New test-taking patterns and their effect on language test validity

Abdullah Al Fraidan[iD]

Correspondence: alfraidan@gmail.
com; afridan@kfu.edu.sa
Department of English Language,
King Faisal University, Hofuf, Saudi
Arabia

Abstract

Educators, especially test creators, are concerned with the construct validity of their tests. Blame has generally been attached to test-wiseness strategies as one source of error in measurement. There is little evidence of any relation between test-taking strategies in general and test validity; thus, it is not known how strategies can affect test validity. In addition, the literature does not report studies that have measured the effectiveness of the test-taking strategies used by learners. In this study, we propose to detail 12 test-taking patterns observed in data gathered from 42 EFL learners on two different vocabulary achievement tests, using think-aloud protocols and interviews. Although there were logical patterns, we also discovered some unusual patterns. In some cases, students managed to arrive at the right answers that were words they did not know using test-wiseness strategies, whereas other students got items wrong although they knew the correct word because of test-unwiseness strategies. In addition to showing how strategies affect test validity, the degree of effectiveness of some strategies was measured. Some factors were addressed that played a role in the effectiveness of certain strategies as well as factors causing the more unusual patterns.

Keywords: Test-taking strategies, Language testing, Test-wiseness, Test-unwiseness, Validity

Overview

Research into test-taking strategies is increasing, with such strategies eliciting special attention. Mainly, researchers are examining how such strategies affect test validity. Although Cohen (2009) notes that we still lack a theory accounting for test-taking strategies (TTSs), it is promising that TTS researchers appear to have reached some type of consensus regarding what TTSs actually are.

Grotjahn (1986) argued that the validation of language tests must include, in addition to the quantitative analysis of test takers' responses, a qualitative analysis of the test-taking process and of the test tasks themselves. Weir (2005) noted that tests can be validated on the basis of the literature as well as of empirical findings, whereas the issue of validity in conjunction with the effectiveness of strategies remains unexplored in the field of TTSs. The best manner in which to explore this issue is to conduct more qualitative in-depth analyses of recordings of the testees' verbalised reasoning, known as think-aloud protocols, and to observe testees during a test.

Previous studies have reported controversial results in different aspects of TTSs, one of which is TTSs' relation to language proficiency. Some studies show that highly proficient students are the greatest users of TTSs, and other studies show the opposite (see Al Fraidan, 2011 and Cohen, 2006, for more).

This paper analyses examples of qualitative data, i.e., verbal protocols, to explore the validity of each test. We are interested in whether the test has accessed the lexical knowledge of a testee or some other type of knowledge, and this is performed by conducting an in-depth analysis of the strategies used by the testee to address the items selected for this qualitative analysis. Occasionally, students may attempt to use tricks or attack the construct being assessed without possessing the necessary linguistic knowledge using what is called test-wiseness (TW) strategies (Cohen, 2009). These strategies are being blamed as a main source of tests rendering invalid. There is another type of strategy, or group of strategies, that also contribute to score invalidity, namely test-unwiseness strategies.

Internal measures of construct validity have become more sophisticated as a means of addressing both test-taking-strategy concerns and other potential issues with regard to language test validity; however, these have not been entirely successful in guarding against the erosion of test validity when faced with test-taking strategies (Goh & Aryadoust, 2010; Lee, 2011; Nikolov, 2006). Test-taking strategies such as clue-word orientation, multiple choice elimination strategies and a variety of other test-taking strategies that are not concerned with ensuring proper knowledge representation but rather with "fooling" or "gaming" the test itself are all but impossible to control for (Lee, 2011; Mohammadi & Abidin, 2012). Paying attention to elements of the test design, from the test type to the duration of the test to the arrangement of specific items and the wording of specific questions, is the only manner with which to create greater validity, and there is no design that could ever hope to achieve a validity so high as to be truly impervious to test-taking strategies (Mohammadi & Abidin, 2012).

This does not indicate, of course, that greater levels of validity in language test design should not be sought or that language tests as a whole must be discarded as valid and accurate measures of language proficiency. Instead, on-going efforts in test design and in educational design—from instruction to testing patterns—must continue to be investigated, adapted and adjusted as necessary to attain the highest possible level of validity. Continuing to adjust all of these elements in response to changing strategies will ensure long-term validity.

There are different sources of invalidity according to Scholfield (1996), ranging from institutional factors (e.g. education policies and restrictions on test makers) to test-taker characteristics (e.g., strategic competence) and test design (e.g., test maker, test formats, test scorers). In this paper, we identify strategies and their relation to test formats and how these strategies affect test validity.

Bachman and Palmer's (1996) framework of communicative testing asserts that students' strategic competence is a key component of the framework. Planning, selection, execution and evaluation of the use of strategies in a testing situation are all important steps in taking tests. Testees go through a cycle of planning and deciding which strategies to use, implementing them and then evaluating the effectiveness of their strategies by judging whether a strategy helped them find an answer or by restarting the cycle and selecting another strategy. This process is called metacognition (see Purpura, 1997, for more). However, research on TTSs has not addressed the issue of how test makers or researchers can assess the effectiveness of strategy use. It is one aim of the current paper to shed some light on how strategies are sometimes used effectively and sometimes used ineffectively. This paper thus seeks to initiate the investigation of a rarely addressed issue, measuring the effectiveness of the use of strategies and the success of their deployment. The effectiveness of strategy use is always determined by the context in which the strategies are used.

The researcher identified some patterns of English as a foreign language (EFL) in Saudi learners while the students were taking two vocabulary achievement tests. These patterns provide a better understanding of how strategies may affect test measurement by showing how strategies interact with test validity.

Methodology

Forty-two male and female EFL Saudi English majors from different study levels participated in this study. They were trained in how to use think-aloud protocols in several training sessions (e.g. matching tasks and crosswords) that lasted 45 min each, after which they were given two vocabulary achievement tests. These tests were teacher-made tests with some flaws; however, the flaws were overlooked to simulate real-life situations because EFL learners are frequently exposed to flawed teacher-made tests in EFL contexts. The first test was a text-based cloze test about "mosquito", a topic with which they were familiar. The test asked the participants to provide an appropriate word for 16 gaps without providing the students with any choices. The text had been studied in the previous semester in a reading course. It appeared that the test maker had used a rational deletion method.

The second test was a multiple choice gap filling (MCGF) test. The test had 16 disconnected sentences with one gap in each sentence and a pool of 21 alternatives presented above the sentences from which students chose the correct answer. According to Al Fraidan's (2011) classification of test types, these two test types are the most frequently used in Saudi Arabia. We chose to use tests that students are frequently exposed to so that we would generate real and natural strategies rather than imposing something new on them that may have generated artificial data. The selection of the two tests from a corpus of vocabulary tests gathered by Al Fraidan (2011) was completely random. Some applied-

linguistics experts had checked the two tests and confirmed their appropriateness for the participants' levels.

Students verbalised their thoughts during both tests; each student had a different test to safeguard against order effect. Interviews were conducted following the tests. The tests took 1 h to complete. The students' vocabulary level was determined by Nation's Vocabulary tests, after which students were classified into better and poorer students. The term "better" was used because no students achieved truly high scores, although their scores were significantly better than those of the poorer students.

Analysis

A detailed transcription of the verbal protocols was performed over four consecutive months. Segments were identified and then assigned a code. The analysis was checked by a second coder (without knowledge of the results of the first coder's analysis), and there was 89% agreement between the two coders, which is a prerequisite to confirming the validity of the coding (Green, 1998). The tests were marked by the researcher, and the cloze test was marked by another scorer, using the acceptable word method and Al Fraidan's (2011) scoring scheme. Some applied linguist experts, including Philip Scholfield, checked the analysis, the coding and the following framework and confirmed the validity and likelihood of the consistency of the framework. This was performed to minimise analysis subjectivity.

Framework of the analysis

A painstaking analysis was conducted on some examples from the think-aloud protocols, supported by interviews with the testees. The following framework was modified several times until it reached a stage that showed that the study data fit quite well.

The analysis was of 96 randomly selected test items (18% of all subtest items), with three items selected from each subtest administered to the 16 subjects.

The data were gathered from the think-aloud protocols, the follow-up interviews and the actual test manuscripts. In a few cases, the information in the actual test manuscripts does not reflect all of the information reported in the verbal protocols. Cohen (1984, p. 72) noted that if verbal protocols are used and respondents are requested to write down some of their thoughts, there will invariably be some loss of data because the student will have stopped verbalising while writing on the test paper. Such cases were excluded because we have no evidence for the reason for the changed answer. Data from the interviews were only added to the examples provided in the "Results" section when they provided more insight into the process of the testee's selecting an answer.

The analysis examined in detail all of the strategies used, describing their role in the testee's reaching an answer. However, in measuring the effectiveness of strategy use, we decided to examine the final strategy used because this strategy is presumed to make the main contribution to the choosing of an answer. Because a subject in any given test is judged by the product, examining the final strategy used to arrive at that product is a sensible manner in which to measure how that product was affected by the use of that strategy. In Tables 1 and 2 below, we will only mention the final strategy or sequence of strategies used. In addition, instances of automatic test processing were avoided as much as possible.

Table 1 Examples of number of passes, strategies and types of pattern used by the students in the cloze test

The Cloze Test				Pass 1		Pass 2		Pass 3	
Subjects		Item	Pattern	No. Strategies	Answer	No. Strategies	Answer	No. Strategies	Answer
Better	MS1	3	2	1	X	1	X	-	-
		7	1	1	X	2	√	-	-
		15	2	1	X	-	-	-	-
	MS2	3	8	3	√	-	-	-	-
		7	9	2	√	-	-	-	-
		11	9	2	√	-	-	-	-
	MS6	2	6	3	X	-	-	-	-
		7	9	2	√	-	-	-	-
		12	5	2	X	2	-	-	-
	MS8	1	10	3	X	-	-	-	-
		6	9	2	√	-	-	-	-
		15	2	2	X	-	-	-	-
	FS1	1	9	2	√	1	-	-	-
		6	9	2	√	1	-	-	-
		15	1	2	X	1	-	-	-
	FS2	3	5	1	X	-	-	-	-
		6	5 & 9	1	X	2	√	-	-
		11	5	1	X	-	-	-	-
	FS6	1	9	2	√	-	-	-	-
		9	9	3	√	-	-	-	-
		14	8	3	√	-	-	-	-
	FS8	6	8	3	√	-	-	-	-
		10	3	3	X	-	-	-	-
		15	3	3	X	-	-	-	-
Poor	MS3	1	9	2	√	-	-	-	-
		8	2	2	X	-	-	-	-
		12	8	2	√	-	-	-	-
	MS4	1	9	2	√	-	-	-	-
		7	9	2	√	-	-	-	
		10	5	2	√	-	-	-	-
	MS5	2	5	2	X	-	-	-	-
		7	5	2	X	-	-	-	-
		13	5	2	X	-	-	-	-
	MS7	4	5	1	X	-	-	-	-
		8	5	2	X	-	-	-	-
		13	5	2	X	-	-	-	-
	FS3	1	6	3	X	-	-	-	-
		5	5	1	X	-	-	-	-
		12	5	1	X	-	-	-	-
	FS4	5	10	1	√	-	-	-	-
		11	1	2	√	2	-		
		14	2	2	X	-	-	-	-
	FS5	1	5	1	X	-	-	-	-
		9	5	1	X	-	-	-	-
		13	5	1	X	-	-	-	-
	FS7	1	4	2	X	-	-	-	-
		8	5	1	X	-	-	-	-
		14	8	3	√	-	-	-	-

Table 2 Examples of number passes, strategies and types of patterns use by the students in the cloze test

MCGF		Item	Pattern	Pass 1		Pass 2		Pass 3	
Subjects				Strategies	Answer	Strategies	Answer	Strategies	Answer
Better	MS1	4	1	1	-	1	√	-	-
		8	2	1	-	2	X	-	-
		16	4	2	X	-	-	-	-
	MS2	1	3	2	-	1	X	-	-
		4	12	1	-	2	√	2	X
		8	8	3	-	1	√	-	-
	MS6	1	9	2	-	2	√	-	-
		5	1	3	X	-	-	-	=
		16	4	2	X	-	-	-	-
	MS8	1	6	-	-	2	-	4	X
		8	9	2	√	-	-	-	-
		15	7	3	-	2	√	-	-
	FS1	1	9	2	√	1	-	-	-
		8	5	1	X	1	-	-	-
		16	9	-	-	2	√	-	-
	FS2	2	11	3	X	-	-	-	-
		4	6	1	X	-	-	-	-
		14	9	2	√	-	-	-	-
	FS6	2	12	2	X	2	X	-	-
		3	12	2	√	1	X	-	-
		9	3	2	-	2	-	1	X
	FS8	1	3	-	-	1	X	-	-
		8	4	2	X	-	-	-	-
		16	9	2	√	-	-	-	-
Poor	MS3	2	9	2	√	-	-	-	-
		4	5	1	X	2	-	1	-
		11	9	2	√	-	-	-	-
	MS4	1	5	3	X	-	-	-	-
		2	5	1	X	-	-	-	-
		-	-	-	-	-	-	-	-
	MS5	5	4	2	X	-	-	-	-
		9	5	1	X	-	-	-	-
		15	5	1	X	-	-	-	-
	MS7	2	5	3	X	-	-	-	-
		8	11	1	X	-	-	-	-
		10	10	2	X	-	-	-	-
	FS3	1	5	1	X	-	-	-	-
		2	4	3	X	-	-	-	-
		3	5	1	X	-	-	-	-
	FS4	1	3	3	X	-	-	-	-
		6	3	3	X	-	-	-	-
		11	5	-	-	-	-	1	X
	FS5	2	5	1	X	-	-	-	-
		8	5	1	X	-	-	-	-
		16	7	1	X	-	-	-	-
	FS7	2	9	3	√			-	-
		8	5	-	-	1	X	-	-
		16	3	-	-	-	-	1	X

The analysis follows a set of categorisations, which identifies first whether the testee knows the test word. Knowing a word here requires a testee to identify the correct meaning and the correct form of the gapped word. In some cases, the testee has partial knowledge of the test word, knowing the meaning of the gapped word but not its proper form. The degree of test word knowledge is determined by what the testees have verbalised during the think-aloud protocol or in their answers in the interviews.

Following the framework, the analysis then examines which strategies were used to arrive at the tested word in addition to what problems and reasons caused the final "product" to be correct or incorrect. The TTS literature agrees that the effectiveness of test strategies is entirely dependent on their proper use in certain situations.

The analysis also examines the number of "passes" required to answer each test item. A *pass* is defined as a phase in tackling a test item, one or more attempts made before answering or giving up and skipping it. The analysis identified up to three passes because this was the maximum number of times an item had been addressed in the data.

The analysis should, therefore, reveal which items present a real indication of the subject's lexical knowledge. In cases in which a testee knows the test word and uses legitimate strategies to answer correctly or incorrectly, the test item is considered to be valid, and the testee achieves the score he deserves. The criterion of legitimacy is that the strategies should adhere to the tested construct and not cause any harm to it. Test makers expect test takers to use legitimate strategies without using any tricks or workarounds. Conversely, using mainly TW strategies will render the test item invalid.

The patterns below were described mainly from the interviews, the researcher's and the applied linguist's judgments.

Because of space limitations, the analysis will provide only a few examples from the verbal protocols, with a comment describing them.

Results

The painstaking analysis of the 96 selected test items from both tests and from all subjects revealed that students used 12 different patterns to answer these test items. The logical patterns indicate that students use strategies whether they know the gapped word or not; however, there is a difference in the outcomes. If students know the gapped word, logic dictates that the answer is going to be correct; and if it is not correct, then the answer has quite likely been guessed at and would normally be incorrect. However, we observed that there is a combined use of legitimate strategies and TW strategies, referred to below by the plus (+) sign in the description of patterns. The plus signs suggest that TW strategies were mainly used to select the answer. In other cases, TW strategies were used to support and confirm the selection of the answer. In these cases, an arrow (→) is used to indicate this. More notably, we observed patterns in which the subject knew the answer but then changed it for one reason or another to an incorrect answer by using a test-unwiseness (TUW) strategy. A TUW strategy is a strategy that loses the testee a deserved mark that he/she would have obtained without

using such a strategy. An example of this is changing a correct answer to an incorrect answer for the wrong reason (for further detail, see Al Fraidan (2014). However, this only applies when the gapped word is known. TUW strategies cannot be applied when the word is partially known because being able to retrieve a meaning for the gap but failing to provide a form in itself indicates a lack of knowledge. However, when a word is known and the testee writes the answer in the gap and then changes it for one reason or another, that would be a TUW strategy. Examples of this are provided in the account of pattern 12. Another interesting pattern is when a subject chooses the wrong answer but in the interview after the test admits that he/she knew the correct answer, as shown in pattern 11. This is coded with a forward slash (/) to indicate that although the answer is incorrect, the testee appears to have known the correct answer. This may have been caused by a flaw in the test item, or the testee was in a hurry when answering, although he/she had enough time. The 12 patterns are presented and described in detail below.

1. Partial knowledge of the gapped word + legitimate strategies + TW strategies + correct = invalid
2. Partial knowledge of the gapped word + legitimate strategies + incorrect = valid
3. Partial knowledge of the gapped word + legitimate strategies + TW strategies + incorrect = valid
4. Unknown gapped word + legitimate strategies + incorrect = valid
5. Unknown gapped word + TW strategies + incorrect = valid
6. Unknown gapped word + legitimate strategies + TW + incorrect = valid
7. Unknown gapped word + TW strategies + correct = invalid
8. Known gapped word + legitimate →TW strategies + correct = valid
9. Known gapped word + legitimate strategies + correct = valid
10. Known gapped word + legitimate strategies + incorrect = invalid
11. Known gapped word + legitimate strategies + correct/incorrect = ?
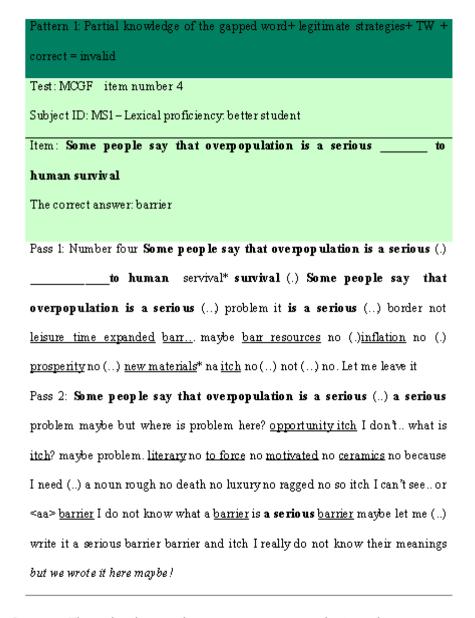12. Known gapped word + legitimate strategies + correct + TUW + incorrect = invalid

A full description of each pattern is presented below with an example followed by a comment. To provide the reader with a full picture, the test type, item number, full item stem, subject ID and his/her proficiency level are provided along with an example from the verbal protocol followed by a comment describing the pattern. The following transcription conventions were used:

Bold the test text

Plain	Students' English verbalisation
Italics	*Students' Arabic verbalisation*
(.)	Pause of less than two seconds
(…)	Pause of longer than two seconds
*	Mispronunciation
Underlined	Attempted or final answer
MS	Male student
FS	Female student

Pattern 1

Pattern 1: Partial knowledge of the gapped word+ legitimate strategies+ TW + correct = invalid
Test: MCGF   item number 4 Subject ID: MS1 – Lexical proficiency: better student
Item: **Some people say that overpopulation is a serious _____ to human survival** The correct answer: barrier

Pass 1: Number four **Some people say that overpopulation is a serious** (.) **_____to human** servival* **survival** (.) **Some people say that overpopulation is a serious** (..) problem it **is a serious** (...) border not leisure time expanded barr... maybe barr resources no (.)inflation no (.) prosperity no (..) new materials* na itch no (..) not (..) no. Let me leave it

Pass 2: **Some people say that overpopulation is a serious** (..) **a serious** problem maybe but where is problem here? opportunity itch I don't.. what is itch? maybe problem. literary no to force no motivated no ceramics no because I need (..) a noun rough no death no luxury no ragged no so itch I can't see.. or <aa> barrier I do not know what a barrier is **a serious** barrier maybe let me (..) write it a serious barrier barrier and itch I really do not know their meanings *but we wrote it here maybe!*

Comment: This male subject made two attempts to answer this item, shown as pass 1 and pass 2 above. On the first attempt, he read the stem for the first time, misreading the last word. It appears that the misreading did not affect his comprehension of the sentence because it appears that he showed he knew a portion of the answer by voicing a second language (L_2) meaning "problem" for the gap IG27b, which is a legitimate strategy; however, he failed in his two attempts to link that meaning to one of the alternatives. He read some of the alternatives and rejected them without giving a reason until he decided to skip it. On his second pass, he applied a sequence of strategies identical to the sequence he used on his first pass, showing a likely preference for a particular sequence of strategies. However, when he wanted to link the meaning he had discerned to a word from this list, he questioned the meaning of one of the alternatives, "itch". He continued searching through and rejecting some of the alternatives. He rejected one of the alternatives because it was not the correct part of speech; he was looking

for a noun. Because, however, he did not verbalize the rejected alternative, we can infer that he had previously worked out that the part of speech of the gapped word was a noun. He continued rejecting some alternatives, most likely on the basis of meaning and part of speech. Eventually, he appeared unable to find an answer, which may have made him lose hope; thus, he then relied mainly on the most commonly used TW strategy, blind guessing. He hit upon the word "barrier", whose meaning he did not know at all, as he verbalized. Notably, he chose a correct answer. Because he did not have full knowledge of the test word, was unable to retrieve the correct form and used blind guessing as his main strategy to select an answer, the score validity of this item is compromised. The overall construct validity of this item is valid because it prompted the use of legitimate strategies related to the meaning of the gapped word and some portion of the lexical item such as its part of speech. However, because those strategies did not help much, he chose to guess; the score is not indicative of the subject's knowledge of the word "barrier". The poor construction of the sentence may have led the student to choose the correct answer, which shows the importance of test makers' writing their test items carefully. The use of TW strategies here partially contributes to the answer, thus invalidating the test score of this item.

Pattern 2

Pattern 2: Partial knowledge of the gapped word +legitimate strategies + incorrect = valid

Test: MCGF   item number 9

Subject ID: FS6 – Lexical proficiency: better student

Item: **The substantial and continuous rise in the price of goods is referred to _____**

The correct answer: inflation

Pass 1: **The substantial and continuous rise in the price of goods is referred to** (.) **The substantial and continuous rise in the price** *increase in the price constant increase* **of goods is referred** *for the reason* **referred to** (..) *literacy maybe civilization culture maybe increases the price* promotion raw materials raw materials! Raw materials *because of them the price will increase because of the expensive price of raw materials basically* **referred to** (.) **The substantial and continuous rise in the price of goods is referred to** (.) **as** (.) raw materials (..) raw materials

Comment: This female subject showed that she understood the stem because she translated a portion of it into first language (L_1) and uttered a correct L_1 meaning for the gap, which shows her partial knowledge of the gapped word. Unfortunately, she

misinterpreted the last word in the stem as "as" although it meant "because". This is shown by her L_1 translation "*for the reason*". This may have caused her not to implement other strategies that may have eventually led her to the correct answer. Apparently, she had another problem in guessing the meaning of the alternative, "literacy". She suspected that "literacy" meant "civilisation" or "culture" as sources for inflation. She also misinterpreted what was required in the gap without being aware of this. She thought the gap required a reason for the increase in prices. She did not realise that the item required a term that was the definition. Subsequently, she was looking for words that may have been the cause of the price increase. She finally selected "raw materials" from among other alternatives she read. The selection was based on her misinterpretation that this word could be a cause of the price increase. Unfortunately, because she failed to select the correct word, the item score is valid because the student's incorrect answer reflects her lack of knowledge of the word "inflation". The overall construct validity of this item is valid because the student used legitimate strategies to arrive at her answer regardless of the quality of the strategies used.

Pattern 3, example 1

Pattern 3: Partial knowledge of the gapped word +legitimate strategies + TW + incorrect =valid
Test: MCGF   item number 3
Subject ID: FS2 – Lexical proficiency: better student
Item: **How can we _____ the spread of air pollution?**
The correct answer: stop

Pass 1: *Third item* **How can we_____ the spread of air pollution?** **air pollution** *we are trying to prevent* air pollution *so it is likely* *something like prevention* (.) *prevention* (..) *to force* maybe *to force delicate*(...) maybe *to force it's the closest word* <sound of pen writing>

Comment: The subject processed the item in L_1, translating a large portion of the stem and voicing the correct meaning in L_1. She suspected two candidates, "force" and "delicate", and decided to use a TW strategy, which she thought was an approximation of the answer, ending up choosing the word "force". She had most likely acquired or learned the word "force" as meaning "to make (someone) do something against their will",[1] which shows that the subject did not know the correct meaning of "force", nor was she able to match the meaning she had figured out to the correct word from among the alternatives. The essence of this strategy is educated guessing, which is mainly used when the subject is not entirely sure of the answer. In this study, educated guessing is considered a TW strategy. Thus, the subject's score for this item is a valid

indication of her particular knowledge of the gapped word because it shows that she did not learn the words "force" or "slow", which would have been the correct answer. The overall construct validity of this item is valid. However, the correct answer for this item comes in the form "to slow". The preceding "to" caused some problems for some subjects but not for the current subject. Several subjects ruled this alternative out, and a few others changed their answers to something else, largely because of the presence of "to", as we will see later in pattern 12. Therefore, we recommend either including the verb form without "to" or ensuring that all students understand this convention, including items such as this to increase the validity for such items. Another interesting example of this pattern is the next one

Pattern 3, example 2

Pattern 3: Partial knowledge of the gapped word+ legitimate strategies + TW + incorrect =valid

Test: MCGF   item number 6

Subject ID: FS4 – Lexical proficiency, better student

Item: **The poor children were embarrassed to come to school in_____ clothes**

The correct answer: ragged

Pass1: **The poor children were embarrassed to come to school in** (..) _____ **to come to school The poor children were embarrassed to come to school in** (..) delicate rough death luxury ragged slow literacy opportunity (..)

Pass 2: **The poor children were embarrassed to come to school in** (..) border leisure time expanded (..) **The poor children The poor children were embarrassed to come to school in** (..) delicate <sound of pen writing>

➢ Interview: *I knew this sentence that it's about bad clothing but I did not know the word so I put delicate*

Comment: This female subject's answer during the interview shows that she knew the correct meaning for the gap, showing her partial knowledge of the test word; however, she was not able, on her first pass, to find a word for that meaning when she read some of the alternatives. On her second pass, she did the same but did not recognise the word "ragged" and decided to blindly guess the word "delicate". A similar conclusion to that presented in example 1 can be drawn here as well.
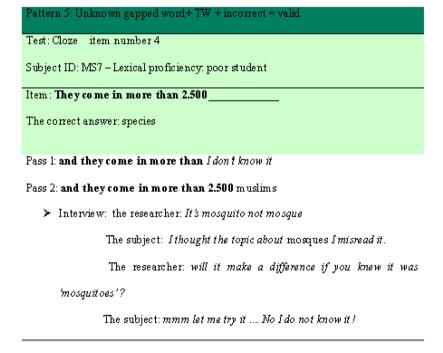
Pattern 4

Pattern 4: Unknown gapped word + legitimate strategies + incorrect = valid
Test: MCGF   item number 5
Subject ID: MS5 – Lexical proficiency: poor student
Item: **Can a society _____ people to use birth control methods?** The correct answer: force

Pass 1: *Item number five* **Can a society Can a society** (..) <u>death</u> *the blank the answer is* death death d e a t h_____ **people to use birth-control methods?**

> Interview: *I chose the word 'death' because there is the word 'birth'*

Comment: This male student, as he noted in the interview, relied on selecting the word "death" because it has a pragmatic meaning in relation to the word "birth" in the stem. This appears to be a good strategy; however, its execution was inappropriate in this context. The gapped word was completely unknown to the subject; thus, he tried to use a strategy to enable him to arrive at an answer. However, his answer was incorrect, reflecting his lack of knowledge of the word "force", which is the correct answer.

Pattern 5, example 1

Pattern 5: Unknown gapped word + TW + incorrect = valid
Test: Cloze   item number 4
Subject ID: MS7 – Lexical proficiency: poor student
Item: **They come in more than 2.500_____** The correct answer: species

Pass 1: **and they come in more than** *I don't know it*

Pass 2: **and they come in more than 2.500** muslims

> Interview: the researcher: *It's mosquito not mosque*

   The subject: *I thought the topic about* mosques *I misread it.*

   The researcher: *will it make a difference if you knew it was* 'mosquitoes'?

   The subject: *mmm let me try it .... No I do not know it!*

Comment: This example shows how poor students have problems not only addressing the gap but with other things on the test, which can be perceived as main causes of answering incorrectly. The male subject had a problem with the topic itself. He thought it concerned "mosques" and that

the gap concerned the number of "Muslims" who had come to a certain mosque. He came up with a word by blind guessing and the answer was, by any measure, incorrect. In the interview, the subject discussed his inability to find an answer for the gap even if had he known what the topic was. This is a clear indication that the student did not have the target word in his vocabulary.

On other test items, and especially on the cloze test, the same subject used words from the text itself to fill in some gaps. The cloze test had several gaps that could be filled by words from the text, although this did not apply to all of the gaps, and this may have led the subject to over-use this strategy. The following is an example of this behaviour from one of the poorer subjects.

Pattern 5, example 2

Pattern 5: Unknown gapped word+ TW + incorrect = valid
Test: Cloze   item number 7
Subject ID: MS5 – Lexical proficiency: poor student
Item: **Unfortunately the mosquito may decide that she loves you. She? Yes she. Did you know that only that the _____ mosquito bites?**   The correct answer: female
Pass 1: *I do not know the meaning of unfortunately* **mosquito may decide that she loves you. She? Yes, she. Did you know that only the that only the** (..) <mm>*I think it is in the beginning of the previous sentence above* **But unfortunately** *the answer is* unfortunatelyu n u n f o r t u n a t e l y

Comment: Similar to pattern 5, example 2, this student also used a word from the text, "unfortunately". He chose a word whose meaning he was having a problem with, applying the identical strategy of blind guessing, which indicates that sometimes a word is chosen because it is unknown to the testee. His guessing was based on reliance on the text. This is shown by his verbalisation that the answer may be in the previous sentence.

Pattern 6

Pattern 6: Unknown gapped word+ legitimate strategies+TW + incorrect = valid
Test: Cloze   item number 1
Subject ID: FS3 – Lexical proficiency: poor student
Item: **No matter who you are or_____ you come from.**   The correct answer: where
Pass 1: **No matter who you are** (..) *I am thinking of the best word* (..) *how* (..) *I chose it because of who you are*   ➤ Interview: *I thought I would reach the answer from 'who you are' so I guessed it was 'how'*

Comment: The female subject tried a good strategy, using contextual clues by relating the answer to a portion of the stem and inferring the meaning from the context; however,

then she could not find the correct word and relied on complete guessing. The item is a valid indication of the student's knowledge.

Pattern 7

Pattern 6: Unknown gapped word + TW + correct= invalid
Test: MCGF   item number 16
Subject ID: FS5 – Lexical proficiency: poor student
Item: **If you touch poison ivy, you will get a terrible _____** The correct answer: itch
Pass 1: **If you touch poison ivy, you will get a terrible** <u>itch</u> ➢ Interview: *I do not know the meaning of itch. All my answers in the MCQ were randomly guessed it always works for me in all subjects and I always get high marks*

Comment: The female subject got a score for answering with a word that she did not know at all. It appears as if she did not bother to tax her knowledge resources and strategies. She used her habit of randomly guessing and apparently her luck brought her more points. This example also shows how TW strategies sometimes help students to select a correct answer and thereby get an unfair mark.

Pattern 8

Pattern 8: Known gapped word +legitimate ? TW + correct=valid
Test: Cloze   item number 3
Subject ID: MS2 – Lexical proficiency: better student
Item: **They can be _____ all over the world** The correct answer: found, seen
Pass 1: **They can be** (.) _____ **all over the world** <aha> so (.) maybe it can be seen or (..) so we have a modal an auxiliary verb and they exist all over the word so we have to use a verb that can be used with can and be, **can be** <u>seen</u> (..)

Comment: The male subject was successful in automatically retrieving a word for the gap from his mental lexicon, and he retrieved the correct form as well on his first attempt. He gave an L_2 meaning, "exist", to confirm his answer. However, he sought help from his L_2 mental grammar to double-check his answer, which is considered a test-wise strategy because he relied on knowledge other than lexical knowledge. (However, note that Alderson and Kremmel ([2013](#)) discuss the (im) possibility of separating grammar from vocabulary

knowledge.) He then used another checking strategy, inserting the retrieved word into the gap and trying it in a portion of the context after the gap. However, the use of the TW strategy did not lead him to the main source for the answer because it was only used for checking and confirmation, which renders the test score of this item valid.

Pattern 9

Pattern 9: Known gapped word +legitimate strategies + correct = valid

Test: Cloze   item number 7

Subject ID: MS2 – Lexical proficiency: better student

Item: **Unfortunately the mosquito may decide that she loves you. She? Yes she. Did you know that only that the _____ mosquito bites?**

The correct answer: female

Pass 1: **Did you know that only the** (.) we have she here so it might be a question of gender so it's **the** <u>female</u> **mosquito** (..) right!

Comment: This pattern is considered the most valid pattern because test makers always try to encourage students to use construct-relevant strategies rather than doing something else. The male subject inferred the gapped word by using the context clue "she", which helped him arrive at the understanding that the gap was for one of the gender types. Thus, his answer was based on the likelihood of a pragmatic meaning in relation to a word in the stem, i.e. "she". The outcome of his attempt was successful, producing a valid score.

Pattern 10, example 1

Pattern 10: Known gapped word +legitimate strategies + incorrect= invalid

Test: MCQ   item number 10

Subject ID: MS7 – Lexical proficiency: poor student

Item: **_____ is the ability to read and write.**

The correct answer: Literacy

Pass 1: <u>opportunity</u> **is the ability to read and write**

➤   Interview: the subject: *I have selected this because it means* the ability to read and write.   Lit... Literacy.

The researcher: *but you have chosen 'opportunity'*

The student: *Oh I did not notice I was on hurry. We have just taken this word in* course "Listening II". Literacy... Literacy

Comment: The male subject appeared to process this item automatically because his protocol did not reveal any strategy; he had learned this word in another course and knew it quite well. What caused the error was that he was in a hurry and most likely mixed up gaps when he was quickly filling in the other gaps, mostly performed by blind guessing as he admitted during the interview. He did not address the items in order, jumping from one item to another, and this may have caused his confusion. It is also possible that he remembered that the correct answer was "literacy" after finishing the test. This is an example of a TUW strategy. Another example of how students lose points when they know an answer in this pattern is presented below.

Pattern 10, example 2

Pattern 10: Known gapped word +legitimate strategies + incorrect= invalid
Test: Cloze   item number 5
Subject ID: FS4 – Lexical proficiency. better student
Item : At some time, you have surely _____ at least one.
The correct answer: seen
Pass 1: **at some time, you have surely** (..) _____ **you have surely at least one. at some time, you have surely** <aa> **you have surely** find **at least you have surely** see (..) **at some time you have surely** see saw **you have surely** saw **at least one.**

Comment: The female student retrieved two competitive candidates for this gap, "find" and "saw", and then decided to select the word "see". She tried to choose the correct form, tense in this case, but failed to get it right. It appears that she was not struggling with grammar here because she knew that she should choose another form of "see" after "have". Instead, her problem mainly concerned the form of the word. According to the scoring scheme for the cloze test, this item earned half a point out of two.

The student's score on this and the previous example does not represent her real knowledge of the tested words.

Pattern 11

Pattern 11: Known gapped word +legitimate strategies + correct/incorrect=?
Test: MCQ   item number 2
Subject ID: FS4 – Lexical proficiency. better student
Item: **That company has the _____ to produce clothing quickly and efficiently**
The correct answer: resources
Pass 1: **That company has the _____ to produce clothing quickly and efficiently.** raw materials <sound of pen writing> *I think it is raw* materials

Comment: As mentioned above, this pattern addresses a subject's giving an answer that appears correct and fits in the context; however, there is something preventing it from being correct. In this example, the correct answer is "resources" Although "raw materials" would also fit perfectly, it cannot be considered correct because the test specifications allow an alternative to be used only once and "raw materials" fits another gap according to the answer key. The reason for reporting this pattern is that there are quite a number of instances of this, mainly linked to this item. A few subjects had given the identical answer, "raw materials". The students who chose this answer most likely did not notice that it fit another gap. This suggests a possible flaw in the test that affected the students' scores. We do not suspect the construct validity of the item because most subjects used inferencing from the context or related the answer to the words "produce" and "clothing" in the stem. However, its appearance with another competitive item, item 12, renders the test more challenging (Cohen, 2009, PC), although in terms of score validity, it is difficult to decide whether to consider it valid:

12- Materials that can be manufactured or prepared to be made more useful are referred to as_____

From one point of view, it is unfair to say that the student does not know the word "raw materials"; however, from another point of view, he/she was only able to recognise one possible context for it and failed to use it in the other context for which it was designed by the test maker. This problem with context could be an indication of the invalidity of this item and the test score.

Pattern 12, example 1

Because this pattern is the most interesting one, we will discuss all of the examples that appeared in the data.

Pattern 12: Known gapped word +legitimate strategies + correct + TUW +incorrect= Invalid

Test: MCQ   item number 4

Subject ID: MS2 – Lexical proficiency: better student

Item: **Some people say that overpopulation is a serious _____ for human survival.**

The correct answer: barrier

Pass 1: **Some people say that overpopulation is a serious** (..) offense? danger? (.) promotion opportunity itch (..) barrier no no motivated (..) luxury ragged promotion raw…opportunity (..) **overpopulation is** danger so (..) **a serious** (..) literacy to force (..) delicate no rough no luxury ragged no (..) promotion no prosperity Doctor *Thinking and talking conflict with each other I suggest you report this in your research* Ok! **Some people say that overpopulation is a serious** (.) motivated ceramics delicate rough death luxury ragged force SKIPPED THE WORD SLOW MAYBE BECAUSE HE ALREADY USED IT itch opportunity (..) promotion border I'll skip this one

Pass 2: Number four **Some people say that overpopulation is a serious** barrier (..) luxury force itch opportunity promotion border (.) inflation (..) here's only one (.) **serious** luxury **serious** ragged **serious** (.) barrier **serious** (.) **serious** death **serious** barrier HE WROTE IT

Pass 3: **Some people (.) say that** (..)probably death? death and survival (..) **a**

Comment: This example is taken from one of the better male subjects, the excellent student who among all of the students answered most often correctly, and when he was scrutinising his answers, he changed his answer from right to wrong. He did quite well in the beginning because he used effective strategies. He started by automatically voicing a couple of L_2 meanings for the gap and then searched for a word from the list of alternatives to match it with the meanings he had inferred. He read a key word from the stem preceding the gap with the L_2 meaning he had figured out. He continued searching the alternatives, and we can deduce that he ruled out the previously used words "slow" and "barrier", which he used to fill other gaps. He rejected "barrier" with a double "no" and skipped verbalising "slow". This can be a justification for rejecting "barrier" as the correct answer because he had previously used it for another gap. He then decided to skip it the first time, and when he tried it for a second time, he dared to try out some of the alternatives in the context. He finally decided to use the word "barrier", which was the correct answer, and wrote it on the test paper. What this subject did thus far was normal practice. He was successful in providing the correct answer; however, he suddenly checked the item for a third time, most likely because he was uncertain of his answer. Then, he decided to change his answer to the word "death"; however, the word "death" does not fit any gap. This most likely deceived him further, and he may have observed that the word "barrier" was a better answer for a different gap; however, he had previously exhausted his options when he changed the answer for this item and placed "barrier" in gap number one.

He did not provide a rationale for his first answer; however, he did for his second answer. He related the word "death" to a word in the stem, "survival". The strategy he relied on was eventually ineffective and unwise. This shows two important things, first, that some good strategies can be used ineffectively and second, that the decision to execute a certain strategy can be unwise; thus, we have an instance of a good strategy that was used unwisely.

The subject knew what was required for the gap. He knew the meanings of both "death" and "barrier". However, his reliance on the final strategy was an unwise attempt to find an answer, especially because the change was made in revision mode and not on the first pass. Because the subject demonstrated knowledge of the gapped word and knew the meanings of the selected competitive words, his incorrect answer is an invalid indication of the student's lexical knowledge.

Pattern 12- example 2

Pattern 12: Known gapped word +legitimate strategies + correct + TUW +incorrect= invalid

Test: MCQ   item number 2

Subject ID: FS6 – Lexical proficiency: better student

Item: **That company has the_____ to produce clothing quickly and efficiently.**

The correct answer: barrier

Pass 1: **That company has the_____ to produce clothing quickly and efficiently. has the** <mmm> *possibly* (..) *it means* company *needs something to produce clothes in a quick way and effeciently possibly it is expanded=ytwsaa no* (..) **has the** prosp...*no this one is not* prosperity (.) resources (.) resources *it has the* resources *= masder or* inflation (.) inflation **to produce clothing quickly and** (.) inflation (.) promotion opportunity itch to force motivated inflation! **The company has the** <u>inflation</u> *I do not know its meaning but I feel it* ... **has the** <u>inflation</u> *possibly means something* civilized *or ability or something* (..) inflation (..)

Comment: This female student did the same as the student discussed in the previous example. She started quite effectively. She translated a bit of the sentence, retrieved a correct L_1 meaning for the gap, translated some of the suspected alternatives into L_1, and then individually rejected them until she arrived at the correct answer, "resources". She translated the answer with a bit of the stem to confirm her answer. However, she succumbed to the temptation to choose a word that she did not know. Her choice was based on a personal feeling, as she mentioned in the interview. Unfortunately, she changed her answer from right to wrong because she was deceived by a "hunch" that did not pay off. It was ineffective and unwise to rely on the last two strategies. It would have been more judicious to stick with her first choice utilising more confirmation strategies, as she did when she translated the answer into L_1, rather than relying on risky strategies such as choosing unknown words or following her feelings, which can invalidate one's own lexical knowledge. Notably, one could claim that the subject tried to be wise by using a TW strategy; unfortunately, that was unwise.

Pattern 12, example 3

Pattern 12: Known gapped word +legitimate strategies + correct + TS +incorrect= invalid
Test: MCQ   item number 3 Subject ID: FS6 – Lexical proficiency: better student
Item: **How can we** _____ **the spread of air pollution?** The correct answer: slow

Pass 1: **How can we** _____ **the spread of air pollution?** (..)

Air pollution *means .. it has to be* (..) *stopping* the pollution (.) *the word will have the meaning of* (.) *stopping or putting an end* (.) border **can we** border barrier resources inflation (.) prosperity promotion opportunity itch to slow to slow! **How can we** to slow **the spread of air pollution?** *But there might be another word but I think* motivated ceramics delicate rough *I think it is* to slow *I think if we can't stop it we can slow the pollution process* to slow <sound of pen writing>

Pass 2: **can we** <u>slow</u> **the spread can we** to slow (.) *I think* **can we** to slow *then it becomes* **how can we** to slow *it is supposed to be* slow *without* **to** delicate *maybe it is* delicate **how can we** *delicate* (.) delicate (.) delicate *the question itself* **how can we** to slow <sound of pen writing> delicate

---

Comment: This pattern is the same as in the previous example. The female subject applied her usual strategies of using L_1 to arrive at the correct answer, "slow", strategies that were effective on the first pass. She was wise to use confirmation strategies such as continuing to read the list of alternatives to ensure there was not a better candidate for the gap and using L_1 translations to confirm her answers. Unfortunately, while she was revising her answers during her second pass, she decided to change her choice to another unknown word, "delicate", repeating what she did in the previous example. Most likely because there is a slight flaw in the alternative "slow" being preceded by "to", she decided to change her answer by identifying another verb. She ended up choosing a word without "to", thinking that "delicate" is a verb. Although relying on her L_2 mental grammar here should be a wise move when scrutinising her answer, this step was instead unwise. She did not continue to properly infer the correct answer and most likely was lazy not to do so, or she thought that the easiest way was to select an unknown answer, i.e. "delicate". The flaw in the test also led to this false answer, rendering it an invalid test item and score.

Pattern 12, example 4

Pattern 12: Known gapped word +legitimate strategies + correct + TS +incorrect= invalid
Test: MCQ   item number 5
Subject ID: FS6 – Lexical proficiency: better student
Item: **Can a society** _____ **people to use birth-control methods?** The correct answer: force

Pass 1: **Can a society** _____ **people to use birth-control methods? Can a society** (.) to force *this word means yojber* = force (.) *since it is a question* (.) society and people *sometimes do not accept new laws and rules so they are forced.* **Birth control methods** (.) **Can a society** <u>to force</u> to force *I will check the other words* border barrier raw materials opportunity to slow literacy to force motivated delicate *possibly to be* to force or delicate (.) to force *I think it is the closest one* to force (.) *because* delicate *I am not sure of its meaning maybe it has the same meaning=yojber* (.)

Pass 2: **Can a society** to force *the same thing* (.) *then it is not* to force ragged **how can a society** ragged (.) **can a society** ragged *I think it is* ragged *because* force *with* to *is wrong* (.) ragged

Comment: The identical subject repeats everything shown in this example, which is nearly identical to the previous one. She relied on L_1 translation and answered correctly with the word "force" but continued checking for other alternatives. She retrieved another word unknown to her and was about to rely on the identical test-unwise strategy, choosing unknown words, but decided to postpone that until her second pass. She justified her rejection of the correct answer as not being appropriate because the word "force" was preceded by "to". Again unwisely, she tried her test-unwise strategy of choosing unknown words, relying on a feeling to choose an answer, as she mentioned in the interview. That cost her some points, which she deserved because she knew the correct answer.

Discussion

The classifications of the preceding patterns range from knowing the word to partially knowing it to not knowing it at all. Logically, knowing a word should lead to choosing

the correct answer and vice versa. However, what we have observed here highlights important issues because there were instances in which that was not the case at all.

The 12 patterns presented here show that the test words were known the majority of the time, confirming that this achievement test covers words appropriate to the lexical level of the subjects. The patterns also indicate that the test items generated valid scores the majority of the time except for some patterns, which will be discussed below.

Patterns 1 and 7 indicate that although the subjects did not have full knowledge of the test word, their strategies enabled them to arrive at the correct answer, which confirms previous research on EFL learners (Addamegh: EFL multiple choice vocabulary test-taking strategies and construct validity, unpublished) (Al Fraidan, 2011; Al Fraidan & Al-Khalaf, 2012) that EFL learners manage to arrive at correct answers using some test-taking strategies.

Table 1 below shows an example of the analysis. The table only presents the pattern used for each test item by 16 subjects, the number of final strategies used to select the answer and the number of "passes" a subject used to answer the item.

Of the 96 items analysed qualitatively, seven items, from both tests, performed by 15 of the "better" students and 12 of the "poor" ones, were answered correctly; the test words were unknown or only partially known to the testees. This finding confirms findings by (Addamegh: EFL multiple choice vocabulary test-taking strategies and construct validity, unpublished), Cohen (1984) and Israel (1982), who observed that between 40 and 50% of unknown items were answered correctly in four multiple-choice tests. It is clearly evident that in our two different test types, the students were nevertheless able to engage in the identical behaviour but with a different frequency.

It appears that in pattern 1, the strategies effectively helped the testee retrieve a correct form for the recognised meanings in each of the five instances. However, in the two examples of pattern 7, random guessing using little help from L_1 was the route to the correct answer.

Notably, the use of test-taking strategies was not effective all of the time because using them did not often help in retrieving a good answer as in patterns 2 and 3, and in guessing unknown words as in patterns 4, 5 and 6. More notably, the use of TTSs sometimes negatively affected some subjects' answers as in pattern 12 because some of these strategies were unwisely applied and resulted in the student's changing the answer from right to wrong. Only the better students, who may have been scrutinising their answers overly intently, used this pattern. This is contrary to what Alhamly and Coombe (2005) observed, namely that the higher a student scores on a test, the more reluctant he/she is to change his/her answer. This study also observed that EFL Kuwaiti and Emirates students, while revising their answers on multiple-choice questions (MCQ) proficiency test, generally changed them from wrong to right and only 19% did the opposite. (Addamegh: EFL multiple choice vocabulary test-taking strategies and construct validity, unpublished) also observed that his subjects changed their answers from right to wrong but did not report a percentage. However, we presume it was a low percentage as well because the percentage of successful answers on his tests was 70%. Our study revealed similar findings. There were instances of students changing their answers from right to wrong, which suggest that this behaviour should be examined rather than ignored. It must be noted that this pattern threatens the validity of a test score and thus should be carefully considered by educators. Surprisingly, in our study, we also observed that students knew the answer and used some effective

strategies but nevertheless selected the wrong answer, for example, in pattern 10, either because they were in a hurry or had a "slip of the pen" while filling in the gap or because they supplied a distorted form of the correct word. This shows that the use of good strategies unwisely or of TUW strategies may affect test results.

Patterns 2, 3, 4, 5, 6, 8 and 9 are the normal, logical mode a testee would follow when performing well on a test that is designed to differentiate between good and poor students because he/she either knows the test words and answers correctly, automatically or with the help of some strategies, or he/she does not know the test words and fails to answer the test items even with the help of strategies.

The patterns used most frequently in both tests were patterns 5 and 9, which as we claimed above, are the two most logical and normal modes a testee would utilise while taking any test. The frequency of use of pattern 5 was associated mainly with the poor students because they utilised it most often (12 of 13 times on the MCGF, 13 of 17 times on the cloze test). Pattern 9 was associated with the better students (6 of 9 times on the MCGF, 9 of 12 times on the cloze test). The MCGF test attracted all types of patterns; however, the cloze test did not attract patterns 7, 11 or 12. The absence of pattern 7 on the cloze test suggests that students were not able to guess an unknown word with the assistance of TW strategies or any other means. The most used TW strategy, blind guessing, was not effective at all on the cloze test because this type of test cannot be negotiated successfully utilising wild guessing as the MCQ tests can and instead requires reading comprehension and solid lexical knowledge. Apparently, patterns 8, 2 and 1 were used more frequently than the others on the cloze test, which suggests that the chance of choosing a correct answer by guessing is less on the cloze test than on the MCGF test. Patterns 9 and 5, which outperformed the others, add to the score validity of both tests.

The analysis of the two tables above reveals two strategies to be most effective in leading to a correct answer on both tests: pragmatic knowledge, meaning likelihood in relation to a word in the stem, PLW; and pragmatic knowledge, meaning likelihood in relation to the entire stem, PLS. Those two strategies are the most frequently used strategies in the Saudi context (Addamegh: EFL multiple choice vocabulary test-taking strategies and construct validity, unpublished) (Al Fraidan, 2011). We argued above that frequent use of strategies does not indicate effectiveness. Cohen, Weaver and Li (1996, p. 27) were concerned that "repeated use of a strategy may just be a sign that the learner is continuing to use a given strategy unsuccessfully. Conversely, it may indicate that the learner has found the strategy useful". Cohen (1998, p. 93) claimed that the evaluation of any test-taking strategy depends on how individual test takers employ the strategies at a given moment on a given task. Students use their particular cognitive styles, degree of cognitive flexibility and their language knowledge while employing a particular strategy. Therefore, we totalled the number of times both strategies were used effectively and observed that strategy PLS was used effectively 12 times and ineffectively eight times on both tests by both types of subjects, whereas PLW was used effectively ten times and ineffectively seven times on both tests by both types of subjects. Again, the better students were the efficient users of both strategies on both tests except on the MCGF test, and strategy PLW was used equally by both types of subjects (three times effectively and three times ineffectively). Notably, PLW was used ineffectively more often on the MCGF test (five times) than on the cloze test (two times). This is likely because of the

alternatives on the MCGF test not all being linked to words in the stem, which confused some students, mostly the poor ones. This renders it clear that a strategy can be used both effectively and ineffectively. It is difficult to be certain how a strategy leads a testee to the correct answer or fails him/her. However, we observed some problems that hindered a successful execution of strategies, and we identified reasons for these problems. Some of these problems and their reasons are presented above in the examples of patterns. We can summarise these problems, which can be meaning-related or form-related, as follows:

1. Knowing the correct meaning of
    - ➢ the gap
    - ➢ the alternatives offered (MCGF)
    - ➢ words in the sentence/text
    - ➢ a word the testee thought of to fill the gap

Sometimes a subject faces a problem with the meaning of a gap, which can lead to skipping or wild guessing. This problem can be worse when a subject has a problem with a false meaning for the gap. Moreover, this can extend to having a problem with the meaning of the alternatives on the MCGF or with a word the testee thought of to fill a gap on the cloze test. On the MCGF, this problem can extend further and render it more difficult to choose between competitive alternatives. Another meaning-related problem can be a problem with words in the stem or the topic of the cloze test, which can make the situation worse, especially for poor subjects. The worst-case scenario is being unable to think of an English word to fill the gap, which is what occurred with male subject 4 when he left the MCGF test untouched.

2. Finding the correct form of the chosen word on the MCGF or the recalled word on the cloze test to fill the gap

Another problem faced by our subjects is a well-known phenomenon in the vocabulary literature, namely lexical attrition of form, a fancy term for word loss or forgetting (Al-Harthi: EFL multiple choice vocabulary test-taking strategies and construct validity, unpublished) (Al-Harthi 2014; Al-Hazemi, 2000; Weltens & Marjon, 1993). It is the inability to retrieve information associated with a form even if clues are given. It was evident in our data that when students retrieved a meaning for a gap and failed to retrieve a corresponding form, they suffered from lexical attrition of form, observed in patterns 1, 2 and 3 and partially in pattern 10.

The reasons for these problems can be summarised as follows:

- Misreading/miscues and misinterpretation of the sentence/text, gap meaning or alternatives are the main cause of strategies' being ineffective and students arriving at an incorrect answer, as occurred in the example of pattern 2 when the subject misinterpreted the meaning of "as", the first step towards an ineffective strategy. An example of misreading can be observed in the example of pattern 5.
- Lack of proper lexical knowledge
  This can be considered the main cause of failure in the majority of the items.
- Overgeneralisation of grammatical rules and collocations

Some subjects overused the "be and -ing" rule. Whenever they saw the verb "to be", they thought of an answer ending with "-ing".

- Not reading after the gap

  This can be a result of subjects' being in a hurry or being careless.

- Flaw in the test

  An example of this can be observed in examples 3 and 4 for pattern 12.

- Cloze-specific: partial knowledge of the topic

  An example of this can be observed in the example of pattern 5.

- Not following the test instructions and supplying only one word for each gap, such as male subject 6 who retrieved a sequence of three words for some gaps on the cloze test.

Conclusion

We have observed that strategies can affect test scores positively or negatively, in favour of the subject or in favour of the test maker. In other words, strategies will benefit the subject if he/she earns high marks wisely, something that test makers do not favour. Strategies can make a difference in test scores, and test makers hope these strategies are used legitimately rather than wisely. Table 1 above shows that whenever an answer is selected in one of the passes, it exhausts the student's resources and strategies to reach an answer. The use of these strategies varied from effective to ineffective. Students should avoid false perceptions, which may skew their answers. On any achievement test, wrong answers show that some elements were not learned or that the words may have suffered from attrition. However, we observed that sometimes subjects use TUW strategies, which does not show their real lexical knowledge. TUW strategies in this study were used mainly while subjects were working in revision mode. However, we are not discouraging scrutinising answers because some researchers see revision as a wise step in test situations (Flippo, 2008); rather, we argue that this step should be taken with great care, executing revisions that further help in securing a correct answer, not the opposite. Alford (1979) argued that second thoughts are likely to result in wrong answers. Based on his own research, Mehan (1974, p. 44) indicated that it may be misleading to draw the conclusion "that a wrong answer is because of a lack of understanding, for the answer may come from an alternative, equally valid interpretation".

Flippo (2008, p. 103) noted that success in changing a choice is the result of being able to recognise the difference between the reason for making the first choice and the reason for making the second choice and that the second choice is made for a good reason. Thus learners should know that they must understand the difference between the two choices and the reason for the choices before making a successful change in their answers. In pedagogical settings, the effectiveness of strategy use can be enhanced by paying attention to the problems and the reasons summarised above.

It is also important to address the issue of relying on a hunch to arrive at an answer, what students called "inside feeling", and where this is in fact coming from. Is the guessing based on partial knowledge, attrition in knowledge or on no knowledge? The classification of types of guessing is important so that we do not regard any type of guessing as random and end up misjudging students' language ability.

Although not thoroughly examined here, one must be cognisant that TTSs are signs of growing proficiency. More research must develop more insights into the real relation between TTSs and language proficiency.

There are different sources of and many reasons for test invalidity, as explained above. A significant possibility is flaws in the test, especially in EFL contexts in which teacher-made tests frequently suffer from flaws. Teacher training assessment is essential and should never be neglected.

The elicited 12 patterns presented here are not intended to be a comprehensive summary but are the patterns observed in the data. More research should be conducted on this matter to increase our awareness and understanding of the issue by confirming or discounting what was observed in our study or perhaps adding more patterns.

Endnotes

[1]According to the Online Oxford dictionary http://www.askoxford.com/concise_oed/force?view=uk

References
Al Fraidan, A. (2011). *Test-taking strategies of EFL learners on two vocabulary tests.* Germany: Lap Lambert Publications ISBN 978-3-8454-7030-6.
Al Fraidan, A. (2014). Test-unwiseness strategies: What are they? *Journal of Applied Science, 14*(8), 828–832.
Al Fraidan, A., & Al-Khalaf, K. (2012). Test-taking strategies of Arab EFL learners on multiple choice tests. *International Education Studies, 5*, 4.
Alderson, C., & Kremmel, B. (2013). Re-examining the content validation of a grammar test: The (im) possibility of distinguishing vocabulary and structural knowledge. *Language Testing. 30*(4), 535–556.
Alford, R. (1979). *Tips on testing: strategies for test taking*. USA: University Press of America.
Alhamly, M., & Coombe, C. (2005). To change or not to change: Investigating the value of MCQ answer changing for gulf Arab students. *Language Testing, 22*(4), 509–531.
Alharthi, T. (2014). Role of vocabulary learning strategies in EFL learners' word attrition. *International Journal of English Language and Linguistics Research. 2*:13–28.
Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
Cohen, A. D. (1984). On taking language tests: What the students report. *Language Testing, 1*(1), 70–81.
Cohen, A. D. (1998). Strategies and processes in test-taking and SLA. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90–111). UK: Cambridge University Press.
Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly, 3*(4), 307–331.
Cohen, A. D. (2009). (Speaker). 11. TTS [12-minute internet video interview]. In G. Fulcher & R. Trasher (Eds.), *Language testing videos* (Retrieved from http://www.languagetesting.info/video/main.html#list. 26 June 2009).

Cohen, A. D., Weaver, S. & Li, T. (1996). The impact of strategies-based instruction on speaking a foreign language. Research Report. Retrieved from https://fdde40f2-a-551982af-s-sites.googlegroups.com/a/umn.edu/andrewdcohen/docments/1994-MkgLrngStratInstrctnaRealityintheFLCurricinKlee.pdf?attachauth=ANoY7cpG8F7Am3_lTxLL2qOuU-HVpTZaeZW88MXG3OQsLEnwlwgN2G51ArhFOY1_nWKlaVAE1UtoHeuS6bUwusN6r8239E2qOUF-1pZlvobZFzw6uTTDdZF4_eYqRETE6jnIO9x2ZFD8oXiK5vO8bWx1JjR50lNbB_r-bJhO0b5bFcNn10ZMPRhfJ0LZPnZpdutvGw0YfJV2FQdO-NmIKyn0XuXlivFSaRrs1M_sFHcQ__oVovCXDrPTsb6R6f-wO3uCMNywJiY6iPedRxH1t_aHytCzghWS9g%3D%3D&attredirects=0. (11 Nov 2011).

Flippo, R. (2008). *Preparing students for testing and doing better in school*. USA: Crowing Press/ Sage Publishers.

Goh, C., & Aryadoust, V. (2010). Investigating the construct validity of MELAB listening test through the Rasch analysis and correlated uniqueness modeling. In *Spaan fellowship working papers in second of foreign language assessment* (Vol. 8, pp. 31–68). Ann Arbor: University of Michigan English Language Institute.

Green, A. (1998). *Verbal protocol analyses in language testing research: A handbook*. Cambridge: University of Cambridge, Local Examination Syndicate.

Grotjahn, R. (1986). Test validation and cognitive psychology: Some methodological considerations. *Language Testing, 3*, 159–185.

Israel, A. (1982). *The effect of guessing in multiple-choice language tests. Course paper*. Jerusalem: School of Education, Hebrew University of Jerusalem.

Lee, J. (2011). *Second language reading topic familiarity and test score: Test-taking strategies for multiple-choice comprehension questions*. USA: University of Iowa [dissertation].

Mehan, H. (1974). Ethnomethodology and education. In D. O'Shea (Ed.), *Sociology of School and. Schooling. Proceedings of Annual Sociology of Education Meetings*. Washington D.C.: National Institute of Education.

Mohammadi, M., & Abidin, M. (2012). Test-taking strategies, schema theory and reading comprehension test performance. *International Journal of Humanities and Social Science, 1*(18), 237–243.

Nikolov, M. (2006). Test-taking strategies of 12-13-year-old Hungarian learners of EFL: Why whales have migraine. *Language Learning, 57*(1), 1–51.

Purpura, J. E. (1997). An analysis of the relationships between test-takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning, 47*, 289–325.a.

Scholfield, P. (1996). *Quantifying Language*. Clevedon: Multingual Matters.

Weir, C. J. (2005). *Language testing and validation*. USA: Pelgrave.

Weltens, B., & Marjon, G. (1993). Attrition of vocabulary knowledge. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon*. Amsterdam: Benjamins.

Publisher's Note