



The power of sound in L2 idiom learning

Journal:	<i>Language Teaching Research</i>
Manuscript ID	LTR-15-0203.R1
Manuscript Type:	Full Research Article
Keywords:	L2 phrase learning, figurative idioms, phonological similarity, recall of forms, explicit instruction
Abstract:	<p>Corpus analyses of learners' dictionaries of English idioms have revealed that 11% to 35% of English figurative idioms show either alliteration (miss the mark) or assonance (get this show on the road), depending on the type considered. Because multiword combinations, particularly idiomatic expressions, present a huge challenge even for advanced learners of English, elaborative techniques that help learners come to grips with this part of the lexicon are welcome. In this study we investigated whether phonological similarity (i.e., alliteration and assonance) facilitates the delayed recall of common L2 figurative idioms which were not known at pretest. In a quasi-experiment involving 50 EFL learners, one group (n = 26) learned significantly more phonologically similar (PhS) idioms than nonPhS (control) idioms after a treatment designed both to raise awareness of phonological similarity (PhS) and to direct attention toward occurrences of it. Learners in a comparison group (n = 24), who experienced no awareness raising or attention direction, recalled more nonPhS idioms. We conclude that the potential of sound repetition needs to be unlocked by pedagogical interventions if learners are to reap mnemonic benefits from it.</p>

SCHOLARONE™
Manuscripts

Abstract

Corpus analyses of learners' dictionaries of English idioms have revealed that 11% to 35% of English figurative idioms show either alliteration (*miss the mark*) or assonance (*get this show on the road*), depending on the type considered. Because multiword combinations, particularly idiomatic expressions, present a huge challenge even for advanced learners of English, elaborative techniques that help learners come to grips with this part of the lexicon are welcome. In this study we investigated whether phonological similarity (i.e., alliteration and assonance) facilitates the delayed recall of common L2 figurative idioms which were not known at pretest. In a quasi-experiment involving 50 EFL learners, one group ($n = 26$) learned significantly more phonologically similar (PhS) idioms than nonPhS (control) idioms after a treatment designed both to raise awareness of phonological similarity (PhS) and to direct attention toward occurrences of it. Learners in a comparison group ($n = 24$), who experienced no awareness raising or attention direction, recalled more *nonPhS* idioms. We conclude that the potential of sound repetition needs to be unlocked by pedagogical interventions if learners are to reap mnemonic benefits from it.

Keywords

L2 phrase learning, figurative idioms, phonological similarity, recall of forms, explicit instruction

Introduction

In order to attain a high level of proficiency in a foreign language learners need to familiarize themselves with very many formulaic sequences, which is a challenging undertaking for most post-childhood learners (Li and Schmitt 2009; Qi and Ding 2011). One type of L2 formulaic sequence that is especially troublesome is that of idioms. Various means have been suggested whereby teachers and materials writers might help learners acquire L2 idioms with greater success (for reviews see Author 2009; 2012). One proposal, directed particularly at the learning of *forms*, is to raise learners' awareness of the prevalence of sound repetition in English idioms. This proposal is based on two assumptions. The first is that given appropriate pedagogical intervention certain patterns of sound repetition, or phonological similarity (PhS), render L2 phrasal expressions relatively memorable. The second is that these patterns of PhS are sufficiently common in the L2 phrasal lexicon for the first assumption to be of interest. In this article both assumptions will be addressed.

Idioms are so multifarious that it is common for researchers to regard the category as fuzzy (e.g., Nunberg, Sag, and Wasow 1994; Zyzkik 2011). Although our proposals are not necessarily limited to any one type of idiom, in this article we focus on so-called figurative idioms—that is, relatively fixed, typically informal conventional expressions whose meanings arise at least in part from a trope such as metaphor, metonymy, hyperbole, and/or irony (Gibbs 1994; Nunberg et al. 1994). In adopting this focus we follow other researchers of L2 idiom learning such as Author (2007); Grant, (2007), and Zyzik (2011).

Sound repetition in English phraseology: alliteration and assonance

The current study is concerned with two patterns of intra-phrase PhS—alliteration and assonance. We define alliteration as the occurrence of the same consonant onset in two or

1
2
3
4 more content words within a phrase (e.g., *miss the mark*) and assonance as the repetition of a
5
6 vowel in a prominent syllable of two or more content words within a phrase, as in *jump the*
7
8 *gun* (basic assonance), *a quick fix* (near rhyme), *go with the flow* (clipped rhyme), and *make a*
9
10 *mistake* (rhyme, which is regarded here as a special case of assonance). A question we first
11
12 need to address is how common alliteration and assonance are in English idioms. Extending a
13
14 previously reported study (Author 2014a), we searched through a learner's dictionary of
15
16 English idioms (Parkinson 2006), examining all defined expressions which include at least
17
18 two content words in addition to any verb or verbs. Of 2906 such expressions (e.g., *Let the*
19
20 *cat out of the bag*, *the name of the game*, *a hornet's nest*), 232 (8%) were identified as
21
22 alliterative and 392 (13.5%) as assonant. Altogether, about 21% showed alliteration and/or
23
24 assonance. Of the 197 so-called 'binomial idioms' (e.g., *high and dry*) not signalled as "old-
25
26 fashioned", about 38% show alliteration and/or assonance. Of 71 'as idioms' (e.g., *as good as*
27
28 *gold*), about 35% alliterate and/or assonate. We also examined alternate five page blocks
29
30 throughout Parkinson (2006), taking note of all defined noun phrases of the structure
31
32 (DETERMINER or POSSESSIVE PRONOUN) + QUALIFIER or NOUN + NOUN (e.g., *top dog*, *hobby*
33
34 *horse*). Of 221 such expressions 28% show either alliteration or assonance. Other counts
35
36 (through different learner's English idioms dictionaries) focusing only on alliteration but
37
38 covering *all* defined expressions have found its incidence to be 13% (Author 2009, p. 114)
39
40 and 17% (Boers and Stengers 2008). If alliteration and/or assonance have a large enough
41
42 positive mnemonic effect on learners' ability to remember L2 phrases, occurrence rates such
43
44 as those just mentioned may mean that these patterns of PhS can facilitate L2 idiom learning
45
46 to some useful degree. We now briefly survey evidence that relations of PhS do influence
47
48 lexical recall.

Experimental evidence for positive effects of phonological similarity on lexical recall

1
2
3
4 In recent decades most experimental investigations of the mental lexicon have taken place
5
6 within a connectionist framework in which it is expected, for instance, that activation of a
7
8 word in memory (the activation being brought about, for example, by the mention of a
9
10 particular topic) may spread to other words. That is, these other words are also activated,
11
12 whereby they too become more available for processing and for production. Within a
13
14 connectionist framework it is expected that the activation of one word by another is
15
16 especially likely to take place if the two words are semantically related and/or phonologically
17
18 similar (e.g., Gupta and MacWhinney 1997; Luce, Goldinger, Auer, and Vitevitch 2000;
19
20 Storkel and Morrisette 2002). The connectionist view accounts well for the results of a vast
21
22 number of investigations of PhS effects on the accessibility of words in memory. Some
23
24 twenty years ago Rubin (1995) noted that hundreds of studies had been carried out to
25
26 investigate PhS effects on participants' ability to learn lists of 'paired associates', that is,
27
28 pairs of items—typically short L1 words or invented L1-like nonwords (see Rubin, 1995, for
29
30 an overview). In these studies the item pairings are stipulated by the researchers. Typically,
31
32 participants go through several cycles of 'practicing' (i.e., trying to memorize) the pairings
33
34 and then being tested on them. A strong finding is that when participants are asked to recall
35
36 practiced word-pairs in any order they like, pairs that show PhS are learned faster and
37
38 recalled better than nonPhS pairs. This holds true for rhyme (e.g., Bower and Bolton 1969;
39
40 Nelson and Garland 1969), for alliteration+assonance (e.g., *hat - ham*) (Bower and Bolton
41
42 1969; Nelson and Garland 1969), and for alliteration and assonance separately (Nelson and
43
44 Garland 1969).

45
46
47
48
49
50 A second extensive subliterature is concerned with learning lists of unpaired words or
51
52 nonwords. When participants are asked to recall these items in any order, they succeed better
53
54 with ones that show PhS, a trend of results which holds for alliteration and rhyme (Gupta,
55
56
57
58
59
60

1
2
3
4 Lipinski, and Actunc 2005) as well as basic assonance (Watkins, Watkins, and Crowder
5
6 1974).
7

8
9 Positive effects of PhS have also been observed in studies of implicit priming, where
10
11 observed response times are shorter when primes and response words alliterate than when
12
13 they do not, with response times being shorter still when primes and response words both
14
15 alliterate *and* assonate (for a summary see Levelt 1999).
16

17
18 An additional relevant stream of research concerns the extent to which relations of
19
20 PhS make primarily oral poetic texts relatively easy to recall. As might be expected, research
21
22 has identified rhyme and alliteration as factors that facilitate ability to recall small-scale
23
24 passages of song lyrics and traditional counting out rhymes: Provided that *one* of a pair or set
25
26 of PhS words is recalled, recall of the other(s) is made more likely (Rubin 1995).
27

28
29 We know of no studies in any of the research streams touched on above which
30
31 directly address PhS effects on the recollectability of conventionalized phrases in L1 or L2.
32
33 One probable reason for this research gap is the extreme difficulty even in a laboratory
34
35 setting of controlling for nuisance semantic variables when the stimulus expressions are real
36
37 phrases: After all, achieving approximate control is hard enough in the case of real *single*
38
39 words. However, in practice-oriented L2 research a more exploratory approach may be called
40
41 for than is normal in psycholinguistics. Accordingly, there have been a number of
42
43 (quasi)experimental paired-design studies of the extent to which alliteration and assonance
44
45 help upper-intermediate and advanced learners recall or recognize short L2 English phrasal
46
47 expressions (usually collocations such as *full force*) as compared to non-phonologically
48
49 similar (nonPhS) expressions (e.g., *full speed*) (Author 2014a; Author 2012, 2014b/c; Author
50
51 2008a/b). In general, a positive effect ranging from large to medium-small has been found
52
53 following treatments that included brief awareness raising about alliteration or assonance
54
55 and/or an intervention to direct participants' attention to the phonological form of the
56
57
58
59
60

1
2
3
4 stimulus expressions. Only small positive effects or even negative ones (Author 2014c) have
5
6 been observed in the absence of such pedagogical interventions. Generally, though, the recall
7
8 advantage shown by phonologically similar (PhS) expressions has been found to dissipate
9
10 with time. In some cases no effect was detected in posttests administered after a delay of
11
12 from 24 hours to one week. It should be noted however that these studies have focused *not* on
13
14 participants' learning of previously unknown expressions but on their ability to recall familiar
15
16 expressions that happened to be encountered on a particular occasion, that is, during the
17
18 experimental treatment. Thus, these studies are much more likely to have probed 'episodic
19
20 memory' than 'semantic memory' (cf., Tulving 1972; Nadel and Hardt 2011). To elaborate,
21
22 episodic memory preserves a link between items of information and the context(s) in which
23
24 the items were encountered and encoded. Episodic memory would be drawn on in following
25
26 an instruction like this, 'Please write down, from memory, as many as you can of the two-
27
28 word phrases that you sorted and then dictated to each other at the beginning of class last
29
30 Tuesday'. Semantic memory, which tends to be more durable and which underpins most
31
32 everyday language use, does not preserve the link between something that is known (e.g., a
33
34 phrase) and the occasion when it was learned (Tulving 1972; Nadel and Hardt 2011). We
35
36 know of five studies which have investigated the storage and relative accessibility in
37
38 semantic memory of previously unknown phrasal expressions (both PhS and nonPhS),
39
40 including figurative idioms (Author 2005, experiments 1-3, Author 2014d, and Author 2008a,
41
42 experiment 3). That is, the tests of recall used in these studies did not require participants
43
44 specifically to remember whether the targeted expressions had or had not been been
45
46 encountered on a particular list or in another stipulated context that was encountered or
47
48 experienced at a particular time and place. In each of these five studies alliteration was found
49
50 to have facilitated participants' ability to recall the forms of phrasal expressions that the
51
52 participants either definitely did not know or else were unlikely to have known before the
53
54
55
56
57
58
59
60

1
2
3
4 study began. However, these studies were either very small in scale or they involved analysis
5
6 of data originally collected for a different purpose than that of investigating PhS effects on L2
7
8 phrase learning. Moreover, none of these five studies explicitly included assonance in its
9
10 focus. For these reasons additional experimental investigation seemed to be in order.

11 12 13 **The research questions**

14
15
16 Our aim in this experiment was to investigate:

- 17
18
19 (1) whether, with respect to recollectability from semantic memory, previously unfamiliar
20
21 PhS idioms have an inherent advantage over previously unfamiliar nonPhS idioms; if not,
22
23
24 (2) whether an experimental treatment including brief awareness raising about PhS patterns in
25
26 idioms and a simple attention direction task leads to higher recall rates of PhS idioms than
27
28 when participants are not made aware of the sound patterns in the target idioms; and if so,
29
30
31 (3) whether this outcome occurs at the expense of recall of the nonPhS idioms in the set of
32
33 idioms to be learned.
34
35

36
37 These research questions were operationalized by comparing the number of idioms
38
39 participants could not produce at pretest with the number that they could produce in posttest
40
41 immediately following the study session and in a delayed posttest administered a week later.
42
43 In the experimental condition the treatment included brief awareness raising about patterns of
44
45 PhS in idioms, following which participants were set a marking task intended to direct their
46
47 attention to instances of PhS among the set of stimulus idioms. This condition is hereafter
48
49 referred to as the AD condition and the group that experienced it, as the AD group. In the
50
51 comparison condition, the treatment consisted simply in the participants being asked to study
52
53 the targeted PhS and nonPhS idioms according to their own preferred personal learning
54
55 strategies. This condition is hereafter referred to as the noAD condition and the group that
56
57
58
59
60

1
2
3
4 experienced it as the noAD group. Both an immediate and a delayed posttest were
5
6 administered to measure learning gains.
7
8

9 **Method**

10 *Participants*

11
12 The 50 participants who completed all stages of the experiment were Dutch-speaking
13
14 students aged 19 to 21 majoring in English and an additional foreign language at the
15
16 Department of Translation, Interpreting and Communication of Ghent University, Belgium.
17
18 They formed two intact classes of respectively 26 and 24 students and their level of
19
20 proficiency in English was estimated as B2 according to the Common European Framework
21
22 of Reference for Languages (CEF), which corresponds to an IELTS score of 5 to 6.5. Each
23
24 class was assigned to one of the treatments ($n_{AD} = 26$; $n_{noAD} = 24$).
25
26
27
28
29
30

31 *Materials*

32
33 The expressions targeted in this experiment were 26 L2 English idioms drawn from two
34
35 learner's dictionaries, as detailed further below. Of these 26 stimulus idioms, 13 show PhS
36
37 (as shortly to be described) and 13 'control' idioms do not (see Table 1). With the exception
38
39 of the delayed post-test, on every occasion that the students encountered the target
40
41 idioms—always in a different random order—there were two filler idioms before the block of
42
43 target idioms and another two at the end. This was to mitigate primacy and recency effects.
44
45 These fillers (*turn back the clock*, *read between the lines*, *reach for the stars*, *bridge the gap*)
46
47 were never presented in the same order twice. Of the PhS idioms, six show alliteration but not
48
49 assonance (e.g., *bite the bullet*). Six show a form of assonance (e.g., rhyme: *pull the wool*
50
51 *over someone's eyes*; near rhyme: *cook the books*; and basic assonance: *jump the gun*). And
52
53 one shows alliteration and assonance (*live off the fat of the land*).
54
55
56
57
58
59
60

1
2
3
4 Attempts were made to control for variables that are known to influence retention but
5
6 which are extraneous to the variables of interest, that is, the presence or absence of PhS and
7
8 of awareness raising followed attention direction. Table 2 summarizes how the two sets of
9
10 idioms compare in terms of length, frequency, and imageability of meaning. The slightly
11
12 lesser average length of the nonPhS idioms was expected to favour their recall very slightly.
13
14 The considerably greater average whole phrase and word frequencies of the PhS idioms
15
16 might be thought likely to favour their recall. However, it has been found that the amount of
17
18 attention paid to a lexical item may correlate *negatively* with familiarity and, therefore, with
19
20 frequency (cf., Tulving and Kroll 1995). Indeed, as will be seen, item frequencies appear to
21
22 have played a negligible role in the outcomes we observed. We were much more concerned
23
24 about the variable ‘imageability of meaning’ (i.e., the degree to which a given idiom conveys
25
26 a sensori-motoric image) since imageability has long been known to be a powerful facilitator
27
28 of lexical recall (Paivio, Yuille, and Madigan 1968). To make the experiment a good test of
29
30 whether PhS can have a positive mnemonic effect, it was essential that the imageability of the
31
32 PhS idioms be equal to or even (to be on the safe side) somewhat less than that of the non-
33
34 PhS idioms. As we know of no published imagery ratings for idioms, we showed 33 L1
35
36 undergraduate Dutch informants a shortlist of 32 candidate idioms and collected from them
37
38 ratings of these idioms on a 7 point scale. Specifically, the informants were each given a copy
39
40 of the study sheet used in the experiment, but with 32 idioms rather than the eventual smaller
41
42 number. It was on this sheet that the informants wrote their imageability ratings. They were
43
44 also given an instruction sheet reading as follows:
45
46
47
48
49
50

51 “Could you give your impression of the imageability of each of the following idioms?

52 Specifically, how much does it suggest to you an image that is visual or which

53
54 involves movement or physical contact. Please give a separate rating for each idiom
55
56
57
58
59
60

1
2
3
4 on a scale of 1 (*difficult* to form an image) to 9 (*extremely easy* to form an image).
5
6 (No half scores, please).”
7

8
9 Informants were asked to consider not only the idioms themselves but also the glosses and
10 the examples of use shown on the study sheet. To minimize overall variation in the amount of
11 attention paid to individual idioms, each informant was randomly assigned a different idiom
12 as a starting point on the study sheet. It turned out that the 16 candidate PhS idioms were
13 rated as more imageable on average than the 16 nonPhS ones. The desired balance was
14 achieved by discarding the three highest rated PhS idioms and the three lowest rated nonPhS
15 ones. Table 1 shows the final list of 26 idioms.
16
17
18
19
20
21
22
23

24
25
26 Table 1 about here.
27

28
29 Table 2 about here
30

31
32 The targeted idioms were drawn from the *Oxford Idioms Dictionary for Learners of*
33 *English* (Parkinson 2006) apart from two drawn from the online version of the *Cambridge*
34 *Advanced Learner's Dictionary and Thesaurus (CALDT)*
35 (<http://dictionary.cambridge.org/dictionary/british/>). The glosses and examples given on the
36 tests and on the study-sheet that participants consulted during the study phase were drawn
37 from Parkinson (2006) except where the definition or example in *CALDT* was clearly
38 superior. For example, Parkinson defines *miss the mark* as “not succeed in achieving or
39 guessing sth”. Because this seemed vague to us, we chose the gloss in *CALDT*: “fail to
40 achieve the result that was intended”. For this idiom we also used the example given by
41 *CALDT*, since Parkinson does not provide one.
42
43
44
45
46
47
48
49
50
51
52
53

54
55 *Procedure*
56
57
58
59
60

1
2
3
4 In a pen and paper pretest administered by one of the authors, participants were asked to
5
6 complete the 26 target idioms along with four fillers. Each of the 30 test items had the same
7
8 format: A single noun was gapped in each targeted idiom, except for that noun's first letter.
9
10 With one exception, the gapped word was always the word following the article *the* (e.g.,
11
12 *REACH FOR THE S _____*). The exception was the idiom *HAVE BIGGER F____ TO FRY*, which
13
14 includes no article. Participants were asked to write their name on their pre-test sheet before
15
16 handing it in.
17
18

19
20 Next, a study booklet was handed out. It consisted of a section for each idiom
21
22 (including the four fillers). Each section presented (a) the complete canonical form of a
23
24 targeted idiom, (b) a definition for it taken from a learner's dictionary, and (c) an illustration
25
26 of usage also taken from a learners' dictionary, for example:
27
28

29
30 **STAY THE COURSE**. Continue doing sth until it has finished or been completed, even
31
32 though it is difficult. "*Very few of the trainees have stayed the course.*"
33
34

35
36
37 **MISS THE MARK**. Fail to achieve the result that was intended. "*Her speech missed the*
38
39 *mark and failed to generate the public support she had been hoping for.*"
40
41
42

43
44 In the noAD condition participants were asked to study the idioms on these sheets. No
45
46 mention was made of sound repetition and students were given no task designed to cause
47
48 them to pay more attention to phonological form than they would if left to their own devices.
49

50
51 In the AD condition the researcher first briefly outlined a common rhetorical function of
52
53 idioms—that of summing up one's attitude to an event or situation (McCarthy 1998, pp. 131-
54
55 149)—and proposed that sound repetition, as in *tear one's hair [out]*, might have rhetorical
56
57 impact over and above any impact of the idiom's meaning alone. The participants were then
58
59
60

1
2
3
4 asked to study the idioms and to mark the keywords in the idiom that share a sound with the
5
6 boxed content word. The researcher recommended to participants that they repeat the target
7
8 idioms subvocally in order to become more aware of any sound repetition. For both groups,
9
10 time was called after about 15 minutes and these sheets were collected.
11
12

13
14 In both conditions the participants were told they would afterwards be tested on their
15
16 knowledge of the form of each of the idioms. The first posttest, which followed immediately,
17
18 consisted of test items in which all the content words of the idiom were gapped except for the
19
20 first letter. The appropriate gloss was repeated from the study-sheet, but not the example of
21
22 usage.
23

24
25 H _____ B _____ F _____ TO F _____. Have more
26
27 important, interesting, or useful things to do.
28
29

30
31 When everyone had finished the test (after approximately ten minutes) the test sheets were
32
33 collected, and the instructor moved on to unrelated matters. A week later, the unannounced
34
35 delayed post-test was given. This test had the same format as the pretest; that is, *one* content
36
37 word in the targeted idiom was gapped except for its first letter, and after the idiom there was
38
39 a gloss. As this was the final test to be given, no fillers were included. When everyone had
40
41 finished the test (after approximately seven minutes) the test sheets were collected.
42
43
44

45 **Design, data analysis, and results**

46
47

48 In our design the outcome variable, recall of form, was measured by the raw scores on three
49
50 tests: (1) a pretest of ability to produce targeted idioms when given a first letter cue for one
51
52 omitted content word; (2) an immediate posttest where multiple content words were omitted
53
54 except for their first letters; and (3) a one-week delayed posttest that was essentially identical
55
56 to the pretest. The independent variable ‘presence / absence of awareness raising and
57
58
59
60

1
2
3
4 attention direction' defined the experimental (AD) and the comparison (noAD) conditions.
5
6 The other independent variable was 'presence / absence of PhS', which was nested under
7
8 'Idioms'. Each participant was exposed to and was tested on each of the 26 targeted idioms,
9
10 meaning that for each participant each test produced two key measures—specifically, one
11
12 aggregated raw score (out of a possible 13) for the PhS idioms and another aggregated raw
13
14 score (out of 13) for the nonPhS idioms. To measure learning from the pretest to a posttest,
15
16 each participant's relevant aggregated raw scores were converted into 'gain scores', which
17
18 were derived from the raw scores according to the standard formula: $GAIN\ SCORE = RAW$
19
20 $POSTTEST\ SCORE - RAW\ PRETEST\ SCORE$. On each posttest each participant had one aggregated
21
22 gain score for the PhS idioms and one for the nonPhS idioms. But it was only the pretest-
23
24 *delayed* posttest aggregated gain scores that were subjected to inferential statistical analysis.
25
26 Finally, for descriptive purposes, each participant's aggregated PhS and nonPhS pretest to
27
28 *delayed* posttest gain scores were converted into a single 'gain difference score', which could
29
30 be positive, negative, or zero, according to the formula: $GAIN\ DIFFERENCE\ SCORE = GAIN$
31
32 $SCORE_{PHS} - GAIN\ SCORE_{NONPHS}$. Descriptive statistics for the AD group are shown in Table 3
33
34 and those for the noAD group are shown in Table 4. Note in particular that while the mean
35
36 pretest scores of the AD and noAD groups are similar, the AD group's total of pretest to
37
38 *delayed* posttest gain scores is higher than that of the noAD group (306 vs 268) and that the
39
40 AD group's gain score subtotal for the PhS idioms is markedly higher than that of the noAD
41
42 group (169 vs 123) while the two groups' subtotals for the nonPhS idioms are much less
43
44 different (137 vs 145).
45
46
47
48
49
50
51

52 Tables 3 and 4 give details for the immediate posttest mainly for completeness:
53
54 Scores from this test were not subjected to inferential analysis because ephemeral learning
55
56 was not the prime concern. However, the pretest to immediate posttest and pretest to *delayed*
57
58
59
60

1
2
3
4 posttest learning trends were basically the same. This is shown by Figures 1 and 2, each of
5
6 which includes one ladder plot for each of the four combinations of +/- attention direction
7
8 (AD) and +/-PhS. In each ladder plot, a diagonal line extending from a pretest score to a
9
10 delayed posttest score represents the learning trajectory of one participant. (There may be
11
12 some superpositioning of lines despite random jittering of lines.) In other words, each ladder
13
14 plot is a profile of gains and each profile in Figure 1 is similar to the corresponding profile in
15
16 Figure 2.¹
17
18

19
20
21 [Table 3 about here]
22

23
24 [Table 4 about here]
25

26
27 [Figure 1 about here beneath its caption]
28

29
30 [Figure 2 about here beneath its caption]
31
32

33
34 As the Anderson-Darling test of distributional normality ($\alpha = .20$) red-flagged a
35
36 number of the pertinent score sets, we tested the difference between participants' PhS and
37
38 nonPhS pretest to delayed posttest aggregated gain scores using Rand Wilcox's (2012, pp.
39
40 406-407) 'yuend' R function which runs a robust version of the paired t -test.² Key results are
41
42 as follows, where all tests were two-sided and $\alpha = .05$.
43
44

45
46 *AD group: MD = 1.23, $t = 2.14_{(25)}$, $p = .042$.*³ The estimated effect size is: $d = 0.59$,
47
48 95% CI [0.03, 1.14]. The latter interval ranges from a trivial positive effect to one that is
49
50 large. However, the point estimate ($d = 0.59$), which is indicative of a strongish medium
51
52 effect, is about seven times more likely than the value at either limit (Cumming, 2012: 99-
53
54 100) and is well above average for effects observed in educational research generally
55
56 (Grissom and Kim, 2012: 127-130).
57
58
59
60

1
2
3
4 *noAD group: MD = -0.92, t = -1.78₍₂₃₎, p = .088, d = -0.51 [-1.09, 0.07].*⁴ The *d* value
5
6 here is consistent with an effect of medium size, albeit in the opposite direction to that seen in
7
8 the AD condition.
9

10
11 Finally, we have not yet mentioned how accurately participants in the AD condition
12 marked the PhS idioms as such or on how well they refrained from marking nonPhS idioms.
13
14 With 26 participants and 13 idioms in each of the two sets, 338 is the maximum possible
15
16 number of correct marks. The number of marks given the PhS idioms was 236 (69% of the
17
18 total possible). NonPhS idioms were wrongly marked as PhS 27 times (8% of the maximum
19
20 possible).
21
22
23
24

25 26 **Discussion**

27
28 Let us now consider the three research questions (RQs) in turn. Regarding RQ 1, our data is
29
30 not consistent with PhS idioms being inherently more memorable since the noAD group
31
32 remembered those idioms markedly less well than they did the nonPhS idioms. RQ 2 asks
33
34 whether PhS idioms can be rendered extra-memorable for learners who experience some
35
36 awareness raising about patterns of PhS and also a task that induces them to look out for
37
38 instances of PhS in idioms that they encounter. Our results suggest that the answer to this
39
40 question is 'Yes'. (In Figure 2, the ladder plot for the AD group, top right, runs higher than
41
42 the plot for any other group.) Finally, RQ 3 asks whether the awareness raising and attention
43
44 direction experienced by the AD group detracted from their learning of the nonPhS idioms.
45
46 Because we did not form the AD and noAD groups by random assignment of participants,
47
48 there is a less than optimal basis for direct comparisons of one group with the other.
49
50
51

52
53 However, the two groups of participants were alike in many respects and, as can be seen in
54
55 Table 3 and 4, their mean pretest scores for both types of idiom are very similar, as are their
56
57
58
59
60

1
2
3
4 gains on the nonPhs idioms (137_{AD} vs 145_{noAD}), where the disparity is far less than on the
5
6 PhS idioms (169_{AD} vs 123_{noAD}). True, the noAD group's mean pretest to delayed posttest gain
7
8 score for the nonPhS idioms is 5.8% higher than that of the AD participants. But let's refer to
9
10 the two plots in the left column of Figure 2: They show very similar gain profiles for the
11
12 nonPhS for both learning conditions; and both profiles are consistent with pretty good
13
14 learning of the nonPhS idioms over the period of the experiment. To approach RQ3 in a
15
16 different way, recall that it was the AD group that made the greatest *total* learning gains (i.e.,
17
18 for PhS and nonPhS idioms together). That is, for the AD group there are 306 gain scores vs
19
20 268 for the noAD group, for a superiority of 114% ($306_{AD} / 268_{noAD} = 1.14$). All in all, there
21
22 is no evidence that the experimental treatment engendered a trade-off effect, let alone one
23
24 large enough to be substantively important.
25
26
27
28

29
30 It may be asked why the noAD participants remembered the nonPhS idioms so much
31
32 better than the PhS idioms, which was a result that we did not expect. Both Figure 1 and in
33
34 Figure 2 it can be seen that the lower-right plot (noAD, +PhS) is different from all the others.
35
36 Looking at Figure 2 we see three rather flat diagonal lines at the bottom of the score range.
37
38 These indicate that three of the noAD participants who had very low pretest scores made
39
40 especially poor gains on PhS idioms at the delayed posttest: There is even one case of
41
42 decline. If the scores of these three noAD group participants are discarded, the point estimate
43
44 of d for the noAD condition leaps toward zero ($-0.51 \rightarrow -0.24$). We cannot, however,
45
46 pinpoint a reason for the overall inferior memorability of the PhS idioms in the noAD
47
48 condition. It may or may not be the case that some of the PhS idioms just happen to be ones
49
50 that learners, perhaps lower proficiency learners in particular, find hard to remember when
51
52 there has been no prior awareness raising and direction of attention.
53
54
55
56
57
58
59
60

1
2
3
4 Further above we referred to Table 2, in which it can be seen that, compared to the
5 nonPhS idioms, the PhS idioms have greater median frequency values for whole idioms, all
6 content words, and words intended as responses in the pretest and delayed posttest (Table 2).
7
8 To get an idea of the roles of these three nuisance variables in our study we checked the
9
10 association between pretest to delayed posttest aggregated gain scores and each of these
11
12 measures of frequency. Like most sets of frequencies of lexical items, the ones at issue here
13
14 show pronounced positive skew. We therefore followed normal practice for frequency data
15
16 and reduced the skew in each data set by applying the logarithmic transformation (Baayen
17
18 2006). When we applied the Anderson-Darling normality test to the transformed frequency
19
20 data, the appearance of distributional nonnormality persisted. Therefore, to calculate
21
22 correlations between the transformed frequencies and the gain scores associated with each of
23
24 the 26 targeted idioms we used the robust percentage bend correlation (r_{pb}), rather than
25
26 Pearson's r (or r_s), since r_{pb} closely tracks r under normality but tends to give a truer
27
28 (potentially far truer) estimate of the population correlation under nonnormality (Wilcox
29
30 2005). We then squared each value of r_{pb} in order to obtain an analog of r^2 , the common
31
32 measure of explanatory power (or variance explained). The results are: whole idiom
33
34 frequency, $r_{pb} = .065$ and $r_{pb}^2 = .004$; logged mean content word frequency: $r_{pb} = .093$ and r_{pb}^2
35
36 $= .009$; and logged response word frequency: $r_{pb} = -.046$ and $r_{pb}^2 = .002$. These values of r_{pb}^2
37
38 are indicative of very small or negligible effects. Moreover, the underlying values of r_{pb} do
39
40 not all have the same sign. It seems rather unlikely then that the greater average frequency of
41
42 the PhS idioms played a telling role in the learning advantage shown by the PhS idioms. This
43
44 is in line with the finding already referred to that people are likely to pay less attention to
45
46 familiar phenomena than to phenomena which are unfamiliar (Tulving and Kroll 1995). The
47
48 means by which the variables of idiom length and imageability were controlled for have
49
50 already been described.
51
52
53
54
55
56
57
58
59
60

1
2
3
4 Finally, in this article so far we have looked at scores matched to participants. But
5
6 such scores can also be matched to language items (idioms in the present case). When this is
7
8 done, each idiom is matched to *one* aggregated gain score that represents the number of
9
10 students who learned the idiom between the pretest and the delayed posttest. Since there were
11
12 two kinds of idioms in the experiment, this way of organizing the gain scores puts them in
13
14 two *independent* sets, one set consists of the 13 gain scores for the PhS idioms and the other
15
16 consists of the 13 gain scores for the nonPhS idioms. Because this alternative design is
17
18 strictly between-items and because the effective sample size is now half what it is in the per-
19
20 participants format (i.e. 13 instead of 26), the statistical power that can now be brought to
21
22 bear is greatly reduced. To elaborate, this alternate design affords a reasonable chance (i.e.,
23
24 80%) of detecting a truly existing effect at $p \leq .05$ only if this effect equates to $d = 1.15$, nearly
25
26 double our highest estimate of $d \approx 0.60$.⁵ Although in L2 research the by-idioms perspective
27
28 is rarely taken into account along with the by-participants perspective (unlike
29
30 psycholinguistic research), this alternative perspective should not be ignored since it co-
31
32 determines the extent to which particular experimental findings are generalizable to different
33
34 learners and to different idioms.⁶ Because inspection of the relevant by-idiom data indicated
35
36 that the assumptions of distributional normality and equal variances cannot in all cases be
37
38 taken for granted, these data were analyzed using an independent samples bootstrap Welch's
39
40 t -test based on the means, with 10000 bootstrap replications for each test (Wilcox 2012, 338-
41
42 343).⁷ The key statistics for the AD group are: $MD = 2.46$, $t = 1.35$, $p = .188$, $d = 0.50$ [-0.30,
43
44 1.30].⁸ If we refer back to the corresponding results from the by-participants analysis, we see
45
46 that although the p value here is over four times as large, the observed effect size can still be
47
48 termed 'medium'. The statistics for the noAD group are: $MD = -1.54$, $t = -0.88$, $p = .383$, $d =$
49
50 -0.33 [-1.12, 0.46]. Here too, compared to the corresponding by-participants analysis, the p
51
52
53
54
55
56
57
58
59
60

1
2
3
4 value is much higher while the observed effect size has remained about the same (i.e.,
5
6 smallish).
7
8

9 **Conclusion**

10
11
12 On the delayed posttest, participants in the AD condition recalled more PhS idioms than
13
14 nonPhS idioms. In the noAD condition the direction of the trend was reversed even though
15
16 the stimulus idioms remained the same. The cause of this reversal is likely to have been the
17
18 absence from the noAD condition of a pedagogical intervention focusing on phonological
19
20 similarity. The fact that the nonPhS idioms were *better* recalled than the PhS idioms in the
21
22 noAD condition, where participants were left to study the stimulus idioms any way they
23
24 wanted, makes it all the more plausible that the outcome in the AD condition is directly
25
26 attributable to the pedagogical intervention. All in all, our findings are very much at odds
27
28 with any possibility that alliteration and assonance may have a practically significant
29
30 mnemonic effect in the absence of a pedagogical intervention such as that seen in the AD
31
32 condition. There remains our third research question, that of whether a pedagogical focus on
33
34 alliteration and assonance may adversely affect the learning of idioms that show no such
35
36 pattern of sound repetition. We found no evidence of such a trade-off effect. A key fact in
37
38 this regard is that when pretest to delayed posttest gain scores for both types of idiom were
39
40 counted together it was found that the AD group made the greater overall gain.
41
42
43
44
45

46
47 In the results section above we mentioned that scores can be averaged both across
48
49 participants and across idioms. We mentioned too that the best basis for generalization from
50
51 the findings of a study such as ours comes from a statistical analysis that takes both of these
52
53 perspectives into account. Although we have given results of an analysis from each
54
55 perspective, both of our analyses are rather traditional. Moreover, the results of the by-idioms
56
57 analysis are less strong than those of the by-participants analysis. Our results, therefore, are in
58
59
60

1
2
3
4 particular need of replication, not just with new participants but also with new idioms. Still,
5
6 our results chime with those of other studies indicating that it can be worthwhile for teachers
7
8 to take a small amount of extra time in class to raise learners' awareness of patterns of sound
9
10 repetition in targeted idioms and perhaps also to stage relevant attention direction exercises
11
12 (e.g., Author, 2005; Author, 2014d).
13
14
15

16 Finally, a question that our study has not addressed is, 'Which is likely to have the
17
18 greater positive effect on memory for form: awareness raising along with attention direction
19
20 or just awareness raising?' Also not addressed, or at least not directly, is the question of
21
22 whether or for how long afterwards learners will autonomously profit from in-class
23
24 awareness raising and attention direction with respect to sound repetition in idioms.
25
26

27 Depending on what might be found out in this regard, it could be fruitful to carry out pairwise
28
29 comparisons of the effectiveness of some of the techniques of awareness raising and attention
30
31 direction that have been suggested in teachers' resource books (e.g., Author 2008c). Suppose,
32
33 for instance, that a class of students propose a list of L2 idioms that they would like to learn.
34
35 Among other things, the teacher could ask the learners to form pairs or threes and sort the
36
37 idioms into four groups—idioms that include alliteration, ones that include assonance
38
39 (including rhyme), ones that include both, and ones that include neither—and following this
40
41 the teacher could lead a plenary discussion of the groupings. In an alternative exercise, the
42
43 class divides into A-B pairs; each A student gets a handout showing half the targeted
44
45 expressions, with each idiom in a short, representative context; each B student gets a handout
46
47 showing the other half of the expressions, each in a short context; on each handout the loci of
48
49 repeated sounds have already been highlighted (e.g., underlined) by the teacher; students A
50
51 and B dictate their phrases to each other; they then check each other's writing and add in the
52
53 highlighting of the sound repetitions; to finish, they discuss the sound repetitions and consult
54
55
56
57
58
59
60

1
2
3
4 the teacher if they have any doubts. An experiment that took into consideration any
5
6 difference in running time between two targeted exercise types could compare their
7
8 effectiveness as means of promoting productive knowledge of targeted expressions.
9
10

11 Notes

- 14 1. In each of these figures there are different orthographic versions of the labels *pre* and
15 *post* (e.g., *Pre* and *pre*). If this were not so, the software would have confused the data
16 that these labels refer to.
17
18
- 19 2. 'R' refers to the statistical freeware (<http://www.r-project.org/>) (R Core Team, 2015).
20
21
- 22 3. To check this result we conducted an even more robust version of this test using the
23 10% trimmed mean instead of the trimmed mean, finding: $p = .036$.
24
25
- 26 4. For all paired measures, calculation of Cohen's d (using pooled SDs) were based on
27 the t -statistic from the robust paired t -test, two-sided. A more conservative method
28 (Dunlap, Cortina, Vaslow, and Burke 1996), yields slightly lower values, i.e., $d = 0.53$
29 instead of 0.59 (AD group) and -0.38 instead of -0.51 (noAD group).
30
31
- 32 5. Wuensch (2009) explains very clearly how to explore this kind of issue using
33 G*Power (<http://www.gpower.hhu.de/en.html>), statistical freeware developed by
34 Faul, Erdfelder, Lang, and Buchner (2007).
35
36
- 37 6. Baguley (2012, pp. 725-730) gives a good account of the issues here; see also Baayen
38 (2006, Chapter 7.)
39
40
- 41 7. Welch's t -test, the default in R, does not assume that variances are equal.
42
43
- 44 8. The same bootstrap based on the 10% trimmed mean found $p = .165$. Incidentally,
45 bootstraps do not involve degrees of freedom.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4 **References**
5
6

7 Author (2005)
8
9

10 Author (2006)
11
12

13 Author (2008a)
14
15

16 Author (2008b)
17
18

19 Author (2008c)
20
21

22 Author (2009)
23
24

25 Author (2012)
26
27

28 Author (2014a)
29
30

31 Author (2014b)
32
33

34 Author (2014c)
35
36

37 Author (2014d)
38
39

40 Author (2014e)
41
42

43 Baayen, R. (2006). *Analyzing Linguistic Data*. Cambridge, UK: Cambridge University Press.

44 Baguley, T. (2012). *Serious stats: A guide to advanced statistics for the behavioral sciences*.

45
46
47
48 Basingstoke, UK: Palgrave Macmillan.

49 Boers, F., & Stengers, H. (2008). A quantitative comparison of the English and Spanish

50
51
52 repertoires of figurative idioms. In F. Boers & S. Lindstromberg (Eds.), *Cognitive*

53
54
55 *Linguistic Approaches to Teaching Vocabulary and Phraseology* (pp. 355–374).

56
57
58 Berlin, Germany: Mouton Degruyter
59
60

- 1
2
3
4 Bower, G., & Bolton, L. (1969). Why are rhymes easy to learn? *Journal of Experimental*
5
6 *Psychology*, 82, 453-461.
7
8
9 *Cambridge Advanced Learner's Dictionary and Thesaurus*. No date. Cambridge, UK:
10
11 Cambridge University Press. Retrieved from
12
13 <http://dictionary.cambridge.org/dictionary/british/>
14
15
16 Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.).
17
18 London, UK: Erlbaum.
19
20
21 Cumming, G. (2011). *ESCI Meta-analysis. Exploratory Software for Confidence Intervals*.
22
23 Retrieved from <http://www.latrobe.edu.au/psy/documents/ESCI-readme.pdf>
24
25
26 Cumming, G. (2012). *Understanding the New Statistics: Effect Sizes, Confidence Intervals,*
27
28 *and Meta-analysis*. New York, NY: Routledge.
29
30
31
32 Davies, M. (2014). *The Corpus of Contemporary American English*. Retrieved from
33
34 <http://www.americancorpus.org>
35
36
37 Dunlap, W., Cortina, J., Vaslow, J., & Burke, M. (1996). Meta-analysis of experiments with
38
39 matched groups or repeated measures designs. *Psychological Methods*, 1, 170-177.
40
41
42 Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical
43
44 power analysis program for the social, behavioral, and biomedical sciences. *Behavior*
45
46 *Research Methods*, 39, 175-191.
47
48
49
50 Gibbs, R. W. (1994). *The Poetics of Mind: Figurative Thought, Language and*
51
52 *Understanding*. Cambridge, UK: Cambridge University Press.
53
54
55 Grant, L. (2007). In a manner of speaking: Assessing frequent spoken figurative idioms to
56
57 assist ESL/EFL teachers. *System*, 35, 169-181.
58
59
60

- 1
2
3
4 Gries S. (2011). Phonological similarity in multi-word symbolic units. *Cognitive Linguistics*,
5
6 22, 491-510.
7
8
- 9 Grissom, R., & Kim, J. (2012). *Effect Sizes for Research: Univariate and Multivariate*
10
11 *Applications* (2nd ed.). New York, NY: Routledge.
12
13
- 14 Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory:
15
16 Computational and neural bases. *Brain and Language*, 59, 267–333.
17
18
- 19 Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect
20
21 in immediate serial recall: The roles of type of similarity, category cuing, and item
22
23 recall. *Memory and Cognition*, 33, 1001–1016.
24
25
- 26 Herrera, H., & White, M. (2010). Canonicity and variation in idiomatic expressions:
27
28 evidence from business press headline. In S. De Knop, F. Boers, & A. De Rycker
29
30 (Eds.), *Fostering Language Teaching Efficiency through Cognitive Linguistics*
31
32 (pp. 1677-187). Berlin, Germany: Mouton de Gruyter.
33
34
- 35 Levelt, W. (1999). Models of word production. *Trends in Cognitive Sciences*, 3, 223-232.
36
37
- 38 Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A
39
40 longitudinal case study. *Journal of Second Language Writing*, 18, 85-102.
41
42
- 43 Littlemore, J. (2001). Metaphor as a source of misunderstanding for overseas students in
44
45 academic lectures. *Teaching in Higher Education*, 6, 333-351.
46
47
- 48 Luce, P., Goldinger, S., Auer, E., & Vitevitch, M. (2000). Phonetic priming, neighborhood
49
50 activation, and PARSYN. *Perception and Psychophysics*, 62, 615–625.
51
52
- 53 McCarthy, M. (1998). *Spoken Language and Applied Linguistics*. Cambridge, UK:
54
55 Cambridge University Press.
56
57
58
59
60

- 1
2
3
4 Nadel, L., & Hardt, O. (2011). Update on memory systems and processes.
5
6 *Neuropsychopharmacology*, 36, 251-273.
7
8
9 Nelson, D., & Garland, R. (1969). Amount and locus of stimulus-response overlap in paired
10
11 associate acquisition. *Journal of Experimental Psychology*, 82, 297-300.
12
13 Nunberg, G., Sag, I., & Wasow, T. (1994). Idioms. *Language*, 70, 491–538.
14
15 Paivio, A., Yuille, J., Madigan, S. (1968). Concreteness, imagery, and meaningfulness values
16
17 for 925 nouns. *Journal of Experimental Psychology*, 76, 1-25.
18
19
20 Parkinson, D., (Ed.) (2006). *Oxford Idioms Dictionary for Learners of English*, 2nd ed.
21
22 Oxford, UK: Oxford University Press.
23
24
25 Philip, G. (2000). An idiomatic theme and variations. In C. Heffer, & H. Sauntson (Eds.)
26
27 *Words in Context: A Tribute to John Sinclair on his Retirement* (pp. 221-233). ELR
28
29 Monograph 18. The University of Birmingham.
30
31
32
33 Philip, G. (2005). Discovering clines of variation: the location and analysis of non-
34
35 canonical forms in general reference corpora. Presented at: Phraseology 2005:
36
37 The many faces of Phraseology. An interdisciplinary conference, 13-15 October
38
39 2005, Louvain-la-Neuve, Belgium. Retrieved from
40
41 <http://amsacta.cib.unibo.it/1132/1/VariationInCorpus.pdf>
42
43
44
45 Qi, Y., & Ding, Y. (2011). Use of formulaic sequences in monologues of Chinese EFL
46
47 learners. *System*, 39, 164-174.
48
49
50 R Core Team. (2015). R: A Language and Environment for Statistical Computing. R
51
52 Foundation for Statistical Computing, Vienna, Austria. Retrieved from [http://www.R-](http://www.R-project.org/)
53
54 [project.org/](http://www.R-project.org/)
55
56
57
58
59
60

- 1
2
3
4 Rubin, D. (1995). *Memory in Oral Traditions: The Cognitive Psychology of Epic, Ballads,*
5
6 *and Counting-out Rhymes.* Oxford, UK: Oxford University Press.
7
8
- 9 Storkel, H., & Morrisette, M. (2002). The lexicon and phonology: Interactions in language
10
11 acquisition. *Language, Speech, and Hearing Services in Schools, 33,* 24-37.
12
13
- 14 Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.),
15
16 *Organization of Memory* (pp. 381-403). New York, US: Academic Press.
17
18
- 19 Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory
20
21 encoding. *Psychonomic Bulletin and Review, 2,* 387-390.
22
23
- 24 Watkins, M., Watkins, O., & Crowder, R. (1974). The modality effect in free and serial recall
25
26 as a function of phonological similarity. *Journal of Verbal Learning and Verbal*
27
28 *Behavior, 13,* 430-447.
29
30
- 31
32
33 Wilcox, R. (2005). *Introduction to robust estimation and hypothesis testing,* 2nd edn.
34
35 Waltham, MA: Academic Press.
36
- 37 Wilcox, R. (2012). *Modern statistics for the social and behavioral sciences: A practical*
38
39 *introduction.* Boca Raton, FL: CRC Press.
40
- 41 Wuensch, K. (2009). An overview of power analysis. Greenville, NC: East Carolina University.
42
43 Available at: core.ecu.edu/psyc/wuenschk/stathelp/PowerAnalysis_Overview.doc
44
- 45 Zyzik, E. (2011). Second language idiom learning: The effects of lexical knowledge and
46
47 pedagogical sequencing. *Language Teaching Research, 15,* 413-433.
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. The 26 targeted idioms

	PHONOLOGICALLY SIMILAR (i.e., ALLITERATIVE / ASSONANT) IDIOMS	NONPHONOLOGICALLY SIMILAR CONTROL IDIOMS
1	<i>bite the bullet</i>	<i>bet the farm</i>
2	<i>bring home the bacon</i>	<i>burn the midnight oil</i>
3	<i>cook the books</i>	<i>feel the pinch</i>
4	<i>get the show on the road</i>	<i>hit the spot</i>
5	<i>have bigger fish to fry</i>	<i>hold the purse strings</i>
6	<i>jump the gun</i>	<i>keep the wolf from the door</i>
7	<i>live off the fat of the land</i>	<i>keep your eye on the ball</i>
8	<i>miss the mark</i>	<i>learn the ropes</i>
9	<i>press the panic button</i>	<i>paint the town red</i>
10	<i>pull the plug on</i>	<i>shoot the messenger</i>
11	<i>pull the wool over sb's eyes</i>	<i>sort out the sheep from the goats</i>
12	<i>set the scene for</i>	<i>stay the course</i>
13	<i>strike the right note</i>	<i>take the bull by the horns</i>

Table 2. Comparison of the targeted phonologically similar (PhS) and nonphonologically similar (nonPhS) idioms

	<i>PhS</i>	<i>nonPhS</i>
Number of words (Mean)	4.31	4.23
Number of content words (Mean) ^a	2.62	2.54
Number of syllables (Mean)	4.77	4.46
Internal <i>somebody</i> / <i>someone</i> ^b	1	0
Frequency of whole idiom (median) ^c	14	9
Total frequency of all content words (median)	44,229	30,641
Frequency of words intended as responses in cued recall (median) ^d	43,829	13,314
Mean imagery rating / Standard deviation	5.84 / 1.24	5.97 / 1.45

Notes.

- a. Includes nouns, verbs, and adjectives.
- b. This was counted as a noncontent word of one syllable.
- c. As given by the online British National Corpus Simple Search facility:
<http://www.natcorp.ox.ac.uk/using/index.xml?ID=simple>
- d. Lemma frequencies are as given by the Corpus of Contemporary American English (<http://corpus.byu.edu/coca/>) (Davies, 2008-2015) in September, 2013.

Table 3. Scores of the AD (experimental) group ($n = 26$) for the PhS (phonologically similar) and the nonPhS idioms.

SCORE SET		SUM / SD	MEAN / MEDIAN	NR OF MAXIMUMS	INTER- QUARTILE RANGE
PRETEST	PhS	98 / 2.16	3.77 / 3.00	0	2.25—4.75
SUM = 200	nonPhS	102 / 1.67	3.92 / 4.00	0	3.00—5.00
IMMEDIATE	PhS	272 / 1.86	10.46 / 11.00	5	9.25—11.75
POSTTEST	nonPhS	266 / 1.88	10.23 / 10.50	3	9.00—11.75
SUM = 538					
DELAYED	PhS	267 / 1.56	10.27 / 10.50	3	9.00—11.00
POSTTEST	nonPhS	239 / 2.00	9.19 / 9.00	0	8.00—11.00
SUM = 506					
GAINS	PhS	169 / 2.40	6.50 / 6.50	0	4.25—8.00
SUM = 306	nonPhS	137 / 2.15	5.27 / 5.00	0	4.00—6.00
GAIN DIFFERENCES		32 / 2.93	1.23 / 1.00	0	-1.00—3.00

Table 4. Scores of the noAD (comparison) group ($n = 24$) for the PhS and the nonPhS idioms.

SCORE SET		SUM / SD	MEAN / MEDIAN	NR OF MAXIMUMS	INTER- QUARTILE RANGE
PRETEST	PhS	100 / 2.41	4.17 / 4.00	0	2.00—5.25
SUM = 198	nonPhS	98 / 1.44	4.08 / 4.00	0	3.00—5.00
IMMEDIATE	PhS	257 / 3.26	10.71 / 12.00	8	10.00—13.00
POSTTEST	nonPhS	277 / 2.08	11.54 / 12.00	10	11.00—13.00
SUM = 534					
DELAYED	PhS	223 / 2.96	9.29 / 10.00	3	7.00—11.25
POSTTEST	nonPhS	243 / 2.13	10.13 / 10.50	3	8.00—12.00
SUM = 468					
GAINS	PhS	123 / 2.72	5.13 / 5.00	0	3.75—7.00
SUM = 268	nonPhS	145 / 1.90	6.04 / 6.00	0	5.00—7.00
GAIN DIFFERENCES		-22 / 2.52	-0.92 / -1.00	0	-3.00—0.25

Figure 1. Gain profiles for the AD condition (top row) and the noAD condition (bottom row) and for the nonPhS idioms (left column) and the PhS idioms (right column), where *Pre* means raw pretest score and *Post1* means raw immediate posttest scores, and where *r* gives the correlation between pretest and posttest scores.¹

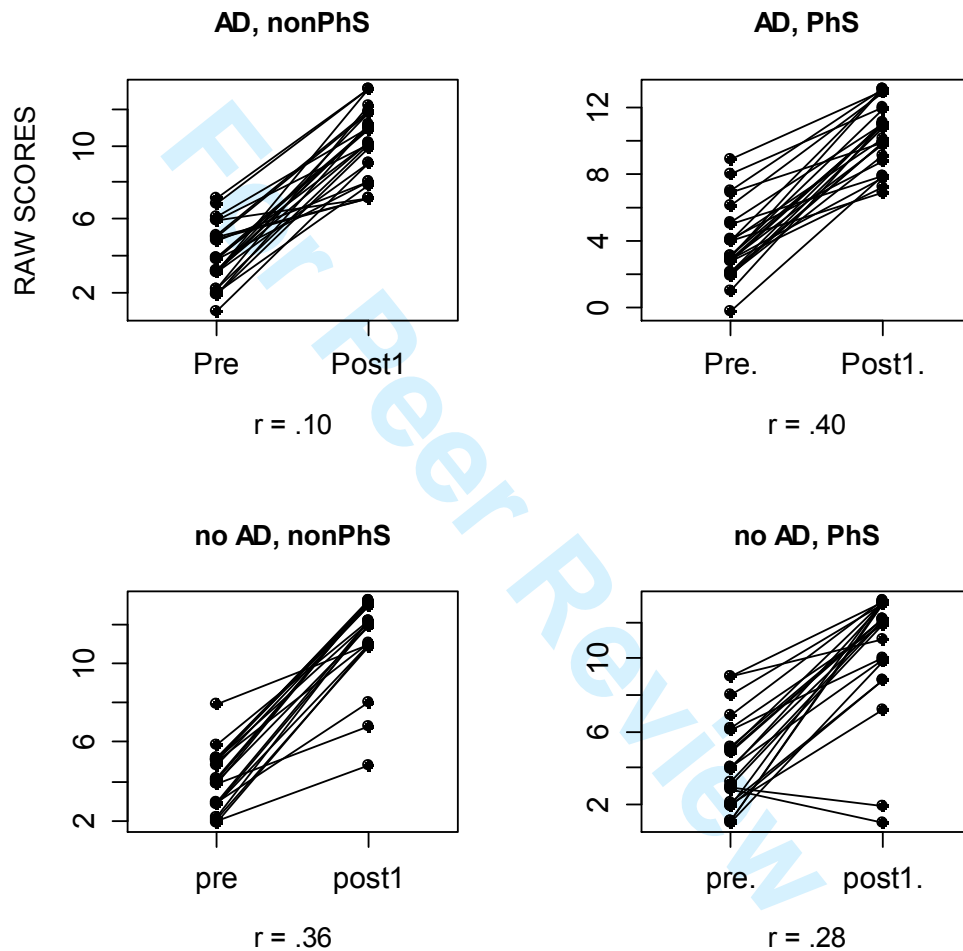


Figure 2. Gain profiles for the AD condition (top row) and the noAD condition (bottom row) and for the nonPhS idioms (left column) and the PhS idioms (right column), where *Pre* means raw pretest score and *Post2* means raw delayed posttest scores.

