A Closer Look at Chinese EFL Learners' Test-wiseness Strategies in Reading Test

Miao Haiyan[1,2,*] & Liu Rilong[3]

[1]National Key Research Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, China

[2]School of Foreign Languages, Jiangxi Normal University, Nanchang, China

[3]School of Engineering, Jiangxi Agricultural University, Nanchang China

*Correspondence: National Key Research Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, Guangzhou, 510420, China. Tel: 1-357-613-7091. E-mail: haiyanm@jxnu.edu.cn

Abstract

This paper reports on an investigation into the relationship of test-takers' use of test-wiseness strategies to Chinese EFL learners' reading test performance. A test-wiseness questionnaire was administered immediately after the final achievement test to probe into how learners thought while completing the reading section of the test. It was found out that the use of test-wiseness strategies bore no positive relationship to the reading test performance; and successful test-takers reported test-wiseness strategy use no more significantly different than unsuccessful test-takers. The results suggested that the bias against Asian EFL learners, especially Chinese EFL learners, in their test-taking process was unfair, since test-wiseness strategies did not contribute to these EFL learners' reading performance and the distinction between successful and unsuccessful test-takers.

Keywords: *test-wiseness strategies; successful test-takers; less successful test-takers; reading test performance*

1. Introduction

One of the major concerns of language testing (LT) research is to provide a model of language ability. Its primary goals are not only to describe and assess individual language ability, but also to develop and empirically validate theories of language test performance and the correspondence between test performance and nontest language use. However, language test performance is not easy to interpret, as there is variability involved. In recent years, many researchers have shown great interest in the sources of variability in performance on language test. According to Bachman and Cohen (2002), there are three major sources of variability central to LT, namely, individual differences in the language abilities that are measured, test-taking strategies and variation in the tasks and context on performance on language test tasks.

In fact, these three aspects are closely inter-related, but the attention they received is not the same. Both the nature and specific aspects of language abilities measured have already been extensively treated (Hymes, 1972; Canale & Swain, 1980; Bachman, 1990; Bachman & Palmer, 1996). Study is also well established that individual performance on language test varies as a function of features in the language use context (Cathcart, 1986; Tarone, 1988; Bayley, 1991, 1994; Young, 1991, 1993; Tarone & Liu, 1995). As for test-taking strategies, though very important, there still remains controversy. It is still unclear whether this strategy use depends more on language use (ability) or more on test-wiseness strategies, and how the influence of test-wiseness strategies is on test performance and test-takers' use of test-wiseness strategies in their usual language testing circumstances. Accordingly, research on the relationship of test-wiseness strategies to language test performance is needed.

2. Literature Review

2.1 Importance of Test-Wiseness Strategies

Language learner strategies encompass both language learning strategies and language use strategies. The former refer to strategies that language learners purposefully use to enhance their language learning and acquisition, whereas the latter strategies they purposefully employ to enhance their performance on language tests. Language use strategies also constitute test-taking strategies when they are applied to tasks in language tests (Cohen, 2000). As for LT researchers, they are much more concerned with test-taking strategies rather than language learning strategies, because these are directly connected with test performance variation, thus influencing the interpretation of individual language ability.

Since the late 1970s, researchers have slowly begun to approach second language testing from the perspective of test-taking strategies. By the 1990s, L2 testing textbooks even acknowledged test-taking strategies as a possible source of test reliability and validity (Bachman, 1990; Cohen, 1994). However, for a long time, people just confuse the two concepts of test-taking strategies and test-wiseness strategies (Millman, Bishop & Ebel, 1965; Dreisbach & Keogh, 1982; Wiersma, 1990). It is not until recently that Cohen (2000) drew a distinction between them. According to him, test-taking strategies consist of both language use strategies and test-wiseness strategies. The former refer to strategies that rely more on language knowledge or ability in the process of producing answers to tasks in testing; the latter may rather be dependent on the respondents' knowledge of how to take tests, such as their hunches, beliefs, or attitudes about what the best way to produce results may be.

To succeed in language tests, learners may need not only language knowledge or ability but also test-takers' beliefs and ways in taking the test, that is, test-wiseness strategies (e.g. time management, and selecting an alternative through deductive reasoning). Then how are test-wiseness strategies related to test performance? Do test constructors have to bear this in mind when preparing the test and interpreting the test results? Actually, test-wiseness strategies have already aroused great interest among Western scholars (Sarig, 1987; Buch, 1994; Cohen, 1991), yet in China they have not received their due attention. In recent years, Asian English learners are biased in their language learning experience (Keck 2014), because their high scores in tests were sometimes believed not to reflect their real language proficiency but their test-wiseness strategies more. The recent denial of Chinese EFL learners' IELTS is a good manifestation of this bias.

2.2 Research Questions

Then it is of urgent need to explore the relationship between Chinese EFL learners' test scores and their test-wiseness strategies. Therefore, the present research attempts to investigate this test-wiseness strategy use in Chinese EFL learners' common English tests such as the English reading achievement. There are three research questions:

1) What is the relationship of test-wiseness strategies to EFL reading test performance?

2) How do successful and less successful test-takers differ in their use of test-wiseness strategies?

3) What are the implications then for test constructors and the interpretation of test performance?

3. Methods

3.1 Subjects

The subjects participating in this study were 80 second-year non-English major students randomly chosen from a key university of China. They were between the ages of 17 and 22, and English was their compulsory subject in their university studies. For every week, they had 4 hours of regular English classroom learning. Their regular English teacher taught them English and conducted a series of reading, writing, listening and speaking activities in class. Based on their reading test performance in the final achievement at the end of the year, the 20 top students were chosen as the successful test-takers and 20 students at the bottom the less successful test-takers.

3.2 Instruments

The research instrument employed in this study was the Chinese EFL test-wiseness strategy questionnaire. Based on Cohen (2002), the questionnaire was comprised of 18 items (see the Appendix). In order to make the questionnaire more understandable, it was written in Chinese so that unnecessary ambiguities were avoided. Items describing test-wiseness in this questionnaire were divided into four types: local strategy (items numbered 7, 8, 15, 16), global strategy (items numbered 1, 2, 5, 6, 13), risk-taking strategy (items numbered 11, 12, 14) and working scheme strategy (items numbered 3, 4, 9, 10, 17, 18). The questionnaire used a five-point scale: 1 (never or almost never true

of me), 2 (usually not true of me), 3 (somewhat true of me), 4 (usually true of me), and 5 (always true or almost true of me). Thus the higher the scores were, the greater the strategy use of test-wiseness was used.

To make the study more meaningful for real-life teaching and testing, the final achievement English test was used for the study of test-wiseness strategies. The reason why such a test was chosen was due to the following two reasons. First, learners treated this type of test very seriously for its importance in their college GPA. Second, this test was typical of the tests Chinese EFL learners take in their English learning. The weighting of reading in the examination was 40, much more than any other items. The reading section consisted of 4 passages, with the aim to measure the test-takers' ability to read English texts for main ideas, details, inferences and vocabulary knowledge. The first three passages were in the multiple-choice form, while the last one was a short-answer type. The topics in this section were relevant to the students' topical knowledge, such as family, campus life and current issues.

3.3 Data Collection and Analysis

The questionnaire was administered to learners immediately after they had finished their final English achievement test for the year. Their English teacher handed out the questionnaires to the learners and required them to complete the questionnaire in fifteen minutes. The students were told that the questionnaire was supposed to explore more about the process of how they took the test, and had nothing to do with their class or test evaluation. The reason why the questionnaire followed immediately the final English achievement test was that the learners could have a fresh memory of their test for the short time interval. There were 80 students who were invited to complete questionnaires, but altogether only 75 valid questionnaires were collected. After the results of their final English achievement test came out, the learners were further divided into successful and unsuccessful test-takers. The top 20 learners were taken as the successful test-takers while the bottom 20 learners were considered as the unsuccessful test-takers.

SPSS was used for further statistical analyses. Both average score and standard deviations were calculated and presented. In order to find out the relationship between test-wiseness strategies and Chinese EFL learners' test scores, a correlational analysis of Pearson Product was conducted to confirm the negative, zero or positive correlation therein. After the correlation between test-wiseness strategies and Chinese EFL learners' test scores were determined, independent T-test was run to explore whether there were any significant differences between the two groups in the use of the test-wiseness strategies.

4. Results and Discussion

4.1 Descriptive Statistics

Descriptive statistics of test-wiseness strategies were presented in Table 1. From Table 1, it could be seen that the five highest mean scores were 4.13, 3.61, 3.59, 3.28, 3.17. The questionnaire items were item 2, item 10, item 1, item 18 and item 6. This indicated that Chinese EFL learners used to a greater degree items 2, 10, 1, 18 and 6. The most frequently utilized test-wiseness strategy was to return to the passage to look for or confirm an answer rather than relying solely on memory of what was in the text. The second frequent strategy was to use the process of elimination. The third one was to read the passage and make a mental note of information location. The fourth one was to estimate the time and assign the time. The fifth frequent one was to make an educated guess.

The five lowest mean scores were 1.47, 1.52, 1.81, 2.04, 2.40. The questionnaire items were item 14, item 12, item 8, item 16 and item 3. The mean scores of these strategies reflected their frequency of usage (Oxford 1990). According to Oxford (1990), the mean score below 1.5 implied that this strategy was never or almost never true, while the mean score between 1.5 and 2.4 indicated usually not true. It followed that Chinese EFL learners never or almost never used the strategy numbered item 14 in the questionnaire, namely taking into consideration the position of the option among the choices. The other four strategies that were usually not true were selecting a choice longer or shorter than the others, looking for the portion of the text that the question refers to and searching for clues, selecting the option containing a word or phrase in the question, and reading the questions first as a guide directed at finding answers.

Table 1. Descriptive Statistics of Test-Wiseness Strategies

Item	Mean	SD	Skewness	Kurtosis
1	3.5867	1.25303	-.393	-.844
2	4.1333	.93481	-1.394	2.302
3	2.4000	1.17404	.762	-.132
4	2.9600	1.21299	.032	-.847
5	2.7200	1.31066	.169	-1.182
6	3.1733	.97777	-.181	-.101
7	3.0133	1.00664	.218	-.527
8	1.8133	1.00933	1.036	-.084
9	2.5200	.96366	-.012	-.516
10	3.6133	.82024	-.677	.664
11	2.6133	.98493	.503	.214
12	1.5200	.92063	2.182	4.886
13	2.7333	.97722	.118	-.137
14	1.4667	.85950	1.811	2.282
15	3.0933	.87261	.066	-.576
16	2.0400	.86117	.183	-1.070
17	2.5067	1.21225	.335	-.923
18	3.2800	1.26875	-.345	-.992

4.2 Test-wiseness Strategies and Reading Test Performance

The total score for the reading section was 40 in these Chinese EFL learners' final achievement test. Students' scores ranged from 15 to 40, and the score distribution was shown in Table 2. It could be seen that there were six students scoring 40, and nine students scoring below 25 but above 20. There were twenty-one students who scored between 30 and 40, twenty-four students between 25 and 35 and fifteen students between 20 and 30. Then what is the relationship between the test-takers' use of test-wiseness and their reading test performance? To answer this question, the correlation coefficient between the two variables was calculated.

Table 2. Scores in the Reading Test

Score Range	≤20	≤25	≤30	≤35	≤40
Number	9	15	24	21	6

However, the correlation between test-wiseness strategies and learners' reading performance was not as high as documented in the literature (Bachman 1996; Cohen), and as expected for Asian learners like Chinese EFL learners (Keck 2014; Lee 2013). Among the eighteen test-wiseness strategies, only one test-wiseness strategy was positively correlated with the students' reading test performance. That particular test-wiseness was item 13 in the questionnaire, that is, students tended to take advantage of clues appearing in other items in order to answer the item under consideration. From this, we can see that test-wiseness was not so important to Chinese EFL college students in their reading test, which was contrary to the commonly-held belief that test-wiseness strategies could greatly increase test scores. This could well be a counter-piece for the disapproval of Chinese EFL learners' test scores in internationally-known high-stake tests, since it was often claimed that Chinese EFL learners' high score was attributed to particular test-wiseness strategies instead of their true language abilities (Keck 2014). This claim was ungrounded and unfair.

Generally speaking, even though the purpose of the reading test was to measure students' ability to use the language, yet there was variability involved, such as guessing, individual differences in time management and working schemes. What was worth pointing out was that students' anxiety and tiredness after the reading test might also contribute to this result that the students' reading test performance was not so positively correlated to their reading test scores.It was undeniable that some variability might be at work in language testing process, but their role was not significantly prominent in this process of language testing. From the above results, it could be concluded that test-wiseness strategies were not so important in college students' reading test performance, and what students employed in the reading test was mainly their ability to use the language. Hence, Chinese EFL learners high scores should better not be exaggerated in such a way as to override their other language use abilities in the testing

procedure.

4.3 Differences between Successful and Unsuccessful Test-takers

To have a clear view of students' use of test-wiseness, the strategies used by successful and unsuccessful test-takers were compared and contrasted through independent-sample T-test. Subjects were grouped into successful and unsuccessful test-takers on the basis of their scores in the reading test. The first 20 top students were taken as the successful test-takers while the bottom 20 students were grouped as unsuccessful test-takers. Table 3 offered a detailed picture of the overall test-wiseness strategy use by the two different groups of top-scoring and bottom-scoring groups.

Table 3. Different Test-Wiseness Adopted by Successful and Less Successful Test-Takers

Strategy number	Successful learners		Unsuccessful learners		T-test		
	Average	S.D.	Average	S. D.	T value	F value	Sig.
1	3.8500	.8751	3.3000	1.4179	1.476	5.395	.026
2	4.2000	.7678	4.2500	.7164	-.213	.081	.777
3	2.3500	1.2258	2.2500	.8507	.300	1.733	.196
4	3.1500	1.3870	3.0000	1.2566	.358	.533	.470
5	2.7500	1.2927	2.6000	1.3917	.353	.510	.480
6	3.1500	1.0400	3.5000	.8272	-1.178	.309	.582
7	3.2000	.8944	3.1000	1.1653	.304	3.646	.064
8	1.8000	.8944	1.7000	1.0311	.328	.501	.483
9	2.4500	.8256	2.5500	1.1459	-.317	2.302	.138
10	3.8500	.6708	3.6500	.7452	.892	1.633	.209
11	2.4500	.8256	3.0000	1.0260	-1.868	.336	.566
12	1.4000	.6806	1.9000	1.0712	-1.762	.697	.409
13	3.1000	.9679	2.5500	1.1459	1.640	2.212	.145
14	1.4000	.7539	1.7000	.9787	-1.086	1.105	.300
15	3.0500	.8870	3.2500	.9105	-.704	.431	.515
16	2.1000	.9679	2.0000	.8584	.346	.640	.429
17	2.5500	1.3563	2.4000	1.3917	.345	.006	.941
18	3.8000	1.0052	3.3000	1.3803	1.310	4.645	.038

The results showed that successful test-takers reported higher mean scores than less successful test-takers in 61% of all of the test-wiseness strategies. In other words, successful test-takers tended to use more test-wiseness in their reading test. However, none of the above T-scores was above the value of 2. The highest absolute T-value was 1.868 for item 11 in the questionnaire that "I will look for an option that seems to deviate from the others, is special, is different, or conspicuous." All this indicated that the differences between successful and unsuccessful test-takers were not as different as enough to reach any statistical significance.

Based on this finding, it could be safely stated that although there did exist differences in the use of test-wiseness strategies between successful and unsuccessful test-takers, the difference was not so significant as to greatly influence their test scores. The scores of reading tests could then not be able to be greatly influenced or easily manipulated by test-wiseness strategies used consciously or unconsciously by learners in their test-taking process. To put it another way, what the reading tests measure was a true reflection of what the learners know about the language and what they could do with the language with their EFL reading comprehension system. There might be differences in the test-wiseness strategies used by successful and unsuccessful test-takers, but these variations did not help in turning a low-achieving test-taker into a high-achieving test-taker. The decisive factor that accounted for the divergence in test scores was the gap of language ability between successful and unsuccessful test-takers.

5. Implications for Test Constructors

The above results demonstrated that the test-wiseness strategies had no positive relation with Chinese EFL learners' reading performance, and successful and less successful test-takers did not show statistically difference in their use of test-wiseness strategies. The findings can well serve as evidence against the bias towards the high scores of Chinese EFL learners. These results have great pedagogical implications to teaching of EFL reading and its testing.

First and foremost, the design of the test itself can help to avoid bias against the use of test-wiseness strategies to a

great extent. Generally speaking, large-scale language tests employ multiple choices, which are known for the high reliability and less demands on human resources (Hughes 1989). However, multiple choices can bring about unpredictability for the test-wiseness strategies, as learners adept with such strategies can get correct answers too. To prevent such cases, other types of tasks can be used in reading tests, for example, summarization tasks, open-ended questions and compositions.

Second, learners' reading performance and ability should be carefully evaluated in terms of learners test scores. Even though there are test-wiseness strategies like guessing, this type of test-wiseness strategies is rather limited and is not significant above chance. Consequently, they can not serve as evidence for the bias against high scores by successful test-takers in certain countries. Undeniably, test scores on reading should not be taken as the sole yardstick of reading ability, and teacher's evaluation and the standard error of measurement should also be included as evaluation criterions.

Last but not least, although test-wiseness strategies does not relate closely to the test scores, the training of test-wiseness strategies can be integrated into daily language teaching. The reason is not only that reading is an important part in language tests, but also that reading plays a significant role in daily learning and working experiences. Some test-wiseness strategies are good for the formation of good reading habits, for instance, time assignment, order of answering, making educated guesses and reading with purposes. There are three ways to improve test-wiseness usage, namely awareness training, one-time strategy training and long-term strategy training (Oxford 1990). For classroom EFL teaching, awareness training is preferred because this can be well integrated into classroom teaching to fit more seamlessly into daily explicit teaching of reading.

References

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. F., & Cohen, A. D. (2002). *Interfaces between Second Language Acquisition and Language Testing Research*. Beijing: Foreign Language Teaching and Research Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*.

Cohen, A. D. (1994). *Assessing Language Ability in the Classroom*. Boston: Newbury House.

Cohen, A. D. (2000). *Strategies in Learning and Using a Second Language*. Beijing: Foreign Language Teaching and Research Press.

Dreisbach, M., & Keogh, B. (1982). Testwiseness as a Factor in Readiness Test Performance of Young Mexican—American Children. *Journal of Educational Psychology*, *2*.

Hymes, D. (1972). On Communicative Competence. In J. B. Pride., & J. Holmes(eds.), *Sociolinguistics*. Harmondsworth: Penguin.

Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing*, *25*, 4-22.

Lee, H. S., & Winke P. (2012). The differences among three-, four-, and five-option-item formats in the context of a high-stakes English-language listening test. *Language Testing*, *30*(1), 99-123. http://dx.doi.org/10.1177/0265532212451235

Lee, J. Y. (2013). Language learner strategy by Chinese-speaking EFL readers when comprehending familiar and unfamiliar texts. *Reading in a Foreign Language*, *27*(1), 71-95.

Yang, H. C., & Plakans L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly, 46*(1), 80-24. http://dx.doi.org/10.1002/tesq.6

Appendix: Questionnaire

Directions: Here is a questionnaire of the test-wiseness strategies that you have probably used in your English test. Please answer each of the items according to what you're actually doing when you take English test, but not what you think or what others will do. Choose 1 (Never or almost never true of me), 2 (Usually not true of me), 3 (Somewhat true of me), 4 (Usually true of me) and 5 (Always true or almost always true of me).

1. I will read the passage first and make a mental note of where different kinds of information are located.

2. I will return to the passage to look for or confirm an answer rather than relying solely on memory of what was in the text.

3. I will read the questions first so that the reading of the text is directed at finding answers to those questions.

4. I will read the questions a second time to make sure that their meaning is clear.

5. I will try to produce my own answer to the question before I look at the options that are provided in the test.

6. I will make an educated guess, for example, use background knowledge or extra-textual knowledge in making the guess.

7. I am ready to change the responses to any given item as appropriate, for example, in the case where new clues are discovered in, say, another item.

8. I will look for the portion of the text that the question refers to and then look for clues to the answer there.

9. I will look for answers to questions in chronological order in the text.

10. I will use the process of elimination--i.e. select a choice not because I am sure that it is the correct answer, but because the other choices don't seem reasonable, because they seem similar or overlapping, or because their meaning is not clear to me.

11. I will look for an option that seems to deviate from the others, is special, is different, or conspicuous.

12. I will select a choice that is longer or shorter than the others.

13. I will take advantage of clues appearing in other items in order to answer the item under consideration.

14. I will take into consideration the position of the option among the choices (a, b, c, or d).

15. I will select the option because it appears to have a word or phrase from the passage in it——possibly a key word.

16. I will select the option because it has a word or phrase that also appears in the question.

17. I will postpone dealing with an item or selecting a given option until later.

18. I will estimate the time needed for completing the items and don't spend too much time on any given item.