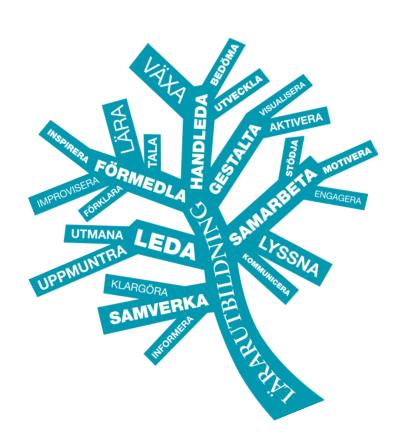# The Effectiveness of Word-Focused Tasks.

How Elaborate Processing Predicts Vocabulary Learning – A Literature Review.

Jane Ekenberg &

Mirjeta Makolli

Ämneslärarprogrammet

Abstract

L2 vocabulary learning is a sizeable challenge for all EFL students. Particularly, the lexical threshold for reading is a substantial learning feat requiring time and effort. Vocabulary research indicates that incidental learning alone is not sufficient for learners to achieve a functional EFL vocabulary. Despite this fact, the Swedish syllabus provides little guidance on the topic and Swedish EFL teachers show little awareness of what teaching methods could complement incidental vocabulary learning such as word-focused tasks. This literature review aims at examining what recent EFL research has found to be important features for making word-focused tasks effective and to discuss resulting pedagogical implications. Research into this field focuses on the predictive abilities of two theoretical models, the involvement load hypothesis and the technique feature analysis. Although empirical studies support the effectiveness of word-focused tasks on vocabulary learning, neither model is found to be superior in its predictive ability. Nevertheless, research shows that task features which incorporate different types of elaborate processing and direct learners' attention to the form-meaning connection are effective.

Table of Contents

1 Introduction

Learners, teachers, and researchers in applied linguistics agree that vocabulary is situated at the very core of language learning. This holds true both for L1 and any subsequent languages learners engage with. However, the way in which words are learnt differ between L1 and L2 in that L1 learners successfully learn large numbers of words by constant exposure and contextual clues (e.g. Laufer, 2020, p. 351; Sonbul & Schmitt, 2010, p. 253). In the case of L2 learners, the necessary level of exposure is rarely available, thus other methods of vocabulary learning are required to facilitate L2 vocabulary development.

The unintentional way of learning words as a by-product whilst focusing on a listening or reading task is referred to in language research as incidental vocabulary learning. The efficiency of incidental L2 vocabulary learning and whether it would lead to sufficient vocabulary gain is a point of contention within the research community. According to Rassaei (2017, p. 77), some researchers argue that extensive reading is the best source for learning L2 words, others hold that retention of vocabulary learnt this way is not of lasting effect, while yet others claim that retention is also weak for explicitly taught vocabulary.

Although the discussion on the effectiveness of different methods to learn vocabulary is still on-going, researchers agree that the challenge of acquiring a functional English L2 vocabulary is a formidable one (e.g. Nation, 2006; Schmitt, 2008). Several researchers have attempted to calculate the lexical threshold at which learners of English could be said to achieve reading comprehension and fluency (Newton, 2020, p. 255). Nation's (2006, p. 22) calculation has achieved wide-spread acceptance within the research community suggesting that knowing 8000-9000 word families would allow learners to read and comprehend a wide range of texts. Accepting Nation's proposed lexical threshold, Schmitt (2008, p. 333) argues that the sheer size of the vocabulary challenge warrants a well-considered and structured approach to vocabulary learning in the EFL classroom.

Research indicates that Swedish upper secondary students may not have acquired the necessary vocabulary for pursuing future studies upon graduation (Lindqvist & Dragemark Oscarson, 2019), despite the fact that preparing students for university studies is a goal for Swedish upper secondary school. In addition, research found a possible knowledge gap in Swedish EFL teachers regarding teaching methods enhancing incidental vocabulary learning (Bergström et al., 2021). The Swedish syllabus for English at upper secondary school (Skolverket, 2020) provides no explicit direction for vocabulary learning and teaching (VLT). Hence, these findings indicate that the Swedish educational system might not have the

structured and well-considered approach to vocabulary learning which Schmitt (2008, p. 333) calls for.

Despite VLT being an active research field over a long period of time, researchers still call for further investigation into effective methods for maximising vocabulary learning episodes in the EFL classroom (e.g. Gohar et al., 2018, p. 867; Laufer, 2020, p. 365). As a reaction to the development of new frameworks to evaluate the effectiveness of word-focused tasks, several empirical studies were published over the last decade to test these. In a recently published systematic review article, Yanagisawa and Webb (2021) present a meta-analysis of the predictive ability of the *involvement load hypothesis* on vocabulary retention. In our literature review, we propose to expand on their study by reviewing recent research measuring the effectiveness of word-focused tasks for the EFL classroom based not only on the involvement load hypothesis, but also on the newer *technique feature analysis*.

In the next sections, we outline how the syllabus and language research impact VLT leading up to the aim of this literature review and the method used to select empirical studies. Subsequently, the current theoretical base for evaluating the effectiveness of word-focused tasks is described followed by an overview of the empirical evidence concerning the predictive ability of the presented theories on effective vocabulary learning. Then, we discuss the findings of the empirical studies as well as the resulting pedagogical implications. Finally, we present our conclusions and propose directions for future research.

1.1 Background

Teaching methods for vocabulary development, as any other pedagogical activity within the classroom, need to be aligned with the syllabus. Furthermore, the Swedish National Curriculum for upper secondary school requires all teaching to be based on scientific research (Skolverket, 2011, p. 1). Correspondingly, this section highlights the role of vocabulary in the syllabus and in research on foreign language teaching.

1.1.1 The Role of Vocabulary in the Syllabus

Traditionally, vocabulary learning played an important part in language teaching in schools. The grammar translation model included a focus on learning lists of vocabulary, whereas the subsequent audiolingual method focused on correct pronunciation (Brown & Lee, 2015, p. 18, 22). The more recent communicative approach strongly influenced the creation of the Common European Framework of Reference for Languages (CEFR) and through that also

EFL syllabuses in several countries (Brown & Lee, 2015, p. 28). The communicative approach incorporates a functional view on languages by breaking down learning into the four skills: reading, listening, speaking, and writing. As a result, CEFR treats vocabulary as an integral part of each skill describing the type of vocabulary needed to achieve each of its six proficiency levels (see Appendix). In addition, the CEFR companion volume (Council of Europe, 2020) explicitly addresses vocabulary range for reception and production as well as vocabulary control, which centres on familiarity of topic for the lower proficiency levels and appropriate and idiomatic use for the higher levels.

Even though CEFR's primary focus is describing the knowledge requirements by proficiency level, a note on the method of learning vocabulary indicates that reading is the most common method (Council of Europe, 2020, p. 131). As mentioned earlier, this is a continuing point of debate among researchers in the field. Despite being influenced by CEFR (Tornberg, 2015, p. 209), the Swedish syllabus for English in upper secondary education mentions 'vocabulary' and 'word' only in general terms in the core content pertaining to receptive and productive skills but offers no directives concerning what type of vocabulary knowledge is implied (Skolverket, 2020). However, the syllabus calls for students to read and write texts of different kinds and for different purposes (Skolverket, 2020, p. 2) connecting it to the formulation in the CEFR companion volume indicating that the higher proficiency levels are characterised by both breadth and depth in vocabulary usage (Council of Europe, 2020, p. 131, 134).

The Swedish syllabus provides no guidance to teachers on how to teach English vocabulary. Furthermore, a qualitative study of Swedish teachers' views on vocabulary learning indicates that they subscribe to the idea of words being learnt as a by-product, but they are not able to name or describe methods by which their teaching could support incidental vocabulary learning (Bergström et al., 2021). As the study only included a small number of interviewees (n=14), generalisations cannot be made. However, their findings indicate a possible knowledge gap concerning word-focused tasks among Swedish EFL teachers meriting further investigation. Moreover, Lindqvist and Dragemark Oscarson's study (2019) shows that a large number of Swedish students do not achieve the CEFR B2 level (see Appendix) for reading after having completed the corresponding level English 6 course (the final mandatory English course in Swedish upper secondary education for students aged 17-19). The authors conclude that the students do not master the required vocabulary for this

level, thus making university studies difficult for these students as that would require them to read and comprehend English course literature and international research.

1.1.2 The Role of Vocabulary in the EFL Classroom

As described in the previous section, the role of vocabulary teaching has changed over time as new teaching approaches have become fashionable. However, vocabulary learning has always remained an integral part of the EFL classroom. Even so, there is room within the communicative approach to integrate vocabulary teaching contextualised in communicative tasks (Brown & Lee, 2015, p. 480).

In his four strands-model, Nation (2007) elaborates on the need to treat VLT as an integral part of language teaching in order for it to be effective. The four strands of the model are: meaning-focused input, meaning-focused output, fluency, and language-focused learning. The first three strands encompass activities in which the learner focuses on meaning and understanding, i.e. context, rather than form. Together the strands encompass both receptive and productive skills as well as incidental and intentional learning. At the core of the model lies the idea that the four strands are complementary and learning opportunities are maximised when combining two or more strands.

In order to clarify the role of VLT, Nation lists the four jobs of the vocabulary teacher (Newton, 2020, p. 256). The vocabulary teacher should: 1. decide what words to teach, 2. help learners to develop effective learning strategies, 3. assess the progress of vocabulary development, and 4. integrate vocabulary learning opportunities into the EFL classroom. The list is constructed with the intent to guide teachers by assigning each job a rank based on the relative amount of time it is meant to be given in VLT.

The apparent issue lies in the divide between the amount of vocabulary required to achieve the set proficiency levels and the time allotted to EFL instruction in school. Furthermore, the many facets of vocabulary knowledge constitute an additional complexity requiring consideration (see section 2.1). This is where language research can make an impact by presenting tools by which teachers can evaluate the effectiveness of word-focused tasks allowing their prioritisation to rest on research.

1.2 Aim and Research Question

This literature review focuses on Nation's fourth job of the vocabulary teacher, namely to provide opportunities for practising vocabulary in the classroom (Newton, 2020, p. 256).

More specifically, it centres on the effectiveness of vocabulary tasks. As mentioned above, the four strands model stipulates that vocabulary learning opportunities are present across all strands and that more than one strand may be integrated in a task to maximise learning (Nation, 2007). This is factored into this review by including three of the strands: meaning-focused input, meaning-focused output, and language-focused learning. Although there is no doubting the decisive impact of learners' motivation and other learner-centred factors on the outcome of vocabulary development (Schmitt, 2008, p. 333), these factors fall outside the scope of this literature review.

Knowing what makes vocabulary tasks effective is a topic of interest in language research, which also benefits practitioners when addressing the sizeable EFL vocabulary challenge and deciding between teaching priorities. Consequently, we formulated the following research question: what do recent empirical studies on EFL indicate as being key features of effective word-focused tasks? To this end, we analysed empirical studies testing the predictive ability of the two elaborate processing models, namely the involvement load hypothesis and the technique feature analysis, and discussed the implications of their findings for VLT.

1.3 Method

As a starting point for this literature review, we aimed to find studies published from 2015 onwards. The reasoning behind this start date is two-fold. First, the research field is an active one with a high publication rate providing a wealth of recent studies for inclusion. Second, we have come across a few literature reviews already covering earlier periods of publication (Huang et al., 2012; Yanagisawa & Webb, 2021; Zou & Yan, 2019), thus our literature review focuses on studies from 2015 to 2021.

Empirical studies assessing the predictive ability of elaborate processing models were collected using two pedagogical databases (ERIC and Education Collection) as well as Gothenburg university library's search engine Supersök. The search words used were combinations of 'vocabulary development', 'vocabulary gain', 'English as a second language' (the databases included EFL under this search word), 'reading', 'involvement load', 'task*', and 'task effectiveness'. In addition, reference lists of pertinent studies were used to identify other studies relevant for inclusion.

For studies to be included in this literature review the following criteria had to be met: 1. English as a second or foreign language was the target language, 2. output tasks such as

cloze exercises, composition, sentence writing, or discussions were used as treatments, 3. at least one of the two elaborate processing models were used for predicting task effectiveness, 4. pre- and post-tests of vocabulary retention were included either concerning recognition and/or recall, and 5. reading was used as input medium for vocabulary exercises. Studies were excluded if their focus lay outside of the assessment of task effectiveness, for example studies centring on learner differences, learning strategies, or other media of input than reading.

All 15 studies included in this literature review are peer reviewed. Two of the studies were published in journals not listed by Register over vitenskapelige publiseringskanaler (Direktoratet for høgare utdanning og kompetanse, n.d.), namely *TESOL International Journal* and *Theory and Practice in Language Studies.* However, these articles are marked as peer reviewed either in Supersök or in ERIC, thus they were deemed to be acceptable for inclusion.

2 Theoretical Framework

Since 2001, two theoretical frameworks have been developed for assessing the effectiveness of word-focused tasks. This has enabled L2 vocabulary researchers to have clearly defined vantage points from which to explore and discuss factors impacting successful vocabulary learning. In this literature review, the terms learning and acquisition are used interchangeably. Moreover, the terms vocabulary and word are used to refer to L2 vocabulary unless otherwise stated. This chapter outlines how the frameworks have developed through expanding on existing theories from a shared understanding of vocabulary knowledge.

2.1 Vocabulary Knowledge

In previous sections, we discussed the size of the vocabulary challenge for EFL learners. There is an added complexity in the many aspects involved in knowing each word. Nation (2013, p. 49) describes this complexity by using three main categories (form, meaning, and use) with three sub-categories for each. Knowing the form of a word entails knowing the written and/or spoken form and identifying parts of the word. To know the meaning of a word, it is necessary to connect form and meaning, make associations, and to understand the concept behind the word. To be able to use a word, grammar, collocations, and socio-cultural restrictions of the word have to be known. Therefore, knowing a word is not a binary phenomenon, but words can be unknown, partially known or entirely known. Moreover, the meaning of a word can be recognised upon encountering the form (receptive knowledge), whilst only at a later learning stage be available to the learner in language production (productive knowledge).

Because the aspects of knowing a word differ from one another, EFL teachers require an array of tools to meet the different didactic needs. As in his four strands model, Nation (2013, p. 60) proposes combining intentional and incidental learning to achieve effectiveness depending on the aspect taught. For establishing meaning, he suggests intentional learning through depth of processing using such activities as elaboration and inferring meaning which may be included in word-focused tasks. In the following sections, the two theoretical frameworks for evaluating such activities are explained starting with their theoretical origins.

2.2 The Origins of the Theoretical Frameworks

As discussed above, vocabulary knowledge is a complex matter and research suggests that elaborating by connecting the different features of this knowledge whilst also connecting it to previous knowledge promotes vocabulary retention (Laufer & Hulstijn, 2001, p. 1). In their formative study, Craik and Lockhart (1972) proposed the cognitive element of depth of processing as decisive in determining whether a new word is retained. Depth of processing implies that using a word and exploring its different meanings through elaboration, thus processing it at a deeper level, is beneficial for learning (Yanagisawa & Webb, 2021, p. 488). Researchers agree that there is a positive impact on retention when learners apply more rather than less elaborate processing for the learning of new words (Laufer & Hulstijn, 2001, p. 6). Another point of agreement in L2 learning is the strong influence of learner motivation (e.g. Schmitt, 2008, p. 333), particularly the effect of intrinsic motivation on retention.

An issue with the depth of processing concept was that it did not include precise enough criteria to be operationalised for empirical study. This, in combination with slow progress in establishing new theoretical frameworks for vocabulary acquisition, spurred Laufer and Hulstijn (2001) to develop a model for operationalising elaborate processing of tasks. They expanded on the depth of processing hypothesis with its existing cognitive component by adding a motivational aspect.

2.3 Involvement Load Hypothesis

Laufer and Hulstijn (2001) proposed the involvement load hypothesis (ILH) as a way to assess the level of processing in L2 vocabulary learning. It consists of one motivational component (*need*), and two cognitive components (*search* and *evaluation*). Need relates to the source of the drive to complete a task. It is said to be moderate when the teacher sets the task, whereas the need in tasks set by the learners themselves is considered strong. Search refers to finding the meaning of an unknown word or having the meaning and searching for the form. The first instance, in which the meaning is given, results in moderate weighting. When the learner has to produce the form based on the concept of the word, the weighting given is strong. Distinguishing between moderate and strong search is a recent modification to ILH (Laufer, 2020, p. 353). Evaluation, the last component of ILH, requires learners to assess how words fit into a context. If the context is given, for example in a gap exercise, evaluation is set at moderate. If learners are to create their own texts including unknown words, evaluation is set at strong.

Involvement is the construct Laufer and Hulstijn (2001, p. 14) use to describe the components implicating successful retention of unknown L2 words. To predict the effectiveness of different word-focused tasks, the components in ILH are weighted (absent = 0, moderate = 1, or strong = 2) and summed up to give each task an involvement load. Thus, their hypothesis is that tasks with higher involvement load result in higher vocabulary retention than tasks with lower involvement load. For instance, a homework assignment where the learner is given a list of words to include in a composition would be assigned the involvement load: moderate need (teacher induced task), moderate search (form is given), and strong evaluation (original production). However, if it were an authentic communicative task, the involvement load would be: strong need (learner induced task), strong search (provided unknown words are included), and strong evaluation (original production). According to ILH, the latter task should result in higher vocabulary retention. It is important to note that ILH does not discriminate between tasks based on input or output orientation nor is it concerned with the mode of the task (oral or written).

2.4 Technique Feature Analysis

Nation and Webb (2011, p. 6-7) identified two potential improvements on ILH regarding the reliability of using it as an evaluation tool and expanding it to also encompass features for designing word-focused tasks. They proposed a model called the technique feature analysis (TFA) which, unlike ILH, allows for the learning of partially known words. TFA is a comprehensive model consisting of five main criteria. The first criterion is motivation, which assesses how motivated learners are to complete the task. The second criterion is noticing, which entails focusing learners' attention on target words, raising awareness of words to be learnt, and the presence of negotiation of meaning. The third criterion is retrieval, outlining if learners are asked to recognise or recall lexical items, the number of retrievals as well as any spacing between them. The fourth criterion is generation, in which the learner encounters and/or uses partially known words in previously unknown contexts. The creation of new contexts for words is known as productive generation. The fifth criterion is retention which concerns the learner's ability to form a correct link between form and meaning. Moreover, this last criterion includes visual support, presenting words in meaningful context, and not using lexical sets due to the risk of confusing the learner. The five main criteria are further divided into subcriteria resulting in 18 questions of which each can receive one point, thus making the maximum score in TFA 18 (for further details see Nation & Webb, 2011).

There is a weak point in TFA, namely the impact of the difficulty in assessing what learners already know (Nation & Webb, 2011, p. 15). Furthermore, the many layers of vocabulary knowledge (see section 2.1) do not contribute equally to knowing a word. The different weighting by which they contribute to vocabulary knowledge is not reflected in the TFA criteria weighting. Nevertheless, TFA presents the possibility to evaluate the effectiveness of word-focused tasks in three different stages: the design phase, during use, and in the evaluation phase (Nation & Webb, 2011, p. 15).

ILH is more widely empirically tested than TFA. However, TFA has been tested in a few empirical studies. In the next chapter, we outline the empirical evidence concerning the predictive ability of the two theoretical frameworks.

3 Findings

In this section, recent empirical studies concerning the predictive abilities of the two theoretical frameworks on elaborate processing tasks for vocabulary development are examined. First, we reviewed studies using ILH outlining the level of support and any related issues. Then, studies concerning TFA were explored to discern its predictive ability as well as assessing the components of the model.

3.1 Involvement Load Hypothesis

Despite ILH being a 20-year-old theory, it continues to attract researchers' attention and a number of studies evaluating its predictive abilities are published each year. The reason for the continued interest in the model may stem from the fact that results remain inconclusive (Gohar et al., 2018, p. 860), thus there is not yet an empirically tested solution to master the EFL vocabulary acquisition challenge. Recent research encompasses voices of support as well as draws attention to components of the model in need of further development.

3.1.1 Empirical Support for Task-Induced Involvement Load

ILH states that higher task-induced involvement load (IL) leads to more effective vocabulary learning than tasks with lower IL (Laufer & Hulstijn, 2001, p. 15). This assumption has been tested in research comparing the outcome of tasks with different IL on immediate uptake as well as on retention. Feng (2015) studied the effects of three tasks with different IL: sentence writing (need 1, search 0, evaluation 2 => IL 3), gap fill (need 1, search 0, evaluation 1 => IL 2), and translation (need 1, search 0, evaluation 0 => IL 1) on a group of low proficiency students. Both the immediate post-test and the delayed post-test confirmed ILH predictions that the effectiveness of tasks on word acquisition was determined by their respective IL. Feng's results (2015) concur with Soleimani and Rahmanian's study (2015) on advanced proficiency learners. However, neither of these studies incorporated a control group, thus not providing a baseline from which the effectiveness of the studied tasks could be evaluated. Alavinia and Rahimi (2019) amended this deficiency in their study of different IL tasks on low proficiency learners and their results were in line with the two previous studies.

A point of dissent in the support for ILH concerns the retention of vocabulary as measured in delayed post-tests. In a study from 2017, Yang et al. found that sentence writing (IL 3) outperformed gap fill (IL 2) and both tasks outperformed the control group for

immediate uptake. However, the results on the delayed post-test did not produce the same ranking, but showed that the effect of the gap fill activity was equal to that of the sentence writing task despite the difference in IL. This deviation from ILH for long-term retention was also found in Douglas' study (2016). However, the sample size in this study was very small (n=7) limiting any generalisation and the research design allowed different time spans between treatments and the delayed post-test which might have skewed the outcome.

3.1.2 Issues With Task-Induced Involvement Load as Highlighted in Research

In addition to the impact of the total IL, ILH also claims uptake of incidentally processed L2 words is contingent upon the weight of the components that constitute IL: need, search, and evaluation (Laufer & Hulstijn, 2001, p. 14). In order to test this assumption, researchers have chosen a research design comparing the effectiveness of tasks with the same IL index, but different configuration of components. The present literature review only found partial support for this assumption as will be outlined below.

Bao (2019) compared the effectiveness of four tasks (choice, combining, definition, and matching) all with IL 2 (need 1, evaluation 1). In line with ILH prediction, three tasks performed equally well in promoting word learning. However, the fourth task, the definition task, outperformed the other tasks, thus indicating that there are factors outside IL influencing vocabulary acquisition. Bao (2019, p. 9) suggests that the definition task required learners to return to the form-meaning connection more often than in the other tasks as they were evaluating the different options. Moreover, the mental effort in this task appears to be higher as meaning had to be inferred from the reading passage and then used in producing the correct form for each given definition, thus reinforcing the form-meaning connection further (Bao, 2019, p. 9). This outcome is indicative of differential processing within the task and concurs with the results of Bao (2015) in which a definition task outperformed a task with higher IL. It should be noted that neither of Bao's studies (2015; 2019) used a delayed post-test, thus his results only pertain to immediate uptake. However, similar results from Yang et al. (2017) indicate that repeated exposure to the form-meaning connection can also lead to more effective long-term retention of L2 vocabulary.

There seems to be an impact of differential processing within evaluation of ILH on vocabulary learning which is not reflected in the current IL index system. Zou (2017) investigated the effectiveness of sentence writing and composition with equal IL (need 1, evaluation 2). Although ILH predicts the outcome to be the same for the two tasks, the

composition task achieved better immediate uptake and better long-term retention than sentence writing. Zou (2017, p. 64-65) argues that there is a difference between the tasks in that the composition task entails a higher complexity requiring more mental effort due to the hierarchical organising of each sentence into a whole. This difference in differential processing is not reflected in the equal evaluation awarded in IL, as the difference in degrees of the evaluation component solely depends on the context being original. Zou (2017, p. 71) suggests an expansion of the scale in IL to include a third degree of evaluation to reflect the complexity and length of writing tasks.

Taheri and Rezaie Golandouz (2021) reached a similar conclusion to Zou (2017) regarding the impact of complexity in writing tasks. They analysed the effectiveness of three equal IL tasks (need 1, evaluation 2) and yet one outperformed the others on immediate uptake. The superior task entailed a large element of creative writing which the authors could not account for in the IL given, thus they too suggested that the current IL component of evaluation is not allowing necessary degrees of complexity to be accounted for. Moreover, Yang and Cao's study (2021, p. 427) on equal IL tasks indicates that evaluation, specifically strong evaluation, may be a more influential component than both need and search, because of the mental effort involved in creating text and content.

Judging from the empirical studies reviewed above, it seems likely that differential processing could provide an additional aspect to IL promoting L2 vocabulary acquisition through word-focused tasks. Notwithstanding these findings, there is also evidence that not all added complexity in tasks is positive for word learning. Research on the effectiveness of the combining task has shown it to be less effective than other tasks with the same IL (need 1, evaluation 1) (Bao, 2015; Bao, 2019; Tahmasbi & Farvardin, 2017). Arguments have been made that the combining task adds a syntactic complexity which diverts learners' attention away from the unknown words rendering it less effective by making word processing a secondary activity (e.g. Bao, 2019). Consequently, it is essential that any complexity in word-focused tasks facilitates the learner's processing of unknown words and does not divert away from unknown words.

3.2 Technique Feature Analysis

TFA was initially created with the intent of more accurately predicting effectiveness of word-focused tasks in order to address a lack of nuances in ILH (see section 2.4). Unsurprisingly, research designs have been developed to compare the predictive ability of the two models

(Gohar et al., 2018; Hu & Nassaji, 2016). Nevertheless, there are also a few studies researching the robustness of the TFA framework for different types of word-focused tasks (Huang et al., 2020; Kamali et al., 2020).

3.2.1 The Predictive Ability of TFA

A common line of critique concerns the inability of ILH to distinguish between tasks such as sentence writing and longer writing assignments. Research has indicated a need to account for differential processing within the evaluation component of ILH (Taheri & Rezaie Golandouz, 2021; Zou, 2017). As a consequence, Gohar et al. (2018) designed their study to assess if the more comprehensive TFA model could better predict differences in vocabulary learning of a sentence writing task and a composition task, a comparison that ILH has not been able to accurately predict (see section 3.1.2). The two tasks were given different weights by the two models. Sentence writing reached an IL of 3 (need 1, search 0, evaluation 2) and a TFA score of 7 (motivation 2, noticing 2, retrieval 0, generation 2, retention 1). As expected, the composition task had an identical IL to the sentence task, but the TFA score was higher at 9 (motivation 2, noticing 2, retrieval 0, generation 3, retention 2). Post-test results demonstrated that composition writing was more effective than sentence writing for word learning and both experimental groups outperformed the control group. Thus, the results are in line with the TFA scores but not reflected in IL. However, since the results did not show a statistically significant difference, the study only partially supports the advantage of TFA over ILH.

In an attempt to explain the superiority of the composition task over sentence writing, Gohar et al. (2018, p. 867) built an argument that composing a longer text required more mental effort as there are more connections to be made for cohesion and coherence of a text than in a stand-alone sentence. Moreover, Gohar et al. (2018, p. 867) claim that retention and generation have a greater impact on the effectiveness of word-focused tasks than the other three criteria in TFA, since the difference in TFA score between the sentence writing and composition task is reflected only in these two criteria. They further detailed this greater impact to be related to the importance of the form-meaning connection (retention) and meeting words in new and different contexts (generation).

In a similar study, Hu and Nassaji (2016) compared four tasks: multiple choice questions, definitions, gap fill, and rewording sentences. The gap fill task was given IL 2 and TFA score 7 while the other three tasks had the same weighting (IL 3 and TFA score 6). This highlights differences between the models as gap fill is considered less effective than the

other tasks in IL, but more effective according to TFA. The post-test showed that the gap fill task, which had the highest TFA score and the lowest IL, was the most effective out of the tasks. This indicates that TFA has a better predictive ability than ILH. However, TFA was not able to predict that the three tasks with a TFA score of 6 would behave differently from each other. This, in combination with the fact that the study did not include a control group, makes it difficult to draw conclusions on the predictive ability of TFA. Moreover, the authors conceded that many participants doing the sentence writing did not complete the task (Hu & Nassaji, 2016, p. 36). Consequently, this study cannot provide guidance on the comparison of the two models, but provides weak partial support for TFA in the instance of the gap fill task.

3.2.2 Assessing TFA for Different Tasks

Kamali et al. (2020) investigated the effectiveness of oral reproduction and summary writing both with a TFA score of 11. The study showed that the two experimental groups significantly outperformed the control group on both immediate and delayed post-tests confirming the effectiveness of word-focused tasks over the reading-only condition. Furthermore, oral production outperformed summary writing on both measurement points, despite having the same TFA score. Hence, the study only partially supports the predictive ability of TFA on vocabulary learning.

The tasks had the same TFA scores in sum and in parts, thus the differences could not be explained by unequal impact of the components of TFA. However, there were differences between the tasks pertaining to length of production and time. The oral reproduction task resulted in more words being used by the learners than in the summary task and the oral task took longer to complete. Hence, a possible explanation for the differences in effectiveness could lie in the time and rehearsal aspects, which might result in a deeper level of processing beneficial for word learning (Kamali et al., 2020, p. 18).

Moreover, Kamali et al. (2020, p. 18) suggest that there might be a hidden difference between the tasks regarding the degree of the generation criterion. The production of original language in the oral reproduction task was thought to be more extensive than in the summary task. As a consequence, oral reproduction necessitated deeper processing in order to allow for the inclusion of collocations and grammatical changes versus the summary which showed grammatical similarities to the original text. These findings suggest that the generation criterion, in its current form, does not offer enough degrees to allow the model to account for

such differences, thus Kamali et al. (2020, p. 18) call for a need to expand the generation criterion of TFA.

Huang et al. (2020) devised a research design to examine three parts of TFA more closely whilst leaving all other criteria the same. The elements investigated were generative use and productive use (both part of the generation criterion), and imaging (part of the retention criterion). There were four different tasks in the study, namely gap fill with pictorial annotations (score in TFA sub-set: 2), sentence writing with textual annotations (score in TFA sub-set: 2), reading comprehension with pictorial annotations (score in TFA sub-set: 1), and gap fill with textual annotations (score in TFA sub-set: 1). The tasks performed according to TFA predictions on both the immediate and delayed post-test resulting in full support for the TFA model. However, the narrow scope of the study to investigate the effectiveness of pictorial annotations using only three elements of TFA as well as the fact that there was no control group limit any generalisations of the study.

4 Discussion and Pedagogical Implications

This literature review aimed at providing an overview of what recent research shows to be key features of effective word-focused tasks. Consequently, empirical studies analysing the predictive ability of two theoretical frameworks developed to this end were reviewed, namely ILH and TFA. In this chapter, we discuss the findings of the empirical studies as well as the resulting pedagogical implications.

4.1 The Empirical Evaluation of ILH and TFA

Articles researching ILH, the first theoretical framework, have shown that there is empirical support for the predictive ability of this model (Alavinia & Rahimi, 2019; Feng, 2015; Soleimani & Rahmanian, 2015). However, there is conflicting evidence of the usefulness of ILH to predict long-term retention of vocabulary (Douglas, 2016; Yang et al., 2017). Furthermore, research into the impact of the three components of ILH (need, search, and evaluation) has outlined a more complex picture of factors contributing to vocabulary learning (e.g. Bao, 2015; Bao, 2019) and only partial support for the effectiveness of the model is presented.

Several studies find that ILH does not reflect all relevant degrees of deeper processing as required to distinguish between tasks (e.g. Bao, 2019; Yang et al., 2017). For instance, the importance for learners to revisit the form-meaning connection repeatedly as well as encountering it receptively and productively are not awarded different IL scores (Bao, 2019, p. 9). Moreover, Zou (2017, p. 70) suggests expanding ILH by introducing an extra level of evaluation including the differential processing associated with longer texts due to the additional need for coherence. Notwithstanding the evidence indicating that differential processing is an important factor in the evaluation of word-focused tasks' effectiveness, studies have also pointed to possible interference from task complexity when it directs learners' attention away from the unknown words (e.g. Bao, 2019).

The second theoretical framework, TFA, is not as thoroughly researched as ILH. Nevertheless, there is some empirical support for its ability to accurately predict the effectiveness of word-focused tasks (e.g. Hu & Nassaji, 2016) and more support for the effectiveness of word-focused tasks compared to control groups (Gohar et al., 2018; Kamali et al., 2020). However, at this time research only provides partial support for the predictive ability of TFA.

As was the case with ILH, research into TFA suggests that the framework does not entail all decisive factors influencing vocabulary learning (e.g. Kamali et al., 2020). In a similar manner to Zou (2017), Gohar et al. (2018) discussed the idea of differential processing and the impact of higher induced mental effort in a task on learners' ability to acquire new words. They suggest that generation and retention, two of the five main criteria of TFA, have a greater impact on vocabulary learning than the other parts of the model. Kamali et al. (2020, p. 18) lend further support to the idea of expanding the criterion of generation. In their study, they propose including additional scoring points for generation to distinguish the level at which a task induces learners to create new context with the unknown words.

4.2 An Evaluative Comparison

Although there is empirical evidence for the predictive abilities of both models concerning vocabulary acquisition of word-focused tasks (e.g. Feng, 2015; Gohar et al., 2018), there is as yet no conclusive evidence of the superiority of one model over the other (Gohar et al., 2018). In addition, research presents three points of concern calling for further expansion of the components in each system to ensure a more accurate evaluation of task features (e.g. Kamali et al., 2020; Zou, 2017).

The first point of concern is related to the inability of either model to clearly depict the effect from repeatedly establishing the form-meaning connection on vocabulary learning (Bao, 2019; Kamali et al., 2020). Moreover, the aspect of including both receptive and productive elements for creating the form-meaning connection is highlighted as promoting vocabulary acquisition. Hence, we find that support is strong for the importance of the form-meaning connection to be included when evaluating word-focused tasks.

The second point of concern pertains to the effects of differential processing as induced by, for example, composition tasks over sentence writing tasks. Concerning ILH, researchers note that the third component of the model, evaluation, does not contain a sufficient number of degrees in its current state to explain the deviating results of the two tasks (e.g. Taheri & Rezaie Golandouz, 2021). This is in line with findings presented by Kamali et al. (2020) for the criterion of generation in TFA. Moreover, previous research into the effects of generative use on vocabulary acquisition supports the need for a more extensive scale when evaluating generation in a task (Joe, 1998). Thus, we conclude that this is a key feature in need of further development.

The final point of concern is the need to balance the aspect of deeper processing, primarily differential processing, with the attentional resources of the learners (e.g. Bao, 2019; Gohar et al., 2018). In this literature review, learner proficiency was not identified as a key factor when determining the effectiveness of word-focused tasks. However, there is evidence that task complexity which distracts learners from the unknown words decreases the effectiveness of tasks. Care should be taken to ensure that the difficulty level of the task is appropriate for the learners, thus ensuring focus remains on vocabulary learning.

Our research question was: what do recent empirical studies on EFL indicate as being key features of effective word-focused tasks? Although neither of the two theoretical frameworks examined was proven to be sufficient for explaining all differences in effectiveness between word-focused tasks, research has demonstrated that some features seem to be more important than others. Particularly, research has found that features pertaining to the creative use of new words as described in the components of evaluation (ILH) and generation (TFA) are powerful indicators of task effectiveness. Moreover, features aimed at reinforcing the form-meaning connection as included in evaluation (ILH) and retention (TFA) have been shown to greatly impact successful vocabulary acquisition. However, this literature review showed that research indicates that neither evaluation nor generation and retention can distinguish between all features of tasks designed to promote creative use and establish form-meaning connection. In the next section, we will discuss the pedagogical implications of these findings.

4.3 Pedagogical Implications

Considering the number of research articles published on the topic of effectiveness of word-focused tasks, it is only logical to ask how these studies can best be put to use in the EFL classroom. As neither of the existing theoretical frameworks were found to conclusively predict the effectiveness of word-focused tasks on vocabulary learning, we cannot recommend that teachers start applying them in their preparatory work without modifications. However, there is sufficient support in the empirical evidence to warrant an awareness of the fact that certain features of tasks better promote vocabulary acquisition than others.

ILH, the first theoretical framework, in its simplicity provides user-friendliness and could easily be used to quickly assess any word-focused task. However, research indicates that the simplicity of the model also constitutes its weakness, as important features could be missed in its application (e.g. Zou, 2017). Thus, should teachers wish to use ILH, they need to

apply extra care when assessing the evaluation component in order to choose between possible tasks.

The second theoretical framework, TFA, allows for the inclusion of more components in the evaluation of tasks. However, the comprehensiveness of the 18 questions could lead to a feeling that the usefulness of the model is not sufficient to compensate for the time invested. Nevertheless, research shows that two of the five main criteria, namely generation and retention, could be more important for vocabulary learning than the others (Gohar et al., 2018, p. 868), thus presenting the possibility to use essential parts of TFA and reduce the time investment.

Contrary to the prevalent perception among Swedish EFL teachers that words are learnt as a by-product of other meaningful tasks (Bergström et al., 2021, p. 11), empirical evidence shows that word-focused tasks outperform reading-only tasks for vocabulary retention (e.g. Kamali et al., 2020). This is in line with Nation's (2013) idea of integrating the four strands of vocabulary teaching and Schmitt's (2008, p. 352) claim that incidental learning is most effective when combined with intentional learning. Therefore, teachers need to be aware of the impact of word-focused tasks as they integrate them into their teaching.

As discussed in the previous section, research supports the positive impact on vocabulary learning from the use of words in new and original contexts (e.g. Taheri & Rezaie Golandouz, 2021). Specifically, there is empirical evidence for the superiority of composition tasks and original sentence writing over more mechanical tasks such as gap fill tasks (e.g. Alavinia & Rahimi, 2019). Moreover, other types of creative tasks were also proven effective for vocabulary acquisition including oral reproduction (Kamali et al., 2020). Thus, teachers should include creative tasks in their teaching to profit from the effectiveness of the creative features. When designing creative tasks of a certain length (longer than isolated sentences), research advises caution to ensure task complexity is focused on the unknown words as to not distract learners from the objective of the exercise (e.g. Bao, 2019).

Notwithstanding the effectiveness of composition tasks, teachers should bear in mind that it is not the length of the task as such which leads to vocabulary learning. A key feature is the number of opportunities afforded to the learner to make the form-meaning connection. Even tasks with a lesser element of creativity and hierarchical organisation, such as definition tasks, were proven effective for vocabulary learning provided that the task directed learners' focus to meaningful contextual clues including new contexts for the words (Bao, 2019, p. 8).

Although not as widely researched as the evaluation component in ILH, there is evidence that the search component has a positive impact on vocabulary learning. This means that allowing learners to use dictionaries themselves to find definitions or L1 equivalents could increase vocabulary retention (Yang & Cao, 2021, p. 428). Hence, teachers should consider this possibility before providing glosses in the form of marginal comments, footnotes, or lists with translations.

To sum-up, there are valuable implications for EFL teachers in recent research on the effectiveness of word-focused tasks. First and foremost, there is empirical evidence of the superior effectiveness of using elaborate processing in post-reading tasks to enhance vocabulary acquisition as opposed to relying solely on incidental learning of new words from reading alone. Secondly, ensuring that learners are given the opportunity to use words in new and original contexts will promote vocabulary retention according to both theoretical models and empirical studies. Thirdly, teachers need to be aware of the focus of any task they include in their teaching, because not all elements of task complexity will lead to deeper processing of words, but could result in learners being distracted from the unknown words. Lastly, the difficulty level of any task needs to be adapted to the intended learners.

5 Conclusions

In light of the fact that the vocabulary challenge in English is so substantial, 8000 – 9000 word families for unconstrained reading comprehension of a wide selection of authentic texts (Nation, 2006, p. 22), Schmitt (2008, p. 333) emphasises the need for well-measured and effective vocabulary teaching in the EFL classroom. Hence, Nation's fourth job of the vocabulary teacher (Newton, 2020, p. 256), namely to provide opportunities for vocabulary learning as an integral part of teaching, is an area worthy of researchers' and teachers' thoughtful consideration. Accordingly, the purpose of this literature review was to examine what recent empirical studies on EFL indicate as being key features of effective word-focused tasks and to discuss resulting pedagogical implications.

Empirical studies support the effectiveness of elaborate processing in tasks over reading-only conditions (e.g. Feng, 2015; Kamali et al., 2020). Moreover, there is strong evidence of the positive impact of differential processing from tasks encompassing the production of new and original contexts for unknown words (e.g. Bao, 2019; Zou, 2017). However, neither of the existing theoretical frameworks (ILH and TFA) could conclusively predict the effectiveness of all word-focused tasks. Consequently, researchers suggest including a more extensive scale in the models when assessing tasks as well as expanding the criteria targeting differential processing to allow for more accurate predictions (e.g. Kamali et al., 2020).

As there is not enough empirical evidence to support the superiority of one theoretical framework over the other, we cannot recommend that these models be used to the letter by EFL teachers. However, research does indicate that the components of generation and retention, and evaluation have a greater impact on vocabulary learning than other components in the models. Thus, including tasks which incorporate differential processing and allow learners to repeatedly establish the form-meaning connection could lead to more effective vocabulary acquisition.

This literature review is subject to limitations. Although there is a continued high publication rate of empirical studies assessing the predictive abilities of theoretical frameworks for effective word-focused tasks, the share of articles analysing the TFA model is considerably smaller than articles investigating ILH. Moreover, the limited scope of two of the TFA-focused studies further restricted possibilities to thoroughly analyse its uses. Consequently, the empirical material used in this literature review may not be representative for TFA and could possibly influence the robustness of the findings for this model.

There is a need for further research as the published empirical data has not yielded conclusive results, despite the large number of articles researching the characteristics of effective word-focused tasks. Further research into the effects of differential processing, for instance in composition tasks, is needed in order to propose a more extensive common framework on which researchers and practitioners alike can base their decisions. Furthermore, we suggest empirical studies investigate the impact on vocabulary learning stemming from the ILH component search building on Yang and Cao (2021). Finally, there is a need for more research into the effectiveness of oral output tasks.

References

Alavinia, P., & Rahimi, H. (2019). Task types effects and task involvement load on vocabulary learning of EFL learners. *International Journal of Instruction, 12*(1), 1501-1516. https://doi.org/10.29333/iji.2019.12196a

Bao, G. (2015). Task type effects on English as a foreign language learners' acquisition of receptive and productive vocabulary knowledge. *System, 53*, 84-95. http://dx.doi.org/10.1016/j.system.2015.07.006

Bao, G. (2019). Comparing input and output tasks in EFL learners' vocabulary acquisition. *TESOL International Journal, 14*(1), 1-12. https://eric.ed.gov/?id=EJ1244083

Bergström, D., Norberg, C., & Nordlund, M. (2021). "Words are picked up along the way" – Swedish EFL teachers' conceptualizations of vocabulary knowledge and learning. *Language Awareness*, 1-17. https://doi.org/10.1080/09658416.2021.1893326

Brown, H.D., & Lee, H. (2015). *Teaching by principles. An interactive approach to language pedagogy*. Pearson.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press. https://rm.coe.int/1680459f97

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment – Companion volume*. Council of Europe Publishing. www.coe.int/lang-cefr

Craik, F., & Lockhart, R. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Direktoratet for høgare utdanning og kompetanse. (n.d.). *Register over vitenskapelige publiseringskanaler*. Retrieved September 29, 2021, from https://dbh.nsd.uib.no/publiseringskanaler/Forside

Douglas, V., Jr. (2016). Enhancing English academic vocabulary acquisition and retention in intensive English programs with the involvement load hypothesis. *Theory and Practice in Language Studies, 6*(12), 2237-2244. https://doi.org/10.17507/tpls.0612.01

Feng, T. (2015). Involvement load in translation tasks and EFL vocabulary learning. *The New English Teacher, 9*(1), 83-101.

Gohar, M. J., Rahmanian, M., & Soleimani, H. (2018). Technique feature analysis or
        involvement load hypothesis: Estimating their predictive power in vocabulary
        learning. *Journal of Psycholinguistic Research, 47*, 859-869.
        https://doi.org/10.1007/s10936-018-9568-5

Hu, H. M., & Nassaji, H. (2016). Effective vocabulary learning tasks: Involvement load
        hypothesis versus technique feature analysis. *System, 56*, 28-39.
        http://dx.doi.org/10.1016/j.system.2015.11.001

Huang, S., Eslami, Z., & Willson, V. (2012). The effects of task involvement load on L2
        incidental vocabulary learning: A meta-analytic study. *The Modern Language
        Journal, 96*(4), 544-557. https://www.jstor.org/stable/23361715

Huang, Y., Zou, D., Wang, F., Kwan, R., & Xie, H. (2020). Investigating the effectiveness of
        vocabulary learning tasks from the perspective of the technique feature analysis: The
        effects of pictorial annotations. *International Journal of Innovation and Learning,
        27*(3), 254-273. https://doi.org/10.1504/IJIL.2020.106810

Joe, A. (1998). What effects do text-based tasks promoting generation have on incidental
        vocabulary acquisition? *Applied Linguistics, 19*(3), 357–377.
        https://doi.org/10.1093/applin/19.3.357

Kamali, M., Behjat, F., & Bagheri, M. S. (2020). Examining the effects of oral reproduction
        and summary writing vocabulary tasks on L2 word learning: Technique feature
        analysis on trial. *Cogent Education, 7*(1), 1-25.
        https://doi.org/10.1080/2331186X.2020.1795966

Laufer, B. (2020). Evaluating exercises for learning vocabulary. In S. Webb (Ed.), *The
        Routledge handbook of vocabulary studies* (pp. 351-368). Routledge.
        https://doi.org/10.4324/9780429291586

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The
        construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26.
        https://doi.org/10.1093/applin/22.1.1

Lindqvist, F., & Dragemark Oscarson, A. (2019). Self-assessment of academic vocabulary
        among Swedish upper secondary school pupils. *Pertanika Journal of Social Science
        and Humanities, 27*(2), 747-772.
        http://www.pertanika.upm.edu.my/pjssh/browse/regular-issue?article=JSSH(I)-0001-
        20

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59-82. https://doi.org/10.1353/cml.2006.0049

Nation, I. S. P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching, 1*(1), 2-13. https://doi.org/10.2167/illt039.0

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139858656

Nation, I. S. P., & Webb, S. (2011). *Researching and analyzing vocabulary* (1st ed.). Heinle, Cengage Learning.

Newton, J. (2020). Approaches to learning vocabulary in the classroom. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 255-270). Routledge. https://doi.org/10.4324/9780429291586

Rassaei, E. (2017). Effects of three forms of reading-based output activity on L2 vocabulary learning. *Language Teaching Research, 21*(1), 76–95. https://doi.org/10.1177/1362168815606160

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329-363. https://doi.org/10.1177/1362168808089921

Skolverket. (2011). *Läroplan för gymnasieskolan* [Swedish National Curriculum for upper secondary school]. https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/laroplan-gy11-for-gymnasieskolan

Skolverket. (2020). *Ämnesplan för engelska* [Syllabus for English]. https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/gymnasieprogrammen/amne?url=-996270488%2Fsyllabuscw%2Fjsp%2Fsubject.htm%3FsubjectCode%3DENG%26tos%3Dgy&sv.url=12.5dfee44715d35a5cdfa92a3#anchor1

Soleimani, H., & Rahmanian, M. (2015). Visiting involvement load hypothesis and vocabulary acquisition in similar task types. *Theory and Practice in Language Studies, 5*(9), 1883-1889.

Sonbul, S., & Schmitt, N. (2010). Direct teaching of vocabulary after reading: Is it worth the effort? *ELT Journal, 64*(3), 253-260. https://doi.org/10.1093/elt/ccp059

Taheri, S., & Rezaie Golandouz, G. (2021). The effect of task type on EFL learners' acquisition and retention of vocabulary: An evaluation of the involvement load hypothesis. *Cogent Education, 8*(1), 1-21.

Tahmasbi, M., & Farvardin, M. (2017). Probing the effects of task types on EFL learners'
    receptive and productive vocabulary knowledge: The case of involvement load
    hypothesis. *SAGE Open, 7*(3), 1-10.

Tornberg, U. (2015). *Språkdidaktik* (5th ed.). Gleerups Utbildning.

Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis
    predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning,
    71*(2), 487-536. https://doi.org/10.1111/lang.12444

Yang, Y., & Cao, X. (2021). Effects of task involvement load on L2 vocabulary acquisition
    and their association with language aptitude. *The Asia-Pacific Education Researcher,
    30*(5), 421-430.

Yang, Y., Shintani, N., Li, S., & Zhang, Y. (2017). The effectiveness of post-reading word-
    focused activities and their associations with working memory. *System, 70*, 38-49.
    http://dx.doi.org/10.1016/j.system.2017.09.012

Zou, D. (2017). Vocabulary acquisition through cloze exercises, sentence-writing and
    composition-writing: Extending the evaluation component of the involvement load
    hypothesis. *Language Teaching Research, 21*(1), 54-75.
    https://doi.org/10.1177/1362168816652418

Zou, F., & Yan, X. (2019). Incidental English vocabulary acquisition through reading: A
    review in the last two decades. *English Language Teaching, 12*(12), 39-45.
    https://doi.org/10.5539/elt.v12n12p39

Appendix – CEFR Global Scale

Common Reference Levels: global scale (adapted from Council of Europe, 2001, p. 24)

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.