

# Teaching Formulaic Sequences: The Same as or Different From Teaching Single Words?

FATIMA A. ALALI  
NORBERT SCHMITT  
University of Nottingham

Formulaic language is an important component of discourse and needs to be addressed in teaching pedagogy. Unfortunately, there has been little research into the most effective ways of teaching formulaic language. In this study, Kuwaiti students were taught words and idioms using the same teaching methodologies, and their learning was measured. The results show that the teaching produced a similar pattern of learning for words as well as idioms, suggesting that at least some of the same types of teaching methodologies we use for individual words can be effective in teaching formulaic sequences. However, the learning of idioms was somewhat lower than that of words. The results also show that reviewing was effective in enhancing the learning to a recall level of mastery, and this was equally true in the learning of words and idioms. Written review was always more effective than oral review. Overall, participants were able to recognize nearly all of the taught words and idioms after 12 days but were able to recall only approximately one-half of them at best, and usually much less.

*doi: 10.1002/tesj.13*

**Formulaic** sequences (multiple-word strings that behave as single units, e.g., realizing a single meaning or function) are a key component of language and are fundamental to the way language is used, processed, and acquired in both the first (L1) and second language (L2; see, e.g., Biber, Johansson, Leech, Conrad, & Finegan, 1999; Meunier & Granger, 2008; Schmitt, 2010; Schmitt & Carter, 2004; Sinclair, 1991; Wray, 2002, 2008). They are regarded as ubiquitous (Carter, 2004), central

(McCarthy, 1998), and “the very centre of language acquisition” (Nattinger & DeCarrico, 1992, p. xv). Martinez and Schmitt (in press) review the reasons why formulaic sequences are so essential in language:

- *Formulaic sequences are widespread in language use.* A number of studies have shown that a large amount of discourse is made up of different kinds of these sequences. Typically, the percentages range from 20% (Sorhus, 1977) to 50% (Erman & Warren, 2000).
- *Meanings and functions are often realized by formulaic sequences.* One reason that these sequences are so widespread is that they communicate a wide number of meanings (e.g., *on the other hand* = conversely) and functions (e.g., *Watch out!* = warning) in discourse. In fact, it has been suggested that for every recurrent communicative need, there is typically conventionalized language (i.e., formulaic sequences) available to realize this need (Nattinger & DeCarrico, 1992, pp. 62–63).
- *Formulaic language has processing advantages.* There is now ample evidence to show that formulaic sequences are processed faster and more accurately than creatively generated language, for example, when embedded in reading texts or when part of grammaticality judgments (Conklin & Schmitt, 2008; Gibbs, Bogadanovich, Sykes, & Barr, 1997; Jiang & Nekrasova, 2007; Underwood, Schmitt, & Galpin, 2004). This processing advantage promotes efficient and effective communication.
- *Formulaic language can improve the overall impression of L2 learners’ language production.* Boers, Eyckmans, Kappel, Stengers, and Demecheleer (2006) show that L2 speakers were judged as more proficient when they used formulaic sequences. The same applies for written discourse (Lewis, 2008; Ohlrogge, 2009).

These reasons show that learners must master formulaic language to reach a proficient level of mastery. However, a word-centered conceptualization of vocabulary has meant that formulaic sequences are seldom taught in any principled manner or tested as part of overall vocabulary knowledge. Consequently, there has been little research on instruction of formulaic sequences or the effectiveness of various teaching techniques in their acquisition. The study presented in this article begins to address this gap by exploring different teaching and review (revision) methods, directly comparing single word and formulaic sequence learning.

## FORMULAIC SEQUENCES AND VOCABULARY TEACHING PEDAGOGY

L2 learners need a large vocabulary to function in English (Nation, 2006; Stæhr, 2008). However, all previous estimates of vocabulary requirements have been based on individual words. Although he does not rely on any empirical count, Jackendoff (1995) argues that the number of formulaic sequences is equal to or even greater than the number of individual words. Given the arguments in the previous section, it seems clear that learners need to know a considerable number of formulaic sequences in addition to individual words. In fact, the difficulty in attaining high levels of English proficiency has been attributed to the lack of formulaic sequence learning and use (Howarth, 1996; Wray, 2000). For example, Irujo (1993) reports that advanced English language learners who made only a few grammatical errors knew little about idioms. Similarly, Farghal and Obiedat (1995) found that Arabic speakers who majored in English had little knowledge about collocations in common topics.

But how are learners to acquire formulaic sequences? Given the lack of much explicit teaching, the unstated assumption seems to be that they will be acquired incidentally from language input. Although this is undoubtedly true to some extent, Schmitt's (2008) review of the incidental acquisition of single words shows that incidental learning is slow and requires numerous exposures, and it usually does not lead to a productive level of mastery. Although research on formulaic sequences is lacking, there is no reason to believe that the same is not true of them. This means that many formulaic sequences need to be taught explicitly.

Unfortunately, there has been little published research on direct teaching of formulaic sequences. One of the few studies to explore whether such sequences can be successfully taught was carried out with students who were studying English just before entering an English-medium university. Jones and Haywood (2004) highlighted formulaic sequences during a 10-week course and found that they were largely successful in raising students' awareness of formulaic sequences, but this awareness did not translate into any substantial increase in the usage of the

sequences in students' output. The researchers did note, however, that although there was no definite improvement in group performance, "there were instances where individual students used phrases accurately and appropriately in their own unsupported writing" (p. 289). This suggests that it may not be easy to increase the number of formulaic sequences produced by students. On the other hand, instruction may have more effect on the accuracy and appropriacy of use of formulaic sequences.

There has been a limited amount of pedagogical advice, such as that given by Nattinger and DeCarrico (1992), Lewis (1997, 2000), and Willis (1990), albeit seldom backed up by empirical research. Boers and his colleagues are among the few scholars who have actually researched the effectiveness of pedagogical techniques for formulaic sequences, and this has resulted in a number of suggested teaching activities (Boers & Lindstromberg, 2008). However, it must be said that there is still very little reliable information about the best ways to teach formulaic sequences and whether the methods typically used to teach individual words are also appropriate for teaching such sequences.

One robust finding from research into individual word learning is the need for recycling and multiple exposures to target words. That is, words need to be recycled to avoid forgetting (Nation, 2001). Research shows that much learning is forgotten quite soon after instruction is finished (Baddeley, 1997), suggesting that the first review needs to occur either during or immediately after the learning session, and then later reviews can be progressively spaced out (i.e., the *expanding rehearsal principle*; Pimsleur, 1967). These findings are for individual word learning, and one would suppose that repetition would also be important for the learning of formulaic sequences. Wood (2002) suggests just this, highlighting the role of repetition in acquiring formulaic sequences and in enhancing their retrieval and ease of access. However, repetition may have differential effects on words and formulaic sequences, and this is explored in this study. Moreover, the type of repetition may also make a difference, so one obvious distinction is also examined here: mode (oral vs. written repetition).

Another important aspect of vocabulary learning is *degree of vocabulary knowledge*, that is how well lexical items are known. If it is accepted that vocabulary knowledge accrues incrementally over time (Schmitt, 2000), then a simple knows/doesn't know dichotomy is untenable, because the important thing is the level of mastery of vocabulary. One way of conceptualizing this is to consider the degrees of knowing the form–meaning link, because it is the most basic aspect of understanding a lexical item (Schmitt, 2010). Laufer and Goldstein (2004) developed the Computer Adaptive Test of Size and Strength (CATSS), which distinguishes four levels of form–meaning knowledge. Although the CATSS was developed for the testing of individual words, we use it to also give a more nuanced assessment of the learning of formulaic sequences in our study.

Overall, this discussion has argued that formulaic sequences are central to language learning and use, and so need to be included in explicit teaching. However, the limited research into the teaching of formulaic sequences has left many pedagogical questions unanswered. This study attempts to shed light on the teaching of formulaic sequences by directly comparing individual word learning with formulaic sequence learning based on a number of different teaching treatments. In particular, the specific research questions that are investigated are as follows:

1. Can we teach formulaic sequences effectively with the same methodologies adopted for single words?
2. Will reviewing newly learned formulaic sequences during the classroom period lead to better retention and retrieval, as it does with single words?
3. Does oral or written review work better for enhancing the learning of single words and formulaic sequences?
4. What degrees of lexical knowledge (recognition/recall) are gained from different teaching and review treatments?

## METHODOLOGY

### Participants

The research took place in Kuwait in a public intermediate girls' school. The participants were 35 female students in one intact class. They all shared the same first language (Arabic). Students

ranged in age between 12 and 13 years. They all started learning English as a school subject from the age of 6, which means they had studied English for 6–7 years. The study was performed as ordinary class lessons. The students were given the Vocabulary Levels Test (VLT; Schmitt, Schmitt, & Clapham, 2001). Almost all of the students scored between 83% and 89% on the 3,000 level of the VLT, indicating that they knew most words up to the 3,000 frequency level and that they were a fairly homogenous group, at least in terms of vocabulary size.

## Materials

**Target vocabulary: Words and idioms.** The study involved the learning of both individual words and formulaic sequences. From the various types of formulaic sequence, we chose to focus on idioms, because they form an essential aspect of language (Wray & Perkins, 2000) and because they are often overlooked by teachers due to their presumed difficulty. Also, Arabic contains many idioms, so they are a class of formulaic language that the participants would be familiar with. However, idioms have seldom been taught in Kuwait as part of the English curriculum, so the participants were highly unlikely to have been exposed to the target English idioms.

A list of candidate idioms was selected from The Free Dictionary (<http://idioms.thefreedictionary.com>), which is based on the *Cambridge International Dictionary of Idioms* and the *Cambridge Dictionary of American Idioms* and claims to include idioms that are current in British, American, and Australian English. The idioms were limited to a maximum of four words to make them as comparable to each other as possible. Because many idioms occur relatively rarely, each idiom's frequency was checked in the British National Corpus to make sure it was relatively common. Because we wished to compare the learning of individual words and idioms in as direct a manner as possible, we selected only idioms that contained one word that we believed would be unknown to the participants. We then extracted these unknown words from the idioms and used them in the single word part of the study. A checklist test with the potential words and idioms was distributed to the students to confirm that they

did not in fact know the target items. The final vocabulary items included the 30 unknown idioms and the unknown single words that were extracted from those idioms (see Appendix A). The 30 idioms and 30 words were divided into three sections for the treatment, that is, 10 idioms and 10 words per class session (discussed in more detail later).

**Test instrument.** In order to answer Research Question 4, we needed a test instrument that could give information on both receptive and productive mastery of the target lexical items. Because the teaching time would be limited in the study, we could not expect learning of a broad range of word knowledge types (e.g., collocation, stylistic appropriateness, derivative forms of a word family). (See Nation, 2001, for a discussion of the various types of word knowledge necessary to truly know a word.) We therefore decided to limit ourselves to the initial stage of learning vocabulary, that is, making a connection between L2 form (i.e., spelling or pronunciation) and meaning. To test the form–meaning link, we created a paper-and-pencil adaptation of the CATSS (Laufer & Goldstein, 2004). The test uses translations to examine both receptive and productive knowledge of meanings and forms of target words and idioms. These categories resulted in four degrees of vocabulary knowledge to be tested (Table 1).

Because the CATSS was developed for single words, a slight modification was necessary to test the idioms. In the idiom multiple-choice tests, one word of the idiom was replaced by another word to form a nonformulaic sequence (see Figure 1).

A series of pilot studies was conducted to check the validity of the test for the participants and purposes of our study. The test was first given to three native English speakers to check for any problems. It was also discussed with the nonnative-English-speaking teachers of the eventual participants to ensure that the items were not too easy or too difficult and to iron out any ambiguous items. The exam was then distributed to 30 students similar to the eventual participants but from another public school. They had not been taught the words previously and had never encountered them before. The test was given to check the test format and the “guessability” of items. Their scores showed that the students could not successfully guess the test items, indicating

TABLE 1. Types of Form–Meaning Knowledge

	Recall	Recognition
Retrieval of form	Form recall	Form recognition
Retrieval of meaning	Meaning recall	Meaning recognition

Source: Laufer & Goldstein (2004).

<p><b>Form recall</b></p> <p>1. ضريبة الامر _____</p>
<p><b>Meaning recall</b></p> <p>2. Pay the piper _____</p>
<p><b>Form recognition</b></p> <p>3. ضريبة الامر</p> <ul style="list-style-type: none"> <li>a. Pay the store</li> <li>b. Pay the customer</li> <li>c. Pay the piper</li> <li>d. Pay the bill</li> </ul>
<p><b>Meaning recognition</b></p> <p>4. Pay the piper</p> <ul style="list-style-type: none"> <li>a. ضريبة التاجر</li> <li>b. ضريبة الامر</li> <li>c. ضريبة الناس</li> <li>d. ضريبة الدولة</li> </ul>

Figure 1. Example of idioms CATSS (L2 English)

that they were well written. They also confirmed that the students did not know the target items, helping to confirm that the participants had no previous knowledge of them in our main study. Finally, another pilot test was carried out with 12 learners, but this time they had a teaching lesson with the target words and idioms. In this piloting, the students scored well on the exam, which also supports the validity of the tests, showing that they could capture learning of the target vocabulary. The time taken to finish the test battery was measured; students took 5 minutes for each test (20 minutes total).

## Teaching Procedure

The study was carried out with one intact classroom group, and included a total of twelve 1-hour class sessions. Six were used for teaching, and another six were used to give a delayed posttest 12 days after the initial teaching treatment. The overall study design involved five stages and is illustrated in Tables 2 and 3. However, before commencing with the research study, the first author (who carried out all stages) took 15 minutes from a previous lesson to introduce the idea of idioms so that all students would be familiar with this concept.

The first stage involved teaching the target words and idioms with translations. First the target item was given in English, and then its Arabic equivalent was presented. We used a PowerPoint presentation to minimize the time spent on writing the translations, giving maximum time to focus on teaching the form and meaning. Exactly the same amount of time was spent on teaching the words and idioms: 10 minutes, with precisely 1 minute for each item. The second stage was a nonrelated distracter task consisting of the teaching of grammar rules from students' curriculum lessons and related grammar activities (10 minutes). This was done to mimic the fact that a number of different linguistic topics are usually taught in the same class (not only vocabulary).

Following this a review treatment was performed (third stage), which varied between three methods: no review, oral review, and written review (see Tables 2 and 3). In the case of no review, the grammar teaching time was doubled (i.e., an extra 10 minutes was given) and no review was given for the target words and idioms, whereas in the oral review students were asked to repeat the

TABLE 2. Procedure for Teaching Single Words

Class 1	Class 2	Class 3
Teaching words (A)	Teaching words (B)	Teaching words (C)
Grammar distracters	Grammar distracters	Grammar distracters
Nonreview	Verbal repetition (1)	Written activity worksheet (2)
Grammar distracters	Grammar distracters	Grammar distracters
Immediate test 1	Immediate test 1	Immediate test 1
Delayed test 2	Delayed test 2	Delayed test 2

TABLE 3. Procedure for Teaching Idioms

Class 4	Class 5	Class 6
Teaching idioms (A)	Teaching idioms (B)	Teaching idioms (C)
Grammar distracters	Grammar distracters	Grammar distracters
Nonreview	Verbal repetition (1)	Written activity worksheet (2)
Grammar distracters	Grammar distracters	Grammar distracters
Immediate test 1	Immediate test 1	Immediate test 1
Delayed test 2	Delayed test 2	Delayed test 2

words aloud 10 times in chorus in the class for 10 minutes. In the written review, students were given a handout (Appendix B) on which they worked in groups to fill in the gaps with the required words and idioms from memory. This made it a written form recall task. The written review was also allotted 10 minutes, making it comparable with the oral treatment. After the review task, the list was taken so that it could not be revised before the subsequent tests.

Next was another distracter stage that involved mostly doing some exercises on the taught grammar. This was included to “flush out” the students’ memory before the vocabulary test so that any answers given on the test would not be merely the result of short-term memorization. Finally, the fifth stage was an immediate test of the target words and idioms. Students were not told that the target items would be tested (see more about this point in the Limitations section). Instructions were read carefully in Arabic to avoid any confusion. In addition, the test was given in four steps, from the form recall format to the meaning recognition format, each separately. To avoid a priming effect, items were randomized from one test to another. The same test battery was also administered 12 days after each classroom session (delayed test). Time for testing was 20 minutes for both immediate and delayed test batteries.

### Data Analysis

The maximum total number for each test level was 10, based on 1 point for each correct answer. The form recall items were strictly scored, with accurate spelling necessary to be judged as correct. Once scoring the data was completed, the scores were analyzed

with an SPSS statistical analysis program. The data were not normally distributed; therefore we used nonparametric tests in conducting the analyses.

## RESULTS AND DISCUSSION

### Descriptive Results

The short-term effects of the teaching, as shown by the immediate posttests, are given in Table 4, and the longer term learning, as shown by the 12-day delayed posttests, in Table 5. Our discussion focuses only on the delayed results, because only they give a good indication of learning that is durable over time (Schmitt, 2010). Also, the results were similar for the immediate and delayed posttests, with some inevitable decline in the delayed posttest scores, so our discussion for the delayed results mostly applies to the immediate results as well.

Overall, the learning gains for words appear to be higher than the learning gains for idioms. The results also suggest that participants gained vocabulary knowledge to different degrees, with form recall level of knowledge having the lowest mean score and the meaning recognition level having the highest mean score. These descriptive results are now explored from a number of perspectives (with statistical analyses) to answer the four questions we posed for the study.

TABLE 4. Vocabulary Gains in the Immediate Posttests for Single Words and Idioms

Treatment conditions	Vocabulary knowledge degrees							
	Form recall		Meaning recall		Form recognition		Meaning recognition	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Words no review	3.20	1.77	5.90	1.83	9.90	0.23	10.00	0.00
Words oral review	5.30	1.61	7.20	0.90	10.00	0.16	10.00	0.00
Words written review	7.40	0.80	8.70	0.87	10.00	0.00	10.00	0.00
Idioms no review	1.40	1.39	4.50	1.91	9.80	0.54	9.90	0.23
Idioms oral review	3.70	1.64	6.20	1.61	9.90	0.23	10.00	0.00
Idioms written review	5.70	1.02	8.10	0.80	10.00	0.00	10.00	0.00

Note. *N* = 35; max score = 10.

TABLE 5. Vocabulary Delayed Posttest Scores

Treatment conditions	Vocabulary knowledge degrees							
	Form recall		Meaning recall		Form recognition		Meaning recognition	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Words no review	0.40	0.60	2.10	1.21	9.70	0.59	9.90	0.23
Words oral review	1.40	0.77	4.00	1.07	9.90	0.23	10.00	0.16
Words written review	3.30	0.71	5.60	0.84	10.00	0.16	10.00	0.00
Idioms no review	0.10	0.23	0.80	1.00	9.60	0.69	9.80	0.45
Idioms oral review	0.50	0.65	2.30	0.92	9.80	0.38	9.90	0.23
Idioms written review	1.50	0.98	4.10	0.90	9.90	0.23	10.00	0.00

Note. *N* = 35; max score = 10.

### Using the Same Methodology for Teaching Formulaic Sequences and Single Words

We first explored whether there were any significant differences between the learning of individual words and the learning of idioms. Wilcoxon signed-rank tests (used because the results were not normally distributed) confirmed there were no significant differences between word and idiom learning for form recognition and meaning recognition for any of the three review conditions (all  $p > .05$ ). However, Wilcoxon signed-rank tests showed significant differences between word and idiom learning for meaning recall for all review conditions (no review:  $Z = 5.02$ ,  $p < .001$ ; oral review:  $Z = 5.19$ ,  $p < .001$ ; written review:  $Z = 5.13$ ,  $p < .001$ ). Likewise, the same tests showed significant differences between word and idiom learning for form recall for all review conditions (no review:  $Z = 3.46$ ,  $p < .01$ ; oral review:  $Z = 4.70$ ,  $p < .001$ ; written review:  $Z = 5.13$ ,  $p < .001$ ).

The results show that the idiom learning was the same as word learning at a recognition level of mastery but was significantly less at a recall level of mastery. That is, the degree of learning depends on the level of knowledge that is assessed. If the ability to *recognize* form or meaning is considered important, then the teaching methodologies in this study can be considered equally effective for single words and idioms, because the learning was virtually 100%

for both types of lexical item. But if the ability to *recall* form or meaning is considered essential, then the teaching methods lead to lower gains for idioms than for words. In essence, these results show that the idiom learning was parallel to word learning in terms of recognition but lagged behind in terms of recall.

It is also important to consider the absolute amount of learning accruing from the teaching treatment. The participants were able to recognize the form and meaning of nearly all of the words and idioms taught after 12 days. However, they were able to recall the meaning of only approximately 20% of the target words when given the form (essentially analogous to receptive knowledge when reading); of idioms this was only 8%. The figures were even lower for recalling the form of the target items when given their meaning, which should correspond to the ability to produce the lexical items (words = 4%, idioms = 1%). Clearly, although learning did occur from the teaching treatment, it was marginal in building a recall level of knowledge, as is illustrated in Figure 2. However, these points pertain to the nonreview teaching treatment only. Research has consistently shown that review generally facilitates vocabulary learning (Schmitt, 2008). The next section examines whether the various types of review enhanced the participants' word and idiom learning in the context of this study.

Overall, the results show that the type of teaching techniques used in this study can lead to good recognition-level lexical gains for words as well as idioms. In terms of recall-level knowledge, idioms are somewhat less well learned, but this difference is small compared to the difference between recognition and recall knowledge. If we think in terms of the big picture (i.e., the absolute gains illustrated in Figure 2), it must be concluded that the teaching resulted in relatively parallel learning of idioms and words. We only have evidence for the methods we used, but the generally parallel results between idioms and words hint that other methodologies may also produce learning gains for formulaic sequences that are similar to the gains for words. (However, see the Implications section for a discussion of specialized teaching techniques for formulaic language.)

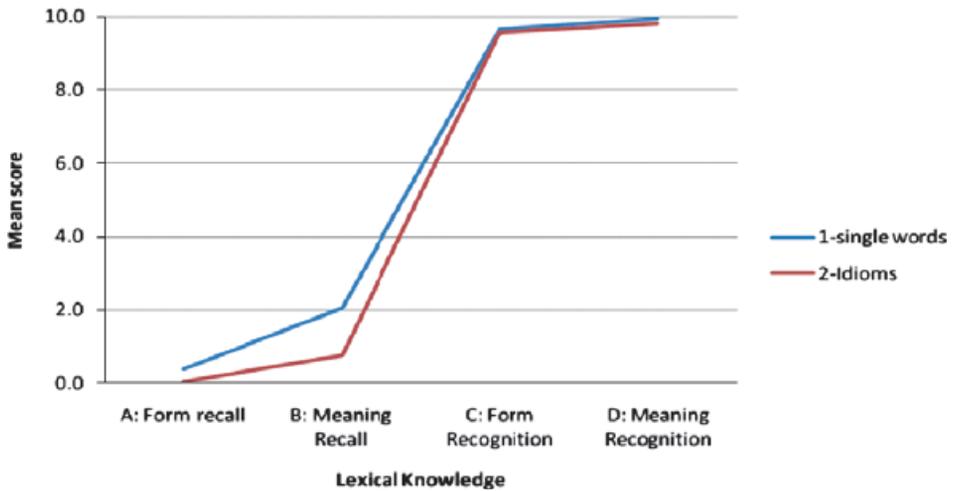


Figure 2. Comparison between single words and idioms in no review treatment (delayed mean scores)

### The Effect of Reviewing

In order to determine whether the review conditions lead to more learning than the teaching treatment alone, we used Wilcoxon signed-rank tests to compare both the oral review scores and the written review scores to the no review scores for words and idioms. These comparisons are given in Table 6.

The single words comparison revealed some statistically significant increases of delayed posttest scores due to both oral and written repetition. Such increases are congruent with most previous research emphasizing the importance of repetition in word learning. Moreover, Table 6 also shows that repetition is equally useful for enhancing idiom learning.

Both oral and written repetition were shown to strongly facilitate learning in terms of recall knowledge, as shown by the extremely large effect size figures. (Effect size figures of  $r > .3$  are considered to be a medium effect,  $r > .5$  are considered a strong effect; Field, 2005.) The teaching treatment by itself led to very high recognition scores, so it is not surprising that repetition did not always improve on these. In terms of meaning recognition, repetition added no statistically reliable extra learning, but in terms of form recognition, repetition usually had a medium effect in benefiting learning. Therefore, if only recognition knowledge is

TABLE 6. The Effects of Review on Learning Single Words and Idioms (Delayed Posttests)

Treatment conditions	Single words			Idioms				
	Median	Z	r	Median	Z	r		
No review–Oral review	No	Oral		No	Oral			
Form recall	0	1	−5.51*	.93	0	0	−3.42*	.58
Meaning recall	2	4	−5.15*	.87	0	2	−4.95*	.84
Form recognition	10	10	−2.89*	.49	10	10	−2.71*	.46
Meaning recognition	10	10	−.577	.10	10	10	−1.27	.21
No review–Written review	No	Written		No	Written			
Form recall	0	3	−5.35*	.90	0	1	−4.80*	.81
Meaning recall	2	6	−5.23*	.88	0	4	−5.23*	.88
Form recognition	10	10	−3.05*	.52	10	10	−2.92*	.49
Meaning recognition	10	10	−1.41	.24	10	10	−2.12	.36

Note.  $N = 35$ ; max score = 10.

\*Wilcoxon signed-rank test  $p < .01$ .

necessary, then the value of repetition after teaching is somewhat questionable. But if recall knowledge is necessary, then repetition was shown to be effective in increasing learning. Considering that a recall level of mastery appears to be necessary to make use of the vocabulary items in reading or listening, then this would suggest that repetition is certainly beneficial in pushing vocabulary knowledge up to the level where it can be useful in real-life contexts. Overall, having a 1-minute review after the 1-minute initial teaching exposure led to stronger learning. This is illustrated in Figure 3.<sup>1</sup>

### The Effectiveness of Oral vs. Written Review

The previous section showed that repetition is effective in increasing the learning of recall knowledge. It is interesting to find out what type of repetition is most effective for idioms and whether it differs from single words. Wilcoxon signed-rank tests were conducted on comparison at the four levels of vocabulary knowledge (Table 7).

<sup>1</sup>It is also useful to note that learning indicated in the posttests is actually the result of three exposures; the immediate posttest must be considered an exposure, and a highly salient one at that, because learners are likely to have a careful focus on the target items during a test.

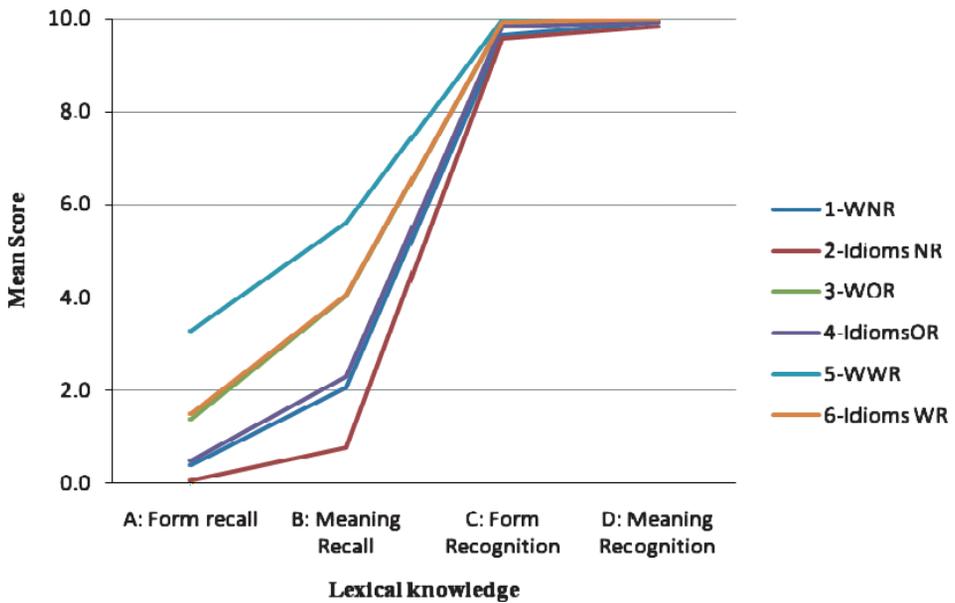


Figure 3. Learning gains in three review conditions

Oral versus written repetition made no difference in terms of recognition knowledge, because the teaching treatment by itself was sufficient for very high scores. However, for recall knowledge, written repetition outperformed oral repetition at both form recall and meaning recall levels of mastery, and the effect size of this advantage was very large ( $r \geq .8$  in all cases). This means that although the oral repetition used in this study was effective in increasing long-term learning, the written repetition was more so. So it seems that making students produce the target items orthographically (i.e., fill in the blank, which is a case of generated repetition; Nation, 2001) led to higher gains than drilling the target items orally. Also, it was found to facilitate the learning of single words as well as idioms (Figure 3).

### Degrees of Lexical Learning

It has been clear from previous analyses that the participants were very successful in learning the lexical items to a recognition level of mastery, but much less so for recall knowledge. The different degrees of learning were formally investigated with a Friedman test analysis. The results show significant differences ( $p < .001$ )

TABLE 7. The Effects of Oral and Written Review on the Learning of Single Words and Idioms (Delayed Posttests)

Treatment condition	Single words				Idioms			
	Median		Z	r	Median		Z	r
	Oral	Written			Oral	Written		
Oral review–Written review								
Form recall	1	3	5.41**	.91	0	1	4.76**	.80
Meaning recall	4	6	5.03**	.85	2	4	5.11**	.86
Form recognition	10	10	-0.58	.10	10	10	-2.00*	.34
Meaning recognition	10	10	-1.00	.17	10	10	-1.41	.24

Note. N = 35; max score = 10.

\*Wilcoxon signed-rank test  $p < .05$ ; \*\*Wilcoxon signed-rank test  $p < .001$ .

across the four degrees of word knowledge. This sanctioned the use of Wilcoxon signed-rank tests as post hoc analyses for single words and for idioms (Table 8). The results can be summarized as the following patterns:

*No review, both words and idioms*

Meaning recognition > Form recognition > Meaning recall > Form recall

*Written review, both words and idioms*

Meaning recognition = Form recognition > Meaning recall > Form recall

*Oral review*

*Words:* Meaning recognition = Form recognition > Meaning recall > Form recall

*Idioms:* Meaning recognition > Form recognition > Meaning recall > Form recall

Although all three review treatments produced nearly maximum recognition learning for both single words and idioms, they differed in recall learning, as shown in Figure 4.

## IMPLICATIONS

This study has a number of pedagogical implications. One of the most obvious is the importance of reviewing. This study confirms a result obtained in virtually all other research into L2 vocabulary: that learners need to engage with words numerous times in order to learn them. For incidental reading (e.g., learning words from

TABLE 8. Friedman Test and Post Hoc Analyses for the Four CATSS Degrees of Knowledge for Single Words and Idioms (Delayed Posttests)

Treatment condition	Median				df	X <sup>2</sup>	Post -hoc analyses <sup>a</sup>
	1	2	3	4			
S. No review	Form recall 0 (1.06)	Meaning recall 2 (1.94)	Form recognition 10 (3.37)	Meaning recognition 10 (3.63)	3	101.64 <sup>***</sup>	4 > 3 > 2 > 1 <sup>**</sup>
S. Oral review	Form recall 1 (1.00)	Meaning recall 4 (2.00)	Form recognition 10 (3.49)	Meaning recognition 10 (3.51)	3	104.67 <sup>***</sup>	4 = 3 > 2 > 1 <sup>***</sup>
S. Written review	Form recall 3 (1.00)	Meaning recall 6 (2.00)	Form recognition 10 (3.49)	Meaning recognition 10 (3.51)	3	104.67 <sup>***</sup>	4 = 3 > 2 > 1 <sup>***</sup>
I. No review	Form recall 0 (1.27)	Meaning recall 0 (1.73)	Form recognition 10 (3.40)	Meaning recognition 10 (3.60)	3	100.05 <sup>***</sup>	4 > 3 > 2 > 1 <sup>*</sup>
I. Oral review	Form recall 0 (1.00)	Meaning recall 2 (2.00)	Form recognition 10 (3.44)	Meaning recognition 10 (3.56)	3	103.83 <sup>***</sup>	4 > 3 > 2 > 1 <sup>*</sup>
I. Written review	Form recall 1 (1.00)	Meaning recall 4 (2.00)	Form recognition 10 (3.47)	Meaning recognition 10 (3.53)	3	104.37 <sup>***</sup>	4 = 3 > 2 > 1 <sup>**</sup>

Note. N = 35; S. = Single words; I. = Idioms; Max score = 10.

\* p < .05; \*\* p < .01; \*\*\* p < .001.

<sup>a</sup>Wilcoxon signed-rank tests, all (>) comparisons are significant at indicated level, all (=) comparisons are nonsignificant (p > .05).

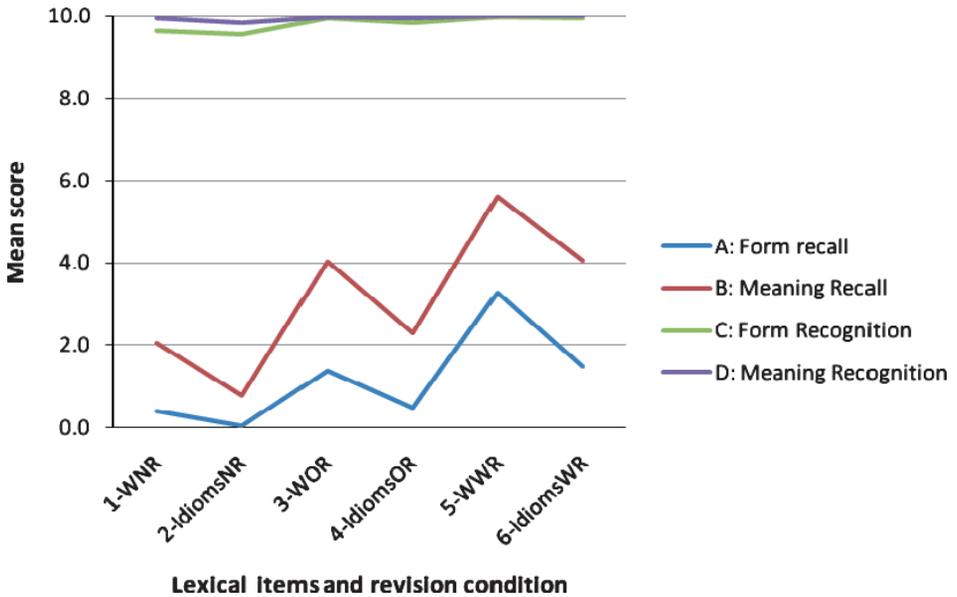


Figure 4. Comparing gains of the delayed test for the four degrees of vocabulary knowledge

pleasure reading), it seems that words must be met in the order of 8–10 times for the initial form–meaning link to be established (Cobb, 2007; Schmitt, 2008). For learning through explicit tasks (as was the case in this study), the number of exposures required is probably lower due to the greater engagement involved, but it is clear that more reviews lead to greater learning. Most of the previous research has focused on the effectiveness of review on the learning of words; in this study we unsurprisingly find that the same beneficial effect is in place for the learning of idioms.

Furthermore, the type of review makes a difference. In this study, written review was far more effective than oral review. This may be due partially to the mode; some research seems to suggest that learners gain more from written exposure than from oral exposure, at least in terms of vocabulary acquisition. But it may also have something to do with the teaching tasks we used. The oral task consisted of choral repetition in the whole class. This was chosen because it is a common teaching technique in Kuwait and many other countries. Nonetheless, it focuses only on form, and not on meaning or the form–meaning link. It also does not require

much mental manipulation of the word form, only rote repetition. Conversely, the written task requires writing the words or idioms into blanks on a handout. This is a productive task and must surely involve more engagement from the student than simple oral repetition. Therefore the written task may be more effective simply because of the greater degree of engagement required. The advantage of the written repetition in this study is probably a combination of both of these causes. Thus, teachers should consider the effectiveness of the various types of review when setting up their vocabulary teaching programs.

This brings up the more general question of what type of instruction is most effective. We looked at only a limited number of types in this study, and they all facilitated learning of words as well as idioms. However, other research, particularly by Frank Boers and his colleagues, has shown that certain types of instruction are particularly effective for learning idioms and other types of formulaic language.

For example, Boers et al. (2006) found that learners who were exposed to considerable listening and reading and made aware of the formulaic language in that input were later judged to be more orally proficient than learners who received the same input but were taught with a traditional grammar + lexis methodology. Likewise, having students consider the underlying categorical orientation of figurative items (e.g., more is up, less is down) helps them learn these items (e.g., *blow up* = inflate; Boers, 2000). (For more on pedagogy related to formulaic language, see Boers & Lindstromberg, 2008; Kovecses & Szabó, 1996; Liu, 2008.) These cognitive/conceptual teaching approaches are likely to be much more effective for teaching formulaic language than individual words. This suggests that there are some areas where the best teaching pedagogy for idioms and formulaic language differs from that for the teaching of individual words.

## LIMITATIONS

The study has three main limitations. The first concerns the testing mode. We used written tests because of practical constraints; however, it is a well-known fact that testing in the same mode as the input (e.g., oral input–oral measurement) leads to better scores

than mixed input and testing modes (e.g., oral input–written measurement). This may have depressed the oral learning scores to an unknown extent.

The second limitation was an artifact of the recursive longitudinal nature of the study. Although the students were not told in advance that they would be tested on the lexical items in the study, by the later classes they may well have seen a pattern and might have started expecting to be tested on the lexical items. This would be especially true by the fourth, fifth, and sixth classes. However, this should have affected the three idiom conditions relatively equally, because any student who had discerned the teaching and testing pattern should have done this by the fourth session. Also, the teaching treatment was completely finished before the first delayed posttest, so any effect on the delayed posttests (from which our analyses come) should have been minimized.

The third limitation concerns the fact that the individual words all came from the idioms that were employed in the study. This means that there was double exposure to the individual words, once as individual words and once again as parts of the idioms, and this may have facilitated the acquisition of those words. However, we think there are good reasons to believe that the effect of this factor is likely to be limited. If the idioms were taught in a manner whereby the component words of the idioms were analyzed for meaning and then put together to derive the meaning of the overall idiom, then surely this would enhance the form–meaning link of the individual target words. But the idioms were not taught this way. They were taught with translations, oral repetition, and written fill-in-the-blank tasks. These tasks focused on the meanings of the overall idioms, not of the component words. So although there might have been an unknown degree of facilitation of form recognition and recall, there should not have been any enhancement of the meanings of the individual target words.

## CONCLUSION

Formulaic language is an important component of discourse, and as such it needs to be addressed in teaching pedagogy. This study showed that our teaching methodology produced a similar pattern

of learning for one type of formulaic language, idioms, as for single words. This suggests that at least some of the same types of teaching methodologies we use for individual words can be effective in teaching formulaic sequences. However, the somewhat lower scores for idioms in this study show that the gains in formulaic sequence learning are likely to be less than that in word learning. But regardless of the teaching method, repetition seems to be an important supplement if lexical items are to be learned to a recall level of mastery. This was equally true in the learning of words and idioms. In this, formulaic sequences seem no different from individual words: They must be engaged with repeatedly in order to maximize learning. In this study, written repetition was more effective than oral repetition, but this may vary for different types of oral and written repetition. That is, it may be that other types of oral repetition (e.g., practicing lexical items in a speech) are more effective than some types of written repetition (e.g., rote writing of lexical items). The study also showed that it is important to distinguish the degree of lexical knowledge. For recognition knowledge (as is typical in multiple-choice tests) there was little difference between any of the treatments, because the participants nearly all received maximum scores. However, for a recall level of knowledge (necessary to use the lexical items in real-world contexts), the learning was much less. This suggests that teachers should carefully consider what levels of knowledge they are trying to enhance and use teaching methodologies that address those levels.

## THE AUTHORS

**Fatima A. Alali** received her MA in applied linguistics from the University of Nottingham. She now teaches in Kuwait.

**Norbert Schmitt** is professor of applied linguistics at the University of Nottingham and is interested in all aspects of second language vocabulary.

## REFERENCES

Baddeley, A. (1997). *Human memory: Theory and practice*. Needham Heights, MA: Allyn & Bacon.

- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow, England: Longman.
- Boers, F. (2000). Metaphor awareness and vocabulary retention. *Applied Linguistics*, 21, 553–571. doi:10.1093/applin/21.4.553
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10, 245–261. doi:10.1191/1362168806lr195oa
- Boers, F., & Lindstromberg, S. (2008). From empirical findings to pedagogical practice. In F. Boers & S. Lindstromberg (Eds.), *Cognitive linguistic approaches to teaching vocabulary and phraseology* (pp. 375–393). Berlin, Germany: Mouton de Gruyter.
- Carter, R. (2004). *Language and creativity: The art of common talk*. London, England: Routledge.
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning and Technology*, 11, 38–63.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 72–89. doi:10.1093/applin/amm022
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20, 29–62.
- Farghal, M., & Obiedat, H. (1995). Collocations: A neglected variable in EFL. *International Journal of Applied Linguistics*, 28, 313–331.
- Field, A. (2005). *Discovering statistics using SPSS*. London, England: Sage.
- Gibbs, R., Bogadanovich, J., Sykes, J., & Barr, D. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language*, 37, 141–154. doi:10.1006/jmla.1996.2506
- Howarth, P. (1996). *Phraseology in English academic writing: Some implications for language learning and dictionary making*. Tübingen, Germany: Max Niemeyer.
- Irujo, S. (1993). Steering clear: Avoidance in the production of idioms. *International Review of Applied Linguistics in Language Teaching*, 31, 205–219.

- Jackendoff, R. (1995). The boundaries of the lexicon. In M. Everaert, E. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and psychological perspectives* (pp. 133–166). Hillsdale, NJ: Lawrence Erlbaum.
- Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *Modern Language Journal*, *91*, 433–445.
- Jones, M., & Haywood, S. (2004). Facilitating the acquisition of formulaic sequences: An exploratory study in an EAP context. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 269–292). Amsterdam, the Netherlands: John Benjamins.
- Kovecses, Z., & Szabó, P. (1996). Idioms: A view from cognitive linguistics. *Applied Linguistics*, *17*, 326–355. doi:10.1093/applin/17.3.326
- Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*, 399–436. doi:10.1111/j.0023-8333.2004.00260.x
- Lewis, M. (1997). *Implementing the lexical approach*. Hove, England: LTP Press.
- Lewis, M. (Ed.). (2000). *Teaching collocation*. Hove, England: LTP Press.
- Lewis, M. (2008). *The idiom principle in L2 English: Assessing elusive formulaic sequences as indicators of idiomaticity, fluency and proficiency* (Unpublished doctoral thesis). Stockholm University, Stockholm, Sweden.
- Liu, D. (2008). *Idioms: Description, comprehension, acquisition, and pedagogy*. London, England: Routledge.
- Martinez, R., & Schmitt, N. (in press). A phrasal expressions list. *Applied Linguistics*.
- McCarthy, M. (1998). *Spoken language and applied linguistics*. Cambridge, England: Cambridge University Press.
- Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam, the Netherlands: John Benjamins.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

- Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, 63, 59–82. doi:10.3138/cmlr.63.1.59
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, England: Oxford University Press.
- Ohlrogge, A. (2009). Formulaic expressions in intermediate EFL writing assessment. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. M. Wheatley (Eds.), *Formulaic language, Volume 2: Acquisition, loss, psychological reality, and functional explanations* (pp. 375–386). Amsterdam, the Netherlands: John Benjamins.
- Pimsleur, P. (1967). A memory schedule. *Modern Language Journal*, 51, 73–75. doi:10.2307/321812
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge, England: Cambridge University Press.
- Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, 12, 329–363. doi:10.1177/1362168808089921
- Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, England: Palgrave Macmillan.
- Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 1–22). Amsterdam, the Netherlands: John Benjamins.
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, 18, 55–88. doi:10.1177/026553220101800103
- Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press.
- Sorhus, H. (1977). To hear ourselves—Implications for teaching English as a second language. *ELT Journal*, 31, 211–221. doi:10.1093/elt/XXXI.3.211
- Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, 36, 139–152. doi:10.1080/09571730802389975
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 53–71). Amsterdam, the Netherlands: John Benjamins.

- Willis, D. (1990). *The lexical syllabus: A new approach to language teaching*. London, England: HarperCollins.
- Wood, D. (2002). Formulaic language in acquisition and production: Implications for teaching. *TESL Canada Journal*, 20, 1–15.
- Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, 21, 463–489. doi:10.1093/applin/21.4.463
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, England: Cambridge University Press.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford, England: Oxford University Press.
- Wray, A., & Perkins, M. (2000). The function of formulaic language: An integrated model. *Language and Communication*, 20, 1–28. doi:10.1016/S0271-5309(99)00015-4

## APPENDIX A

TABLE A1. Single Words<sup>a</sup> and Idioms Used in the Study

Make no <b>bones</b> about it	Put the <b>cart</b> before the horse
Mend <b>fences</b>	In the <b>lap</b> of luxury
Life and <b>limb</b>	At full <b>tilt</b>
<b>Ivory</b> tower	Starting from <b>scratch</b>
Down in the <b>dumps</b>	Through <b>gritted</b> teeth
<b>Dwell</b> on the past	
<b>Cast</b> a long shadow	Clean as a <b>whistle</b>
Run of the <b>mill</b>	Birds of a <b>feather</b>
Off the <b>hook</b>	Pay the <b>piper</b>
Add <b>insult</b> to injury	Kick the <b>bucket</b>
Fit as a <b>fiddle</b>	Nose to the <b>grindstone</b>
Turn <b>belly</b> up	Against the <b>grain</b>
Take for <b>granted</b>	All the <b>rage</b>
Mad as a <b>hatter</b>	<b>Crunch</b> time
The tip of the <b>iceberg</b>	<b>Slip</b> of the tongue

<sup>a</sup> The words in **bold** are the single words used in the study.

## APPENDIX B

### SAMPLE EXERCISE SHEET DISTRIBUTED TO LEARNERS DURING THE WRITTEN REVIEW TREATMENT

#### SINGLE WORDS EXERCISE WORKSHEET

Please fill in the space with the appropriate words:

1. A: What did your brother do when he found about his broken bicycle?  
B: He shouted with \_\_\_\_\_.
2. A: What is your mother's job?  
B: She is a \_\_\_\_\_.
3. A: How did they grind the wheat in the past?  
B: They used a \_\_\_\_\_.
4. A: What else do you need to bake this bread?  
B: I need 200 pounds of whole \_\_\_\_\_.
5. A: What caused him to fall?  
B: He blamed his \_\_\_\_\_ on the ice.
6. A: What cause the Titanic to sink?  
B: It hit a big \_\_\_\_\_ in the ocean.
7. A: The king \_\_\_\_\_ 100 KD to all citizens.  
B: Really, that's amazing.
8. A: What kind of an instrument does Fatima play?  
B: She is very good playing the \_\_\_\_\_.
9. A: What is that sound?  
B: It is the \_\_\_\_\_ of stones under the car wheel.
10. A: Is she fat or pregnant?  
B: I do not know, but she has a big \_\_\_\_\_.

#### IDIOMS EXERCISE WORKSHEET

Please fill in the space with the appropriate idiom:

1. A: Fake tiger print was so fashionable in the seventies.  
B: It's \_\_\_\_\_ again now.
2. A: Ali seems different lately.  
B: From what I can see he is \_\_\_\_\_.
3. A: I need to pass this exam.  
B: I'm sure if you keep your \_\_\_\_\_ you'll have a good grade.
4. A: She never thought about her parents.  
B: She really goes \_\_\_\_\_ compared to her mom.

5. A: Why did you tell her about my illness?  
B: I didn't mean to tell her that. It was a \_\_\_\_\_.
6. A: These difficulties are just \_\_\_\_\_.  
B: Do you mean that there is more?
7. A: Didn't you check the weather before you packed to go to the beach?  
B: No, it's always sunny this time of the year. I'll \_\_\_\_\_ the beautiful weather will continue and just go.
8. A: I feel sick and really cold.  
B: You may feel sick now, but after a few days of rest and plenty of liquids, you'll be \_\_\_\_\_.
9. A: What happened to you? You look disturbed.  
B: Yes, a little. I have to hand in this project this week. It's \_\_\_\_\_ and I'm not sure I'll finish.
10. A: Is there something wrong?  
B: Yes, the computer \_\_\_\_\_ in the middle of an important job.