

## **Confidence, Accuracy and Strategies in Answering Multiple Choice Vocabulary Questions**

**Ghurmallah A. Al-Ghamdi\* and David R. Davies\*\***

*\*Assistant Professor and \*\*Lecturer.*

*School of English, King Faisal Air Academy,  
Riyadh, Saudi Arabia*

(Received 17/10/1415; accepted 28/1/1416)

**Abstract.** This is a report of a two-part study of whether guessing strategies can be effectively employed by second language learners in responding to multiple-choice vocabulary tests. For the first experiment, forty-eight cadets were randomly selected at King Faisal Air Academy. Each subject responded three times to five different multiple-choice vocabulary items. At each of the successive presentations, the item was embedded in an increased amount of 'context': none, one sentence and two sentences. The analysis showed that confidence increased as a function of context. In the second experiment, five additional cadets from the same institution were asked to provide think-aloud protocols as they worked through eighteen vocabulary items in one-sentence contexts. Our conclusion was that guessing may be an effective strategy, but individual personality factors appear to control its frequency of use.

### **Context and Purpose of the Study**

The King Faisal Air Academy (KFAA), Riyadh, is an institution which trains young Saudis to be pilots. An important part of their initial training is the acquisition of English language skills, as all subsequent training is in English. In recent years, staff at the KFAA School of English have focused on the development and validation of communicative oral tests. At the same time the staff have been careful not to neglect the more traditional type of multiple choice tests, since these play a significant role in the evaluation of students' English language skills at the Academy. Cadet assessment is based on scores obtained on the American Language Course Placement Test (for a detailed appraisal see Al-Ghamdi).<sup>(1)</sup>

(1) Ghurmallah Al-Ghamdi, "English Proficiency in the Saudi Air Academy: Validating a New Test Battery," unpublished Ph.D. thesis, University of Edinburgh, 1987.

The construction and validation of multi-choice language items is a subject with an extensive literature. However, rather less attention has been devoted to elucidating and analyzing the strategies employed in responding to multiple-choice tests. The role of contextual guessing strategies has been described in seminal works by van Parreren and van Parreren<sup>(2)</sup> and Clark and Nation.<sup>(3)</sup> Clearly, it is possible for students to apply these strategies to multiple-choice vocabulary items. This possibility has led us to investigate this issue. Two questions suggested themselves: firstly, to what extent and with what effectiveness are contextual guessing strategies employed by students at KFAA? Secondly, would it be possible to deduce the nature of these strategies by observing the cadets in controlled conditions?

### Types of Guessing Strategies

The research cited above yielded a repertoire of contextual guessing strategies which may be summarized as follows:

1. a. Recognize the type of text which contains the unknown word.
- b. Summarize the content of the passage.
- c. Make any necessary inferences from the text.
- d. Look at the immediate context of the target word. (One would assume that while synonyms, hyponyms and opposites might exist in the immediate environment of the target the most frequent clue would be pragmatic collocations.)
- e. Determine the part of speech of the target word.
- f. Use any clues contained in the word form of the target

Clearly, the first two strategies do not apply to multiple choice items, since the target will be contained in only one or perhaps two sentences. The other strategies are, however, relevant. It is important to know if guessing strategies are effective when answering multiple choice questions since, as Davies<sup>(4)</sup> points out, examinations often have an impact on the content and methods of classroom teaching. The realization that guessing strategies are useful in examinations will motivate both teachers and students to practice these strategies in reading lessons. Furthermore, in

---

(2) C.F. van Parreren and S. van Parreren, "Contextual Guessing: A Trainable Reader Strategy." *System*, 9 (1981), 235-41.

(3) D.F. Clark and I.S.P. Nation, "Guessing the Meaning of Words from Context: Strategy and Technique," *System*, 8 (1980), 211-20.

(4) A. Davies, "Follow My Leader: Is That What Language Tests Do?" in *New Directions in Language Testing*, ed. Y.P. Lee, Angela Fok, Robert Lord and Graham Low (London: Pergamon, 1985), p. 13.

sections of lessons devoted to vocabulary, it should encourage the teaching of more pragmatic aspects of meaning and collocation.

One objection here could be based on fairness; that is, students gain marks when they have not actually "learned" a vocabulary item. This objection may be easily dismissed, since an element of guessing forms part of real-life communication at all levels. Even in the mother tongue, people sometimes guess incorrectly and communication breaks down with embarrassing or, occasionally, disastrous consequences. Thus, the inculcation of systematic L2 guessing strategies is a legitimate educational goal.

The point of this research then is to examine the nature and effectiveness of guessing strategies among a representative sample of cadets at KFAA. Assuming rational behavior, we may rule out the possibility that respondents will knowingly choose an incorrect answer. In other words, guessing behavior is intimately linked to confidence or "belief in one's accuracy" and must be examined in terms of this variable. Following related studies of guessing strategies in classroom settings, some researchers have argued that the use of such strategies shows a willingness to take risks.<sup>(5)</sup> According to Chastain<sup>(6)</sup> and Rossier,<sup>(7)</sup> risk-taking contributes positively to L2 proficiency. Not surprisingly, risk-takers tend to be extroverts.<sup>(8)</sup> In the context of the present study these findings are likely to be misleading for two reasons. Firstly, with multiple-choice tests participation is obligatory. While even an uninformed guess has a one-in-four chance of being rewarded, it is failure to participate which carries the greatest risk. Secondly, it seems counter-intuitive to argue that the use of contextual guessing strategies is indicative of a risk-taking personality. Quite the contrary; employing all possible clues would seem to be the safest strategy available.

Attempts to enumerate the communicative strategies available to the learner have been carried out by Tarone<sup>(9)</sup> and Poulise.<sup>(10)</sup> Pribakht<sup>(11)</sup> could find no link

- (5) See C. Hosenfeld, "A Preliminary Investigation of the Reading Strategies of Successful and Unsuccessful Second-language Learners." *System*, 5 (1977), 110-23.
- (6) K. Chastain, "Affective and Ability Factors in Second Language Acquisition." *Language Learning*, 25 (1975), 153-61.
- (7) R.E. Rossier. "Extraversion-intraversion as a Significant Variable in the Learning of English as a Second Language." unpublished Ph.D. thesis, University of Southern California. 1975.
- (8) C.M. Ely, "An Analysis of Discomfort, Risk-taking, Sociability and Motivation in the L<sub>2</sub> Classroom," *Language Learning*, 36, No. 1 (1986), 1-25.
- (9) E. Tarone, "Communication Strategies, Foreigner Talk and Repair in Interlanguage." *Language Learning*, 30, No. 2 (1980), 417-21; *idem.*, "Some Thoughts on the Notion of Communicative Strategy," *TESOL Quarterly*, 15, No. 3 (1981), 285-89.
- (10) N. Poulise, "Problems and Solutions in the Classification of Compensatory Strategies," *Second Language Research*, 3, No. 2 (1987), 141-53.
- (11) T. Pribakht, "Strategic Competence and Language Proficiency," *Applied Linguistics*, 6, No. 2 (1985), 132-46.

between the proficiency of the learner and the selection of strategies. However, these studies are concerned with learner strategies in oral contexts. Our main interest is in the phenomenon of guessing in terms of reading at an inter-clausal level.

A psychological description of the guessing process is needed to supplement the language strategies given earlier. Such a description is offered by Peterson and Pitz.<sup>(12)</sup> When students encounter an unfamiliar word, they will typically search the context for a clue. On finding one, a hypothesis will be formed and a level of confidence generated. The student will then search for a second clue and, on discovering one which confirms this hypothesis, the level of confidence will rise. It may be, however, that the second clue will force the student to change hypotheses. In this case confidence will continue to rise and should be the product of both clues. This process will continue until clues are exhausted or the subject decides to abandon guessing. This gradual process of increasing clues and confidence seems much more plausible than one in which the subject amasses large amounts of evidence before forming a hypothesis since, as Sperber and Wilson<sup>(13)</sup> suggest, the working memory tries to use an optimally small processing effort.

Using this description of guessing it should be possible to design experiments to test its effectiveness. This can be done by recording the confidence and accuracy of students who know the target item and are therefore not guessing and contrasting this record with the confidence and accuracy of students who are guessing. As clues accumulate, the confidence and accuracy of the latter group should approach that of the former. To this effect, the following **null hypotheses** are postulated:

- Accuracy and confidence will not increase as the amount of context increases.
- Confidence and accuracy will not be higher for known targets than for unknown targets.
- Subjects will not employ contextual strategies to derive unknown targets.

### Method

Two experiments were carried out. The first was concerned with assessing the effectiveness of guessing but was somewhat artificial in that it involved a “benevolent” guessing environment where clues were presented to the subjects one after another. Also, these clues were convergent and did not point to differing hypotheses,

---

(12) D. Peterson and G.F. Pitz, “Effect of Input From a Mechanical Model of Clinical Judgment,” *Journal of Applied Psychology*, 7, No. 1 (1986), 163-67.

(13) D. Sperber and D. Wilson, *Relevance: Communication and Cognition* (Oxford; Basil Blackwell, 1986).

and there was ample information to form a hypothesis. If guessing does not prove to be effective in this environment, it is clearly a suspect strategy. Also, most important, in this experiment subjects were placed in a situation where they were forced to guess unknown words; the consequence is that although we may be able to evaluate the strategy as effective we will not know if this effectiveness is sufficient to encourage students to use guessing in real life. The second experiment was more natural and asked to what extent guessing would be employed in a more realistic situation where subjects had more freedom of choice. Again, we tried to evaluate the effectiveness of guessing in such a situation.

### **Experiment 1**

#### **Purpose**

The purpose of this experiment was to evaluate the effectiveness of the process of guessing unknown target words described above by contrasting it with a situation where students are familiar with the target word.

#### **Subjects**

Forty-eight subjects took part. All were Saudi males studying intermediate-level English at the Academy. Their average age was eighteen. Before joining the Academy, they studied English as a subject in intermediate and secondary schools for an average of three periods a week for a total of six years. The criteria by which cadets are selected by the Academy are as follows:

- a. An individual has to be a graduate of high school.
- b. He has to pass a flying aptitude test.
- c. No English proficiency level is demanded as a condition of acceptance, because cadets are trained in English; cadets represent various levels of English proficiency ranging from novice to advanced.
- d. They have to pass a medical checkup.

The subjects were selected at random from the upper intermediate level.

#### **Procedures**

Seven multiple-choice test items were constructed. In each item, the target word was underlined and the subjects were instructed to select one of four choices closest in meaning to the target. Of the seven target words, five were genuine lexical items and two were nonsense words constructed in accordance with English phonology. The genuine items were chosen because the subjects were expected to have at least a passive acquaintance with each.

The forty-eight subjects were divided into two equal groups, A and B. They were presented with the target words in the sequence shown below:

	<b>Group A</b>	<b>Group B</b>
<b>Item 1</b> <sup>(14)</sup>	intelligent	intelligent
<b>Item 2</b>	tall	tall
<b>Item 3</b>	<i>tarc</i>	tiny
<b>Item 4</b>	envious	<i>tarc</i>
<b>Item 5</b>	<i>vond</i>	careful
<b>Item 6</b>	take-up	<i>vond</i>

Each item was presented in three forms: with no contextual information in the stem; embedded in one sentence; presented with two sentences of context. The target word 'careful' is illustrated in all three forms in 3 below.

### 3. *Form 1*

*careful*

- a. interested
- b. careless
- c. cautious
- d. enjoyable

*Form 2*

The driver was *careful* because the traffic was heavy.

- a. interested
- b. careless
- c. cautious
- d. enjoyable

*Form 3*

The driver was *careful* because the traffic was heavy. He didn't want to have an accident.

---

(14) Item 1 was included for practice only; it was not scored.

- a. interested
- b. careless
- c. cautious
- d. enjoyable

After seeing each form, the subject was asked to give a level of confidence by underlining one of the categories in 4.

- 4. not confident
- a little confident
- reasonably confident
- absolutely confident

These responses were graded on a scale of 0 to 3. Thus, for each item, the subjects recorded a total of six responses: three which were marked for accuracy (right or wrong), and three which indicated confidence level (graded on a scale of 0 to 3). This provided 288 responses for each item.

### **Predictions**

The subjects would be expected to score higher on both confidence and accuracy for known (genuine lexical item) than for unknown (nonsense) targets. Both confidence and accuracy would be expected to increase as a function of context. The important result, however, is the interaction. Here we would expect confidence and accuracy for known targets to be high on form 1 and to remain so on subsequent forms. For unknown targets we would expect confidence and accuracy to begin very low on form 1, but to increase and be comparable to known targets on form 3. If this is the case, then guessing will be an effective strategy not only in terms of accuracy but also subjects will have a real measure of confidence in their hypotheses; that is, they would be ready to utilize not only the hypotheses but also the strategies used to gain them in more realistic situations.

### **Criticisms**

Given that the forms of each target differ only in contextual cues, the subjects' attention will naturally be drawn to the importance of this variable. In non-experimental situations they might fail to spot such cues. Also, the experiment dictates an order in which the subjects always see the target forms in the same order: consecutive forms providing increased amounts of context. This would not happen in real test situations. Similarly, with the unknown target, subjects are forced to guess in that

they are given a target and an answer is demanded from them. In real life they may decline to hazard a guess.

The experiment was run in a language laboratory taking six subjects at a time. Subjects were assured this was not a test of their abilities and were taken through a set of instructions containing an example which was worked through, with subjects filling in their individual answers and confidence score. The first item on the test paper was an example which was not counted in the results.

### Results and Discussion

Two-way independent measures ANOVAS were run on the confidence and accuracy data. The results are in Tables 1 and 2. Parametric statistics were used after checking that the data approximated normal distribution. The scores fell within 2 standard deviations of the mean.

#### Discussion

There are significant differences for both confidence and accuracy on known and unknown targets and amount of context. More importantly the interaction of target types and context is significant for both dependent variables. The means for these interactions are given in Tables 3 and 4. The results are generally in keeping with what we expected. There is a small discrepancy for known targets which we expected to be very high for both confidence and accuracy in all three forms. Rather both dependent variables begin reasonably high on form 1 and increase on forms 2 and 3, so the context is actually helping objects on the known targets to improve both

Table 1. ANOVA for gains in confidence related to target type and amount of information.

Source	Sum of squares	Degrees of freedom	Mean square	F-test
Between groups				
Target A	18.5	1	18.5	28.906**
Info. B	114	2	57	89.063**
A x B	15.55	2	7.8	12.188**
Within groups				
Within groups	180.45	282	.64	
Total	328.5	287		

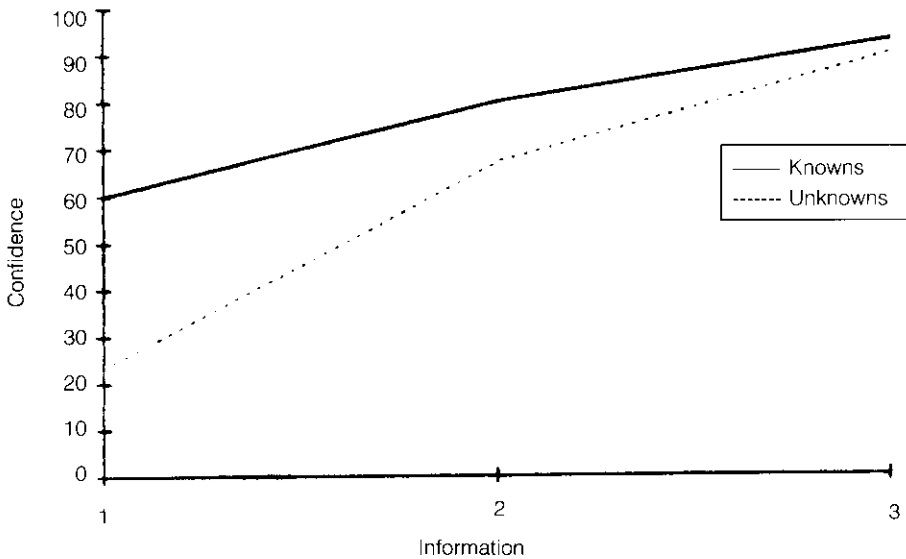
\*\*p < .01.



**Table 2.** ANOVA for gains in accuracy related to target type and amount of information.

Source	Sum of squares	Degrees of freedom	Mean square	F-test
Between groups				
Target A	9	1	9	30.928**
Info. B	23	2	11.5	39.519**
A x B	5.8	2	2.9	9.966**
Within groups				
Total	78	143	.291	

\*\*p &lt; .01.

**Graph 1.** The interaction of target by information for confidence.

accuracy and confidence. Perhaps also the level of confidence on form 1 was suppressed since subjects were aware that further forms were coming. For the unknown targets both confidence and accuracy scores begin very low on form 1, but rise to meet the dependent variable scores on both forms 2 and 3. The confidence interaction scores are displayed on Graph 1 as percentages.

**Table 3. Confidence means for the interaction of target type x information.**

	Form 1	Form 2	Form 3
Known target	1.8	2.4	2.8
Unknown target	0.7	2	2.7

**Table 4. Accuracy means for the interaction of target type x information.**

	Form 1	Form 2	Form 3
Known target	1.46	1.83	1.92
Unknown target	0.42	1.46	1.83

The biggest gain in confidence for unknown targets is on form 2, and by form 3 the level of confidence is very similar to that of the known targets. The accuracy interaction is very similar; by the time form 3 is reached, guessing is not only an effective strategy in terms of accuracy, but subjects are also highly confident in their guesses, almost to the same degree as subjects operating in the known condition. They view their guessing favorably and will be ready to employ the hypotheses gained. Guessing, then, seems to be an effective strategy if there is a reasonably benevolent environment containing a reasonable number of forms.

A final point is that for unknown targets on form 1 we would expect a measure of accuracy which we get due to lucky answers. But we would expect confidence in these answers to be around zero since the targets are completely unfamiliar and contain no morphological clues. However, we can see in Graph 1 that subjects at this point are expressing some confidence in their answers. This suggests that some subjects are "false alarming", i.e., thinking that they recognize a word form which does not exist. This point is supported by van Parreren and van Perreren<sup>(15)</sup> who noted that

(15) Van Parreren and Van Parreren, p. 237.

their subjects made guesses from the word form and that these guesses were risky and unreliable, yet most subjects did not feel the unreliability. The word form then is an attractive source of information, but it is a problematical one in that it can lead to error yet generate confidence.

Concerning correlations between confidence and accuracy, the ANOVAs give a general picture, but correlations should indicate how subjects are acting as individuals. Using the Pearson  $r$  correlation the overall coefficient was .31. The coefficients for known and unknown targets were also taken and they are given in Table 5.

**Table 5. Pearson  $r$  coefficients between confidence and accuracy for known and unknown targets.**

Known	Unknown
.47	.20

These coefficients are low. However, they are not negative, so subjects are not being led into a situation where they are confident but inaccurate. Rather the coefficients reveal a picture where, when a subject makes a correct guess, that subject's confidence is not as high as it should be and vice versa. Part of the reason for this is explained by Davis,<sup>(16)</sup> who points out that some subjects have naturally high confidence whilst others are rather more naturally cautious. The result is that a high-confidence subject who is unsure about a hypothesis will still give a confidence level of 1, while low-confidence subjects may be reluctant to give a confidence of 3 in hypotheses they consider good and use 2 as their maximum.

The result is a tendency for confidence scores to cluster in the central region across the range of accuracy, thereby depressing the correlation coefficient. The effect of this on subjects' ability to use the hypothesis formed and their perception of guessing as a valuable strategy may not be great. The fact that low-confidence subjects do not attach very high levels of confidence to hypotheses for which there is a lot of evidence does not mean that they will not use them. They are used to working in a low-confidence environment where few, if any, hypotheses get full confidence marks; thus, they recognize a good hypothesis for what it is and use it. Conversely, high-confidence subjects will recognize weak hypotheses in which they have expressed some level of confidence.

Table 6 reports the coefficients for know targets across amounts of information. It reveals a gradual improvement in the coefficients until at form 3 the unknown

(16) D.R. Davies, "Guessing Unknown Words in English," unpublished Ph.D. thesis, University of North Wales, 1991.

**Table 6. Pearson r coefficients between confidence and accuracy for unknown targets across information.**

Form 1	Form 2	Form 3
-.11	.03	.33

items are reasonably similar to those of the knowns. The interesting point, however, is at form 1, which produces a low negative correlation. Here guesses are made solely from the word form and this coefficient shows how dangerous such guesses are at least when no morphological evidence is present in the form. Some subjects get an answer correct by accident but generate no confidence. Others seem to be able to generate confidence through perceiving “something” in the word form, but their assumptions are completely mistaken.

### Experiment 2

As we have seen, guessing strategies can generally produce a high level of accuracy and a high level of confidence suggesting that they are effective. However, a limitation of Experiment 1 is that all subjects are forced into making a guess in the unknown target condition by the methodology of the experiment. Thus a second follow-up experiment was needed to discover if the levels of confidence and accuracy generated by guessing in Experiment 1 will replicate in a more realistic situation and to reveal to what extent subjects will employ guessing strategies when given freedom of choice. On the basis of the results of Experiment 1, we would expect subjects to use guessing since it appears to be effective. On the other hand, there is evidence in the literature to suggest that this may not be the case. Hosenfeld<sup>(17)</sup> suggests, for example, that subjects termed “risk takers” are more inclined to use strategies like guessing than subjects who are not “risk takers.” Personality, then, might be a factor and could lead to certain students not using guessing strategies even though they are effective. Also, this experiment should delineate which strategies, mentioned above in 1, subjects tend to favor and how effective these strategies are.

### Method

A practice paper for the English Comprehension Level examination was chosen at random. According to the program of instruction that comes with these tests, they are highly reliable. Internal consistency of these tests ranges between .91 and .96.

(17) Hosenfeld, p. 112.

This paper contained eighteen multiple-choice questions similar to that in Experiment 1, although no attempt was made in this second experiment to add more clues. Five subjects similar to those in Experiment 1 were selected at random.

The method of investigation used was think-aloud protocols. Subjects were asked to answer the eighteen questions and to verbalize their thoughts as they tried to discover the answers to the questions. They were also asked to state their confidence in their answers. There are limitations to this technique. Afflerbach and Johnston,<sup>(18)</sup> for example, point out that the characteristics of the task can influence the data obtained. Our main worry here was that while the attention of subjects was not directed to forms as in Experiment 1, engaging in verbalization might encourage subjects to process more information than they normally would. However, we felt that think-alouds were best suited to our needs since, as Block<sup>(19)</sup> points out, "they are most informative about the reading process when readers have problems understanding what they are reading" because the kind of processing used to solve these problems is not automatic and can be verbalized.

### **Procedure**

The experiment was conducted in a private office free from interruption. Subjects did the experiment one at a time. Each subject was given an orientation to the task which involved working through two examples. They were told they were free to use English or Arabic or both. They were also told their responses would be recorded on tape. After this preparation they seemed able to respond with relative ease. They were also free to work through the questions in their own time, and no attempt was made to present forms in a sequence or draw attention to particular forms. One researcher remained in the office throughout the experiment, and if there were any pauses during the protocol the researchers did not prompt the subject as it was felt that this might encourage the processing of data which would not otherwise be used. The tapes were then transcribed by both researchers, each one at a time, to ensure reliability of transcription. In addition, we asked one colleague whose native language is Arabic and who is also proficient in English to choose one tape at random and check its content against our transcription.

### **Results and Discussion**

The typical method of answering a question where no difficulty was encountered

---

(18) P. Afflerbach and P. Johnston, "The Use of Verbal Reports in Reading Research," *Journal of Reading Behavior*, 16 (1984), 307-21.

(19) E. Block, "The Comprehension Strategies of Second Language Readers," *TESOL Quarterly*, 20, No. 3 (1986), 464.

was as follows. The subject read the question stem in English and sometimes translated the key and or part of the stem into Arabic. The distractors were read in English and were sometimes translated. The distractors were then compared to the target word and evaluated in terms of synonymy. Here the target and distractors were usually pronounced in English with the evaluation being carried out in Arabic. The target and distractors were again translated sometimes during this evaluation.

Where a target proved to be unfamiliar this technique was abandoned and the following strategies listed as 5 were employed.

5. a. *Transformed repetition*

When the subject failed to recognize a target, the target was repeated several times using a different pronunciation each time. For example, *latter* was pronounced /'lætar/, /'le:ter/ and /'leiter/. The subject seemed to be trying out different phonetic forms in the hope of finding a recognizable one. Sometimes this strategy also showed up on distractors.

b. *Untransformed repetition*

Here a target was repeated, but this time using the same pronunciation each time. Some possible reasons for this strategy will be discussed later. It is possible that the above strategies may be a product of the think-aloud protocols, but they could be carried out in real life either silently or by the student verbalizing text.

c. *Use of the context*

This involved strategies given above at 1c, d, e, f. Here distractors were fitted to the context and evaluated in terms of available clues. Target, distractors and clues tended to be pronounced in English while the evaluation was done in Arabic. One example was the stem: "I'm going to *take up* a foreign language." The target is unfamiliar, but the subject fastens on to the distractor *carry* and notes that it is impossible to "carry languages." The subject seems to have realized that the distractors are verbs using strategy 1e and that verbs have selection restrictions. The subject then searches for a verb with an appropriate selection restriction and finds *study* which fits *language* by using strategy 1d.

The first area of investigation is the extent to which subjects will use these guessing strategies. A description of the strategic behavior of subjects together with their confidence and accuracy totals is given in Table 7. Looking at this table, it is subject 4 who is clearly the most strategic in behavior, employing over 50% more guessing strategies than the nearest rivals. Also, by looking at the accuracy scores in Table 7, it is clear that subject 4 is only marginally different from subjects 2, 3 and 5. Subject 4 scores 11 from 18 and the others score 13, 13 and 10 respectively. So sub-

Table 7.

Questions	Subjects				
	S1	S2	S3	S4	S5
1				C	BC
2				AC	
3	C			C	
4		C	A	C	B
5					
6			C	B	
7				AC	
8					
9					B
10					
11		B	C	BC	
12					
13					B
14					
15					
16	C		C	C	
17					
18				AC	
Confidence from a total of 54	46	42	41	42	31
Accuracy from a total of 18	17	13	13	11	10

Key: A-transformed repetition; B- untransformed repetition; C - use of context

jects 2, 3, 4 and 5 all experience not only a measure of difficulty but a similar measure of difficulty. However, it is only subject 4 who tries to resolve this difficulty by a widespread use of guessing, employing as we have noted 50% more strategies than the nearest subject. The other subjects do use strategies, but generally Hosenfeld's<sup>(20)</sup> finding does appear to be validated. Guessing may be an effective strategy as we saw in Experiment 1, but individual personality factors appear to control its frequency of use.

The second factor to be investigated is how successful subject 4's strategic behavior is. Given the situation subject 4 is in, it is unlikely that strategies will help this subject to get higher accuracy and confidence than other members of the group. What must be determined is whether guessing strategies help subject 4 remain in contact with the rest of the group.

For the confidence scores it is possible to perform an independent measures ANOVA to see if there are any differences between the subjects. The result is given in Table 8. What we see is that, while the value of *F* exceeds 1, it does not exceed the ratio of 2.48 needed for *F* to be significant at the .05 level of probability. In terms of confidence then, subject 4 is not different from other members of the group.

For accuracy we cannot perform an ANOVA since the data is nominal. Neither can we perform a chi square as this data is not independent. Looking at the data, however, there seems to be very little difference in accuracy between subject 4 and subjects 2, 3, and 5. Only subject 1 might be seen as really more successful. Therefore, subject 4 appears to be very similar to the majority of subjects in the study in terms of confidence and, reasonably so, in terms of accuracy. However, what is untypical about this subject is that difficulty is encountered on nine out of the eighteen test items and attempts to guess are made. This, as noted, is more than twice the guessing activity of the other subjects. This tends to show how effective guessing strategies can be both in terms of generating accuracy (subject 4 got six out of nine guesses correct) and in generating a level of confidence which is "normal" for the group and would allow the subject to feel comfortable with the hypotheses gained.

Table 8. ANOVA for confidence on the five subjects.

Source of variance	Sum of squares	Degrees of freedom	Mean square	F-test
Between groups	7.76	4	1.94	1.946
Within groups	83.08	85	.977	

(20) Hosenfeld, p. 120.



One final point is to look at the correlation for confidence and accuracy. The coefficient this time is .79, considerably higher than in Experiment 1. There could be a reason for this. Experiment 1 was clearly an artificial experiment, whereas Experiment 2 appeared to the subjects to be more of a genuine language test (even though subjects were assured they were not being evaluated) and tests of course are taken very seriously by students. It could be that in a situation like a test, which is perceived by the student to be serious, a more careful evaluation of forms is made than in an experimental situation where the results do not carry any consequences. This more careful evaluation tends to close the gap between high and low confidence subjects. Furthermore, this correlation encompasses not just guessing behavior and answers which were known but also unknown answers which were not guessed. This last condition was not present in Experiment 1 because subjects were forced to guess. Possibly, this factor has improved the coefficient by producing a set of low-accuracy, low-confidence answers not present earlier.

Thus, this correlation tends to reveal a more balanced guessing behavior than that in Experiment 1. More appropriate levels of confidence are attached to hypotheses which are either known or supported by evidence which can be processed. Low levels are given where difficulty is encountered and guessing is rejected as a strategy. Also the stronger coefficient illustrates that "wild" guesses based on the word form were rarely made. We would expect its kind of guessing to come off the strategy of transformed repetition where in changing the pronunciation the subject might mistakenly hit on a form which sounded familiar. This never manifested itself. Transformed repetition was usually supported by use of the context or if used alone it either met with success or was abandoned, usually the later. One subject, after repeating a target in several different versions stated in Arabic, "My guess is random. I give up." It did not appear to be a very effective strategy in its own right.

As for untransformed repetition, the repetition of a word form in the same pronunciation seemed to signal that the subject was in considerable difficulty and there was no recorded instance of this strategy being supplemented by a move to transformed repetition and only one instance where it was supplemented by the use of context. This example is worth describing. The question stem was "You must be *careful* when in deep water." Subject 5 was not familiar with the target and moved on to the distractors finding *asleep* which was pronounced repeatedly as /slɪp/ from which the word *slip* was derived. This word was then fitted to the context and rejected on the basis that a person must not *slip* in deep water. This is a *very risky strategy* since it leads to an uncorrected understanding of a distractor. Also the subject in repeating /slɪp/, seemed to be trying to talk up the level of confidence held in the hypothesis until it became *slip* and was used.

This kind of fixation on a word form might account for the guesses made from the word forms recorded by Van Parreren and Van Parreren<sup>(21)</sup> and in Experiment 1 here. Subjects repeat the form until they think they see something familiar and they are able to “talk up” their confidence. Ironically here, this strategy led to a correct move which resulted finally in the item being answered correctly. We would emphasize, however, that only one instance of this was found. It may be again that when the task the subjects perform is regarded as highly serious they are less likely to use the hypotheses formed on the basis of word form with no morphological forms present as opposed to an experimental situation. Apart from this single instance, untransformed repetition led to the abandoning of guessing.

### Conclusions

Within the limitations of the study, it seems then that the kind of strategies used by the subjects appear to be very effective in that they generate both accurate hypotheses and considerable confidence in these hypotheses in an environment where there are a reasonable number of convergent forms. Correlations in Experiment 1 were weak, suggesting that levels of accuracy were not awarded with an appropriate level of confidence. We have argued, however, that this in itself would not inhibit subjects from using a hypothesis and that, in fact, it may have something of an experimental effect. Some strategies, most noticeably guessing from the word form, are risky, producing poor accuracy. The important question is whether these strategies will also produce a level of confidence which would enable the guess to be used. In a more realistic situation we feel that the answer would be no. The strategies of transformed and untransformed repetition which would lead to guesses from the word form, were either supplemented by guessing from the context or were followed by the abandonment of guessing.

In summary, guessing the meaning of unknown words can be an effective strategy when answering multiple choice vocabulary items which are read. The realization that such strategies can be useful in tests should motivate students to develop them. Finally, a key point which should be emphasized is that while in Experiment 1 all subjects were forced into guessing behavior, in Experiment 2, where they had more freedom, all subjects used guessing to some extent, but only 4 used the strategy extensively. This was not because guessing does not produce good confidence levels, nor was it because the other subjects had such good marks that they did not need the strategy. Certainly subjects 2, 3 and 5 could have raised their marks by guessing more extensively. Personality and cultural factors seem to be at work here. These factors

---

(21) Van Parreren and Van Parreren, p. 237.

may well be related to a general feeling in Saudi society that guessing entails the risk of a loss of face attendant upon error. They may also be a reflection on the nature of military training. Risk-taking may, therefore, be affected by some psychological or social discomfort which might need to be reduced in order to encourage cadets to take a more active role in language classrooms as well as in test-taking situations.

## الثقة وصحة الاختيار وخطط الإجابة على اختبارات المفردات اللغوية ذوات الاختيار المتعدد

غرم الله بن عبدالله الغامدي\* و ديفيد ريسي ديفز\*\*

\*أستاذ مساعد و\*\*محاضر، مدرسة اللغة الإنجليزية، كلية الملك فيصل الجوية، الرياض، المملكة العربية السعودية

ملخص البحث. تتألف هذه الدراسة من جزئين وتهدف إلى معرفة ما إذا كان بإمكان دارسي اللغة الإنجليزية - كلغة أجنبية - استخدام خطط التخمين بصورة فاعلة عند الإجابة عن اختبارات المفردات اللغوية من ذوات الاختيار من متعدد، كما تهدف الدراسة أيضاً إلى التعرف على ماهية هذه الخطط وأنواعها وتكرار استخدامها. ولهذا الغرض قمنا بتصميم تجربتين؛ تتضمن الأولى اختبار ثمانية وأربعين طالباً من طلاب كلية الملك فيصل الجوية يمثلون مستوى كفاية لغوية واحداً (المتوسط الأعلى) بطريقة عشوائية، أجاب كل طالب ثلاث مرات عن كل بند، في المرة الأولى يظهر البند منفرداً (دون سياق لغوي)، وفي المرة الثانية يكون السياق اللغوي عبارة عن جملة، ثم يصبح السياق اللغوي جملتين في المرة الثالثة، على التوالي، بالإضافة إلى ذلك فقد طلب من كل دارس أن يختار المستوى الذي يبين ثقته في الإجابة عن كل سؤال وفق معيار يتألف من أربع نقاط. أشارت نتائج الدراسة إلى أن الثقة والاختيار الصحيح للإجابة يزدادان مع زيادة السياق اللغوي للكلمات المألوفة وغير المألوفة على السواء إلا أن معدل الثقة والاختيار الصحيح كان أعلى عند الإجابة عن المفردات المألوفة. أما التجربة الثانية فكانت عبارة عن اختبار خمسة طلاب آخرين من الكلية نفسها وبالمستوى اللغوي نفسه. طلب منهم الإجابة عن ١٨ بنداً لغوياً تمثل اختباراتاً للمفردات اللغوية ضمن سياق لغوي لا يتعدى الجملة، ولكن هذه المرة (مع التفكير بصوت مرتفع عند الإجابة عن كل سؤال). سجلت إجابات الطلاب على أشرطة سماعية، وكما في التجربة الأولى، فقد طلب من الدارسين إعطاء مستوى ثقته بعد الإجابة عن كل سؤال. تهدف التجربة الثانية إلى التعرف على أنواع الخطط عند الإجابة عن مثل هذه الأسئلة. أمكن التعرف على ثلاثة أنواع رئيسة من الخطط بالإضافة إلى أخرى ثانوية. أشارت نتائج التجربة إلى أن التخمين قد يكون استراتيجياً فاعلة، ولكن هناك عوامل تتعلق بالدارسين أنفسهم تؤثر على تكرار استخدامه.