

Test-taking strategies and item specifications: Focusing on a high-stake test

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: 2243-7703
Online ISSN: 2243-7711

OPEN ACCESS

Received: 12 October 2014

Revised: 24 November 2014

Accepted: 19 December 2014

Available Online: 3 January 2015

DOI: 10.5861/ijrse.2015.1012

Abstract

The present study was conducted to investigate the test-taking strategies that respondents used when completing two different item specifications (Factual Information and Inference questions) of a high-stake reading comprehension test. 130 randomly chosen undergraduate EFL learners preparing themselves for graduate studies were divided into three groups based on the results of a proficiency test. Based on their scores from the proficiency test, each ability group had 40 participants. All groups completed two reading passages of Iranian National University Entrance Exam for MA (INUEMA) and two Test-taking Strategy questionnaires which contained four types of strategies. 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 after completing Inference and Factual Information questions. 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 occurred. The results revealed that the high-ability and the intermediate groups, compared with the low-ability group, employed more evaluation strategies when completing Inference items of INUEMA. Moreover, low-ability test takers used test-wiseness strategies for Inference items more than the high-ability and the intermediate groups. With respect to the Factual Information items, the intermediate group used more monitoring strategies than the high-ability test takers. Focusing on the validity of the above high-stake test, the findings of the present study are likely to be of great interest to EFL material designers, instructors and high-stake test developers.

Keywords: test-taking strategies; reading comprehension; item specification; factual information; inference

Test-taking strategies and item specifications: Focusing on a high-stake test

1. Introduction

Investigating construct validity of a test which was defined according to Anderson, Bachman, Perkins, and Cohen (1991) as the relationship between the performance of the participants on a test and the construct which the test purports to measure, with a focus on the respondents' behavior in taking the tests has been the missing part in the studies related to test validation; there is little knowledge about what test takers actually do when completing test items and whether this reveals to the ability under examination. Some language testing researchers (e.g. Phakiti, 2003), however, have been concerned with the identification and characterization of individual properties that influence one's performance on language tests. There have been two types of systematic variation in how the test takers perform a test (Bachman, 1990, p. 180):

- difference among individuals in their communicative language ability (CLA), processing strategies and personal characteristics; and
- difference among tests in terms of method or task

According to Phakiti (2003), *psychological and social characteristics such as strategy use, motivation, and attitude* are some differences of the individuals which affect their performance on a test (p. 28). With respect to strategy use, though, there are two types of learning strategies; cognitive and metacognitive (Cohen, 1984). A cognitive strategy has been defined as "*the strategy that involves mental manipulation or transformation of materials or tasks and is intended to enhance comprehension, acquisition, or retention*" (O'Malley & Chamot 1990, p. 229). Phakiti (2003, p. 29), further, suggests that metacognitive strategies involve "*active monitoring and consequent regulation and orchestration of cognitive processes to achieve cognitive goals*". Metacognitive process is commonly believed to be deliberate and self-controlled (Cohen, 1984).

One type of metacognitive strategies is test taking strategies which test takers consciously use to "*direct and control their cognitive strategies for successful performance on the test*" (Phakiti, 2003, p. 29). Cohen and Upton (2006) defined test-taking strategies as the processes which test takers select consciously to perform the language task. In fact, selection is an essential element in calling a process strategy. Otherwise, the processes cannot be called as strategies. In some cases, these strategies include dealing with the language task the test taker is performing. In other cases, the test takers may use *test-wiseness* to compensate for the need to use their actual language knowledge (Cohen, 2007). This is in line with Franson's (1984) claim that respondents may not go through the text but rather around it.

An area which has increasingly attracted researchers in validation studies is to use findings from test-taking strategy research on the process through which learners arrive at their test responses in different contexts. These studies have specially focused on validating tests of reading comprehension through the strategies used by test takers when answering them. For instance, Barati (2005) conducted a validation 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 that study, strategies did not always promote test performance but rather there were cases in which they affected the test results negatively. The finding of the 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 determine the reading and test-taking strategies which test-takers used with different item types on the Reading section of the LanguEdge Courseware materials aimed at familiarizing prospective test takers with TOEFL iBT.

The results of the above study revealed that the Reading section of the TEOFL 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 totally include an academic reading task but a test-taking task with some aspects similar to academic (Cohen, 2007).

Regarding the Entrance Exams in Iran, Razmjoo and Heidari Tabrizi (2010) conducted a content analysis of the TEFL Iranian MA held in 2007. The purpose of this study was two-fold. First, it aimed at analyzing the content of the MA Entrance 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 the most recent study conducted in Iran, Kashkouli, Barati, and Nejad Ansari (2015) investigated the test-taking strategies that test takers employed to answer the Iranian National University Entrance Exam for MA in TEFL (INUEMA). 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-wiseness strategies.

A review on test validation studies indicates that few investigators have paid significant attention to the process in which a test taker engages when taking a test and the effect of item specifications on the use of strategies by test takers. Therefore, it seems vital that one dimension of test validation should take account of the processes (i.e. strategies) that students may use while completing different item specifications of a reading comprehension test. Such an approach can provide insights into the way test tasks are processed, and may consequently provide test constructors with invaluable information about test performance before actual test results are used for decision making purposes. Moreover, a glance at the studies on test-taking strategies indicates that very few have focused on the processes test takers undertake when completing two different item specifications; Inference and Factual Information questions. The present study, therefore, investigated the test-taking strategies used by participants when completing Inference and Factual Information questions of a high-stake test, INUEMA. The idea is that uncovering the type and frequency of such strategies would provide validity evidence for this highly influential test. Therefore, the present study attempted to answer the following research question:

- Is there any significant difference in the types of test-taking strategies used by various ability group test takers when completing different item specifications in INUEMA?

2. Method

2.1 Participants

The population from which the participants of the study were selected included 130 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, each ability group consisted of 40 participants.

2.2 Instruments

The instruments in this research were 1) TOEFL reading paper as proficiency indicator, 2) the reading subtest of INUEMA, 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 INUEMA included 2 passages of 300-400 words with a total of 11 questions. The item specifications of INUEMA were checked by a group of expert judges. Table 1 presents test item specifications based on the expert judges' decisions.

Table 1

INUEMA Item Specifications according to Expert Judgment

Item Type	No. of Items	Passage No.
Word Meaning	1	2
Factual information	5	2
Not Given Information	1	1
Inference/Implied Meaning	6	1

Items measuring *Word Meaning* and *Not Given Information* were excluded from data analysis because expert judges decided only one item in the test were measuring each.

The third instrument used in this study was a Test-taking Strategy Questionnaire (see appendix) obtained from Barati (2005). This questionnaire consisted of 27 items each of which presented 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 overviewing 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:

- To measure the reliability of the instruments (both the TOEFL passages and the Reading section of INUEMA)
- 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:

- 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.

- 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 for answering 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 ability level which lasted about 30 minutes. In the second stage of data collection, the participants completed two reading comprehension sub-tests of INUEMA and a test-taking strategy questionnaire. The questionnaires were administered after each reading comprehension sub-test. This was because they completed two reading passages differing in their item specification and therefore the two test-taking strategy questionnaires were to address the items in each. Each time the participants delivered the completed questionnaire to the researcher, they received a new one for the next passage. It goes without mentioning that the two questionnaires for both passages were exactly the same so the researcher asked the test takers to write their names or codes on both questionnaire as well as the test to which the questionnaires belonged. This session took 50 minutes of the participants' time.

3. Results & 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 for Inference and Factual Information questions. 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 section consists of two parts. In the first part (3.1), the use of different types of strategies by test takers when completing Inference questions is presented. The second part (3.2) discusses strategies used by test takers when completing Factual Information questions.

3.1 Inference questions

To investigate whether there was any significant difference in the types of test-taking strategies used by various ability group test takers when completing Inference items in INUEMA, a series of Kruskal-Wallis tests was conducted. When the difference between the three ability groups was significant, Mann-Whitney U test was conducted as a follow-up significance test to reveal where the difference occurred.

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

Table 3

Kruskal-Wallis for three ability performance on planning when completing Inference items

	Score
Chi-Square	.807
Df	2
Asymp.Sig	.668

As Table 3 showed, the p value is .668 ($p > .05$); therefore the difference among the three ability groups on the use of planning was not significantly different. Next section will present the results of Kruskal-Wallis test for the

performance of the three ability groups on monitoring when completing Inference items

Monitoring - In order to find out if there was significant difference in the use of monitoring by the three ability group test takers when completing Inference items of INUEMA, a Kruskal-Wallis test was conducted. Table 4 shows the results:

Table 4

Kruskal-Wallis for three ability performance on monitoring when completing Inference items

	Score
Chi-Square	4.124
Df	2
Asymp.Sig	.127

Table 4 indicated that the p value is .127 ($p > .05$); therefore the difference among the three ability groups on the use of monitoring was not significantly different. The next section provides the results of Kruskal-Wallis test for the performance of the three ability groups on evaluation when completing Inference items.

Evaluation - This section provides the results of Kruskal-Wallis test for the performance of the three ability groups on evaluation when completing Inference items. Table 5 indicates the results:

Table 5

Kruskal-Wallis for three ability performance on evaluation when completing Inference items

	Score
Chi-Square	19.049
df	2
Asymp.Sig	.000

Table 5 showed that p value is .000 ($p < .05$); therefore, the difference among the three ability groups on the use of evaluation when completing Inference questions of INUEMA was significantly different. To see where this significance occurred, Mann-Whitney U test was conducted on the related data. Table 6 presents the result of Mann-Whitney U test to seek the difference between high-ability and intermediate test takers:

Table 6

Mann-Whitney U on high & intermediate performance on evaluation when completing Inference questions

	Score
Mann-Whitney U	760.000
Wilcoxon W	1580.000
Z	-.391
Asymp.Sig (2-tailed)	.696

As Table 6 indicated, the p value of the Mann-Whitney U test is .696 ($p > .05$) which means that the difference between the high-ability and the intermediate groups' use of evaluation strategy was not significant. Table 7 presents the result of Mann-Whitney U test to seek the difference between high-ability and low-ability test takers:

Table 7

Mann-Whitney U on high & low-ability performance on evaluation when completing Inference questions

	Score
Mann-Whitney U	403.000
Wilcoxon W	1223.000
Z	-3.841
Asymp.Sig (2-tailed)	.000

Table 7 revealed that the p value of the Mann-Whitney U test is .000 ($p < .05$) which means that the difference

between the high-ability and the low-ability groups' use of evaluation strategy was significant. To find out which group outperformed the other, the mean ranks of the high-ability and the low-ability were observed. Table 8 revealed the mean ranks:

Table 8

Mean ranks of high-ability & low-ability use of evaluation for Inference items

Group	N	Mean Rank	Sum of Ranks
high	40	50.41	2016.50
score low	40	30.59	1223.50
Total	80		

Table 8 indicated that the high-ability test takers' mean rank was more than that of the low-ability group. The conclusion was that the former used more evaluation strategies than the latter when completing the Inference items of INUEMA. Moreover, to find out the difference between the intermediate and the low-ability test takers, another Mann-Whitney U test was conducted. Table 9 shows the results:

Table 9

Mann-Whitney U on intermediate & low-ability performance on evaluation when completing Inference questions

	Score
Mann-Whitney U	421.500
Wilcoxon W	1241.500
Z	-3.668
Asymp.Sig (2-tailed)	.000

As Table 9 indicated, the p value of the Mann-Whitney U test is .000 ($p < .05$) which means that the difference between the intermediate and the low-ability groups' use of evaluation strategy was significant. To see which group employed more evaluation strategies for inference question, the mean ranks of the intermediate and the low-ability test takers were observed. Table 10 revealed the mean ranks:

Table 10

Mean ranks of intermediate & low-ability use of evaluation for Inference items

Group	N	Mean Rank	Sum of Ranks
intermediate	40	49.96	1998.50
score low	40	31.04	1241.50
Total	80		

Table 10 indicated that the intermediate test takers' mean rank was more than that of the low-ability group. The conclusion was that the former used more evaluation strategies than the latter when completing the Inference questions of INUEMA. Next section provides the results of Kruskal-Wallis test for the performance of the three ability groups on test-wiseness when completing Inference items of INUEMA.

Test-wiseness - This section provides the results of Kruskal-Wallis test for the performance of the three ability groups on test-wiseness when completing Inference items. Table 11 indicates the results:

Table 11

Kruskal-Wallis results for three ability performance on test-wiseness when completing Inference items

	Score
Chi-Square	27.358
Df	2
Asymp.Sig	.000

Table 11 showed, p value is .000 ($p < .05$); therefore, the difference among the three ability groups on the use of test-wiseness when completing Inference questions of INUEMA was significantly different. To see where this

significance occurred, Mann-Whitney U test was conducted on the related data. Table 12 presents the result of Mann-Whitney U test to seek the difference between high-ability and intermediate test takers:

Table 12

Mann-Whitney U on high & intermediate performance on test-wiseness when completing Inference questions

	Score
Mann-Whitney U	687.000
Wilcoxon W	1507.000
Z	-1.093
Asymp.Sig (2-tailed)	.274

As Table 12 indicated, the p value of the Mann-Whitney U test is .274 ($p > .05$) which means that the difference between the high-ability and the intermediate groups' use of test-wiseness strategy was not significant. Table 13 presents the result of Mann-Whitney U test to seek the difference between high-ability and low-ability test takers:

Table 13

Mann-Whitney U Test on high & low-ability performance on test-wiseness when completing Inference questions

	Score
Mann-Whitney U	268.000
Wilcoxon W	1088.000
Z	-5.156
Asymp.Sig (2-tailed)	.000

As Table 13 indicated, the p value of the Mann-Whitney U test is .000 ($p < .05$) which means that the difference between the high and the low-ability groups' use of test-wiseness strategy was significant. To see which group employed more test-wiseness strategies for inference question, the mean ranks of the high-ability and the low-ability test takers were observed. Table 14 revealed the mean ranks:

Table 14

Mean ranks of high & low-ability use of test-wiseness for Inference questions

Group	N	Mean Rank	Sum of Ranks
high	40	27.20	1088.00
low	40	53.80	2152.00
Total	80		

Table 14 indicated that the low-ability test takers' mean rank was more than that of the high-ability group. The conclusion was that the former used more test-wiseness strategies than the latter when completing the Inference items of INUEMA. Moreover, to find out the difference between the intermediate and the low-ability test takers, another Mann-Whitney U test was conducted. Table 15 shows the results:

Table 15

Mann-Whitney U on intermediate & low-ability performance on test-wiseness when completing Inference questions

	Score
Mann-Whitney U	423.500
Wilcoxon W	1243.500
Z	-3.648
Asymp.Sig (2-tailed)	.000

As Table 15 indicated, the p value of the Mann-Whitney U test is .000 ($p < .05$) which means that the difference between the intermediate and the low-ability groups' use of test-wiseness strategy was significant. To see which

group employed more test-wisness strategies for inference question, the mean ranks of the intermediate and the low-ability test takers were observed. Table 16 revealed the mean ranks:

Table 16

Mean ranks of intermediate & low-ability use of test-wisness for Inference questions

Group	N	Mean Rank	Sum of Ranks
intermediate	40	31.09	1243.50
score low	40	49.91	1996.50
Total	80		

Table 16 indicated that the low-ability test takers' mean rank was more than that of the intermediate group. The conclusion was that the former used more test-wisness strategies than the latter when completing the Inference items of INUEMA. Next section discusses the results of Kruskal-Wallis tests concerning the difference in the use of strategies by ability groups when completing factual Information

3.2 Factual Information

To investigate whether there was any significant difference in the types of test-taking strategies used by various ability group test takers when completing Factual Information items in INUEMA, a series of Kruskal-Wallis tests was conducted. When the difference between the three ability groups was significant, Mann-Whitney U test was conducted as a follow-up significance test to reveal where the difference occurred.

Planning - In order to find out if there was significant difference in the use of planning by the three ability group test takers when completing Factual Information items of INUEMA, a Kruskal-Wallis test was conducted. Table 17 shows the results:

Table 17

Kruskal-Wallis for three ability performance on planning when completing Factual Information items

	Score
Chi-Square	.807
Df	2
Asymp.Sig	.668

Table 17 indicated that the p value is .668 ($p > .05$); therefore the difference among the three ability groups on the use of planning when completing Factual Information questions was not significantly different. Next section provides the results of Kruskal-Wallis test for the performance of the three ability groups on monitoring when completing Factual Information items.

Monitoring - This section provides the results of Kruskal-Wallis test for the performance of the three ability groups on monitoring when completing Factual Information items. Table 18 indicates the results:

Table 18

Kruskal-Wallis for three ability performance on monitoring when completing Factual Information items

	Score
Chi-Square	8.628
Df	2
Asymp.Sig	.013

Table 18 showed, p value is .013 ($p < .05$); therefore, the difference among the three ability groups on the use of monitoring when completing Factual Information questions of INUEMA was significantly different. To see where this significance occurred, Mann-Whitney U test was conducted on the related data. Table 19 presents the result of Mann-Whitney U test to seek the difference between high-ability and intermediate test takers:

Table 19*Mann-Whitney U on high & intermediate performance on monitoring when completing Inference questions*

	Score
Mann-Whitney U	487.000
Wilcoxon W	1307.000
Z	-3.021
Asymp.Sig (2-tailed)	.003

As Table 19 indicated, the p value of the Mann-Whitney U test is .003 ($p < .05$) which means that the difference between the high-ability and the intermediate groups' use of monitoring strategy was significant. To see which group employed more monitoring strategy for Factual Information questions, the mean ranks of the high and the intermediate test takers were observed. Table 20 revealed the mean ranks:

Table 20*Mean ranks of high & intermediate use of monitoring for Factual Information questions*

Group	N	Mean Rank	Sum of Ranks
high	40	32.68	1307.00
intermediate	40	48.33	1933.00
Total	80		

Table 20 indicated that the intermediate test takers' mean rank was more than that of the high-ability group. The conclusion was that the former used more monitoring strategies than the latter when completing the Factual Information items of INUEMA. Moreover, to find out the difference between the high-ability and the low-ability test takers, another Mann-Whitney U test was conducted. Table 21 shows the results:

Table 21*Mann-Whitney U on high-ability & low-ability performance on monitoring when completing Factual Information questions*

	Score
Mann-Whitney U	746.000
Wilcoxon W	1566.000
Z	-.521
Asymp.Sig (2-tailed)	.602

As Table 21 indicated, the p value of the Mann-Whitney U test is .602 ($p > .05$) which means that the difference between the high-ability and the low-ability groups' use of monitoring strategy was not significant. Table 22 presents the result of Mann-Whitney U test to seek the difference between the intermediate and low-ability test takers:

Table 22*Mann-Whitney U on intermediate & low-ability performance on monitoring when completing Factual Information questions*

	Score
Mann-WhitneyU	603.000
Wilcoxon W	1423.000
Z	-1.903
Asymp.Sig (2-tailed)	.010

As Table 22 indicated, the p value of the Mann-Whitney U test is .057 ($p > .05$) which means that the difference between the intermediate and the low-ability groups' use of monitoring strategy was not significant. Section 3.2.3 provides the results of Kruskal-Wallis test for the performance of the three ability groups on evaluation when completing Factual Information items.

Evaluation - This section provides the results of Kruskal-Wallis test for the performance of the three ability groups on evaluation when completing Factual Information items. Table 23 indicates the results:

Table 23

Kruskal-Wallis for three ability performance on evaluation when completing Factual Information items

	Score
Chi-Square	1.731
df	2
Asymp.Sig	.421

Table 23 indicated that the p value is .421 ($p > .05$); therefore the difference among the three ability groups on the use of evaluation when completing Factual Information questions was not significantly different. This means that none of the ability groups (high, intermediate, or low) employed evaluation strategies significantly more or less than the others. Next section provides the results of Kruskal-Wallis test for the performance of the three ability groups on test-wiseness when completing Factual Information items.

Test-wiseness - This section discusses the results of Kruskal-Wallis test for the performance of the three ability groups on test-wiseness when completing Factual Information items. Table 24 indicates the results:

Table 24

Kruskal-Wallis for three ability performance on test-wiseness when completing Factual Information items

	Score
Chi-Square	1.407
df	2
Asymp.Sig	.495

As table 24 showed, the p value was .495 ($p > .05$); therefore the difference among the three ability groups on the use of test-wiseness when completing Factual Information questions was not significantly different. This means that none of the ability groups (high, intermediate, or low) employed test-wiseness strategies significantly more or less than the others for Factual Information questions.

4. Conclusion

The present study aimed at investigating any significant difference in the type of strategies used by three ability groups of test takers when completing Inference and Factual Information questions. The motivation for conducting research in such field came from the insight obtained 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, the difference among the three ability groups on the use of evaluation and test-wiseness for Inference items was significant. To see which ability group used more of these strategies than others, Mann-Whitney U test was conducted between each two groups; high/intermediate, high/low, and intermediate/low. The results showed that the high-ability and the intermediate groups employed more evaluation strategies than the low-ability test takers when completing Inference items of INUEMA. Moreover, low-ability test takers used test-wiseness strategies for Inference items more than the high-ability and the intermediate groups. With respect to the Factual Information items, the intermediate group used more monitoring strategies than the high-ability test takers.

The low proficiency group in the present study did not significantly utilize monitoring and evaluation 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.... (p. 62)

The frequent use of monitoring and evaluation 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)

In other words, as Jamil, Aziz, and Razak (2010) mention, this use of monitoring and evaluation 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 INUEMA 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). This is in line with Nevo's assert that strategies of guessing or any kind of non-contributory strategies cannot be validly used for items that are difficult in nature (1989).

4.1 Pedagogical Implications

Investigating what successful language learners do to promote their learning was the starting point for doing research on 'strategies', in general. In fact the idea was that the way successful language learners behave can be taught to unsuccessful learners and will help them enhance their learning ability (Rubin 1975). Further, it was believed 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.

In this respect, the present study investigated the test-taking strategies that respondents used for two different item specifications: Inference and Factual information questions. The fact that monitoring and evaluation were 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 high-ability and intermediate test takers on INUEMA which can prove the construct validity of it. 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, 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, the number and scope of the items in it might have been insufficient to investigate the real nature of test-taking strategies in EFL learners. Moreover, 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., and 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.
- 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. (2007). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307-331. <http://dx.doi.org/10.1080/15434300701333129>
- Cohen, A. D., & Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading tasks* [Monograph No. 33]. Princeton, NJ: ETS.
- 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.
- Fransson, A. (1984). Cramming or understanding? Effects of intrinsic and extrinsic motivation on approach to learning and test performance. In Alderson J. C. and Urganhart A. H. (eds.), *Reading in foreign language* (pp. 86-121). London: Longman.
- 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.
- Kashkoui, Z. & Barati, H., & Nejad Ansari, D. (2015). An investigation into the test-taking strategies employed for a high-stake test: Implications for test validation. *International Journal of Research Studies in Language Learning*, 4(3), 61-72. <http://dx.doi.org/10.5861/ijrsl.2014.852>
- Nevo, N. (1989). Test-taking strategies on a multiple-choice test of reading comprehension. *Language Testing*, 6(2), 199-215. <http://dx.doi.org/10.1177/026553228900600206>
- O'Malley, J., M. and Chamot, A., U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139524490>
- 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>
- 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>
- 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.

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 (never), 2 (sometimes), 3 (often), 4 (usually), and 5 (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.