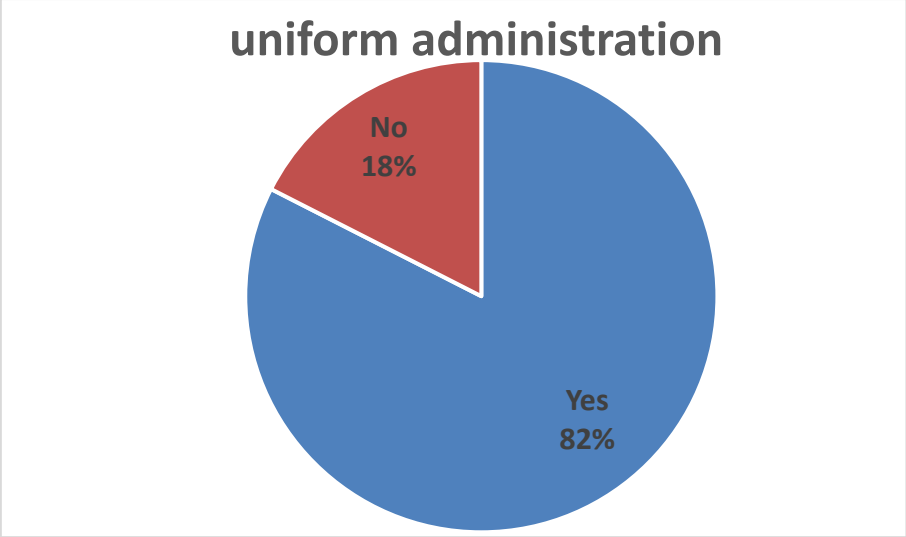

Figure 1: Uniform administration

Among all the 40 interviewees, eighteen just indicated their positive or negative response without any explanation, while the other 22 provided their reasons for their judgment. Further analysis of these test takers' interview data shows us how test takers perceived whether there was consistent administration across different testing sites, which is presented in Table 4.

Table 4: Reasons for judgment on administration

Judgment	Reasons	Frequency
Students in different testing sites/rooms take the test under the **same** condition.	Related to test and testing practice	6
	Related to test administration method	7
	Related to test takers' belief in the authority of large-scale tests	8
Students in different testing sites/rooms take the test under **different** conditions.	Different equipment	3
	Different ways of arranging seats	2
	Different outer environment	2
	Different invigilators	1

According to Table 4, interviewees' reasons for equal test taking condition can be categorized into three types. First, the reasons are related to test and testing practice. When giving this kind of reason, interviewees would mention the same test time period all test takers had (85 minutes for elementary level, 105 minutes for intermediate level, and 130 minutes for advanced level), the same test tasks for test takers to do, and the same procedures all test takers went through. During the interview, interviewee 9 were aware that she and her other classmates did not take the test at the same time, but she knew the the procedures were the same for them, this is why she made the conclusion that students in different testing sites/rooms took the test under the same condition. The second type of reason is related to the test administration method, referring to the fact that VETS is administered via computer and that all test tasks have to be completed on computer. In interviewees' opinion, the equipment is the same, and all the test takers finished the test on computer, so the condition must be the same for all test takers. The following interviewee's statement can best illustrate this point:

> *'I think all test takers undergo the same condition. You know, to do VETS, you have to use computer. And now, basically speaking, all schools are equipped with computers in computer rooms and test takers take the test there. So I think they are under the same condition.'*
>
> *Interviewee 5*

The most frequently mentioned reason is related to test takers' belief in the authority of large-scale tests. Test takers have great faith in large-scale tests like VETS. In their mind, as long as it is a large-scale test, it is fair, therefore, the condition for every test taker must be the same. Interviewee 15 said: "Although I have never observed how students in other colleges take the test, I just believe that this kind of large-scale test organized by colleges across the country would follow the same rules and regulations", and interviewee 16 showed similar opinion. Interviewee 24 responded to the question with a rhetorical question. "Aren't tests fair nowadays? Everything is the same."

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

Some other interviewees showed different opinions: they thought that students in different testing sites/rooms took the test under different conditions. Although the reasons were not as frequently mentioned as those listed above, they were worth taking notice of, since they provided potential directions for how to improve the test and test practice. Among the four reasons, the most frequently mentioned one is about test equipment. In these interviewees' opinion, although all candidates took the test via computer, the equipment in different areas might not be the same, which led to different conditions. The following two interviewees' statement can help us better understand this reason.

'...might be different. For example, the equipment in this test room is much better, and the Internet service is also better. However, the Internet service in the next room is not as good, and the equipment is terrible.'

Interviewee 13

'I think test takers sit the test under different conditions, which is related to the equipment. The equipment for each student is different. If you are lucky, you can have a well-functioned computer, if you are unlucky....'

Interviewee 14

According interviewees 13 and 14, even in the same test site, the equipment could be different. Therefore, interviewee 14 considered it to be kind of luck if a test taker is assigned a good computer. From these explanations, it can be inferred that administering a test via computer would bring greater stress to test takers than administering it via paper format. The other reasons for interviewees' negative judgment included different ways of arranging seats, different outer environment, and different invigilators across different test rooms.

When asked, in their opinion, 'whether there was cheating during VETS administration?', seven of them (17.5%) responded with 'yes' and 33 of them (82.5%) told us that there would not be cheating while candidates were taking the test, which is clearly presented in Figure 2.

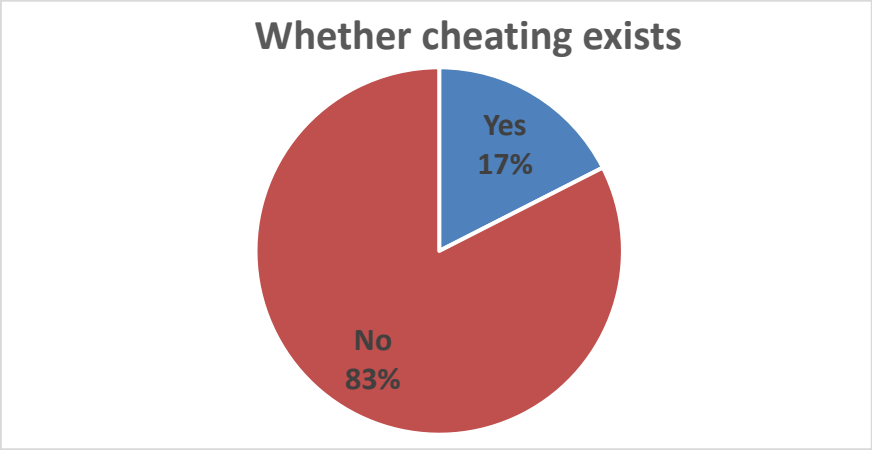

Figure 2: Whether cheating exists

In addition to indicating 'yes' or 'no' to the interview question, 33 of the interviewees further provided reasons for their judgments on whether there's cheating in VETS, which are classified and summarized in Table 5.

Table 5: Reasons for judgment on cheating

Judgment	Reasons	Frequency
absence of cheating	Strict examination regulations	14
	Computer-based test	9
	Belief in test takers' conduct	8
	Seat arrangement	7
	Low difficulty level of the test	3
existence of cheating	Seat arrangement	5
	Computer-based speaking test	2

From Table 5, we know that there are five types of reasons given by interviewees for their judgment about why there didn't exist any cheating while candidates were taking VETS. The most frequently

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

mentioned reason is that there were strict examination regulations. About the strict regulations, interviewees mentioned: 1) Before entering the examination room, they had to turn in their mobile phones and go through rigorous procedures like facial recognition check; 2) in the test rooms, the invigilators were strict in that they would walk around the examination room; and 3) there were surveillance cameras installed in the examination room. The second most frequently mentioned reason is about the test taking method, that is, VETS is administered via computer. Some interviewees thought that because test takers took the test via computer, and during the test, they could not exit from the test page, which meant that they didn't have any opportunity to search the Internet for answers (Interviewees 11 and 26). According to some interviewees' experience and understanding, administering the test via computer makes it possible to provide test takers with random test tasks, therefore, even some candidates wanted to peek into others' answers, they could not succeed. And because the speed and pace of doing tasks were different from person to person, the tasks shown on the screen would be different, which effectively prevented test takers from copying others' answers. The third most frequently mentioned reason is interviewees' belief in test takers' good conduct. Interviewee 24 stated that college students are adults, they would not do anything against the rule like cheating. And interviewee 37 held similar opinion, in her opinion, students taking part in the test were all aware of the serious consequences of cheating, and would not get involved in that. The fourth reason mentioned is the seat arrangement in the examination room. When talking about the seat arrangement, interviewees said that test takers were seated far from each other and there was a partition between two students who sat next to each other (interviewees 3, 18, 21 and 30). Such a way of arrangement prevented test takers from peeking into others' answer, because if they did so, their behavior would be so obvious that the invigilator would notice him/her immediately (interviewee 4). The least frequently mentioned reason for no cheating in VETS is the low difficulty of the test. Interviewee 2 frankly told us that it was really unnecessary to cheat in VETS because it was so easy for them (English majors) to pass the test, and this opinion was shared by interviewees 36 and 39.

Table 5 shows us that there are only two reasons for interviewees' judgment that there existed cheating in VETS. The first one is about the seat arrangement. In the discussion above, interviewees mentioned this aspect as reasons for no cheating. At that time they just mentioned the partition between neighbouring students. When students considered the seat arrangement as the reason for cheating, they had two focuses. Some interviewees said that there was no partitions set between neighbouring students in their test room, so some test takers could easily see the answers of the students who sat next to him/her. Both interviewees 7 and 8 were concerned about this problem. Interviewee 8 said:

'As long as you have good eyesight, you can easily see the answers of the candidates in front of you., or the one on your left or on your right'
Interviewee 8

These interviewees' responses indicated the inconsistency of test administration across different test sites, which is worth further investigating. Some interviewees said that in their test sites, there were two types of seat arrangements. Students were either seated in rows or in a circle. And according to interviewee 15, if students were seated in a circle, it was easier for them to see another person's answers. The other reason for potential cheating in VETS is the speaking tasks. These interviewees pointed out the disadvantage of many test takers in a test room doing speaking tasks together almost at the same time. It was loud and if some test takers didn't know the answer, they could easily copy others' answer (interviewees 7,8).

5 Conclusion

In summary, informed by analysis results of the questionnaire data and interview data, we can know that VETS was perceived to be well administered. It revealed that test takers were satisfied with the administration in their testing sites, that they strongly believed in the the same condition across different test sites, and that they strongly believed in absence of cheating during test administration. Some problems raised by a small number of test takers are also worth noticing, including unsteady equipment, seat arrangements in test rooms across different test sites, and the potential cheating in the speaking section of VETS. On the whole, the current research showed that examining test fairness from the perspective of administration is a necessity. No matter how good the quality of test items are, without good administration, the fairness of the test is undermined.

References

[1] Bridgeman, B. & Schmitt, A. (1997). Fairness issues in test development and administration. In W.

ISSN 2522-6398 Vol. 8, Issue 12: 35-43, DOI: 10.25236/FER.2025.081206

Willingham & N.S.Cole (Eds.), Gender and fair assessment (pp.185-225). Mahwah, New Jersey: Lawrence Erlbaum.
[2] Chen, Xiangming. (2000). Qualitative research in social sciences. Education Science Press.
[3] Cheng, L., & DeLuca. (2011). Voices from test-takers: Further evidence fro language assessment validation and use. Educational assessment, 16:2, 104-122.
[4] Fan, J. & Ji, P. (2014). Test candidates' attitudes and their test performance: the case of the Fudan English Test. University of Sydney Papers in TESOL, 9, 1-35.
[5] Fox, J. & Cheng, L. (2007). Did we take the same test? Differing accounts of the Ontario Secondary School Literacy Test by first and second language test-takers. Assessment in Education: Principles, Policy & Practice, 14:1, 9-26.
[6] Haladyna, T.M., & Downing, S.M. (2004). Construct-irrelevant variance in high-stakes testing. Educational Measurement: Issues and and Practice, 23, 17-27.
[7] Jang, E. (2002). Folk fairness in language testing. Paper presented at the Southern California Association for Language Assessment Research conference (SCALAR 5).
[8] Karami, H. (2013). The quest for fairness in language testing. Educational Research and Evaluation. 19 (2-3): 158-169.
[9] Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), European language testing in a global context: Proceedings of the ALTE Barcelona Conference (pp.27-48). Cambridge, UK: Cambridge University Press
[10] McCabe, D.L., Trevino, L.K., & Kenneth, D.B. (2001). Cheating in academic institutions: A decade of research. Ethics & Behavior, 11, 219-232.
[11] Song, X. (2014). Test fairness in a large-scale high-stakes language test. Unpublished doctoral dissertation. Canada: Queen's University.