The Review of the Article `Test-taking Strategies for a High-stakes Writing Test: An Exploratory Study of 12 Chinese EFL Learners` Published in Assessing Writing Journal, (2012), 17, 174–190

ROZANA SHAMSABADI
Ph.D. candidate of TEFL, University of Isfahan, Isfahan, Iran

ABSTRACT

This review, grew from the original article by Yun Xu and Zunmin Wu, aims to report on a qualitative research study into the test-taking strategies employed in completing two picture prompt writing tasks—Situational Writing and Interpretational Writing in the Beijing Matriculation English Test. The research reveals a variety of test-taking strategies involved in completion of the two tasks. Based on the strategies employed, the results proved the presence of similarities in having the same starts according to the first picture, having parallel narrations, applying equal effort to each picture, not missing essential information of the pictures, and adding an ending to their narration in all participants’ writings. Students strived to guess test-developers’ intent and figure out the “best” theme instead of freely expressing their own interpretations. What the results of this study reveal is that they might fail to achieve the crucial goal of measuring creative thinking. The researcher found that the original intention of Interpretational Writing had been badly distorted. The present review of this article has attempted to explore the extent to which this paper fulfils the stated aims by having a cautious look through the papers’ methodology in general and participants’ choices in details.

Keywords: EFL learners, Test-taking Strategies, High-stakes Writing Test
1. Introduction

There are ample empirical researches confirming the positive interaction between test performance and use of certain test-taking strategies (e.g. Rupp et al, 2006). Some researchers appreciated substantial contribution of test-taking strategies to the process of construct validation of language tests (e.g. Bachman, 1990; Cohen, 1998). Cohen believed that deficiency of test takers in using appropriate test taking strategies depreciates the validity of language tests and attenuate the test results. Cohen & Upton’s (2007) classification defines test-taking strategies as “those test-taking processes which respondents have selected and which they are conscious of, at least to some degree”. Several studies have been conducted to explore test-taking strategies in second/foreign language tests. As an example, Pour-Mohammadi (2010) reviewed studies on test-taking strategies in second/foreign language reading comprehension tests focusing on history of research on test-taking strategies, discussion of its development, the significant role of schemata in reading comprehension and the effects of testing method on the respondents’ and their test performance. In another study by Pourdana et al. (2012), the possible relationship between test-taking strategies and successful performance on EFL reading comprehension and also EFL learners’ level of language proficiency was investigated. The findings in this study were interpreted as the low knowledge of test-taking strategies in Iranian EFL context and the importance of attending to the cognitive processes effective in taking language tests. Goudarzi & Ghonsooly (2014) also studied the relationship between metacognitive awareness and test-taking strategies used by learners studying English as a foreign language (EFL) and the consequent effects of such a relationship on their language test performance. The results of this study emphasized the importance of metacognitive awareness and test-taking strategy use in learning process and learners’ performance.

Despite the positive effect of implementing appropriate test-taking strategies on the process of test taking, these strategies are often ignored in many EFL/ESL teaching curricula. This may be due to some reasons. Most teachers may not have realized the effectiveness of teaching these strategies, or they are so much involved in teaching English content that they do not devote any time to teaching the strategies. This leads to the existence of too many students knowing a great deal of English yet unable to perform satisfactorily on language tests (Ghafournia, 2013)
2. An Exploratory Study by Yun Xu and Zunmin Wu (2012)

The authors of this article stress at the outset that research on test-taking strategies assists test developers in finding out how test takers approach and interact with testing tasks. If the actual test-taking process does not coincide with what test developers expected, the test might fail to measure what it intends to measure (Bachman, 2000). Referring to literature on test-taking strategies, Rogers and Harley (1999) define them as those techniques used by the test takers in order to increase scores in a test-taking situation. These strategies include: reading the instructions carefully, scheduling the allocated time appropriately, making use of clue words in the questions, delaying answering difficult questions, reviewing the work in order to check the answers, etc. According to Skehan (1991, p. 290), “all learners use strategies; what good learners do is to choose the right strategy for the right occasion”. This attitude toward the importance of understanding the ways language learners interact with tests tasks by the test developers is really relevant with context of this study. The results of the work under review might serve as an anchor to a work of reference to many educational centres in China where senior graduates in Beijing are supposed to take part in the fore mentioned Beijing Matriculation English Test annually, so that qualified candidates for admission into institutions of higher education could easily be selected. The MET has become an extremely powerful high-stakes test which not only evaluates students, but also “influences how society evaluates the schools, and how schools, parents, and students evaluate the teachers” (Cheng, 2008, p. 19).

I will begin by giving an overview of the organization of the paper and the methodology the author has applied. Following the introduction, there are background, method and results sections. The introduction, as it is evident from the above quote from Bachman, usually sets out the authors’ underlying beliefs about the main concern of the research he is conducting. Two research questions of “what are the test-taking strategies involved in completion of Situational Writing and Interpretational Writing?” and “do Situational Writing and Interpretational Writing measure what they intend to measure?” are also raised by Yun Xu and Zunmin Wu to be examined and investigated in this study. It then goes on to give a short introduction to the national Beijing MET. A final note by the authors in the conclusions section defines and summarizes the strategies students have used while taking the test.
Under the title “background”, the authors have introduced the Beijing MET as a norm-referenced test developed by the Beijing Education Examinations Authority. Regarding the test format, it consists of four parts of listening, knowledge application, reading and writing. The allotted time is two hours and the test is regarded as the gatekeeper for universities in the highly competitive education system (Cheng & Qi, 2006; Qi, 2005). Due to MET’s great pressure, teachers spent a large amount of class time teaching to the test content and automatizing test performance through mock tests. Authors have mentioned the inclusion of large number of mock tests before running the experiment. One would like to add ‘choosing participants from schools with no or less mock tests and less concern for MET test may have changed the results fundamentally’. Writing section of the Beijing MET comprises 23% of the total score and includes two picture prompt writing tasks. The first is Situational Writing in which test takers should finish a letter addressed to an American boy. Interpretational writing is the other task in which they are supposed to express how they understood a picture. Interpretational Writing allows some latitude for different interpretations of the cartoon and encourages creative thinking. When giving a score, raters would consider whether the essay focuses on the subject matter; whether the content is rich; whether creative and logical thinking is reflected. The essay length is what the authors have mentioned as one of the factors that would affect scoring.

The “Test-taking strategies” section explores how understanding the actual test-taking process needs to coincide with what test developers expected in order to make sure that testing tasks measured what they intended to measure. This fact is considered in this paper as the key point where validity requirements of a test can be guaranteed. What the paper tries to highlight is the necessity of conducting the study of test-taking strategies to support the outcomes of test development processes. There is a discussion of test-taking strategies which draws on Cohen, (2006) and The taxonomy of test-wiseness and its classifications proposed by Millman et al. (1965) represented the most notable achievement in this field. Overview of test taking strategies and their frequency criteria as possible sources of variance in test scores also are presented in this section. Three main categories of test-taking strategy research for the purpose of validation, language proficiency related to test-taking strategies, and strategy instruction for standardized tests (Cohen, 2006) are also discussed by reminding the fact that these have been the dominant area of research on test-taking strategies after 1980s. Despite its focus on the main
categories of test-taking strategy research, this article does not include a rich literature by presenting the previous works done on test-taking strategies for other national or international English tests done by other authors. Thus, 'Introduction’ section of this review paper tries to review the literature on such issues as complementary section to the original work by Yun Xu and Zunmin Wu (2012).

The “methodology” section of the article reviews the approach, instrument and data collection procedures applied by the two authors. Regarding the participants, twelve Chinese students representing two key and two ordinary senior high schools in both urban and suburban areas of Beijing have got involved in this study. The second group of participants includes four English teachers from four senior high schools in Beijing. In my reading, the choice of participants is logical since they are all senior graduates who need to get admission into institutions of higher education. Including both male and female participants would guarantee gender effect in the results of the study. The cognitive ability and development of male and female students differ in many respects. They use different learning strategies and enjoy different cognitive abilities. For example, the "difference is greatest in mental rotation tasks" (Spenser, 2007). Additionally, care has been taken not to place less necessary details about the participants and the information about their background. It is easy to find, and clearly presented in the table (p. 4). Whereas a reviewer has to read the article cover to cover, it might take a conventional reader a whole career to do so by spending some time to look for specific information over time.

“Instruments and data collection” section provides statement of the approach the study has taken to accomplish the authors’ goals. It is tried to demonstrate that a qualitative, exploratory research study based on think-aloud Protocol and interviews has been conducted. In the think-aloud phase, participants were supposed to use Chinese, English, or mixed code or whatever language in their mind. To adjust the think-aloud procedures and the interview guidelines to suit teenage participants, a pilot study of a ten minute- practice on a topic was conducted. To be a positive point in this study, not a picture prompt writing task was included, in the fear that the training might contaminate the verbal protocols. Retrospective telephone interviews and discourse analysis were other tools of data collection conducted among their English teachers to explore both the issues raised by their students and focusing on reoccurring organizational patterns. Positively, using interview to elicit participants’ feedback on using test-taking strategies could lead into more in-depth results. There is, however, little or no reference to
ways and reasons for which the authors have conducted telephone interviews with teachers since it is not considered as a reliable tool for data collection. The author’s intention of including this phase is not clear. As for data analysis, the authors have found reoccurring patterns and put them into categories based on their internal connections. After the finalization of coding schemes, the protocols were coded using the software Winmax 98.

To be considered as one of the merits of the study, the coding was done by the first author and double checked by the second author to guarantee its validity. For discourse analysis, the narrative structure initiated by Labov (1972) and revised by Berman (1997) and the generic structure of expository essays (Callaghan & Rothery, 1988) were adopted to analyze the macro-structures of Situational Writing and Interpretational Writing. One of the strong points of this article is the use of famous and reliable taxonomies for discourse analysis, learning strategies, writing strategies and macro-structure of Situational Writing by known linguists in language teaching and testing especially evident in data analysis and test-taking strategies sections (pp.3&5).

Additionally, one essential feature of this study is its focus on individual test-takers performance. However, the authors should be forgiven for including the phrase “12 Chinese EFL Learners”, the exact specification of individuals engaged in the study, in the title of the study and listing the number of the participants in a nonprofessional way.

The following sections of the study suddenly turn into reporting on the most frequently used strategies by the participants. In the results section, twenty two writing strategies under four broad categories based on Oxford (1990)’s learning strategy framework were identified. Analyzing students’ interview protocols think-aloud protocols and written products, researchers found twenty one test-management strategies and categorized them into seven categories. Concerning the organizational pattern for Situational Writing, all participants showed similarities in having the same starts according to the first picture, having parallel narrations based on the four pictures, applying equal effort to each picture, not missing essential information of the pictures and score loss according to the scoring rubrics, adding an ending to their narration, having similar structures and the existence of temporal order in their narrations. The authors examined the macro-structure of Interpretational Writing and found that two parts were shared by essays of most students except Daisy and Denise, picture description and theme representation. One issue raises here by noticing two non-Chinese names of “Daisy and Denise”. What is extremely surprising to my appeal is the question of why of
such misnomers. The authors may have forgotten to mention that they have had some non-native participants in this study or used substitutions or widely recognizable nick names for their participants. Yun Xu and Zunmin Wu subsequently address the question, crucial in the context of China’s higher education, whether Situational Writing and Interpretational Writing measure what they intend to measure. As a starting point, they investigated the result of participants’ experience of suffering from severe score loss in mock tests. Students strived to guess test-developers’ intent and figure out the “best” theme instead of freely expressing their own interpretations. What the results of this study reveal is that they might fail to achieve the crucial goal of measuring creative thinking. All the students were concerned about essay length. Although in the experiment no time limit was imposed upon students, they rushed to finish the tasks. Revision was completely ignored by six students who handed in their essays immediately after they stopped writing since there is no time for revision on the real tests. In the last part of the “data analysis” section, the authors go into more detail about how three sources of officially publicized scoring rubrics, information from “insiders” who used to participate in the actual scoring process and teachers’ own accumulation of coaching experiences have been identified from the current research.

How is this article to be valued? This depends largely on what kind of readership one has in mind. For somebody who is familiar with common trends of MET Beijing test, commonly used strategies by test takers and the test’s prerequisites, there is not that much news. In contrast, to a newcomer in the field or to someone who did not know much on the test properties and school practices in Beijing, I imagine this must be a very stimulating and thought-provoking article. There are, however, a number of problems with this work which I feel mainly arise precisely from not having placement test to make the participants’ level of proficiency homogeneous. Homogeneity is regarded as an integral part of every successful study investigating learners’ linguistic features and capabilities in different learning contexts. In addition, the long process of data collection and elicitation in this study seems to be demotivating both for participants and researchers and may affect the results of the study so the possible influence of data collection procedure and method is one of main sources of problem in this study. One point that is evidently lacking in his paper is the quantitative methods alongside with making efforts to describe the strategy and search for reasons underlying certain strategy use to report research findings. The final point I wish to include is the fact that due to
practical constraints and method effects, unfortunately the experiment was conducted under a *non-testing* situation. This brings us back to statement of the aims of the present review stated at the very beginning of this review: ‘whether the article by Yun Xu and Zunmin Wu fulfils their stated aims ‘To which the answer must be ‘Up to a point.’ Yes, although this article suffers from repetitious claims, it surveys the exact concern of the Chinese community who are willing to participate in institutions of higher education by going through a national English test (MET). Whether as a primer, or a reference work, this is an article no serious English teacher at Chinese language institutes and schools can afford to be without. Mentioning one seemingly redundant procedure of “focusing on reoccurring organizational patterns “(p.5) gives pause for thought. Since nowhere in the study the findings from this stage have been presented and applied. It might be of limited value and can represent an invaluable tool for data analysis.

3. Conclusion

The Yun Xu and Zunmin's study enjoys certain advantages and suffers from some disadvantages as delineated above. The following summery reiterates those and adds some further characteristics that have not been mentioned above. To sum up, the present study draws on qualitative approaches to collect data using a variety of instruments such as think aloud protocols and interviews. As discussed in the previous sections of this review paper, the researchers are interested in individual responses and eliciting data through interviews has empowered them with in-depth results. As well, the emphasized research questions attempt to account for all the viable aspects of test-taking strategies applied by Chinese test takers and the results of the researchers’ investigation are clearly providing comprehensive response to them. On the other hand, inappropriate use of the phrase “12 Chinese EFL Learners” has caused the title to be somehow senseless. Research has shown that students’ school attendance, performance, and test-taking strategies are significantly affected by students’ gender (e.g. Madigan, 2011; Schramm, 2011). Although the participants are chosen from both males and females, the study provides no specifications as to how this might have affected the choice of strategies among students. There is also another *problem* with this work which is related to not having placement test to make the participants’ level of proficiency homogeneous.
Despite the positive effect of implementing appropriate test-taking strategies on the process of test taking, these strategies are often ignored in many EFL/ESL teaching curricula. This may be due to some reasons. Most teachers may not have realized the effectiveness of teaching these strategies, or they are so much involved in teaching English content that they do not devote any time to teaching the strategies. This leads to the existence of too many students knowing a great deal of English yet unable to perform satisfactorily on language tests (Ghafournia, 2013). Overall, this paper surveys the exact concern of the Chinese community who are willing to participate in institutions of higher education by going through a national English test (MET). To a newcomer in the field or to someone who did not know much on the test properties and school practices in Beijing, I imagine this must be a very stimulating and thought-provoking article.
References


Behavioral Sciences, 70, 90 – 94.


