Exploring factors affecting active recall of technical terminology

Elżbieta Lesiak-Bielawska
Fryderyk Chopin University of Music
elzbieta0606.lb@gmail.com

Abstract

Computer-based language instruction has been extensively researched in EAP and a series of studies have investigated the impact of technologies on the acquisition of discipline-specific vocabulary. Most of them are product-oriented. This study adopts both the product and process perspective on the issue and explores differences in active recall of technical terminology between two student groups learning in different blended learning modes – the computer lead mode and face-to-face mode. The differences explored not only relate to the students' post-test score means but also the scores obtained by them on the *Foreign Language Aptitude Test*. The results obtained suggest that recall of technical vocabulary was more difficult for the group acquiring discipline-specific terminology in the computer lead mode. They also indicate that some learner variables related to foreign language aptitude had an impact on the students' post-test scores.

Key words: technical terminology, vocabulary recall, blended learning, different study conditions, foreign language aptitude

1.Introduction

Research into factors influencing success in second/foreign language (S/FL) learning goes back to the 1960s and first focused on "big two" individual difference (ID) variables – language aptitude and motivation (Ellis, 2004: 536). Researchers who view language aptitude as a major determinant of successful language learning emphasize its relative autonomy in relation to other ID factors affecting S/FL learning achievements, such as motivation, attitude, anxiety or learning styles (Carroll, 1981; Skehan, 1989). Others point to the fact that successful language learning is also affected by a series of non ID factors, such as sex, age, family background, learners' first language (L1) or teaching methods (Oxford, 1999).

This paper reports research findings related to selected factors affecting vocabulary recall in English for Specific Academic Purposes (ESAP). The objective of the study conducted in an academic setting was to explore the impact of two different blended learning modes on ESAP

vocabulary recall and a series of underlying individual difference variables affecting learners' retention and recall of technical terminology. It was thus an attempt at investigating both the product- and process-related differences between the groups as reflected in the post-test score means received by the learners and the results obtained by them on the *Foreign Language Aptitude Test* (FLAT, Rysiewicz, 2012).

2.Research background

Although the phenomenon of ESP is a modern one, the study of languages for specific purposes (LSP) is centuries old and dates back to ancient travel guides, dictionaries of medical and maritime terms, as well as military and nautical glossaries (Grosse and Voght, 1991). Its modern origins, however, go back to the early 1960s and are associated with J. M. Swales' (1988) *Episodes in ESP*.

During the early phase of ESP evolution, research focused on English for Science and Technology (EST) in academic settings (Hutchinson and Waters, 2010). At the onset of its development, ESP remained under the influence of the teaching of English for Academic Purposes (EAP) and research conducted in the area (Dudley-Evans and St John, 2009). Alongside EST, which is considered to be the main area in EAP, other important specialisms include English for Medical Studies (EMS) and English for Legal Purposes (ELP), let alone a growing interest in the academic study of business, finance, banking, economics and accounting (Dudley-Evans and St John, 2009).

Since the turn of the twentieth century, ESP instruction has been considerably influenced by the development of new technologies. Although not originally designed for education, their current everyday use has led to their incorporation into the foreign language classroom. The Internet has had a profound influence on the process of foreign language learning and information technology has been applied in all types of contexts, both as a tool for language learning and space for generating new forms of discourse (Bloch, 2013). Internet-based instruction has been incorporated into many EAP courses for several reasons (Krajka, 2007). The most important of them relates to the limited availability of ESP teaching materials, which is often conditioned by marketing interests and predicted sales[1]. Thus in many EAP teaching

[1] Master and Brinton (1998) divide published ESP materials into two categories – major and minor ESP areas, with the former being exemplified by English for Business and Economics (EBE). As reported by Bocanegra-Valle (2010), a decade later the situation remained the same. The majority of the published ESP books (215 titles)

contexts, coursebook instruction is frequently either completely replaced or simply supplemented with Web-based lessons, since the Internet offers a limitless repository for authentic ESP materials as well as opportunities for active interaction across the Web.

In the context of EAP, computer-based language instruction and the use of Web 2.0 technologies have been researched extensively, focusing, for example, on the use of wikis (e.g. Kuteeva, 2011), blogs (e.g. Murray et al., 2007), networking (e.g. Warschauer, 2002), the use of corpora (e.g**.** MacDonald et al., 2013) or blended learning (cf. Tomlinson and Whittaker, 2013). As shown in those studies, new technologies undoubtedly exert a positive influence on the development of foreign language skills, let alone vocabulary, which **–** after years of neglect – has obtained its central status in the process of foreign language acquisition (Laufer, 1997; Nation, 2001).

The priority of foreign language lexis over grammar that is frequently emphasized in the area of Foreign Language Teaching (FLT) (Lewis, 1993; Komorowska, 2005) seems to result from the growing significance of foreign languages, for both general and specific purposes. In the ESP field, it stems from our recognition that ESP vocabulary is closely linked to the specialized language of a particular discipline and thus related to specific-purpose needs of target learners. Aspiring to become fully-fledged members of a particular community, they need to understand and use that specific-purpose vocabulary to show their understanding of a given discipline and demonstrate their ability to ”make meaning and engage with disciplinary knowledge” (Woodward-Kron, 2008: 246).

The present study comes close to other studies that tried to explore the impact of blended learning approach on the acquisition of discipline-specific vocabulary (e.g. Ono and Ishihara, 2012; Jung and Lee, 2013; Vasbieva et al., 2016). It thus focused on the product-related differences between the student groups acquiring ESAP vocabulary in different learning modes (i.e. on the students’ post-test score means). However, in contrast to previous studies, it also attempted to incorporate process-related differences between the learners as reflected in their scores on six foreign language aptitude subtests (FLAT; Rysiewicz, 2012).

focused on “Banking, Business and Finance”, which is still considered to be the top ESP area. This practical distinction reflects the policy of many publishers who “are naturally reluctant to produce materials for very limited markets” (Hutchinson and Waters, 2010: 106).

The origin of language aptitude research goes back to the 1920s and 1930s. It was initiated with the objective of increasing the cost effectiveness of language teaching in American schools (Spolsky, 1995). Thirty years later, the same thinking gave rise to the second wave of language aptitude test development often referred to as the "golden period" of scientific language aptitude testing (Rees, 2000). It resulted in the development of two test programs – the *Modern Language Aptitude Test* (MLAT; Carroll and Sapon, 1959) and the *Pimsleur Language Aptitude Battery* (PLAB; Pimsleur, 1966), which by far have been the most widely used and referred to.

In the 1970s and 1980s, research into foreign language aptitude somewhat languished as the concept was criticized for its anti-egalitarian nature (i.e. low test score identifies a language loser who has no chance to learn a foreign language) and indecent origin closely associated with the outdated audio-lingual methodologies (Dörnyei and Shehan, 2003). Since the 1990s, a notable reawakening of interest in language aptitude has been observed. According to Wen and Skehan (2011: 18), most of the post-Carroll aptitude research can be grouped into the following three categories:

1. psychometric in nature measurement oriented research that targets the development of aptitude tests themselves (e.g. *CANAL-FT*, Grigorenko et al., 2000);

2. research based on the different components of aptitude as conceptualized by Carroll (e.g. Sparks and Ganschow, 1991; Skehan, 1986; Sasaki, 1996), which is quite scare, especially into the working memory component which is now receiving more attention;

3. research which sets out to address aptitude-treatment interactions (e.g. Wesche, 1981; Robinson, 1995; Robinson, 2002).

The growing interest in the issue of foreign language aptitude has led to adapting the batteries, especially the MLAT, considered to be "the best overall instrument for predicting language-learning success" (Parry and Child, 1990: 52), to foreign language contexts. The *Foreign Language Aptitude Test* (FLAT; Rysiewicz, 2012) employed in the present research exemplifies an attempt at adapting the MLAT created for second language contexts to a foreign language educational context.

It is rightly contended that empirical investigation into the effectiveness of ESP teaching is limited (Dudley-Evans and St. John, 2009) and mainly restricted to English for Academic Purposes (EAP) (Basturkman, 2010). The present study exemplifies the contention, since it was

conducted in an academic setting. To the best of my knowledge, no study attempted to explore differences between learner groups acquiring ESAP vocabulary in two different blended learning modes. What is more, no study tried to investigate how the product-related differences between the groups (i.e. the post-test score means) are reflected in the process-linked learner profile (i.e. the scores on foreign language aptitude subtests) (Blake, 2013).

3. Key research categories

The teaching procedure referred to as "**blended learning**" combines traditional face-to-face interaction with computer-mediated activities. It is one of the most frequently employed options in the process of foreign language teaching and is the result of combining two different learning environments with a view to optimizing the conditions for learning and increasing its efficiency. According to Garrison and Vaughan (2008), each type of learning brings with it a distinct atmosphere and entails a unique learning experience. Face-to-face interaction promotes social presence and allows learners to get a sense of belonging to a community. It is thus important in building group identity and prepares learners for a more individualistic style of online learning. It can also be added that many students are accustomed to this kind of instruction and find it difficult to successfully cope with less traditional approaches to learning, especially those involving a higher proportion of computer-based activities, which require more autonomy and independence on the part of the students.

According to the definition of blended learning offered in the publications of the Sloan Consortium, the online component can come within the 30-79% range of the course (Maciaszczyk, 2011). When developing blended instruction, course designers frequently refer to Driscoll's (2002) framework for the possible blends of instructional delivery methods consisting of Web/Computer-Based Training, Web/Electronic Performance Support System, Web/Virtual Asynchronous Classrooms, and Web/Virtual Synchronous Classrooms. However, this framework views blending as a mix of web-based technologies with other kinds of technology-based instruction and does not include diverse instructional methods used in classroom settings. Neither does it consider such facets of learning modes as self-paced learning, collaboration, assessment and learning support materials which are conducive to students' learning in the process of blended learning instruction (Carmen, 2002).

Consequently, for the purpose of the present research, blended learning is defined as any combination of face-to-face, computer and self-study modes (cf. Whittaker, 2013). The research

reported in the paper used two blended learning modes: the face-to-face lead mode, which employed more face-to-face tasks and fewer computer-mediated activities and the computer lead mode, which consisted of fewer face-to-face assignments and more computer-based activities (also see section 5.3.).

Vocabulary is considered a core component of language proficiency inasmuch as it provides much of the basis for how well learners speak, listen, read, and write. It is of great significance in ESP courses, which aim at equipping students with discipline-specific discourse and thus need to focus on specialized vocabulary of a given field. A particularly important issue related to **ESP vocabulary** is its size. As observed by Nation (2008: 10), "we do not know a lot about technical vocabularies but they probably range in size from around 1,000 words to 5,000 words depending on the subject area." This implies that ESP learners are likely to face an extremely challenging task if they are to fully develop their understanding and use of specialized vocabulary in a given subject area – be it university or a professional context.

In the literature, ESP vocabulary is referred to by different names – special-purpose, specialized, technical, sub-technical and semi-technical vocabulary (Coxhead, 2013). These terms usually relate to the vocabulary of a particular area of study or profession. Given the fact that the range of a word is important in ESP, a specialized word can be defined as having a narrow range of use within a particular subject area. This means that specialized words belong to a particular academic area or professional discipline and can be "identified by referring to specialists who have a good knowledge of the subject area" (Chung and Nation, 2004: 252).

As far as vocabulary knowledge is concerned, learners usually know more words than they use. For that reason, it is customary to distinguish between receptive and productive vocabulary knowledge (e.g. Meara, 1990; Nation, 2001). Passive knowledge is associated with receptive language skills, i.e. reading and listening, and implies that one is able to understand the input, i.e. to perceive the word form and retrieve its meaning or meanings from memory. Active knowledge, on the other hand, is related to productive skills, i.e. speaking and writing, and involves one's ability to retrieve a particular spoken or written word form appropriately to the meaning that is to be expressed.

Owing to the confusion related to the passive-active vocabulary knowledge distinction (for more details see Laufer and Golstein, 2004), Laufer and Golstein (2004) distinguish four degrees of vocabulary knowledge:

1. **Active recall** involves supplying the target word in a foreign language, e.g. the student provides the target word in its written form being given the number of blanks corresponding to the number of letters and one extra letter in the appropriate blank;

2. **Passive recall** refers to showing the understanding of the target word meaning, e.g. the student is to match the target word with its L1 translation choosing from a few options;

3. **Active recognition** involves choosing the target word from a few options, e.g. the student has to match the target word L1 translation with the target word by choosing from a series of lexical items, some of which are semantically unrelated;

4. **Passive recognition** relates to choosing the target word meaning from a series of options provided in L1.

The above distinction, with its emphasis on active/passive recall of vocabulary, is relevant to the purpose of the present study. Its objective was to investigate the subjects' success on a language task involving active recall of technical terminology and relate their scores to a series of underlying language aptitude-linked IDs.

Thus, last but not least is **the *Foreign Language Aptitude Test*** (FLAT, Rysiewicz, 2012), which is a Polish adaptation of the MLAT created in and for the Polish educational context. However, in contrast to its prototype, which employs five tasks to test language aptitude, the battery employed in the research consists of six subtests, each of which measures different abilities related to the construct.

The ***Phonetic Alphabet*** **(PhA)** subtest measures phonetic coding ability, or the ability to form a link between speech sounds and their graphemic representations. It also checks the capacity and duration of echoic memory. Before performing the task related to the subtest, students are provided with a printed phonetic script, presented in quite simple and regular symbols. The task includes 30 sets of four words each. First, students follow their scripts and listen to a set of short nonsense words. Then they hear one word at a time and have to decide on one of the four printed alternatives.

The subtest is important for the purpose of the present study, since it measures phonetic coding ability, which is well supported in aptitude research. As demonstrated by Carroll (1981) and Sparks et al. (1997), it predicts general second language proficiency and is of great significance

in the early stages of learning (Skehan, 1999). All the major aptitude batteries include tests of phonetic coding ability as a determiner of the degree to which learners can make use of both oral and written linguistic input. These sections of aptitude batteries require that testees perceive orally presented nonsense syllables, hold these sounds in their phonological working memory for the duration of the task, simultaneously consciously segmenting the syllables into smaller phonological units so as to finally identify the relationships between these phonological units and their written representations. Consequently, the *Phonetic Alphabet* subtest measures phonological working memory[2], which functions as a gateway for storing linguistic knowledge into long-term memory (Baddeley et al. 1998) and plays a pivotal role in vocabulary learning (e.g. Masoura and Gathercole, 1999).

Another related to the above-discussed section of the battery is **the *Hidden Words* (HidWs)** subtest, which assesses phonetic coding ability and vocabulary in students' L1 in the situation of semantic ambiguity. The subtest consists of 30 L1 words, whose spelling is distorted, for example, by the word vowels being removed or erroneous word orthography employed. The order of the word letters is not changed. The task related to the subtest consists in recognizing the distorted words by ticking from the four printed alternatives the word meaning that best corresponds to a given hidden word. Accordingly, the subtests addresses two issues – knowledge of native vocabulary and 'grapheme-to-meaning' dimension, which functions in basically the same way as Carroll's 'phonemic coding' does on the 'sound-to-meaning' plane.

Not only is the sub-test significant for predicting vocabulary acquisition connected with the role of phonetic coding ability, but it is also linked to research findings that relate the learner's native language skills to the capacity to master a second/foreign language. This latter view is supported by Skehan (1989), who observes that foreign language aptitude is, to some extent, a residue of first language learning ability. Also, as posited by the authors of the Linguistic Coding Differences Hypothesis (LCDH) (cf. Sparks, 1995; Sparks and Ganschow, 1991; 2001), the learner capacity to learn a second/foreign language is closely related to the individual native language learning skills, and most particularly to 'linguistic coding', which refers to L1 literacy

[2] Working memory has its origin in unitary short-term memory which is the site of temporary storing of small amounts of material in short periods of time (Baddeley, 1992). According to Baddeley (1992), this 'brain system' is needed to hold information (the storage function of working memory) and to manipulate new information with known material (processing function).

skills such as phonological/orthographic processing and word recognition, its coding and decoding, or – in other words – single-word reading.

The ***Artificial Language*** **(ArtLg)** subtest assesses inductive language learning, which refers to the ability to perceive the change in language forms as expressed by grammatical endings, function words, word order, etc., and recognize accompanying changes in grammatical and lexical meaning. Performing the task related to the subtest, learners are requested to analyze words and expressions from a fictitious language, Doran, presented to them. Following that, they have to choose the right Polish translation of each of the 10 sentences presented to them in Doran. In addition to that, they are asked to translate Polish fragments provided in three sentences into Doran.

This is another subtest relevant to vocabulary learning which assesses how skillfully learners can cope with form changes of a language unknown to them and how successfully they can relate these changes to appropriate meaning. Thus its objective is to illustrate the strengths and weakness of their working memory as well as their capacity to segment language input in its written form and impose some structure on it.

The ***New Words*** **(NWs*) subtest** measures rote, or short-term memory, for words. In this test, learners have a total of one minute to memorize the most of 24 Kurdish/Polish word pairs. Retention is tested by means of a multiple choice test in which students are requested to choose the proper equivalent for each Kurdish word from five Polish alternatives. All the distractors are selected from the 24 Polish words contained in the original list.

This subtest is important for the purpose of the study, since it tests rote memorization of vocabulary word pairs in a short period of time. The use of the subtest is based on the assumption that learners' memory varies and this very fact accounts for differences between them in this kind of vocabulary learning task.

The recently included subtest ***Number Learning*** **(NL)** measures the already-discussed components of the construct treated separately in artificial language (inductive language learning), new words (rote memory for words) and phonetic alphabet (phonetic coding ability). It tests the ability to infer language rules through the auditory analysis of linguistic input. In this test, subjects hear some numbers in a new language (only numbers 1-4, 10-40, 100-400)

and are provided with some auditory practice to learn them. Then they must write 20 numbers between 1 and 400 and their combinations as the numbers are read in a new language.

Last but not least is the ***Words in Sentences*** **(WSs)** subtest, which estimates grammatical sensitivity by testing learners' ability to recognize grammatical functions of words in sentences without directly referring to their grammatical knowledge. Subjects are first presented with a key sentence in which a word or phrase is underlined. In the sentence or sentences following the key sentence, four alternative words or phrases are underlined. Subjects must select the one that performs the same function as the word or phrase underlined in the key sentence. There are altogether 23 key sentences in the subtest.

Since this subtest assesses learners' ability to recognize grammatical functions of words and phrases presented in sentences, it is also relevant to the purpose of the study. However, in contrast to other sections of the FLAT which address intentional learning[3] of decontextualized vocabulary items, this subtest explores testees' extracting skills related to functions and also meaning of words and phrases embedded in context. In this sense, it is related to incidental vocabulary acquisition, which is generally defined as the "learning of vocabulary as the by-product of any activity not explicitly geared to vocabulary learning" (Hulstijn, 2001: 271).

[3] Intentional vocabulary learning is defined as "any activity geared at committing lexical information to memory" (Hulstijn, 2001: 271).

4. Research problem

The present study was designed to explore the impact of two different instructional procedures on active recall of ESAP vocabulary. It also aimed at investigating foreign language aptitude-related differences between the learner groups acquiring technical terminology in different study conditions, or two different blended learning modes. It was assumed for the purpose of the present study that statistically significant differences could be observed between the groups in their active recall of ESAP terminology. They were hypothesized to be accounted for by a series of foreign language aptitude-related differences between the students and the learning mode in which ESAP terminology was acquired.

The selected aspect of ESAP was defined as the British political system versus the American political system, and their institutions. The topics discussed in class focused on such issues as the British Monarchy, British Parliament, executive, judiciary, two-party system, general elections in the UK, as well as the American Presidency, Congress, political parties, elections and the judiciary in the USA.

4.1. Research hypotheses and operationalization of variables

The present study set out to answer the following research questions:

(1) Are there any statistically significant pre-test score differences between the groups related to their entry ESAP vocabulary knowledge?

(2) Are there any statistically significant post-test score differences between the groups acquiring technical vocabulary in different study conditions in the area of active recall of ESAP terminology?

(3) Are there any statistically significant differences between the groups related to their results obtained on various subtests of the FLAT?

To explore the differences between the groups related to the above questions, the Independent T-Test was employed. In keeping with the assumptions of the test, three sets of hypotheses were formulated pertaining to the above research questions. The null hypotheses for the study (H_0) stated that the score means of the students from the groups related to the pre- and post-treatment scores and various subtests of the FLAT were not significantly different. The alternative hypotheses (H_1) assumed that the score means of the students from the groups displayed

statistically significant differences in their pre- and post-test results as well as in the scores obtained by them on various FLAT subtests.

Accordingly, the students' results on the pre-test constituted an independent variable, the FLAT subtests constituted measures of moderator variables and their post-treatment scores – a dependent variable. The variables thus operationalized included:

- pre-treatment test – independent variable
- post-treatment test – dependent variable
- FLAT (total score) – moderator variable
- PhA – moderator variable 1
- HidWs – moderator variable 2
- WSs – moderator variable 3
- NWs – moderator variable 4
- ArtLg – moderator variable 5
- NL– moderator variable 6

Different instructional conditions, i.e. face-to-face lead mode and computer lead mode, represented the grouping variable (independent variable).

5. Methods

5.1. Respondents

The study was conducted in a group of 113 intermediate adult learners studying EFL at the University of Cardinal Stefan Wyszynski in Warsaw in the academic year 2014-2015. The group consisted of 90 females and 23 males. All the students studied international relations at the Faculty of Law and Administration. The respondents had the same teacher and came from 6 student groups.

The selection of students followed a non-probabilistic sampling strategy referred to as convenience, or opportunity sampling (Dörnyei, 2011). All the subjects studied in the researcher's own institution and qualified for the B1+ level of proficiency in English for General Purposes (EGP). Their ESAP vocabulary knowledge was measured on a test with the maximum score of 20 points. The pre-test scores obtained by the students ranged from 0 to 2 points and the differences between the groups were not statistically significant (see section 6).

5.2. Research design

To investigate differences between two learner groups acquiring ESAP vocabulary in different instructional conditions, a comparison study employing experimental research design was conducted. It had both the pre-treatment and post-treatment tests, two student groups acquiring technical terminology in different instructional conditions and no random assignment of subjects (Nunan, 2004).

Before each group was provided with the appropriate teaching procedure, the respondents' knowledge of ESAP terminology was assessed by means of the pre-treatment test. It consisted of 19 English sentences, in which 20 words were blanked. The students were required to supply missing words, drawing on the number of blanks, each of which corresponded to one word letter and using only one letter that was provided for each word in one of the blanks. Following the treatment that was different for each group, the subjects received the same test, i.e. the post-treatment test. The maximum score for the pre-/posttest was 20 points (see Appendix 1).

5.3. Teaching procedures

Two disparate blended learning modes employed in the study represent two different study conditions. Both were designed for eighteen-lesson-unit instruction (with each lesson unit lasting 45 minutes) and contained the same five elements:

- face-to-face material presentation;
- face-to-face exercises;
- computer activities;
- self-study;
- pre- and posttest.

However, depending on the group, the proportion of some elements was either higher or lower and thus affected the lead mode of each group. The lead mode employed in group A is referred to as the face-to-face mode, whereas the blended learning procedure that was employed in group B is referred to as the computer lead mode. After the presentation of the lexical material in each group, the students acquired technical terminology in their group lead mode. Table 1 presents differences between the groups when it comes to the learning/teaching procedures employed.

Elements of instruction	Group A: Face-to-face lead mode	Group B: Computer lead mode
Pre-treatment test	½ lesson unit	½ lesson unit
Face-to-face presentation of the core materials	4 lesson units	4 lesson units
Face-to-face consolidation exercises	**7 lesson units**	**2 lesson units**
Computer consolidation exercises	**2 lesson units**	**7 lesson units**
Self-study	4 lesson units	4 lesson units
Post-treatment test	½ lesson unit	½ lesson unit

Table 1. Differences between the groups in the teaching/learning procedures employed

As can be seen from it, the core material presentation in the face-to-face mode took 4 lesson units. Following that, group A – working under the teacher's supervision – spent 7 lesson units doing pencil and paper exercises and a series of topic-related communicative activities in class, while 2 lesson units of their time were devoted to computer activities designed to consolidate the lexical material. In group B, the time devoted to these activities was reversed. It took one lesson unit to administer the pre-treatment and post-treatment tests, i.e. ½ lesson unit for the pre-test and ½ lesson unit for the post-test.

The exercises conducted in the face-to-face mode, computer mode and self-study mode supplemented the syllabus and provided the students with controlled practice and extension activities, giving them the opportunity to review and recycle in the group lead mode all that was presented in the face-to-face mode. For the face-to-face material presentation, a blend of customized texts with reading comprehension questions was employed in both groups. They were based on *Lexical Compendium: Politics* by K.A. Luto and M. Ganczar (2007). The texts were supplemented with a selection of short YouTube videos (e.g. the British Constitution, principles of US Constitution, etc.) followed by comprehension questions. The materials presented in class were also embedded on an open-source e-learning platform MOODLE (UKSW, 2011) so that the students could refer to them when learning.

Additionally, a series of interactive quizzes to selected texts and videos as well as a range of vocabulary exercises (e.g. crosswords, gap-filling, matching tasks, etc.) were created for group B by means of the free lesson construction software *Hot Potatoes* Version 6.0. These activities were designed for 7 lesson units. Group A had access to the same activities but in the face-to-face mode. Following the consolidation of the lexical material in the group appropriate mode, the students were requested to give mini presentations on, e.g. the principles of the British/American Constitution, British/American Legislature, Judiciary, etc., prepared by them either in class or at home.

5.4. Research procedures

Attempting to explore differences between the groups related to active recall of ESAP vocabulary and the students' results on the Foreign Language Aptitude Test, the following procedure was launched:

- administering the ESAP vocabulary pre-test which aimed at collecting data on the subjects' pre-treatment knowledge of technical terminology;
- purposive sampling of the respondents to different instructional conditions; three student groups of 56 respondents were classified to group A (face-to-face mode, whereas group B (computer lead mode) consisted of 57 subjects; the groups were matched in terms of size, area of study, the results obtained on the EGP placement test and ESAP pre-test;
- administering the *Foreign Language Aptitude Test* (FLAT; Rysiewicz, 2012) to collect data related to their language aptitude;
- launching the teaching procedure different for each group;
- administering the ESAP vocabulary post-test which was designed to collect data on the subjects' post-treatment active recall of technical terminology;
- employing IBM SPSS Statistics 21.0 to conduct both the preliminary processing of the collected data and their statistical analysis.

6. Results

In the present study, IBM SPPS Statistics 21.0 was employed to conduct the preliminary processing of the collected data and their analysis. To describe general tendencies in the research findings and the overall spread of the scores, descriptive statistics was used. It constituted the basis of further inferential statistics. Subsequently, the Independent T-Test was employed since there were no considerable dissimilarities between the groups when it comes to the number of learners. Also, after running the Kolmogorov-Smirnov test, the data collected in the respondent group tended to be normally distributed. In addition to that, the results of Levene's Test for the assumption of homogeneity of variance showed that population variances were equal.

The Independent T-Test was used to explore differences between the groups related to the students' pre-treatment ESAP vocabulary knowledge, their post-treatment scores in the area of active recall of technical terminology and the results obtained by them on various subtests of the FLAT. The threshold value for the significance level, or the alpha level (α), was established at 0.05.

As regards **the first research question** related to the students' entry knowledge of ESAP terminology, the obtained results indicated that there was no statistically significant difference between the groups when it comes to their scores obtained on the pre-treatment test (M = 0.00; SD = 0.00 for group A and M = 0.05 SD = 0.29 for group B t (111) = 1.34; p > 0.01). Thus the null hypothesis was accepted.

As for **the second research question**, on the basis of the results obtained in the course of the study, the alternative hypothesis was accepted, since the post-treatment results showed a statistically significant difference in scores between the groups (M = 13.09; SD = 4.66 for group A and M = 8.03; SD = 4.13 for group B *t* (111) = 6.12; $p < 0.01$). Table 2 summarizes the results obtained for the groups at the pre- and post-treatment stages of the study.

Comparison of pre- and post-test results	**Group A**	**Group B**	**Differences in the score means**
pre-treatment test	M = 0.00; SD = 0.00	M = 0.05; SD = 0.29	t (111) = 1.34; p > 0.01
post-treatment test	M = 13.09; SD = 4.66	M = 8.03; SD = 4.13	t (111) = 6.12; p < 0.01

Table 2. Comparison of the pre- and post-test results

In addition to the statistically significant difference in scores between the groups at the post-treatment phase, the results of the present study showed statistically significant differences between them in three moderator variables tested, i.e. New Words, Words in Sentences and Artificial Language subtests (**the third research question**). Table 3 summarizes statistically significant results for all the tested variables.

Tested variables with statistically significant results	**Differences in the score means**	**Cohen's effect size**
post-treatment test	t (111) = 6.12; $p < 0.01$	$d = 1.16$
subtest ***New Words***	t (111) = 3.425; $p < 0.01$	$d = 0.65$
subtest ***Words in Sentences***	t (111) = 2.14; $p < 0.05$	$d = 0.40$
subtest ***Artificial Language***	t (111) = 1.81; $p = 0.073$	$d = 0.34$

Table 3. Statistically significant results for the tested variables

As can be seen from it, the post-treatment results show a statistically significant difference in scores between the groups t (111) = 6.12; $p < 0.01$, with Cohen's effect size $d = 1.16$. The d value > 0.8 points to a large effect size thus indicating that the magnitude of the difference in the means was large, with the learning mode explaining 25% (eta squared = 0.25) of the variance in the post-test results.

Also, the results obtained on the subtest *New Words* show a statistically significant difference in score means between the groups t (111) = 3.425; $p < 0.01$, with Cohen's effect size $d = 0.65$. And here again, the score means of the group with the face-to-face lead mode are higher than score means of the group acquiring vocabulary in the computer lead mode (M = 52.34; SD = 9.99 for group A and M = 46.28; SD = 8.84 for group B). The d value suggests a medium effect size thus indicating that the magnitude of the difference in the means was medium, explaining 9% (eta squared = 0.09) of the variance on the subtest within the group.

Likewise, the results obtained on the subtest *Words in Sentences* show a statistically significant difference in score means t (111) = 2.14; $p < 0.05$, with Cohen's effect size $d = 0.40$. Also here, the score means of the group acquiring ESAP vocabulary in the face-to-face lead mode are higher than score means of the group working in the computer lead mode (M = 53.96; SD = 9.71 for group A and M = 50.10; SD = 10.60 for group B). The d value suggests a small effect size and indicates that the magnitude of the difference in the means was small, explaining 4% (eta squared = 0.039) of the variance on the subtest within the group.

Finally, the results obtained on the subtest *Artificial Language* show a trend towards significance t (111) = 1.81; $p = 0.073$, with Cohen's effect size $d = 0.34$. The score means of

the group with face-to-face lead mode are also higher than the score means of the computer lead mode group (M = 52.36; SD = 9.1 for group A and M = 49.24; SD = 9.21 for group B). The *d* value suggests a small effect size and indicates that the magnitude of the difference in the means was small, explaining 3% (eta squared = 0.028) of the variance on the subtest in the group. Table 4 summarizes the score means for the statistically significant results.

Tested variables with statistically significant results	**Group A**	**Group B**	**Eta squared**
post-treatment test	M = 13.09; SD = 4.66	M = 8.03; SD = 4.13	0.25
subtest *New Words*	M = 52.34; SD = 9.99	M = 46.28; SD = 8.84	0.09
subtest *Words in Sentences*	M = 53.96; SD = 9.71	M = 50.10; SD = 10.60	0.039
subtest *Artificial Language*	M = 52.36; SD = 9.1	M = 49.24; SD = 9.21	0.028

Table 4. Comparison of the score means for the statistically significant results

7. Discussion and conclusions

The study set out to investigate differences between two learner groups acquiring technical terminology in different study conditions. The research focused on active recall of ESAP vocabulary and was an attempt at exploring both the product- and process-related differences between the groups as reflected in the learners' posttest score means and the results obtained by them on the *Foreign Language Aptitude Test*.

The results obtained show a statistically significant difference in the area of ESAP vocabulary recall, with group A (face-to-face lead mode) scoring higher than group B (computer lead mode). They also indicate that the magnitude of the difference in the score means was large. Consequently, it can be concluded that the learning mode considerably affected the results of

the groups in the area of active recall of technical terminology. However, it was not the only factor accounting for higher scores of the learners studying in the face-to-face mode.

The findings of this study also point to other variables being responsible for the more successful performance of these learners on the post-treatment test. They relate to their foreign language aptitude in general, and more specifically to the capacity of their short-term memory, their ability to learn vocabulary both intentionally and incidentally as well as their ability to cope with linguistic input and make sense of it**.** Accordingly, it is possible to say that the results of the study point to the combined effect or interrelationship of the learning mode and a series of ID variables related to foreign language aptitude. The learners acquiring vocabulary in the face-to-face lead mode not only manifested specific abilities for vocabulary learning, but also worked in the instructional mode promoting more traditional didactic solutions. They worked under the teacher's supervision and received indispensable support when learning, which to some extent might have affected their post-test scores. This approach to language instruction is familiar to the majority of students and enables them to function more efficiently and successfully when learning. In contrast to them, the students whose contacts with the instructor were limited might have felt unsupported. They might as well have experienced an increased workload owing to a series of computer-based vocabulary consolidation activities they were required to complete on their own. Consequently, the learning mode might have to some extent influenced their post-test scores.

Research findings presented in this paper are to be treated as suggestive rather than definite for at least two reasons. The most important of them relates to the reliance on convenience, or opportunity sampling to select the target population meeting certain practical criteria pertinent to the objective of the study. The very strategy makes general relevance of the study findings less significant since the extent of generalizability of this type of sample is rather negligible (Dörnyei, 2011).

Another equally important reason is linked to the fact that the research design of this comparison study included only some individual variables and neglected others (such as strategies of vocabulary learning, vocabulary retrieval strategies, motivation, etc.), the exploration of which might have provided more insight into the processes of lexical retention and recall in two student groups. Neither has it incorporated personality-related variables that might have had an impact on the learners' preferences and their participation in a range of learning assignments. Owing to the above mentioned limitations of the study, one cannot hastily

generalize the research findings to the entire population or unambiguously conclude that traditional didactic procedures – in contrast to didactic solutions involving a higher percentage of online elements and/or tasks – produce better results when it comes to active recall of technical terminology.

It seems that more studies are needed to explore the impact of various blends of online and offline activities on vocabulary acquisition in ESAP. It appears advisable that these research designs rely on probability sampling, and not only opportunity sampling strategy. Also, these studies cannot ignore the specificity of learner IDs, since as shown in this paper different methods of teaching, or didactic procedures, are efficacious in relation to different people (also see Leith, 1974 in Eysenck and Eysenck, 2003). Thus it seems indispensable to complement further research into the issue with additional individual variables (e.g. strategies of lexical acquisition and retention, personality variables) that might considerably affect the process of lexical acquisition. In addition to learner-related variables, it is advisable that such studies incorporate instructional variables into their design (e.g. learners' perception of instruction and content quality, study workload, learning support) (cf. Lim and Morris, 2009). As yet, a first tentative attempt at explaining the combined impact of the learning mode and a series of ID variables related to foreign language aptitude has been made.

References

Baddeley, A. D. (1992). Working memory. *Science* 255, pp. 556–559.

Baddeley, A. D., Gathercole, S. E., Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* 105 pp. 158–173.

Basturkman, H. (2010). *Developing Courses in English for Specific Purposes*. New York: Palgrave Macmillan.

Blake, R.J. (2013). "Technologies for language learning". In M. Byram & A. Hu (eds.), *The Routledge Encyclopedia of Language Teaching and Learning* (pp. 712-715). London: Routledge.

Bloch, J. (2013). Technology and ESP. In B. Paltridge, S. Starfield (eds.) *The Handbook of English for Specific Purposes* (pp. 429-447). Chichester: John Wiley and Sons.

Bocanegra-Valle, A. (2010). Evaluating and designing materials for the ESP classroom. In M.F. Ruiz-Garrido, J.C. Palmer-Silveira, I. Fortanet-Gómez (eds.) *English for Professional and Academic Purposes*. Utrecht Studies in Language and Communication 22 (pp. 141-165). Amsterdam/New York: Rodopi.

Carmen, J. M. (2002). *Blended learning design: Five key ingredients*. http://www.agilantlearning.com/pdf/Blended%20Learning%20Design.pdf. [August 18, 2017]

Carroll, J.B. (1981). Twenty-five years of research on foreign language aptitude. In K.C. Diller (ed.) *Individual Differences and Universals in Language Learning Aptitude* (pp. 83-118). Rowley,Mass.: Newbury House.

Carroll, J.B., Sapon, S. (1959). *The Modern Language Aptitude Test*. San Antonio, TX: Psychological Corporation.

Chung, T.M., Nation, P. (2004). "Identifying technical vocabulary". System 32, 251-263. https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/2004-Chung-Identifying - Tech-Voc.pdf. [June 29, 2017].

Coxhead, A. (2013). Vocabulary and ESP. In B. Paltridge, S. Starfield *The Handbook of English for Specific Purposes* (pp. 141-158). Chichester: Wiley-Blackwell.

Dörnyei, Z. (2011). *Research Methods in Applied Linguistics.* Oxford: OUP.

Dörnyei, Z., Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty, M. Long (eds.) *The handbook of second language acquisition* (pp. 559-630). Oxford: Blackwell.

Driscoll, M. (2002). *Web-based training: Creating e-learning experiences.* San Francisco: Jossey-Bass.

Dudley-Evans, T., St John, M.J. (2009). *Developments in English for Specific Purposes: A Multi-disciplinary Approach.* Cambridge: CUP.

Ellis, R. (2004). Individual differences in second language learning. In A. Davies, C. Elder (eds.) *The Handbook of Applied Linguistics* (pp. 525-551). Oxford: Blackwell Publishing.

Eysenck H. & Eysenck M. (2003). *Mind Watching. Why We Behave the Way We Do?* Gdańsk: Gdańskie Wydawnictwo Psychologiczne. [In Polish]

Garrison, D. R., Vaughan, N. D. (2008). *Blended Learning in Higher Education: Framework, Principles and Guidelines*. San Fransisco: John Wiley and Sons.

Grigorenko, E., Sternberg, R., Ehrman, M. E. (2000). A theory based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *Modern Language Journal* 84 pp. 390–405.

Grosse C.U., Voght G.M. (1991). "The evolution of languages for specific purposes in the United States". *Modern Language Journal* 75, 181-195.

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity". In P. Robinson (ed.). *Cognition and Second Language Instruction* (pp. 258-287). Cambridge: Cambridge University Press.

Hutchinson, T., Waters A. (2010). *English for Specific Purposes: A Learning-centred Approach*. Cambridge: CUP.

Jung, R., Lee, C.H. (2013). Using Internet video clips for university students' vocabulary development in blended learning. *Multimedia-Assisted Language Learning* 16 pp. 67-96.

Komorowska, H. (2005). *Methodology of Foreign Language Teaching.* Warszawa: Fraszka Edukacyjna. [In Polish].

Krajka, J. (2007). *English Language Teaching in the Internet-Assisted Environment. Issues in the Use of the Web as a Teaching Medium.* Lublin: Maria Curie-Skłodowska University Press.

Kuteeva, M. (2011). Wikis and academic writing: Changing the writer–reader relationship. *English for Specific Purposes* 30 pp. 44-57.

Laufer, B. (1997). "What's in a word that makes it hard or easy? Some intralexical factors that affect the learning of words". In N. Schmidtt, M. McCarthy (eds.) *Vocabulary: Description, Acquisition and Pedagogy* (pp. 140-155). Cambridge: CUP.

Laufer, B., Golstein, Z. (2004). Testing vocabulary knowledge: Size, strength and computer adaptiveness. *Language Learning* 54 pp. 399-436.

Leith, G.O. (1974). Individual differences in learning: Interaction of personality and teaching methods. In *Personality and Academic Progress, Conference Proceedings* (pp. 14-25). London: Association of Educational Psychologists.

Lewis, M. (1993). *The Lexical Approach: The State of EFL and a Way Forward*. Language Hove: Teaching Publications.

Lim, D. H., Morris, M. L. (2009). Learner and instructional factors influencing learning outcomes within a blended learning. *Educational Technology and Society 12*, pp. 282–293.

MacDonald, P., García-Carbonell, A., Carot-Sierra, J. M. (2013). Computer learner corpora: Analysing interlanguage errors in synchronous and asynchronous communication. *Language Learning and Technology* 17 pp. 36-56.

Maciaszczyk, S. (2011). The process of e-learners' learning. In H. Komorowska (ed.) *Learning a Foreign Language – the Learner's Perspective* (pp. 192-207). Warszawa: Oficyna Wydawnicza Łośgraf.

Masoura, E. V., Gathercole, S. E. (1999). Phonological short term memory and foreign language learning. *International Journal of Psychology* 34 pp. 383–388.

Master, P., Brinton, D.M. (eds.) (1998). *New Ways in English for Specific Purposes*. Alexandria, VA: TESOL.

Meara, P. (1990). A note on passive vocabulary. *Second Language Research* 6 pp. 150-154.

Murray, L., Hourigan, T., Jeanneau, C. (2007). Blog writing integration for academic language learning purposes: Towards an assessment framework. *Ibérica* 14, pp. 9-32.

Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nation, P. (2008). *Teaching Vocabulary: Strategies and Techniques*. Boston: Heinle Cengage.

Nunan, D. (2004) *Research Methods in Language Learning.* Cambridge: CUP.

Ono Y., Ishihara M. (2012) Integrating mobile-based individual activities into the Japanese EFL classroom. *International Journal of mobile learning and Organisation* 6 pp. 116-137.

Oxford, R.L. (1999). Second language learning: Individual differences. In B. Spolsky *Concise Encyclopedia of Educational Linguistics* (pp. 552-560). Oxford: Elsevier Science Ltd.

Parry, T. S., Child, J. R. (1990). Preliminary investigation of the relationship between VORD, MLAT and language proficiency. In T. S. Parry, Ch. Stansfield (eds.) *Language Aptitude Reconsidered* (pp. 30-66). Englewood Cliffs, NJ: Prentice-Hall Regents.

Pimsleur, P. (1966). *The Pimsleur Language Aptitude Battery*. New York: Harcourt, Brace, Jovanovic.

Rees, J. (2000). "Predicting the future of foreign language aptitude". In S. Cornwell & P. Robinson (eds.) *Individual Differences in Foreign Language Learning: Effects of Aptitude, Intelligence and Motivation* (pp. 187-197). Tokyo: Aoyama Gakuin University.

Robinson P. (1995). Attention, memory and the "noticing" hypothesis. *Language Learning* 45 pp. 283-331.

Robinson, P. (2002). Effects of individual differences in intelligence, aptitude and working memory on incidental second language learning: A replication and extension of Reber, Walkenfield and Hernstadt (1991). In P. Robinson (ed.), *Individual Differences and Second Language Instruction* (pp. 211–266). Philadelphia: Benjamins.

Rysiewicz, J. (2012). *Foreign Language Aptitude Test* (*FLAT*). *Overall Description, Analysis, Statistics and Test Procedures*. Online at http://www.academia.edu/1743794/Test_Uzdolnien_do_Nauki_Jezykow_Obcych_TUNJO_._Charakterystyka_ogolna_opis_analiza_statystyki_i_procedury_testowe. [In Polish] [April 29, 2017].

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence*. New York: Peter Lang.

Skehan, P. (1986). Where does language aptitude come from? In P. Meara (ed.) *Spoken language* (pp. 95-113). London: Center for Information on Language Teaching.

Skehan, P. (1989). *Individual Differences in Second Language Learning*. London: Edward Arnold.

Skehan, P. (1999, March). *Aptitude*. Paper presented at the conference on Individual differences in foreign language learning: Effects of aptitude, intelligence and motivation, Tokyo, Japan.

Sparks R., Ganschow, L. (1991). Foreign language learning differences: Affective or native language aptitude differences? *Modern Language Journal* 75 pp. 3-16.

Sparks, R. (1995). Examining the Linguistic Coding Differences Hypothesis to explain individual differences in foreign language learning. *Annals of Dyslexia* 45 pp. 187-214.

Sparks, R., Ganschow, L. (2001). Aptitude for learning a foreign language. *Annual Review of Applied Linguistics* 21 pp. 90-111.

Sparks, R., Ganschow, L., Patton, J., Artzer, M., Siebenhar, D., Plageman, M. (1997). Prediction of foreign language proficiency. *Journal of Educational Psychology 89*, pp. 549-561.

Spolsky, B. (1995). Prognostication and language aptitude testing, 1925-1962. *Language Testing* 12 pp. 321-340.

Swales, J.M. (1988). *Episodes in ESP: A Source and Reference Book on the Development of English for Science and Technology*. New York: Prentice Hall.

Tomlinson, B., Whittaker, C. (eds.) (2013). *Blended Learning in English Language Teaching: Course Design and Implementation*. London: British Council.

Vasbieva, D.G., Klimova, I.I., Agibalova, E.L., Karzhanova, N.V., Bírová, J. (2016). Enhancement of students' vocabulary learning through a blended learning approach. *IEJME –Mathematics Education* 11 pp. 1195-1203.

Warschauer, M. (2002). Networking into academic discourse. *Journal of English for Academic Purposes* pp. 45-58.

Wen, Z., Skehan, P. (2011). A new perspective on foreign language aptitude research: Building and supporting a case for "working memory as language aptitude." *Ilha do Desterro: A Journal of English Language, Literatures and Cultural Studies* 60 pp. 15–44.

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. C. Diller (ed.) *Individual Differences and Universals in Language Learning Aptitude* (pp. 119-154). Rowley, MA: Newbury House.

Whittaker, C. (2013). "A military blend". In B. Tomlinson, C. Whittaker (eds.) *Blended Learning in English Language Teaching: Course Design and Implementation* (pp. 175-184). London: British Council.

Woodward-Kron, R. (2008). More than just jargon – the nature and roles of specialist knowledge in learning disciplinary knowledge. *Journal of English for Academic Purposes* 7 pp. 234-49.

Appendix 1

I. Supply correct words to complete the sentences below.

1. The **_ _m_** _ _ _ _ _ _ _ _ is not a political union, but an intergovernmental organization of fifty-four independent member states that were once part of the British Empire.
2. New **l** _ _ _ _ _ _ _ _ _ measures were taken to stop the flow of drugs into the country.
3. The country has entered a new era of democracy with the inaugural session of its democratically elected **a** _ _ _ _ _ _ .
4. Since Bolivia has vast reserves of lithium some wonder if the country can become a green energy _ _ _ _ _ _ _ _ **r** ?
5. The **J**_ _ _ _ _ _ _ interprets laws and decides on cases arising out of the laws.
6. Ministers invited by the Prime Minister to attend regular meeting to discuss policy are known as the _ _ _ _ _ **_t** .
7. Governments make laws and the police _ _ **f**_ _ _ _ them.
8. Last year the Supreme Court finally **_v**_ _ _ _ _ _ _ _ the decision.
9. A law-making body which consists of two parts is referred to as **_i**_ _ _ _ _ _ _ .
10. A former Russian spy is convicted of high **t**_ _ _ _ _ _ for helping the US uncover Anna Chapman.
11. In some countries, for example in France, a member of the lower house of parliament is referred to as _ _ _ _ **_y** .
12. This **_ _n**_ _ _ _ _ _ _ _ _ needs a Labour MP once again, after losing the last by-election to Tory Edward Timpson.

13. The Prime Minister is warning the **_l_ _ _ _ _ _ _ _** that the alternative voting system is unfair.

14. Protection for the consumer is guaranteed by **s_ _ _ _ _ _** .

15. The **_x_ _ _ _ _ _ _** is responsible for making sure that the new laws and other decisions are done in the way they have been planned.

16. The members of the House of **_ _ _ _ _ _s** are elected by citizens.

17. Under UK electoral laws, the doors of all **_ _ _ _ _ _g** stations must close at 10pm and no one can be issued with a ballot **_ _ _ _r** after that time.

18. It was the lowest by-election **_ _ _ _ _ _t** for 11 years, with only 28.8 per cent bothering to take part.

19. These two countries signed a peace **_r_ _ _ _** last year.