

An investigation into the test-taking strategies employed for a high-stake test: Implications for test validation

Kashkouli, Zohreh ✉

Department of English, University of Isfahan, Iran (Kashkouli_z@yahoo.com)

Barati, Hossein

Department of English, University of Isfahan, Iran (h.barati@gmail.com)

Nejad Ansari, Dariush

Department of English, University of Isfahan, Iran (nejadansari@gmail.com)



ISSN: 2243-7754
Online ISSN: 2243-7762

OPEN ACCESS

Received: 13 July 2014

Revised: 18 November 2014

Accepted: 30 November 2014

Available Online: 2 December 2014

DOI: 10.5861/ijrsl.2014.852

Abstract

Many L2 testing researchers considered test-taking strategies as a possible source of insight concerning test validity. In this regard, the present study was conducted to investigate the test-taking strategies that test takers employed to answer the Iranian National University Entrance Exam for MA in TEFL (INUEMA). 260 randomly chosen undergraduate EFL learners preparing themselves for INUEMA were divided into three groups based on the results of a proficiency test. 80 students were grouped as high-ability group, 78 participants as low-ability and 88 students as the intermediate group. All groups completed the reading section of INUEMA and the Test-taking Strategy Questionnaire which contained four types of strategies. The three groups were compared with respect to the four types of test-taking strategies they employed. The findings revealed that from among all participants, intermediate group used test-taking strategies more than others. The results also showed that monitoring and evaluation were used significantly more than other strategies, which means that takers relied on their academic reading skills for both specific and general comprehension of the texts using neither their background knowledge nor test-wisness strategies. Signaling the validity of the above high-stake test, the findings of the present study are likely to be of great interest to EFL material developers, instructors and testing organizations.

Keywords: test-taking strategies; validity; reading; monitoring; evaluation; test-wisness

An investigation into the test-taking strategies employed for a high-stake test: Implications for test validation

1. Introduction

For more than four decades, scholars have focused on second language (L2) testing from the viewpoint of the strategies that respondents use for performing a language test (e.g., Cohen & Aphek 1980, Homburg & Spaan 1981, Cohen 1984; etc). Many L2 testing textbooks (e.g. Bachman 1990; Cohen 1994) considered test-taking strategies as a possible source of insight concerning test reliability and validity. For instance, Bachman (1990), with reference to Messick (1989), Cohen (1984) and Grotjahn (1986), mentioned that test taking processes provide evidence for construct validity of a test and experimental approaches are not enough for validating a test: *"a more critical limitation to correlational and experimental approaches to construct validation is that these examine only the products of the test taking process-the test scores- and provide no means for investigating the processes of test taking"* (p. 269). Cohen (1998) mentioned that that test taking strategies can be used for validation purposes *"while there is nothing new in pointing out that certain instruments used in SLA research are lacking in validity, it is a relatively new undertaking to use data on test taking strategies to validate such tests"* (p. 92).

In this regard, a call has been for the development of language tests that properly indicate the language ability of the test taker to manage the language task being assessed, while "guarding against opportunities for selecting the right answers for the wrong reasons" (Cohen & Upton, 2006, p. 32). It is obvious that test takers have learned various techniques to find correct answers to reading tests *"without fully or even partially understanding the text"* (Cohen, 1986, p. 132). In other words, in a test which claims to evaluate academic reading ability, the questions should be in a way that the test takers be obliged to use reading skills in responding to items, rather than relying on test-wiseness strategies.

1.1 Test-Taking Strategies

Test taking strategies are metacognitive strategies that test takers consciously use to *"direct and control their cognitive strategies for successful performance on the test"* (Phakiti, 2003, p. 29). In other words, test-taking strategies are introduced in the literature as those test-taking processes that the test takers have consciously selected; in other words, *"the notion of strategy implies an element of selection"* (Cohen & Upton, 2006, p. 34). They range from decisions made by the test takers for the language tasks to creating shortcuts to arrive at answers to the questions. For example, the test taker may utilize *"surface matching of identical information in the passage and in one of the response choices"* or may not read *"the text as instructed" but simply look "immediately for the answers to the given reading comprehension questions"* (Cohen, 2007, p. 305). *"In such cases, the respondents may be using test-wiseness to circumvent the need to tap their actual language knowledge or lack of it"* (Cohen, 1998, p. 93). In this regard, Fransson (1984) says that respondents may not proceed via the text but rather around it (p. 90).

There have been a number of studies related to test-taking strategies in language testing since 1990 such as Anderson, Bachman, Perkins, and Cohen (1991), Block (1992), and Purpura (1997), Purpura (1998), and Phakiti (2003). For instance, Phakiti's research (2003) investigated the cognitive and metacognitive strategies to reading test performance in three language ability groups. This study employed both quantitative and qualitative data analysis. Its 384 participants enrolled at a Thai university took an 85-item multiple-choice reading comprehension achievement test. Eight of the students (4 highly successful and four highly unsuccessful) were selected for further investigation on the nature of cognitive and metacognitive strategy use. His findings demonstrated that the use of cognitive and metacognitive strategies had a positive relationship to the reading test performance; and highly successful test-takers reported significantly higher metacognitive strategy use than the

moderately successful ones who in turn reported higher use of these strategies than the unsuccessful test-takers.

Further Barati (2005) conducted another study on test-taking strategies and adult EFL learners. In that study, he integrated the quantitative and qualitative research to examine test-taking strategies effect on adult EFL learners' reading test performance. His findings showed that test-taking strategies affected the reading skills test performance of all groups of participants significantly. In fact, according to the finding of this study, strategies did not always promote test performance but rather there were cases in which they affected the test results negatively. The finding of this research also revealed that less able test takers deployed test-wiseness strategies significantly more frequently than other participants.

Another study (Cohen & Upton, 2006) on test-taking strategies consisted of a process-oriented effort to describe the reading and test-taking strategies that test-takers used with different item types on the Reading section of the LanguEdge Courseware materials developed to familiarize prospective respondents with TOEFL iBT. The findings of the above study revealed that the Reading section of the TOEFL iBT does call for the use of academic reading skills for passage comprehension. It was also obvious through the verbal protocol that "*the Reading section of the LanguEdge test did not fully constitute an academic reading task but rather a test-taking task with academic-like aspects to it*" (Cohen 2007, p. 99).

Regarding the situation in Iran, Razmjoo and Heidari Tabrizi (2010) conducted a content analysis of the INUEMA (Iranian National University Entrance Exam for MA) for TEFL held in 2007. The purpose of this study was two-fold. First, it aimed at analyzing the content of the MA exam in order to see if any pattern was at work in the process of devising such exams. The second aim of that study was to pinpoint and describe the problems with this exam and to offer some suggestions to remedy the problems. The findings supported the idea that the validity of the exam was not strongly established due to the exclusion of or de-emphasis over the content categories given significant credit in the B.A. program. The problems found during the analysis showed that the exam was not a standard one; still some of the basic principles of language testing were not observed in the process of constructing the exam.

In another study, Salehi and Yunus (2012) explored the effects of the Iranian University Entrance Exam on the high school English teachers. To investigate the teachers' perceptions of the University Entrance Exam, a survey questionnaire was administered to a stratified random sample of 132 high school English teachers. The descriptive data analysis revealed that little attention was given to three language skills of speaking, writing, and listening in the classroom as these skills were not tested by University Entrance Exam. Moreover, the researchers found that University entrance Exam negatively and implicitly influenced English teachers to teach to the test format.

Although several studies have been conducted on the University Entrance Exam (Konkour) in Iran, there is currently a paucity of research on validating its MA counterpart. This paucity becomes more acute in the area of language testing for no comprehensive research has been conducted on INUEMA in TEFL to validate this test regarding the test-taking strategies that prospective candidates use at the time of answering such test. Therefore, as the above discussion indicates, research on constructively validating the high-stakes test through test-taking strategies is still in its beginning ages and needs more work on different participants in different learning situations.

Moreover, since vital decisions are made based on the examinees' performance on such exams; decisions that do have decisive impacts directly on the examinees' future, contributing to INUEMA prospective candidates' success on such a competition is worth researching. Farhady (2003, p. 85) mentioned that "*one of the important usages of tests in education is to select people for different educational levels*". If it is used at a higher level, it will be much more significant and complex. The specific importance of this issue is due to highly financial and mental investment on selected people. It should be noted that in order to make sound judgments about any objections on these tests, they should be validated by researchers in this field. In other words, language tests are designed to measure an individual's language ability, but if the test taker's employment background (for

example) influences his or her performance on the test, then it is generally considered that construct irrelevant variance has been introduced into the measurement of the individual's language ability (Wanger, 2006).

1.2 Research Question

As it was mentioned, this study seeks to answer the following research question:

- Is there any significant difference in the types of test-taking strategies used by three proficiency groups of test takers when completing the Reading section of INUEMA?

2. Method

2.1 Participants

The population from which the participants of the study were selected included 260 randomly chosen students aged from 21 to 26, majoring in English Language and Literature, as well as Translation studying at different universities of Iran. They were all adult undergraduate EFL learners in their last semester of BA. Two TOEFL reading passages, each with 7 questions, were administered as the proficiency indicator before the main phase of the study to divide participants into three proficiency group based on their scores. Based on their scores from the proficiency test, 80 students were grouped as high-ability group, 78 participants as low-ability and 88 students as the intermediate group.

2.2 Instruments

The instruments in this research were 1) TOEFL reading paper as proficiency indicator, 2) the reading subtest of NIUEMA, and 3) Test-taking Strategy Questionnaire. The TOEFL reading paper contained two reading passages of about 300 words each with 7 questions. The Reading subtest of NIUEMA contained 3 passages of 300-400 words with a total of 20 questions. The item specifications of NIUEMA were checked by a group of expert judges. Table 1 presents test item specifications based on the expert judges' decisions.

Table 1

NIUEMA Item Specifications according to Expert Judgment for the second research question

Item type	No. of items	Passage nos.
Word meaning	1	3
Subject pronoun reference	1	2
Factual information	3	3
Not given information	4	1, 2, 3
Inference/implied meaning	11	1, 2

The third instrument used in this study was a Test-taking Strategy Questionnaire (see appendix) obtained from Barati (2005). This questionnaire consists of 27 items each of which presents a statement about the use of one strategy. According to the results of Barati (2005), from the total of 27 items, 6 items asked for planning strategies, 13 items asked about test takers' use of monitoring strategies, 4 items address evaluation strategies, and 4 items ask about test-wiseness strategy. The test-taking strategy questionnaire was translated into Persian to be in the participants' native language and avoid any ambiguity. In this instrument the Likert scale was used: 1 = never; 2 = sometimes; 3 = often; 4 = usually, and 5 = always. Participants were asked to mark the questionnaire in a way that it indicates how frequently they used each strategy when they were completing items of the reading paper. The structure of this questionnaire is presented below:

Table 2

The structure of the Test-taking Strategy Questionnaire

Strategy	No. of items	Task description
1. Planning	6	previewing or overviews tasks in order to determine what actions to be done
2. Monitoring	13	checking comprehension, accuracy and/or appropriateness of action which is taking place
3. Evaluation	4	checking comprehension after completion of receptive language activities
4. Test-wiseness	4	using the knowledge and experience of how to take the test in answering the items

Before conducting the main study, the instruments (two TOEFL reading passages, the reading subtest of INUEMA, and the test-taking strategy questionnaire) were piloted on a sample of the target population including 30 participants studying in their last year of undergraduate studies. The pilot study was carried out for the following reasons:

1. To measure the reliability of the instruments (both the TOEFL passages and the Reading section of INUEMA)
2. To observe the amount of time the participants need to answer the two TOEFL texts, the reading comprehension passages of INUEMA, and the test-taking strategy questionnaire

The pilot study informed the main phase of research in the following aspects:

- A. The Cronbach Alpha for the reliability estimate of the TOEFL tests was, .87 suggesting that the test was quite acceptable as an indicator of proficiency. Moreover, the Cronbach Alpha for the reliability estimate of the INUEMA was, .79 suggesting that the test was reliable in terms of its internal consistency.
- B. The amount of time needed for the participants to answer the instruments was established. In other words, the pilot study showed that the reasonable amount of time was 30 minutes TOEFL passages, 45 minutes for three INUEMA reading subtests, and 10 minutes for test-taking strategy questionnaire.

2.3 Procedure

Data collection was carried out in two separate stages. The first stage was a session devoted to participants' answering the two TOEFL passages as indicators of proficiency level which lasted about 30 minutes. In the second stage of data collection, the participants answered three reading subtests of INUEMA, and the test-taking strategy questionnaire which lasted approximately 55 minutes.

3. Results and discussion

To answer the research question, a series of Kruskal-Wallis tests was conducted to investigate whether the three ability group's measures of performance on four types of test-taking strategies revealed significant difference. In cases the difference among ability groups was significant; Mann-Whitney U test was conducted as a follow-up significance test in order to indicate where the significant difference occurs. The following sections will present the result of Kruskal-Wallis tests for the performance of the high, intermediate, and low-ability test takers on each type of test-taking strategies.

3.1 Planning

In order to investigate if there was significant difference in the use of planning by various ability group test takers when completing INUEMA, a Kruskal-Wallis test was conducted. Table 3 shows the results:

Table 3*Kruskal-Wallis results for the three ability group's performance on Planning*

Statistics	Score
Chi-Square	3.028
df	2
Asymp.Sig	.220

As Table 3 showed, the p value is .220 ($p > .05$); therefore the difference among the three ability groups on the use of planning was not significantly different. In fact, the three ability groups did not employ planning strategy significantly different from each other. Next section will present the results of Kruskal-Wallis test for the performance of the three ability groups on monitoring.

3.2 Monitoring

Table 4 presents the results of Kruskal-Wallis to reveal whether high, intermediate, and low-ability groups differ significantly on the use of monitoring when completing INUEMA:

Table 4*Kruskal-Wallis results for the three ability group's performance on Monitoring*

Statistics	Score
Chi-Square	12.200
df	2
Asymp.Sig	.002

As Table 4 showed, p value is .002 ($p < .05$); therefore, the difference among the three ability groups on the use of monitoring is significantly different. To see where this significant difference occurred, Mann-Whitney U test was conducted on the related data. Table 5 presents the result of Mann-Whitney U test to seek the difference between high-ability and the intermediate test takers:

Table 5*Mann-Whitney U Test on the high and the intermediate groups' performance on monitoring*

Statistics	Score
Mann-Whitney U	2468.000
Wilcoxon W	5708.000
Z	-3.346
Asymp.Sig (2-tailed)	.001

As Table 5 indicated, the p value of the Mann-Whitney U test is .001 ($p < .05$) which means that the difference between the high and the intermediate group's use of monitoring strategy was significant. To see which group outperformed the other, mean ranks of both groups were considered. Table 6 reveals the mean ranks of the high and the intermediate group.

Table 6*Mean ranks of the high-ability and the intermediate group's use of Monitoring*

group	n	Mean Rank	Sum of Ranks
high	80	71.35	5708.00
intermediate	88	96.45	8488.00
Total	168		

Table 6 indicated that the intermediate group's mean rank was more than that of the high-ability group. The conclusion was that the intermediate group used more monitoring strategies than the high-ability group when completing the INUEMA. To investigate the difference between the high and the low-ability group's use of monitoring strategies, another Mann-Whitney U test was conducted, this time between the high and the

low-ability test takers. Table 7 presents the results:

Table 7

Mann-Whitney U Test on high and low-ability group's performance on Monitoring

Statistics	Score
Mann-Whitney U	3030.000
Wilcoxon W	6270.000
Z	-.314
Asymp.Sig (2-tailed)	.754

As Table 7 indicated, the p value of the Mann-Whitney U test is .754 ($p > .05$) which means that the difference between the high and the low-ability group's use of monitoring strategy was not significant. To investigate the difference between the intermediate and the low-ability group's use of monitoring strategies, another Mann-Whitney U test was conducted, this time between the intermediate and the low-ability test takers. Table 8 presents the results:

Table 8

Mann-Whitney U Test on the intermediate and low-ability group's performance on Monitoring

Statistics	Score
Mann-Whitney U	2634.000
Wilcoxon W	5715.000
Z	-2.586
Asymp.Sig (2-tailed)	.010

As Table 8 indicated, the p value of the Mann-Whitney U test is .010 ($p < .05$) which means that the difference between the intermediate and the low-ability group's use of monitoring strategy was significant. To see which group outperformed the other, mean ranks of both groups were considered. Table 9 reveals the mean ranks of the intermediate and the low-ability group.

Table 9

Mean ranks of the intermediate and the low-ability group's use of monitoring

group	n	Mean Rank	Sum of Ranks
intermediate	88	92.57	8146.00
low	78	73.27	5715.00
Total	166		

Table 9 indicated that the intermediate group's mean rank was more than that of the low-ability group. The conclusion was that the intermediate group used more monitoring strategies than the low-ability group when completing the INUEMA. Next section presents the results of Kruskal-Wallis test for the performance of the three ability groups on evaluation.

3.3 Evaluation

Table 10 presents the results of Kruskal-Wallis to reveal whether high, intermediate, and low-ability groups differ significantly on the use of evaluation strategy when completing INUEMA:

Table 10

Kruskal-Wallis results for the three ability group's performance on evaluation

Statistics	Score
Chi-Square	3.425
df	2
Asymp.Sig	.180

As Table 10 showed, the p value is .180 ($p > .05$); therefore the difference among the three ability groups on the use of evaluation was not significantly different. Next section will present the results of Kruskal-Wallis test for the performance of the three ability groups on test-wiseness.

3.4 Test-wiseness

In order to investigate if there was significant difference in the use of test-wiseness by various ability group test takers when completing INUEMA, a Kruskal-Wallis test was conducted. Table 11 shows the results:

Table 11

Kruskal-Wallis results for the three ability group's performance on test-wiseness

Statistics	Score
Chi-Square	2.543
df	2
Asymp.Sig	.280

As Table 11 showed, the p value is .280 ($p > .05$); therefore the difference among the three ability groups on the use of test-wiseness was not significantly different. Next section will present the results of Kruskal-Wallis test for the performance of the three ability groups on the four types of test-taking strategies when completing different item contents.

4. Conclusion

The present study aimed at investigating any significant difference in the type of strategies used by three proficiency groups of test takers when completing the reading subtest of NIUEMA. The inspiration for conducting such research came from the insight gained from theoretical underpinnings as elaborated by Shohamy, (2001, p. 7), when she states,

"...in the testing literature test takers are often kept silent; their personal experiences are not heard or shared. It seems that the testing profession ... is not interested in such accounts...listening to the voices of test takers provides testers with a new and unique perspective and a deep insight into tests and their meanings".

As the results revealed, from among all participants, intermediate group used test-taking strategies more than others. The low proficiency group in the present study did not significantly utilize test-taking strategies. This is in line with previous researches in this field (Anderson, 2002; Dhieb-Henia, 2003; Dreyer & Nel, 2003; Eskey, 2005; Steinagel, 2005).The reason, according to Zhang and Seepo (2013), may be that

Low proficiency students have poor monitoring skills during reading which is vital for the reading achievement. The explanation for this could be the low proficiency students' weak metacognitive awareness in applying the strategies and their poor linguistic knowledge which led to the further discussion in the following section. (p. 62)

The findings of the present research also showed that monitoring was used significantly more than other strategies, which means guessing or test-wiseness strategies were not used by test takers on this test. If this happened, then the construct validity of the test could be under question. Although their focus was on FCE Reading paper, Kashkouli and Barati's (2013) findings showed that monitoring strategy was more used than other types by the three proficiency groups of test takers. The frequent use of monitoring strategies instead of test-wiseness disclosed that respondents were in reality focused on the reading passages and finding the appropriate answers to the questions through the use of appropriate, but not counter valid, strategies because as Cohen and Upton assert.

"A test claiming to evaluate academic reading ability would be expected to include tasks calling

for test takers to actually use academic reading skills in responding to items, rather than being able to rely on test-wiseness tricks" (2006, p. 117)

Moreover, the number and scope of the items in the test-taking strategies questionnaire might have been insufficient to investigate the real nature of test-taking strategies in EFL learners. Any generalization based merely on the findings of the present study should, therefore, be made with care. Further similar empirical studies are needed with more items in their questionnaires so that each strategy is addressed by a few more items and more strategies overall are examined than those studied in the present investigation.

In other words, as Jamil, Aziz, and Razak (2010) mention, this use of monitoring strategies "*signals that [the test takers] were serious in selecting the correct answers and serious about the test because it displayed their worries should they make the wrong decision ... they were conscious of what they were doing and did care when it came to selecting their answers in a test*" (p. 120). This proof for the validity of NIUEMA can be considered the most important finding of this study since as Cohen (1984) mentions, "*the main conclusion in is that a closer fit should be obtained between how the test constructors intend for their tests to be taken and respondents actually take them*" (p. 70).

4.1 Pedagogical Implications

The present study investigated the test-taking strategies that respondents used for answering National Iranian University Entrance Exam (NIUEMA) for TEFL. Research on 'strategies', in general, began with the idea that what successful language learners do to promote their learning can be taught to unsuccessful learners and will help them enhance their learning ability (Rubin, 1975). Further, the belief was that if effective learning strategies were recognized, teachers and curriculum developers could incorporate the development of these strategies in their teaching/learning methods, thereby improving the learning ability of poor (ineffective) learners. This would, in turn, reinforce strategy training. Thus, the pedagogical implication of much strategy research has been one of the main concerns for learning strategy researchers.

The fact that monitoring was used significantly more than other types of test-taking can be a crucial reason for field practitioners to notice their importance and include them in their teaching practice and course materials especially in MA Entrance Exam preparation courses. Moreover, this study revealed that guessing or test-wiseness strategies were not frequently used by test takers on this test which can prove the construct validity of NIUEMA. Because validity and the consequent decisions based on the results of high-stake tests are of great importance for all people involved including practitioners and the test developers, the findings of the present study are likely to be of great interest to EFL material developers, instructors and Testing Organizations. Moreover, language learners and candidates for post graduate studies could benefit from the way high proficient test takers approached MA Entrance Exam.

In spite of the effort made in the present study, more research should be conducted to investigate the effect of such factors as age, gender, level of text difficulty, type of and other relevant factors which may affect the use of test-taking strategies by language learners. Although the questionnaire employed in the present study investigated the test-taking strategies in detail, protocol analysis (think-aloud procedures) may shed more light on the process of taking a test by EFL learners.

5. References

- Anderson, N. J. (2002). *The role of metacognition in second language teaching and learning*. ERIC Digest EDO. Washington, OC: ERIC Clearinghouse on Languages and Linguistics.
- Anderson, N. J., Bachman, L. F., Perkins, K., & Cohen, A. (1991). An exploratory study into the construct validation of a reading comprehension test: Triangulation of data sources. *Language Testing*, 8(1), 41-66. <http://dx.doi.org/10.1177/026553229100800104>
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

- Barati, H. (2005). *Test-taking strategies and the assessment of reading skills: an approach to construct validation*. Unpublished Doctoral Dissertation. University of Bristol. UK.
- Block, E. L. (1992). See how they read: comprehension monitoring of L1 and L2 readers. *TESOL Quarterly*, 26(2), 119-143. <http://dx.doi.org/10.2307/3587008>
- Cohen, A. D. & Upton, T.A. (2006). *Strategies in responding to the new TOEFL reading tasks* [Monograph No. 33]. Princeton, NJ: ETS.
- Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing*, 1(1), 70-81. <http://dx.doi.org/10.1177/026553228400100106>
- Cohen, A. D. (1986). Mentalistic measures in reading strategy research: Some recent findings. *The ESP Journal*, 5(2), 131-145.
- Cohen, A. D. (1998). *Strategies in learning and using a second language*. London: Longman.
- Cohen, A. D. (2007). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307. <http://dx.doi.org/10.1080/15434300701333129>
- Cohen, A. D., & Aphek, E. (1980). Retention of second-language vocabulary over time: Investigating the role of mnemonic associations. *System*, 8, 221-235. [http://dx.doi.org/10.1016/0346-251X\(80\)90004-4](http://dx.doi.org/10.1016/0346-251X(80)90004-4)
- Dhieb-Henia, N. (2003). Evaluating the effectiveness of metacognitive strategy training for reading research articles in an ESP context. *English for Specific Purposes*, 22(4), 387– 417. [http://dx.doi.org/10.1016/S0889-4906\(03\)00017-6](http://dx.doi.org/10.1016/S0889-4906(03)00017-6)
- Dreyer, C., & Nel, C. (2003). Teaching reading strategies and reading comprehension within a technology-enhanced learning environment. *System*, 31(3), 349–365. [http://dx.doi.org/10.1016/S0346-251X\(03\)00047-2](http://dx.doi.org/10.1016/S0346-251X(03)00047-2)
- Eskey, D. E. (2005). Reading in a second language. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 563–580). Mahwah, NJ: Lawrence Erlbaum Associates.
- Farhady, H. (2003). Review of TEFL MA Entrance exams. *Research in Foreign Languages*, 13, 79-106.
- Fransson, A. (1984). Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. In J. C. Alderson & A. H. Urguhart (Eds.), *Reading in foreign language* (pp. 86-121). Longman, London.
- Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing*, 3, 159-186. <http://dx.doi.org/10.1177/026553228600300205>
- Homburg, T. J., & Spaan, M. C. (1981). ESL reading proficiency assessment: testing strategies. *TESOL Quarterly*, 4(2), 21-57.
- Jamil, A., Aziz, M. S., A. & Razak, N., A (2010). The utilization of test-taking strategies among female students in a tertiary institution. *GEMA Online™ Journal of Language Studies*, 10(3), 105-125.
- Kashkouli, Z., & Barati, H. (2013). Type of test-taking strategies and task-based reading assessment: A case in Iranian EFL learners. *Procedia-Social and Behavioral Sciences*, 70(25), 1580-1589. <http://dx.doi.org/10.1016/j.sbspro.2013.01.226>
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education, Macmillan Publishing Company.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20(1), 26-56. <http://dx.doi.org/10.1191/0265532203lt243oa>
- Purpura, J. E. (1997). An Analysis of the relationship between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 42(2), 289-325. <http://dx.doi.org/10.1111/0023-8333.91997009>
- Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high and low-ability test takers: A structural equation modeling approach. *Language Testing*, 15(3), 333-379.
- Razmjoo, S. A., & Heidari Tabrizi, H. (2010). A content analysis of the TEFL M.A. Entrance examinations (Case study: Majors courses). *Pan-Pacific Association of Applied Linguistics*, 14(1), 159-170.
- Rubin, J. (1975). What the "good language learner" can teach us. *TESOL Quarterly*, 9, 41-51. <http://dx.doi.org/10.2307/3586011>
-

- Salehi, H., & Yunus, M. Md. (2012). University entrance exam in Iran: A bridge or a dam? *Journal of Applied Sciences Research*, 8(2), 1005-1008.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. London: Longman.
- Steinagel, L. O. (2005). *The effects of reading and reading strategy training on lower proficiency level second language learners*. Unpublished doctoral dissertation, Brigham Young University.
- Wanger, E. (2006). Can the search for "fairness" be taken too far? Columbia University Working Papers in *TESOL & Applied Linguistics*, 6(2).
- Zhang, L., & Seepho, S. (2013). Metacognitive strategy use and academic reading achievement: Insights from a Chinese context, *Electronic Journal of Foreign Language Teaching*, 10(1), 54-64.

Appendix

Test-taking Strategies Questionnaire

Dear Student:

Thank you very much for your contribution to this study. Please write your name (optional), age, and term of study below and then fill this questionnaire.

Name Age Term of study

After taking a test, there are a number of sentences that the test takers may use to describe how they answered the questions and what processes or strategies they used. What did you do? What were your strategies in answering the items in today's tests? Please read the following strategies and choose your answers from the given (1-5) scale: 1(never), 2(sometimes), 3(often), 4(usually), and 5(always).

What you did during the test	1	2	3	4	5
	(never)	(sometimes)	(often)	(usually)	(always)
1. I was aware of the need to plan a course of action.					
2. Before beginning the test, I tried to identify easy and difficult parts of the test.					
3. Before I started the test I decided to leave difficult questions for later.					
4. I looked for the points for each sub-test before starting the test.					
5. I looked for the sub-tests which I thought were more important before starting the test.					
6. I read the test items before reading the texts in each section to search for their answers in the text.					
7. I answered shorter text's items before longer ones.					
8. Before answering the items, I planned how to complete the test and followed my plan throughout.					
9. I made short notes and underlined main ideas while completing the test.					
10. I translated the texts and the items into Persian.					
11. I spent more time on difficult questions.					
12. I read the texts and questions several times.					
13. I thought carefully about the meaning of the test items before answering them.					
14. I used my background knowledge to answer the questions.					
15. During the test, I was well aware of what I was doing and how I was doing it.					
16. I checked my answers to pervious questions while completing the test.					
17. I corrected my mistakes immediately after I found them.					
18. To find clues to the responses I did not know, I asked the tutor for clarification.					
19. At any time during the test, I was aware of how much of the test remained to be completed.					
20. I tried to understand the questions very well before attempting to answer them.					
21. I answered some items by finding clues in other items.					
22. If no choice (in multiple-choice items) appeared correct to me, I had a pre-determined choice to mark.					
23. I made sure I understood what had to be done and how I was to do it.					
24. I carefully watched my progress to complete the test on time.					
25. I checked the accuracy of my responses as I progressed through the test.					
26. At the end of the test, I answered the unanswered items randomly (without referring to the texts).					
27. I carefully checked my answers before submitting the test.					