

2017

Comparing the Effectiveness of Phrase-Focused Exercises. A Partial Replication of Boers, Demecheleer, Coxhead, and Webb (2014)

Frank Boers
fboers@uwo.ca

Tu Cam Thi Dang
Hue College of Foreign Languages, Viet Nam

Brian Strong
Victoria University of Wellington

Follow this and additional works at: <https://ir.lib.uwo.ca/edupub>

 Part of the [Education Commons](#)

Citation of this paper:

Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises: A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, 21(3), 362-380.

Comparing the effectiveness of phrase-focused exercises. A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014)

Frank Boers, Tu Cam Thi Dang and Brian Strong

Abstract

In a recent article, Boers, Demecheleer, Coxhead, and Webb (2014) deplored the lack of effectiveness for the learning of verb-noun collocations of a number of exercise formats which they sampled from EFL textbooks and put to the test in a series of quasi-experimental trials. The authors called for further investigations into possible improvements to such exercise formats. The present article is a response to that call. It also addresses methodological issues which may have affected Boers et al.'s (2014) findings and which rendered their conclusions tentative. In the quasi-experiment reported here, EFL learners were given fill-in-the-blank exercises on verb-noun phrases in one of three formats: (1) choose the appropriate verb, (2) complete the verb by using a first-letter cue, and (3) choose the appropriate intact phrase. A delayed post-test gauged the learners' ability to recall the meaning of the phrases as well as their verb-noun partnership. In both regards the exercise where learners worked with intact phrases generated the best results. We then evaluate the extent to which exercises for phrase learning in ten recent EFL textbooks accord with recommendations that follow from the quasi-experimental findings.

Keywords:

collocations, idioms, lexical phrases, textbook exercises, errorless learning, trial-and-error, interference.

I Introduction

The past three decades have witnessed a growing interest in the phraseological or formulaic dimension of language (e.g., Polio, 2012; Sinclair 1991; Siyanova-Chanturia & Martinez, 2015; Wray 2002) and its relevance for second language learners (e.g., Barfield & Gyllstad, 2009; Lewis, 1993; Meunier & Granger, 2008; Nattinger & Decarrico, 1992; Schmitt, 2004; Wood, 2010). Indeed, language abounds with a panoply of conventional word strings (e.g., Erman & Warren, 2010), which have in the literature been labelled variously as lexical phrases, multiword units, formulaic sequences, prefabricated chunks, expressions, idioms, word partnerships, collocations, and more, and for which we shall adopt the umbrella term ‘phrases’ in the present article. It is undeniable that the challenge of second or foreign language learning includes the challenge of mastering this phraseological dimension (Pawley & Syder, 1983). Several studies have furnished evidence that a good command of phraseology helps learners come across as native-like speakers (e.g., Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006) and writers (e.g., Dai & Ding, 2010). Familiarity with a large repertoire of phrases is also strongly associated with receptive fluency (e.g., Ellis, Simpson-Vlach and Maynard, 2008; Siyanova-Chanturia, Conklin, & Van Heuven, 2011; Sonbul, 2015).

Unfortunately, in the absence of massive amounts of exposure to the target language (which is typical of many foreign-language-learning contexts as compared to immersion contexts), learners tend to be slow at acquiring its phraseological dimension (e.g., Laufer and Waldman 2011; Li and Schmitt 2010). Diverse interventions intended to accelerate phrase learning have therefore been examined in recent years (see Boers & Lindstromberg, 2012, for a review). These range from the manipulation of texts so as to ensure repeated encounters with the same phrase (e.g., Pellicer-Sanchez, 2015; Webb, Newton & Chang, 2013) and/or to

make selected phrases more visually salient (e.g., Sonbul & Schmitt, 2013; Szudarski and Carter, 2014) to explicit phrase-focused language study (e.g., Boers, Eyckmans & Stengers, 2007; Eyckmans, Boers & Lindstromberg, 2016; Laufer & Girsai, 2008; Peters, 2016).

The present article investigates a type of intervention with direct relevance for the mainstream language classroom – the use of phrase-focused exercises of the kind one finds in contemporary course books. These are exercises on worksheets where learners are required, for example, to assemble phrases by matching jumbled-up constituent parts, to supply missing constituents of phrases in gapped sentences, or to match phrases with single-word synonyms. In essence, these kinds of exercise formats require the learner to distinguish between correct and incorrect word combinations and/or between correct or incorrect form-meaning mappings. It seems to be assumed by the designers of these materials that doing such exercises will lead learners to make a mental note of which associations are correct and thus to be retained in memory, and which associations are wrong and thus to be dismissed and forgotten.

Considering that exercises on various sorts of phrases are now commonplace in course books, it is surprising how little empirical research has so far been conducted to test the effectiveness of these exercise. One recent attempt to do this is Boers, Demecheleer, Coxhead and Webb (2014), a study which focused on verb-noun collocations. In a series of four trials with different cohorts of ESL students, students' knowledge of collocations was first gauged in a pre-test, which consisted of gapped sentences where the verb (e.g., *make*, *commit*) was missing before a given noun phrase (e.g., *a suggestion*, *a crime*). As part of their coursework in class, the students were later given exercises on the same verb-noun collocations. The exercises mimicked exercise formats (e.g., matching verbs and nouns to form word partnerships) which the authors had found in various course books and teacher manuals. After finishing each exercise, the students were given corrective feedback. A post-test, which

was identical to the pre-test, was administered two to three weeks later. In all four trials, the comparisons of pre-test and post-test performance revealed only marginal learning gains. Post-test scores were typically only between five and ten per cent better than the pre-test scores. While most students learned a few new collocations from the exercises, the exercises also appeared to create confusion in the students' minds about collocations which they had actually shown correct intuitions about in the pre-test. The study also found that, when learners made a wrong choice in the exercise, they were highly likely to make a mistake also in the post-test, and this on items where their pre-test response had been correct. This supports arguments in favour of learning practices in which the rate of error is deliberately kept minimal, an approach which is in keeping with a strand of memory research that has found errorless learning to be superior to learning through trial-and-error (Baddeley & Wilson, 1994; Warmington, Hitch, and Gathercole 2013; Warmington and Hitch 2014).

To further explore the possibility that trial-and-error might not be the most judicious procedure when it comes to learning verb-noun partnerships, Stengers and Boers (2015) set up a pre-test – post-test experiment with L2 learners of Spanish who were assigned either to a trial-and-error or an errorless exercise condition. In the trial-and-error condition, the learners were asked to supply the missing verbs before their noun partners in gapped sentences without any assistance. On completion of the exercise, they were given a list of exemplars of the collocations and asked to correct any mistakes they had made. In the errorless condition, by contrast, the learners were given that list of exemplars alongside the exercise sheet, and so they could consult the list to avoid making mistakes. The latter procedure actually resembles the approach found in McCarthy and O'Dell's (e.g., 2002, 2005) books for independent study of phrases, where users can consult explanations and examples (on the left-hand page) as they tackle the exercises (on the opposite, right-hand page). The former, trial-and-error procedure, on the other hand, resembles the approach taken by many other materials writers (see further

below), where learners are expected to rely on prior knowledge, to make ‘educated’ guesses and to use elimination strategies before seeking feedback from the teacher or from an answer key. While this probably raise learners’ awareness of the challenging nature of phraseology in general and of gaps in their knowledge of the targeted phrases in particular, it also invites them to temporarily ponder word combinations (or form-meaning mappings) which – if they turn out wrong – subsequently need to be overridden by some form of corrective feedback.

However, Stengers and Boers (2015) found very little evidence for the effectiveness of corrective feedback where participants in their trial-and-error condition made mistakes – only 15% of the corrected exercise responses were followed by correct responses in a two-week delayed post-test (which had the same, gapped-sentence format as the exercise). The gains from pre-test to post-test were better (albeit not significantly so) under the exemplar-given, errorless procedure, but were far from spectacular either – 18%. The authors argue this is probably due to the lack of cognitive investment required by this procedure (participants could simply copy the right responses from the list of exemplars). They therefore call for further research that examines ways of keeping phrase-focused exercises sufficiently challenging while at the same time minimizing the risk of error. Part of the quasi experiment reported further below is a response to that call. Before moving on to that report, however, we need to explain in somewhat more detail why the aforementioned studies – Boers et al. (2014) in particular – require replication and in what ways the present study is different.

II Motives for a partial replication of Boers et al. (2014)

Apart from the general need for more replication research (Porte, 2012), there are several reasons why the quasi-experimental trials reported in Boers et al. (2014) invite partial replication. One is that some of the sample sizes were extremely small (e.g., $n < 10$), and so it

is not so surprising that inferential statistics failed to detect significant between-group treatment effects.

A second reason is that their quasi-experiments were preceded by a pre-test which may have influenced the students' subsequent performance both at the exercise stage and then the post-test stage. In the pre-test, students were asked to supply the missing verbs of verb-noun collocations in gapped sentences. This is, essentially, a trial-and-error exercise, but without provision of feedback. If Boers et al. (2014) and Stengers and Boers (2015) are right in arguing that trial-and-error carries the risk that erroneous choices linger in memory, then a pre-test of this kind may well exacerbate this undesirable effect. In other words, some of the confusion which the authors attribute to the exercises may actually be attributable to the pre-test experience. This may at first sight seem implausible, given the long history of memory research that shows the usefulness of test-taking (e.g., Roediger & Karpicke, 2006, for a review). The evidence in favour of testing is strongest when learners supply the correct response and then receive confirmation that it is correct (Allen, Mahler, and Estes 1969; Karpicke and Roediger 2008), although benefits of corrective feedback have also been attested (Bahrick and Hall 2005; Potts and Shanks 2014). Crucially, however, no feedback is given on a pre-test. Also, the evidence for the benefits of test-taking comes mostly from experiments on paired-associate, form-meaning mapping, where participants learn to match unfamiliar words with distinct meanings. The phrase-learning challenge may be somewhat different, however, as it includes learning which already familiar words form partnerships. Moreover, among these already familiar words figure ('de-lexicalised') high-frequency words that lack semantic distinctiveness (e.g., *make* rather than *do* in *make an effort*) and/or that have near-synonymous competitors (e.g., *tell* rather than *say* in *tell lies*). It may therefore be harder for the learner to suppress infelicitous word choices when it comes to remembering collocational patterns than when it comes to remembering distinct word-meaning mappings.

In any case, while the jury may still be out on whether those in favour of errorless learning or those in favour of trial-and-error are ‘right’ in the context of phrase learning, it seems undeniable that pre-testing has the potential to influence subsequent learning – be it negatively or positively –, and so an experimental set-up which avoids pre-testing is desirable.

There are two alternatives for the use of a pre-test as a way of controlling for prior knowledge of target items and of ascertaining that treatment groups are comparable in that regard. One is to use pseudo-words, which ensures that no participant has any prior knowledge of the items. In the case of phrases, however, the learning challenge often lies in remembering the partnerships forged by already known words rather than remembering new words. For instance, a post-beginner learner of English is likely to be familiar with the verbs *make* and *do* and with the nouns *effort* and *homework*, but may nonetheless fail to combine these appropriately (resulting in **do an effort* and **make homework*). Given this peculiarity of the phrase-learning challenge, we opt in the present study for another alternative for pre-testing. This other alternative is to recruit a group of students from the same population as those who take part in the actual treatment study, and administer the test to this group for ‘norming’ purposes. This use of the test then allows for the identification of target items that are unknown to all the test-takers in the ‘benchmark’ group and thus almost certainly unknown also to their same-profile peers in the treatment groups.

A third motive for a partial replication of Boers et al. (2014) is that in their study only knowledge of the lexical make-up of the target phrases was tested, i.e., knowledge of the *form* of the phrases. There was no examination of the impact of the exercises on learners’ comprehension or retention of the *meaning* of the target phrases. To be fair, the study focused on ‘collocations’ and the authors may accordingly have assumed that the meaning of the target phrases was transparent, and that the only learning challenge concerned form, not

meaning. However, a study by Boers and Webb (2015) has demonstrated that teachers' intuitions about the semantic transparency of multiword expressions do not coincide well with how learners experience them. In any case, the repertoire of conventional verb-noun expressions that a learner may wish to develop will likely include 'non-compositional' or 'non-literal' expressions, that is, expressions that in the phraseological tradition would be called idioms rather than collocations (e.g., Cowie, 1981; Moon, 1998). In the present study, we thus recognise that also the semantics of phrases (whether they are called collocations or something else) can pose problems, and we therefore examine learners' post-treatment ability to recall not only the form (or lexical composition) of the phrases but also their meaning.

Boers et al. (2014) cautioned that, in the case of collocation exercises, trial-and-error procedures that generate a high error rate enhance the risk of confusion. On the other hand, Stengers and Boers (2015) demonstrated that a procedure that is void of challenge (such as copying words from example sentences) cannot be expected to work wonders either. This calls for the design of exercise formats that reduce error rates while preserving a sufficient degree of cognitive engagement on the part of the learner. In the quasi experiment we report here, we put one such potential alternative to the test – the provision of first-letter cues to help learners complete the missing words.

Finally, Boers et al. (2014) tested the effects of exercise formats they had encountered in a random sample of books, including some not-very-recent ones and some not-well-known ones. It seems worth carrying out a more systematic evaluation of widely distributed textbooks against the backdrop of the quasi-experimental findings.

III Research questions

In the quasi-experiment, we compare the effectiveness of three exercise formats intended to foster knowledge of verb-noun phrases. The first format presents learners with gapped

sentences from which the verb is missing and with a list of verbs to choose from to complete the blanks. The second format does not provide a list of options to choose from, but instead gives the first letter of the missing verbs as a cue in the gapped sentences – a way of constraining guesses. The third format presents learners with gapped sentences from which the whole verb-noun expression is missing, preceded by a list of the phrases to choose from to complete the blanks. In all three conditions, comprehension of the sentences and the target expressions is supported by L1 translations. Each of the exercises is followed by (corrective) feedback.

Like Boers et al. (2014) and Stengers and Boers (2015), we have chosen to focus on verb-noun combinations in this study rather than, say, adjective-noun combinations, because studies have shown that (other things being equal) verb-noun partnerships tend to be particularly problematic for language learners (e.g., Laufer & Waldman, 2011; Nesselhauf, 2003; Peters, 2016) and they thus seem worthy candidates for instruction.

The research questions we address by means of the quasi-experiment are these:

1. Do the three aforementioned exercise formats bring about different learning gains as measured by a delayed test on *form* recall, i.e., recall of the composition of the phrases?
2. Do these three aforementioned exercise formats bring about different learning gains as measured by a delayed test on recall of the *meaning* of the phrases?

An additional question we address further below is:

3. To what extent do phrase-focused exercises in contemporary textbooks chime with recommendations that distilled from the available quasi-experimental evidence?

IV The quasi experiment

1 Method

1.1 Participants

Four parallel groups of students were involved in the study (n 30, 35, 25, and 27). The first group was used for ‘norming’ purposes. The other three groups took part in the actual treatment study. All the students were second-year English majors in a College of Foreign Languages at a university in Vietnam. They had been learning English as a foreign language for seven years. At the end of their previous semester, they had passed a B1-level exam (CEFR – Common European Framework Reference) in the four skills (listening, speaking, reading and writing), indicating they had an intermediate level of proficiency in English (Taylor & Jones, 2006). According to the students’ mean percentage scores in that exam, the four groups were comparable in level of proficiency: 62.9 (SD 7.5), 61.1 (SD 6.8), 61.6 (SD 8.5), and 58.7 (SD 8.5). The group with the highest mean exam score was chosen for the norming test. Of the actual treatment groups, the third group (which was given the ‘select-the-phrase’ exercise format – see below) appears slightly weaker than the others, but a one-way ANOVA for independent samples reveals no significant difference: $F(2, 84) = 1.13, p = .33$.

1.2 General design

We first pre-selected, from sources such as McCarthy and O’Dell (2002; 2005), 20 English verb-noun phrases which we thought stood a good chance of being unfamiliar to our learner-participants. To verify that these 20 verb-noun combinations constituted strong word partnerships, we looked up the mutual information (MI) scores of the verb-noun combinations in the *Corpus of Contemporary American English (COCA)*. All 20 were found to have MI scores > 3 (see Table 1), indicating that these verb-noun combinations indeed qualify as collocations according to the threshold proposed by, for example, Hunston (2002). Note that the items vary in likely degree of semantic transparency and that some (e.g., *cut corners*; *call someone’s bluff*) can on the basis of their non-compositional nature be

considered idioms – and are indeed listed in, for example, the *Collins Cobuild Dictionary of Idioms* (2002).

<Table 1 around here>

A norming test was then administered to the first group ($n = 30$) in order to identify the phrases in the pre-selected set which Vietnamese students at their level of proficiency were highly unlikely to be familiar with.

The other three groups ($n = 35, 25,$ and 27) were randomly assigned to one of three treatment conditions, where they were given a fill-in-the blank exercise (of a format that differed between the groups) focusing on the selected phrases. Comprehension of the phrases and the sentences in which they were embedded was assisted by means of L1 translations. Two weeks after the treatment (i.e., after doing the exercise), the students took a post-test to assess whether they had retained the lexical composition and the meaning of the phrases. The exercises and the tests were integrated in the students' regular English classes.

1.3 Norming test

As mentioned, the purpose of the norming test was to find out which of the pre-selected 20 verb-noun phrases these second-year English majors were still unfamiliar with. The students were given 20 gapped sentences, each targeting one of the phrases. In each gapped sentence, the verb was left out. A Vietnamese translation of the sentence was added. For example:

The country organized a solemn ceremony to _____ tribute to soldiers who died in the war.

Đất nước đã tổ chức một buổi lễ long trọng để tri ân các chiến sĩ đã hy sinh trong cuộc chiến.

The verbs to be supplied were all in the infinitive, so the students did not need to attend to inflectional morphology. The students were given 15 minutes to complete the test.

The results of the norming test revealed that 12 of the 20 phrases (see Table 1) were unfamiliar to all the students, and were thus highly likely to be unfamiliar also to the actual treatment groups. Post-treatment successes (if any) on these 12 items would thus almost certainly be attributable to treatment rather than prior knowledge.

1.4 The exercises

Three exercise formats (select the verb, 1st letter given, and select the phrase) were used for the treatment study, with each format to be used by one of the treatment groups in a between-participant study design. The same sentential contexts and their Vietnamese translations were used as in the norming study. The eight collocations which had been shown to be familiar to some of the students in the norming test were retained in the exercise, but the principal focus of the between-group comparison will be on the 12 items that were found to be unfamiliar to all the students who took the norming test.

The teachers of the three treatment groups followed identical procedures. First, they handed out the exercise worksheet to their respective groups and gave the students 15 minutes to complete their exercise. Then they handed out the answer key, which showed the same but now completed sentences. The students were asked to check their answers. They were told to put a tick (✓) after each right response on their worksheet. For wrong answers, they were told to use a different-color pen to cross out the wrong response and to write down

the correct response instead. When the students had finished making these corrections, the teacher collected the work sheets and the answer key.

As mentioned, the exercise formats differed between the three treatment groups. One group ($n = 35$) received a worksheet where a list of the missing verbs was given at the top of the sheet, and the students were required to choose the appropriate verbs to complete the gaps. This is a format included in Boers et al. (2014) and was found in that study not to be particularly effective. It is a very common format in textbooks, however (see further below), and for that reason alone merits further evaluation.

The second group ($n = 25$) received a worksheet with the same gapped sentences, but instead of a list of verbs to choose from, the 1st letter of the missing verb was given in the gap as a cue. This format was absent from Boers et al.'s (2014) study. We include it as a potential alternative worth putting to the test because, as argued by Stengers and Boers (2015), a 1st letter cue can block potential erroneous substitutes (e.g., the cue *t__* should prevent the learner from writing down *make*, for instance, in *t_____ a photo*).

The third group ($n = 27$) received a worksheet with the same sentential contexts but with larger gaps, where the whole verb-noun expression was missing. A list of the missing expressions was given at the top of the sheet. This format may not draw learners' attention specifically to the verb, but, as argued – but not actually tested – by Boers, et al. (2014), it may engage learners more with (figuring out) the meaning of the expressions (as they need to evaluate which sentential context is compatible with the meaning of the verb-noun expression). This is the format which Boers et al. (2014) tentatively concluded was the more judicious one of the formats they examined, because it appeared less prone to engendering erroneous verb-noun associations in a learner's memory. Their evidence was far from conclusive, however. Besides, there are grounds for expecting that matching intact expressions with sentential contexts stimulates acquisition of the meaning of the expressions

relatively well, but acquisition of their formal properties less well. This, at least, would be consistent with Barcroft's work on the first stages of word learning (Barcroft, 2015, for a comprehensive review), which suggests that attention to the meaning of a new word ('semantic elaboration') results first and foremost in meaning retention whereas attention to its form ('structural elaboration') results first and foremost in form retention, and typically creating a trade-off effect between the two types of attention. If we consider the lexical composition of a multiword expression to be a formal property of the expression, then the same trade-off might occur when learners' do exercises which direct their attention to the makeup of an expression versus those which require learners to engage with its meaning.

1.5 The post-test

Two weeks after the treatment (i.e., after the exercise session), a post-test was administered to the three treatment groups. The post-test consisted of gapped sentences where the students were asked to fill in the blanks with suitable verbs. No list of options to choose from was given and neither were first-letter cues given. Also the Vietnamese translation of the sentences was removed. In order to examine if the students remembered the meaning of the phrases, they were asked to write a Vietnamese translation of the verb-noun phrase in a space below the sentence. The sentential contexts given in the post-test were the same as in the exercises, but they appeared in a different order. The students were given 15 minutes to complete the test.

Each correct verb supplied in the gapped sentences counted for one mark. Two Vietnamese-English bilinguals collaboratively assessed the translation responses. When responses diverged from the translation that accompanied the sentences on the exercise worksheets, an agreement was reached on which of these were acceptable.

2 Results

We shall first focus on the results pertaining to the 12 items which, according to the norming test, the participants in the treatment groups almost certainly lacked knowledge of. Table 2 sums up the descriptive statistics of the three treatment groups' performance on the part of the post-test where the participants were required to supply the missing verb, i.e., the part testing participants' recollection of the composition (or form) of the phrases. The mean score obtained by the group which had worked with intact phrases was the highest. The mean score of the group which had been asked to select the right verb from a set of options was the lowest.

<Table 2 around here>

Because the Shapiro-Wilk test indicated that the scores in the 1st-letter-given group were not normally distributed, we resort to non-parametric tests to further examine the between-group differences. The Kruskal-Wallis test signals there is a difference among the three groups' scores: $H(2,84) = 8.85; p = .012$. Pairwise comparisons using the Mann-Whitney test show a significant difference between the select-the-verb condition and the select-the-phrase condition: $z(60) = -3.34; p = .0008$.¹ The effect size is medium-large: $d = .81$.² No additional significant between-group differences were revealed.

The data are consistent with the thesis that increasing the success rate at the exercise stage benefits retention of the correct verb-noun combinations more than reliance on corrective feedback after trial-and-error. The select-the-verb condition yielded fewer correct exercise responses for the 12 target items (mean = 3; SD = 1.47) than the 1st-letter-given condition (mean = 3.84; SD = 1.82), which in turn yielded fewer correct exercise responses than the select-the-phrase condition (mean = 5.11; SD = 2.33). The number of correct

exercise responses supplied by the 87 students taken together correlated positively with the number of correct verb recalls in the post-test: $r = .29$ ($p = .006$).

Supplying the right response in the exercise is not the only factor that matters, however. Of the 105 correct exercise responses in the select-the-verb exercise, only 33% were followed by a correct post-test response. This suggests that many of the correct exercise responses were the outcome of lucky guesses when students matched nouns with verbs from the list of options. If so, seeing the lucky guesses confirmed by the answer key was clearly not often sufficient to entrench the correct verb-noun association in long-term memory. In comparison, of the 96 correct responses in the exercise where the first letter of the missing verbs was given, 48% were followed by correct post-test responses, and in the case of the exercise in which students worked with intact phrases, 55% of the 138 correct responses were followed by correct post-test responses. Pairwise comparisons of these different proportions using Chi Square tests yields, for select-the-verb vs. 1st-letter-given: Yates $\chi^2 = 4.42$; $p = .036$, and for select-the-verb vs. select-the-phrase: Yates $\chi^2 = 11.49$; $p = .0007$.

The next question we need to ask is whether correcting responses by consulting an answer key helps to establish the correct verb-noun associations in long-term memory. Of the 314 failed exercise responses under the select-the-verb procedure, only 38% were followed by correct post-test responses. To compare, of the 214 and the 186 failed exercise responses under the 1st-letter-given and the select-the-phrase procedure, respectively 60% and 59% were followed by correct post-test responses. Chi Square again shows the difference between these proportions to be significant: Yates $\chi^2 = 5.68$; $p = .017$ and Yates $\chi^2 = 4.70$; $p = .030$, respectively.

Let's now turn to the part of the post-test where the participants were required to provide the meaning (L1 translation) of the phrases. Table 3 sums up the results.

<Table 3 around here>

Also on this part of the test, the exercise condition where students worked with intact phrases produced the best outcome. However, the condition where 1st letter hints were given now generated the poorest results. The Shapiro-Wilk test signals that the scores in that group are not normally distributed, and so we opt for non-parametric tests again to further examine the between-group differences. The Kruskal-Wallis test yields $H(2,84) = 19.2; p < .0001$, and pairwise comparisons using the Mann-Whitney test show two significant differences: (1) between the select-the-verb and the 1st-letter-given conditions: $z(58) = 3.23; p = .001; d = .90$, and (2) between the select-the-phrase and the 1st-letter-given conditions: $z(50) = 4.0; p < .0001$, where the effect is particularly large: $d = 1.46$.

The test data regarding meaning recall need to be interpreted with more caution than those concerning form recall, however, because, strictly speaking, the norming test only probed knowledge of form. In other words, while we can be pretty confident, thanks to the norming test, that the students in the treatment groups lacked prior knowledge of the lexical composition of the phrases, we cannot be as confident when it comes to prior knowledge of their meaning. Still, given their comparable levels of proficiency, it would be a remarkable coincidence if the 1st-letter-given group had much poorer prior knowledge of the meaning (but not the form) of the target phrases than the other two treatment groups.

For completeness' sake, we also report the test results for all 20 items together, i.e., including the eight phrases which, according to the norming test, students were more likely to be familiar with than the 12 items we have focused on so far. As shown in Tables 4 and 5, these results show the same trends as reported above.

< Tables 4 and 5 around here >

As regards the form-recall test, the Kruskal-Wallis test confirms there is a between-group difference: $H(2,84) = 6.11$; $p = .047$. Pairwise comparisons using the Mann-Whitney test show that the select-the-phrase group significantly outperformed the select-the-verb group: $z = 2.69$; $p = .007$; $d = .64$. The between-groups differences are more pronounced when it comes to the meaning-recall test, where Kruskal-Wallis yields $H(2, 84) = 16.6$; $p = .0002$. Mann-Whitney shows that both the select-the-phrase and the select-the-verb groups significantly outperformed the 1st-letter-given group, with: $z(50) = 3.75$ ($p = .0002$; $d = 1.37$) and $z(58) = 3.13$ ($p = .0017$; $d = 1.01$), respectively.

3 Discussion

While the exercise condition in which the first letter of the missing verbs is given as a hint looked relatively effective as far as later recall of those verbs is concerned, it appears much less effective when it comes to retention of meaning. While the correlations between the students' scores on the form-recall part and the meaning-recall part of the test (for the 12 target items) suggest the two aspects of knowledge developed roughly in parallel in the select-the-verb group ($r^2 = .329$) and in the select-the-phrase group ($r^2 = .395$), the correlation is much weaker in the 1st-letter-given group ($r^2 = .136$). This lends support to the idea (consistent with Barcroft, 2015) that the 1st-letter-given format stimulated engagement with the formal makeup of the phrases relatively well, but did not at the same time stimulate as much engagement with the meaning of the phrases.

The overall picture, which considers both form and meaning retention, therefore suggests that the exercise format in which students are asked to select intact expressions to fit sentential contexts is the most beneficial of the three formats this quasi-experiment set out to evaluate. It may be worth recalling at this point that the group which was given the intact-

phrases exercise format was the group with slightly lower English exam grades than the other groups, which lends additional credibility to the interpretation that their better performance on the phrase test is to be attributed to the exercise condition they were assigned to.

Like Boers et al. (2014), we found that the select-the-verb format, which essentially requires learners to reassemble broken-up phrases, carries the risk of engendering wrong verb-noun associations. Post-test responses in the select-the-verb condition included malformed collocations such as ‘*talk volumes*’ (instead of *speak volumes*), ‘*give tribute to*’ (instead of *pay tribute to*), ‘*pay his bluff*’ (instead of *call his bluff*) and ‘*cast fault with*’ (instead of *find fault with*). These are all instances where students made wrong substitutions from the set of verb options provided to them on the exercise worksheet. It may be worth reiterating here that the students received corrective feedback on their exercises – they crossed out their wrong responses and copied the correct responses from the answer key. As we saw above, this correction procedure was seldom followed by a correct post-test response in the select-the-verb group, which suggests that it is more judicious to minimize the risk of error at the exercise stage than to rely on the benefits of corrective feedback.

We hypothesized that a gap-fill format where the 1st letter of the missing verb is given as a cue would constrain the possibility of writing wrong guesses, and the rate of correctly supplied verbs was indeed higher in this condition than in the select-the-verb condition, both at the exercise and the post-test stage. It is also possible that the challenge of generating (part of) the verb in the exercise was better preparation for the post-test – because in the post-test no list of options was available for the students to choose from and so it was perhaps slightly more congruent with the 1st-letter-given exercise format.

As far as meaning recall is concerned, the better performance under the select-the-phrase exercise condition corroborates the hypothesis that this exercise format invites learners to engage with the meaning of the phrases – in order to match the phrases with

semantically compatible sentential contexts. This did not appear to detract from uptake of the lexical composition of the phrases, given that students in this condition managed to recall the verbs of the phrases (i.e. form) better than those in the select-the-verb condition. By contrast, students in the 1st-letter-given condition performed relatively well on the part of the post-test which required them to recall the verbs, but they performed very poorly on the part that required them to supply the meaning of the verb-noun expressions. It is plausible that their focus on generating verbs in the exercise and then checking these against the answer key usurped attention which they might otherwise have given to the sentential contexts and the accompanying translations.

In sum, the results of the quasi-experiment point to the conclusion that a trial-and-error exercise where students are asked to reassemble broken-up collocations is less effective than a procedure that leaves phrases intact from the start. This supports Boers et al.'s (2014) tentative recommendation that, *in the event* textbook writers decide to create phrase-focused exercises, then they should give precedence to exercises which present the expressions as holistic units. And, *if* it is at all deemed necessary to channel learners' attention to the collocational makeup of multiword expressions by asking them to re-assemble broken-up expressions or supply missing constituents, then it is judicious to design and implement the exercise such that the risk of undesirable cross-associations is minimized. A straightforward way of doing the latter is to provide learners with examples of the target phrases beforehand. This, then, raises the question to what extent the exercise formats and procedures for phrase learning that are included in contemporary textbooks chime with these recommendations.

V Textbook analysis

1 Sample

We selected ten general EFL textbook series that according to the publishers' statements are used around the world (see Table 6). Because we wished to align the textbook analysis to the proficiency level of the participants used in the quasi-experimental studies on the matter available to date, we chose to focus on the student books intended for intermediate students. We manually screened each textbook for exercises with a focus on phrases. In order to ensure a sizeable sample, we did not confine the search to exercises exclusively targeting verb-noun combinations. To be included in the inventory of phrase exercises, the exercises did need to carry a label such as "expressions" that indicated the focus was on multi-word items. The screening produced a bank of 323 phrase-focused exercises. More than 65% of these use the term "phrases". In comparison, only 4% use the term "collocations", which may reflect an effort on the part of the authors to avoid linguistic jargon. The mean number of phrase-focused exercises per book is 32, but there is marked variation in the number of such exercises, ranging from 17 to 58.

2 *Assisted or trial-and-error procedures? Intact vs. broken-up phrases?*

Against the backdrop of the quasi-experimental findings discussed above, we discerned three broad implementations of phrase-focused exercises (see Table 6). In the first kind of implementation, the textbook users are presented first with the intact target phrases embedded in some context (sentences or longer passages) that illustrate their form and meaning. In the case of non-transparent items (such as idioms and phrasal verbs), this implementation may also include explicit explanations of the meaning or function of the phrases. The learner is then required to do exercises on these intact phrases, assisted by the examples (and explanations) given. This resembles the presentation in McCarthy and O'Dell's (e.g., 2002; 2005) books for independent study, and also the exemplar-guided, errorless condition in Stengers and Boers (2015) belongs to this category. In Table 6 we refer to this category as

“assisted work on intact phrases”. This makes up almost half of our sample of exercises (160 exercises, or 49.5%).

In a second kind of implementation, the textbook users are not first provided with contextualized examples or other assistance. The phrases are presented as prompts intact, but it is through doing the exercise that learners are expected to try and work out their meaning or function. About one quarter (87 exercises, or 27%) of our sample of exercises belongs to this category of trial-and-error practice.

The third implementation of phrase-focused exercises is also of the trial-and-error type in the sense that no contextualized examples or explanations are given to guide learners’ exercise responses. It differs from the second implementation, however, in that the phrases are not presented intact. Instead, the prompts are parts of broken-up phrases that have to be re-assembled or incomplete phrases that have to be completed. This practice, which according to the research findings to date is not the most advisable, characterizes 76 (or 23.5%) of the exercises in our sample.

It is worth mentioning that very few – only 29 – of the ‘unassisted’ exercises, i.e., of the 163 exercises that rely on trial and error, request students to check the accuracy of their responses by referring to an answer key (to be looked for in an appendix). This means that altogether 133 exercises (or 41%) in our sample neither provide students with input to help them avoid error nor refer them to feedback to help them assess their responses. Many textbook authors must presume it is the teacher’s role to ensure their students realize which responses are right and which are wrong.³

3 *The popular exercise formats*

We further categorized exercise formats by type of action required on the part of the learner.

By far the most common type of action involves ‘matching’. This subsumes a range of

variants, whereby students rejoin broken up sentences, match phrases to definitions or with single-word substitutes, choose among a list of options which word collocates with a given prompt, or re-assemble phrases from jumbled up constituents. Altogether 197 (or 61%) of the exercises in our sample engage learners in one or the other form of matching. A feature that is common to all these matching formats is that students can indicate their responses without actually having to write the target phrases. Instead, they can draw a line between associated items or circle their choice in a multiple-choice exercise, for instance.

The second most frequent type of exercise is ‘gap-filling’, of which our sample contains 77 instances (24%). This category subsumes sentence-level exercises similar to the ones we have evaluated in the present article, but it also includes text-based formats, such as completing blanks in a transcript after listening to an audio recording. Of the sentence-level gap-fill exercises, 53 are of the non-assisted, trial-and-error type, and 31 of these present incomplete phrases for the learner to supply missing constituents. To our surprise we found only one example of a 1st-letter-cued gap-fill exercise, and this instance also provided the last letter of the missing phrases.

One may argue that the mental operation in gap-fill exercises also involves matching, because the learner needs to associate phrases (or parts of phrases) with their compatible contexts. The only difference with the aforementioned matching formats is that the learner is expected to actually write down (parts of) phrases in the blanks reserved for them. Together, matching and gap-fill exercises clearly make up the bulk (almost 85%) of the phrase-focused exercises in contemporary EFL textbooks.

Other exercise types are much less common. The third most common type, which does not come close in frequency to matching and gap-filling, is ‘sentence composition’ (20 instances, or 6% of the sample). This exercise instructs learners to generate sentences incorporating given phrases. Oddly enough, this exercise type also occurs occasionally (nine

instances) in the non-assisted category, i.e., without any examples or explanations to help learners use the given phrases felicitously. A small number of exercises (6) require students to determine whether errors occur in given sentences. One may wonder if learners might not find it hard later on to suppress the memory left by the erroneous forms they have been asked to contemplate.

VI Conclusions

Overall, the findings from our quasi-experimental study support earlier assertions that, when it comes to designing and implementing exercises on L2 multiword expressions such as verb-noun collocations, it is advisable (a) to minimize the rate of error at the exercise stage so as to reduce the risk of creating undesirable cross-item associations that learners may find hard to suppress later on, and (b) to present the multiword units as intact wholes from the start rather than asking learners to re-assemble broken-up units or supply missing parts as a way of getting to know the target phrases. The former recommendation is consistent with research that favours errorless learning over learning through trial-and-error (e.g., Warmington & Hitch, 2014). The second recommendation is in accord with the view that the acquisition of formulaic language in L1 comes more naturally than in L2 precisely because L1 multiword expressions are encountered, processed and stored as holistic units during naturalistic L1 learning (e.g., Wray, 2002).

While we are not arguing that the conditions for adult second language acquisition should necessarily mimic those of L1 acquisition, the research findings to date do lead us to question the efficacy of practices whereby learners are introduced to new phrases by requiring them to experiment with different word combinations before it is revealed to them which combinations are the ones to be retained in memory and which are the ones to be suppressed in future. And yet, according to our analysis of 323 phrase-focused exercises

included in ten recent EFL textbooks, this is precisely what almost a quarter of these exercises ask learners to do.

Notes

1. All p-values in this article are two-tailed.
2. Cohen's *d* effect sizes were computed on the basis of group means, standard deviations and group size. A *d* value of $> .80$ is generally considered to indicate a large effect.
3. One may argue that students are free to seek assistance from other sources beyond the textbook itself, such as dictionaries. The fact remains, though, that these exercises are not accompanied by explicit prompts to do so.

References

- Allen, G. A., Mahler, W. A., & Estes, W. K. (1969). Effects of recall tests on long-term retention of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 8, 463–470.
- Baddeley A., & Wilson B.A. (1994). When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*, 32, 53–68.
- Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term memory: A metacognitive explanation for the spacing effect. *Journal of Memory and Language*, 52, 566–577.
- Barcroft, J. (2015). *Lexical input processing and vocabulary learning*. Amsterdam, Netherlands: John Benjamins.

- Barfield, A., & Gyllstad, H. (Eds.) (2009). *Researching collocations in another language: Multiple perspectives*. Basingstoke, UK: Palgrave Macmillan.
- Boers, F., Demecheleer, M., Coxhead, A., & Webb, S. (2014). Gauging the effects of exercise on verb-noun collocations. *Language Teaching Research, 18*, 54–74.
- Boers, F., Eyckmans, J., & Stengers, H. (2007). Presenting figurative idioms with a touch of etymology: More than mere mnemonics? *Language Teaching Research, 11*, 43–62.
- 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.
- Boers, F., & Lindstromberg, S. (2009). *Optimizing a lexical approach to instructed second language acquisition*. Basingstoke, UK: Palgrave Macmillan.
- Boers, F., & Lindstromberg, S. (2012). Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83–110.
- Boers, F., & Webb, S. (2015). Gauging the semantic transparency of idioms: do natives and learners see eye to eye? In R. Heredia & A. Cieslicka (Eds.), *Bilingual figurative language processing* (pp. 368-392). Cambridge, UK: Cambridge University Press.
- Collins Cobuild Dictionary of Idioms* (2002, 2nd ed.). Glasgow, UK: HarperCollins Publishers.
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics, 2*, 223–235.
- Dai, Z., & Ding, Y. (2010). Effectiveness of text memorization in EFL learning of Chinese students. In D. Wood (Ed.) *Perspectives on formulaic language: Acquisition and communication* (pp. 71-87). New York, NY: Continuum.

- Ellis, N.C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42, 375–396.
- Erman, B., & Warren, B. (2001). The idiom principle and the open choice principle. *Text*, 20, 87–120.
- Eyckmans, J, Boers, F, & Lindstromberg, S. (2016). The impact of imposing processing strategies on L2 learners' deliberate study of lexical phrases. *System*.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.
- Karpicke, J. D, & Roediger, H. L. III (2008). The critical importance of retrieval for learning. *Science*, 319, 966–968.
- Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, 29, 694–716.
- Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning*, 61, 647–672.
- Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 22–46). New York. NY: Continuum.
- Lewis, M. (1993). *The Lexical Approach*. Hove, UK: Language Teaching Publications.
- McCarthy, M. & O'Dell, F. (2002). *English idioms in use*. Cambridge, UK: Cambridge University Press.
- McCarthy, M., & O'Dell, F. (2005). *English collocations in use*. Cambridge, UK: Cambridge University Press.

- Meunier, F., & Granger, S. (2008) *Phraseology in foreign language learning and teaching*. Amsterdam, Netherlands: John Benjamins.
- Moon, R. (1998). *Fixed expressions and idioms in English: A corpus-based approach*. Oxford, UK: Clarendon Press.
- Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford, UK: Oxford University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223-242.
- Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191–226). London, UK: Longman.
- Pellicer-Sanchez, A. (2015). Learning L2 collocations incidentally from reading. *Language Teaching Research* (Online First).
- Peters, E. (2006). The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, 20(1).
- Polio, C. (Ed.) (2012). *Topics in formulaic language. Annual Review of Applied Linguistics*, 32.
- Porte, G. (Ed.) (2012). *Replication research in applied linguistics*. Cambridge: Cambridge University Press.
- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, 143, 644–667.
- Roediger, H. L. III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1, 181–210.
- Schmitt, N. (Ed.) (2004) *Formulaic sequences*. Amsterdam, Netherlands: John Benjamins.

- Sinclair, J. (1991). *Corpus, Concordance and Collocation*. Oxford, UK: Oxford University Press.
- Siyanova-Chanturia, A. & Martinez R. (2015). The Idiom Principle revisited. *Applied Linguistics*, 36, 549–569.
- Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. B. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 776–784.
- Sonbul, S. (2015). Fatal mistake, awful mistake or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, 18, 419–437.
- Sonbul, S., & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning*, 63, 121–159.
- Stengers, H., & Boers, F. (2015). Exercises on collocations: A comparison of trial-and-error and exemplar-guided procedures. *Journal of Spanish Language Teaching* (online early view).
- Szudarski, P., & Carter, R. (2014). The role of input flood and input enhancement in EFL learners’ acquisition of collocations. *International Journal of Applied Linguistics* (online early view).
- Taylor, L., & Jones, N. (2006). Cambridge ESOL exams and the Common European Framework of Reference (CEFR). *Cambridge ESOL Research Notes*, 24, 2–5.
- Warmington, M., & Hitch, G. J. (2014). Enhancing the learning of new words using an errorless learning procedure: Evidence from typical adults. *Memory*, 22, 582–594.

- Warmington, M., Hitch, G. J., & Gathercole, S. E. (2013). Improving Word learning in Children using an Errorless Technique. *Journal of Experimental Child Psychology*, *114*, 456–465.
- Webb, S., Newton, J., & Chang, A. C. S. (2013). Incidental learning of collocation. *Language Learning*, *63*, 91–120.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, UK: Cambridge University Press.
- Wood, D. (Ed.) (2010). *Perspectives on formulaic language: Acquisition and communication*. New York, NY: Continuum.

The EFL textbooks

- Clandfield, L. (2011). *Global: Intermediate coursebook: Student's book*. Oxford: MacMillan Publishers Limited.
- Clare, A., & Wilson, J. (2011). *Speakout: Intermediate student's book*. Essex: Pearson Education Limited.
- Cunningham, S., & Moor, P. (2005). *New cutting edge intermediate: Student's book*. Essex: Pearson Education Limited.
- Hancock, M., & McDonald, A. (2009). *English result: Intermediate student's book*. New York: Oxford University Press.
- Kay, S., & Jones, V. (2009). *New inside out: Intermediate student's book*. Oxford: MacMillan Publishers Limited.
- Kerr, P., & Jones, C. (2005). *Straightforward: Intermediate student's book*. Oxford: MacMillan Publishers Limited.
- McCarthy, M., & O'Dell, F. (2007). *English Phrasal Verbs in Use: Advanced*. Cambridge; New York: Cambridge University Press.

Oxenden, C., & Latham-Koenig, C. (2006). *New English file: Intermediate student's book*.

New York: Oxford University Press.

Richards, J., & Bohlke, D. (2011). *Four corners*. Cambridge: CUP.

Roberts, R., Clare, A., & Wilson, J. (2011). *New total English: Intermediate student's book*.

Essex: Pearson Education Limited.

Soars, L., & Soars, J. (2003). *New headway: Intermediate student's book*. New York: Oxford

University Press.

Table 1: The 20 verb-noun phrases and their MI scores.

Verb-noun phrases	MI score	Verb-noun phrases	MI score
make a contribution	4.14	pay tribute*	8.38
cut corners*	7.21	take effect*	4.05
take a toll*	4.94	cause casualties	4.37
find fault*	4.94	take a picture	3.51
bear the brunt*	10.6	take notes	3.39
turn the tide*	5.92	do homework	4.57
call one's bluff*	5.36	buy time*	3.83
cast doubt	7.46	make progress	3.99
give chase*	4.24	talk nonsense	3.97
speak volumes*	7.61	move mountains*	3.63

Notes: (1) No students who took the norming test showed knowledge of the items

marked in the table with an asterisk. (2) MI scores collected from COCA in May 2015.

Table 2: Descriptive statistics for form recall of the 12 target items

Condition	Mean	Median	St. dev.	Min.	Max.
A. Select the verb (n 35)	3.49	3	2.79	0	10
B. 1 st letter of the verb given (n 25)	4.96	6	3.45	0	10
C. Select the intact phrase (n 27)	5.37	5	1.67	2	9

Table 3: Descriptive statistics for meaning recall of the 12 target items

Condition	Mean	Median	St. dev.	Min.	Max.
A. Select the verb (n 35)	4.34	3	2.96	0	12
B. 1 st letter of the verb given (n 25)	1.92	0	2.36	0	6
C. Select the intact phrase (n 27)	5.30	5	2.35	2	10

Table 4: Descriptive statistics for form recall of all 20 items

Condition	Mean	Median	St. dev.	Min.	Max.
A. Select the verb (n 35)	9.17	9	3.58	3	17
B. 1 st letter of the verb given (n 25)	10.68	11	4.34	5	18
C. Select the intact phrase (n 27)	11.11	11	2.34	6	16

Table 5: Descriptive statistics for meaning recall of all 20 items

Condition	Mean	Median	St. dev.	Min.	Max.
A. Select the verb (n 35)	9.11	8	4.11	0	18
B. 1 st letter of the verb given (n 25)	4.60	3	5.05	0	14
C. Select the intact phrase (n 27)	10.48	10	3.67	5	17

Table 6: Phrase-focused exercises in contemporary EFL textbooks

Textbook	Assisted ¹ work on intact phrases			Unassisted work on intact phrases			Unassisted work on broken-up phrases			Total
	Matching	Gap-fill	Other	Matching	Gap-fill	Other	Matching	Gap-fill	Other	
English Result	4	1	3	2	0	0	2	4	0	17
Four corners	13	0	0	3	0	0	2	0	0	18
New Headway	5	0	1	2	6	1	5	0	0	20
Straightforward	5	1	2	5	1	0	2	4	1	21
Cutting Edge	0	0	1	16	1	0	5	3	0	26
New English File	10	2	6	0	1	0	2	8	0	29
Global	7	1	10	7	0	4	6	2	1	38
New Inside Out	15	8	1	4	5	0	7	2	0	42
Speakout	16	4	8	12	3	0	9	2	1	54
New Total English	21	7	8	8	5	1	2	6	0	58
Total	96	24	40	59	22	6	42	31	3	323
	160 (49.5%)			87 (27%)			76 (23.5%)			

¹Note: “Assisted” = preceded at least by contextualized examples to guide the exercise responses.