

# Vocabulary Notebooks: Balancing Learner Autonomy and Vocabulary Acquisition

Christopher SULLIVAN

## 1. INTRODUCTION

A recent study asserts that “most researchers and teachers collectively agree that the recording of new words in vocabulary notebooks of one form or another should be promoted” (McCrostie, 2007: 246). This statement seems incontrovertible enough, as there are several reasons why vocabulary notebooks could be considered useful.

First and foremost, they utilize deliberate language learning, which can be an effective method of acquiring large amounts of L2 vocabulary in a short time (Nation & Webb, in press). Secondly, they can promote learner autonomy (Schmitt & Schmitt, 1995), especially when learners are allowed to choose which words are recorded in the notebooks. Thirdly, with guidance from the teacher, they can expose learners to a wide variety of vocabulary learning strategies (Fowle, 2002; Schmitt & Schmitt, 1995). Finally, they are not dependent on technology or expensive resources, and are therefore easy to implement in classrooms and schools (Fowle, 2002).

For reasons such as these, the inclusion of vocabulary notebooks is increasing in ESL classrooms, and some ESL publishers are even starting to package blank vocabulary notebooks along with their textbooks (e.g. *Communication Spotlight*, 2006). It is a great surprise, therefore, that three empirical studies on the subject of vocabulary notebooks has much to say about their *ineffectiveness*, especially in regards to the words that learners choose, and their effect on learner autonomy.

## 2. LITERATURE REVIEW

In 2002, Moir and Nation conducted a study on ten adult language learners using vocabulary notebooks in an ESL course. The program required the participants to choose 30–40 words each week, and record them in their notebooks. Moir and Nation’s rationale in allowing the learners to choose their own vocabulary was two-fold: “first, self-selecting vocabulary allows individual learners to focus on vocabulary that meets their own needs and, second, it is believed that selecting their own words would result in increased motivation to learn” (2002: 18). Unfortunately, through a series of interviews with the participants, Moir and Nation discovered that:

- the words participants selected were not taken from a wide variety of sources, and were mostly chosen from texts introduced in class
- the words were generally selected at random, and chosen because they were “unknown”
- the words were of low frequency, and limited distribution, and even the participants believed these words to be of limited use (2002: 22)

In a 2007 study, McCrostie examined the notebook use of 124 first-year university students. McCrostie’s findings echoed those of Moir and Nation, as he found that:

- 82% of all the words selected by the participants were from textbooks and class handouts
- 43.25% of the words were nouns, and 28% were verbs, with less than 29% accounting for all the other parts of speech
- 58% of the words were chosen from the 3000 and higher frequency levels
- 34% of the words were chosen because they were unknown, as opposed to only 8% being chosen because they had been seen or heard frequently (2007: 250–251)

In addition, both studies noted that the participants tended to view words as isolated units, and focused on direct L1 translations at the expense of other aspects of word knowledge, such as collocations and word families (McCrostie 2007: 253; Moir & Nation 2002: 23).

Having a teacher predetermine the contents of learner notebooks, i.e. providing students with a list of words to be recorded and learned, would be an obvious solution to the problems outlined above. In fact, this is the solution that Walters and Bozkurt implemented in a 2009 study of a class of 20 preparatory school students. At the beginning of their study, they selected 80 target words from the course textbook, and in the subsequent four weeks, had the learners record 20 of these words in their notebooks per week. Each week, the learners performed various manipulations on the words, such as listing collocations and creating example sentences, as recommended by Schmitt & Schmitt (1995), and Schmitt (2000: 137).

Walters and Bozkurt created an experimental design comparing the class which used the notebooks to two other classes who explicitly studied the same vocabulary without notebooks. They then used pre- and post-tests to measure the differences in gains between the experimental and control groups. They concluded that the experimental group not only out-performed both control groups on the receptive and productive vocabulary post-tests, but that it also demonstrated “more receptive and productive knowledge of target words, in contrast to words that were also included in the lessons, but were not recorded in the vocabulary notebooks” (2009: 417). However, these significant gains in vocabulary knowledge came at the expense of learner autonomy. Interviews conducted at the end of the study revealed that the use of the notebooks did *not* promote independent vocabulary study. “The students almost unanimously agreed,” they write, “that they would only continue their use of vocabulary notebooks if it were required” (2009: 418).

A conundrum now emerges: allowing the teacher to choose a relevant, useful, and balanced word list appears to result in loss of

autonomy and motivation; but giving students free reign to determine their own lists results in very poor word selection.

In the conclusions of their respective articles, both Moir & Nation (2002) and McCrostie (2007) suggest a third solution: learner-determined but teacher-guided contents. Perhaps the most obvious way for teachers to guide learners' choices would be to encourage the learners to consult frequency lists. Training learners to choose higher-over lower-frequency words, especially when a higher-frequency word has additional meanings a learner is not familiar with, would be one possible way to increase the effectiveness of vocabulary notebooks. This is the solution that I would propose to test.

As any sort of experiment involving self-selected vocabulary would result in no two learners having the same set of target words, administering pre- and post-tests in order to measure vocabulary gains would be extremely difficult, if not impossible. Meara, & Rodríguez Sánchez (2001) suggest that certain issues aside, learner-self assessment can provide reliable measures of vocabulary knowledge, and so I would therefore recommend using a form of Paribakht and Wesche's (Wesche & Paribakht, 1996) *Vocabulary Knowledge Scale* (VKS) on two occasions, to have the learners measure their own vocabulary gains.

As well as being evidence of learning, gains in word knowledge could also be interpreted as evidence of a balanced word list. If learners successfully choose high-frequency words which are relevant to their own learning situations, then gains in knowledge should result as the learners continue to re-encounter these words. Additionally, the usefulness and balance of the word choices would be assessed by categorizing each word in terms of frequency, and part of speech.

Learner attitudes towards vocabulary notebooks, in terms of learner autonomy and motivation, would then be measured through a survey and interviews at the end of the study.

### **3. RESEARCH QUESTIONS**

Since the studies listed in Section 2 above have concluded that

100% teacher- and 100% learner-determined vocabulary notebooks are ineffective at simultaneously promoting the two goals of useful vocabulary choice and learner-autonomy, the comparison which I propose would be between a word list whose contents would be: 50% teacher determined, and 50% learner-determined but teacher-guided (Condition A); versus 100% learner-determined but teacher-guided (Condition B). Condition A is hypothesized to provide more balance in word choice at the expense of autonomy; while Condition B would provide the reverse. As Walters and Bozkurt's (2009) study has already shown that the use of vocabulary notebooks provides significant gains in learning when compared to no vocabulary use, a control group would not be required for this experiment.

The research questions for this experiment would therefore be:

- 1) Which condition would promote greater gains in learner vocabulary knowledge as measured by two self-assessment scales?
- 2) Which condition would promote greater balance and usefulness of word choice as measured by frequency, and part of speech?
- 3) Which condition would promote greater learner autonomy as measured by a survey and interview?

## **4. DESIGN**

### **4.1 Participants**

The design of this study is ideally suited to lower-intermediate to intermediate level, first- or second-year university students. Lower level students suit the study better because they may still have not acquired all of the higher-level items from the frequency lists; university students suit the study because they possess the higher levels of discipline required to update and maintain vocabulary notebooks. With certain adjustments, however, this study could undoubtedly be adapted for use with different learners in different learning situations. Two intact classes should be used: one for Condition A (50% teacher determined/50% teacher-guided content), and one for Condition B (100%

teacher-guided content).

## 4.2 Method

Ideally the study would begin at the start of a course, or semester. As mentioned in Section 2 above, a pre-test would not be used. The study calls for 5 weeks of recording words into a vocabulary notebook, and a self-assessment measure 3 weeks afterward, for a total timeline of roughly 2 months.

At the beginning of the study, the concept of vocabulary frequency levels would be explained to the participants. The participants would then be given a diagnostic test, such as Nation's *Vocabulary Levels Test*, and they would make note of their own level of mastery. The vocabulary lists would be made available for the participants to consult, preferably as spreadsheet files or through an online link.

The concept of a vocabulary notebook would be introduced. The participants would be told that they are responsible for recording 20 words a week into their notebooks (for a total of 100 words at the end of the study). As the later measures of self-assessment are receptive (see below), words would be recorded under their respective word families (Nation & Webb, in press). The participants in Condition A would be supplied with 10 words each week by the teacher/researcher, and told to choose another 10 on their own. The words chosen by the teacher could be taken from a variety of sources: high-frequency words with multiple meanings; words appearing frequently in the course materials; lower-frequency words which are common to that particular learning situation; information from a pilot study etc. Reasonable proportions of the different parts of speech should also be chosen for the learners.

The participants in Condition B would be told to select 20 words on their own.

Participants in both groups would be trained in the first week to select words in the following manner. When an unknown word (or known word with an unknown meaning) is encountered the participants should check it against the vocabulary lists. If the word is

at their level of mastery, or one level above, they should record it in their notebooks. If the word is two or more levels above, they should carefully consider the need to record the word. The “need” to record a word will, of course, depend on the particular participant, however, some guidelines can be observed. Lower-frequency words which are encountered in several sources should take precedence over lower-frequency words which are encountered several times in the same source. Consequently, words encountered several times in the same source would take precedence over a word which only appears once. Choosing very infrequent or rare words (e.g. Level 6 and above) should be discouraged unless a convincing reason for inclusion can be provided.

After recording a word in their notebook, but *before* checking the meaning in a dictionary, the participants would be told to score it on a modified version of the *Vocabulary Knowledge Scale* (VKS). This score would be recorded in the notebook, next to the word itself. The five scores are:

0. I don't know this word.
1. I have seen this word before, but I don't know/remember what it means.
2. I have seen this word before, and I think it means \_\_\_\_ .
3. I know this word. It means \_\_\_\_ .
4. I know different meanings for this word. It means \_\_\_\_ , \_\_\_\_ ... \_\_\_\_ .

The scale has been modified at Level 4: the original productive item “I can use this word in a sentence” has been changed to the receptive “I know different meanings for this word”. This change has been introduced for two reasons. Firstly, it reinforces the importance of learning multiple meanings of a word. As the studies in Section 2 illustrated, many learners will choose to record a completely unknown low-frequency word over an only partially learned higher-frequency word with multiple meanings. The second reason for the change is that the removal of the productive knowledge element insures that the scale

only measures one thing-receptive knowledge- and is therefore closer to a true scale (Nation & Webb, in press). In order to obtain honest measures of self-assessment it is imperative that the teacher inform the participants that vocabulary knowledge scores as indicated by the scale will not be used as grades for the course. The initial scores recorded by the participants will, however, be used in lieu of a pre-test.

Over the course of the five weeks, both groups should spend an equal amount of class time working with the words in their notebooks. It is beyond the scope of this proposal to prescribe a definite syllabus or set of activities that the participants should use, however, attention should definitely be paid to such areas as alternate meanings, collocations, word families, synonyms and antonyms etc. Schmitt and Schmitt (1995) and Walters and Bozkurt (2009) both contain excellent ideas for a schedule of classroom activities involving vocabulary notebooks. Participation in these activities can be used for course assessment.

One of the main learner complaints noted in Walters and Bozkurt (ibid.) was that vocabulary notebooks require much time and effort to maintain. For this reason it is recommended that the information recorded in the notebooks be restricted to: a) the L2 word; b) the initial VKS score; c) the L1 meaning(s); and d) 1-2 example sentences. This restriction means that information such as reason for word choice, and source of word (information that was collected and examined in the Moir and Nation, 2002, and McCrostie, 2007, studies) would unfortunately be excluded from the final analysis. Any classroom activities requiring additional information such as keywords, collocations, pronunciation guides etc. should utilize separate handouts. If the vocabulary notebooks are loose-leaf paper kept in a binder, then these classroom handouts can be added in afterwards.

At the end of the fifth week, all participants should submit their final vocabulary notebook word list as a spreadsheet file. The file would only need to contain two sets of data; the words themselves in one column; and the initial VKS score in another. The teacher should make



a duplicate copy of each participant's file. The original (File A) would contain the word list and VKS scores. The teacher would then delete the VKS scores from the duplicate (File B) so that it only contained the word list. After three weeks, the teacher would then give each participant their own personal File B. In lieu of a post test, the participants would be asked to rescore their word list in File B according to the VKS *without* referring to their original vocabulary notebooks. Again, it is imperative that the teacher emphasize to the students that the second set of VKS scores would be used for research and *not* assessment purposes (although classroom activities involving the word lists may be assessed). If the teacher still feels that dishonest responses are a possible problem, nonsense words could randomly be added to the word lists in File B as a precaution.

The teacher would then collect each participant's File B, and merge it with File A in order to create a third file. This file (File C) would contain three columns: the word list; the initial VKS score, and the second VKS score. A fourth column would be created by subtracting the first VKS scores from the second, and this column would show the direction and strength of learning (positive numbers) and forgetting (negative numbers) for each word. As the VKS is not a true interval scale, however, these raw scores must be interpreted with caution. Nevertheless, a rough idea of the amount of learning achieved by each participant could be obtained by totaling up all of the numbers in Column 4. The first research question, "Which condition would promote greater gains in learner vocabulary knowledge?" could then be answered by calculating and comparing the mean scores for the participants in the two conditions.

Additionally, all of the words from all of the participants in one condition could be combined into a single master spreadsheet file. The data in that file could give additional information about the ease or difficulty of learning and remembering each word. For example, a participant may have scored a word like "explore" as a "0" initially, and then as a "3" on the second assessment, for a total of "+3" as a final score.

By consulting the master list, we may find that the word “explore” was scored as “+4” once, “+3” five times, “+2” three times, and “+1” or less zero times. This would suggest that everyone who recorded the word “explore” made some knowledge gains with the item, and may also indicate that the word appears frequently in the L2 environment. In contrast, a word like “perpendicular” may have received no scores above “+1”, two “0” scores, three “-1” scores, two “-2” scores etc. This would suggest that the word is more difficult to remember and/or encountered less in the L2 environment of the participants.

In order to measure the balance of the word lists, a simple census of the various frequencies and parts of speech for the master lists of each condition would be performed. This would answer the second research question: “which condition would promote greater balance and usefulness of word choice?”.

Finally, in order to answer the third research question, “which condition would promote greater learner autonomy as measured by a survey and interview?”, a Likert scale survey and follow-up interview would be administered. Again, it is beyond the range of this proposal to define the exact questions asked, however, Walters and Bozkurt (2009), Fowle (2002), and Moir and Nation (2002) all provide useful examples.

## **5. INTERNAL VALIDITY**

Nation and Webb (in press) list several validity considerations that must be taken into account when designing and implementing vocabulary research. These will each be addressed in turn.

### **5.1 Subjects**

The subjects should be at the same level, or roughly equivalent for Condition A, as each participant will receive the same 50 target words from the teacher. As most universities stream classes into approximately equal groups, this should not be a major issue.

## **5.2 Materials**

Classroom materials should be the same in both conditions. Any additional sources (CD's, DVD's, websites, books and magazines) that participants draw vocabulary from will obviously differ from participant to participant. However, one could argue that each participant has equal access to these materials (the same CD's, books, and magazines are sold in many shops; websites are free to visit etc.).

The target words will obviously differ from person to person, but the two self-assessment scales have been implemented to control for this.

## **5.3 Treatment**

The same treatments will be applied across participants, and across groups, with the sole difference of Condition A being provided with 50 words from the teacher. The surrounding conditions (time on task in classroom activities, other courses, school, EFL environment) will also be the same across groups.

## **5.4 Measures**

Measures will be both administered and scored the same across participants and groups. See Section 4.2 above for details. As the treatment extends between, and not within, groups (the self-assessment scales controlling for within group differences), there is no need to control for order effects.

## **6. ECOLOGICAL VALIDITY**

Several issues of ecological validity must also be considered.

### **6.1 Texts**

The classroom materials used in the study should be typical for learners in terms of content and length. The additional materials that the participants draw vocabulary from (books, CD's etc.) will be self-selected, and should therefore be appropriate (teacher guidance on

material selection may be necessary, however).

## 6.2 Words

With teacher guidance and instruction, and access to word frequency lists (see Section 4.2 above), the unknown words should be appropriate for all the participants. The words that the participants select will obviously be taken from situations of context, and the words that the teacher selects (in Condition A), should be as well. As the words are all real words of an appropriate level, no ethical concerns would be raised.

## 6.3 Treatment

As vocabulary notebooks are now becoming more and more common in EFL classrooms, the treatment in both conditions can be considered part of a normal learning activity. The participants can be made aware that they are taking part in an experiment, and in fact, informing them that they are may be necessary in order to elicit honest self-assessment responses.

## 6.4 Measures

The type of measures will obviously be relevant to the learning goal, as self-assessment is an important part of the selection of words in a vocabulary notebook. In fact, the vocabulary notebook that comes bundled with the textbook *Communication Spotlight* comes with a guide to the *Vocabulary Knowledge Scale* in the instructions to the user, and recommends that each word recorded be scored on the VKS.

## 7. PILOT STUDY

Several potential problems may be avoided if the study is first given a pilot-test:

- 1) Is 20 words a week too much (or not enough) for the participants to record in their notebooks?
- 2) Is the information recorded in the notebooks (the L2 word; the

- initial *Vocabulary Knowledge Scale* score; the L1 meanings; example sentences) too much or not enough for the participants?
- 3) Can the participants be relied on to provide honest VKS scores?
  - 4) What is the best way to provide access to vocabulary frequency lists?
  - 5) Will the participants be able to manipulate their data using spreadsheet software correctly?
  - 6) Which questions/items should be included on the survey and interview measuring student autonomy?
  - 7) Which activities should be employed, and which areas should be covered, when working with the vocabulary notebooks in class?

In addition, running a pilot study may provide the researcher/teacher with useful information about which 50 words to provide in Condition A (50% teacher determined/50% teacher-guided content).

## 8. CONCLUSION

In Section 2, the author demonstrated that what little research exists on vocabulary notebooks shows their ineffectiveness at promoting both learning and learner autonomy at the same time. Despite these findings, their use in EFL classrooms is only continuing to rise. It is hoped that the design proposed in this paper will lead to a way in which vocabulary knowledge gains *and* independent study of vocabulary can both be increased through the use of these notebooks, to the benefit of EFL learners everywhere.

## REFERENCES

- Communication Spotlight* (2006) Tokyo: Abax
- Fowle, C. (2002) "Vocabulary Notebooks: Implementation and Outcomes". *ELT Journal* 56 (4), pp. 380-388
- Schmitt, N., & Schmitt, D. (1995) "Vocabulary Notebooks: Theoretical Underpinnings and Practical Suggestions". *ELT Journal* 49 (2), pp. 133-143

- Schmitt, N. (2000) *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press
- McCrostie, J. (2007) "Examining Learner Vocabulary Notebooks". *ELT Journal* 61 (3), pp. 246–255
- Meara, P., & Rodríguez Sánchez, I. (2001) "A Methodology for Evaluating the Effectiveness of Vocabulary Treatments". Retrieved November 14, 2009, from <http://www.lognostics.co.uk/vlibrary/index.htm>
- Moir, J., & Nation, I. S. P. (2002) "Learners' Use of Strategies for Effective Vocabulary Learning". *Prospect* 17 (1), pp. 15–35
- Walters, J., & Bozkurt, N. (2009) "The Effect of Keeping Vocabulary Notebooks on Vocabulary Acquisition". *Language Teaching Research* 13 (4), pp. 403–423
- Wesche, M., & Paribakht, T. S. (1996) "Assessing Second Language Vocabulary Knowledge: Depth Versus Breadth". *Canadian Modern Language Review* 53, pp. 13–40

#### Keywords

Vocabulary Acquisition, Learner Autonomy, Learning Strategies