MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Teamwork makes the dream work? The effects of collaborative writing on the development of learners' vocabulary knowledge in the EFL classroom"

verfasst von / submitted by

Katja Resch, BEd

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Education (MEd)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 199 507 525 02
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Lehramt Sek (AB) UF Englisch, UF Psychologie und Philosophie
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Ute Smit

Acknowledgements

First of all, I would like to thank my supervisor Prof. Ute Smit - who not only created a supportive, appreciative and safe space for writing a master thesis, but also helped me with her mindful and encouraging words to overcome my inner perfectionist voice, so that I actually managed to complete my thesis in 2022.

I am also highly grateful for my workplace and the support I received from my headmaster, colleagues, and of course my dear third grade students of 2021/22 a.k.a. the lovely participants in this study.

My friends have been a great source of help and support. They always had a sympathetic ear and found comforting and uplifting words for me. I also feel extremely lucky to have a boyfriend on my side, who believed in me, helped me push myself when my motivation was low, and reminded me to have a break when I became too absorbed in work.

A special thanks also goes to my former colleague and still highly appreciated “German-English exchange partner” Claire Nicholls - who I could always turn to when I was puzzled by the English language, and, who devoted some of her precious free-time to proof-reading parts of this thesis.

In addition, I would like to thank Ms. F., who helped me to put things into perspective, to get to know myself better, and to embrace the various parts within me, so that I can now also thank myself for this accomplishment.

Finally, I am immensely thankful for my family and everything they have done for me. Each of them has supported me in their own way and left a mark on myself and this resulting master thesis. Thank you for the home you created, that also served as a space for recharging my energy throughout this process.

Table of contents

Appendices

List of tables

List of figures

List of abbreviations

CW	Collaborative writing
CLT	Communicative language teaching
IW	Individual writing
EFL	English as a foreign language
GR_{CW}	Group collaborative writing
GR_{IW}	Group individual writing
L2	Second language
LP	Less proficient
LRE	Language related episode
MP	More proficient
MTLD	Measures of textual lexical diversity
SCT	Sociocultural theory
SLA	Second language acquisition
VKS	Vocabulary knowledge scale
ZPD	Zone of proximal development

In statistical testing

p	p value, level of statistical significance
t	t-value (in a t-test)
N	total number of units in a sample
n	number of units in a subgroup of the sample
SD	standard deviation
M	mean
r	correlation coefficient
U	Mann-Whitney U
F	f-ratio value in an ANOVA
ANOVA	analysis of variance

1. Introduction

Writing is one of the four skills that learners of the English language need to develop throughout their schooling. This skill is inevitably linked to the development of vocabulary knowledge, in particular to the development of productive aspects of word knowledge, such as knowing how to use words correctly in context. Such productive vocabulary knowledge, however, takes more time and effort to build up. While learners find it easier to enlarge their vocabulary size by studying words to establish form-meaning pairings, they usually find it more challenging to deepen this knowledge by making use of the words meaningfully and correctly in context. Although communicative language teaching (CLT) has arrived in Austrian EFL classrooms – promoting learner interaction and the development of communicative skills – it appears that writing is still widely considered as a solitary activity that is assigned as home exercise. This common practice, however, bears the risk to put a particular group of students – mostly those who are less proficient and receive little support at home – at a disadvantage for their vocabulary learning. It was during my first months of teaching that I started questioning this common practice and remembered a concept I had encountered during my studies: the *zone of proximal development* (Vygotsky 1978), in which learners can advance from one stage of development to the next with the help of peer interaction. I thought: "Why not use the linguistic resources my students have to offer, to help those who struggle in producing meaningful and accurate texts? Can teamwork – during a collaborative writing task – make the dream work, and support vocabulary learning?"

Therefore, I set the objective for my Master thesis to explore the effects collaboration throughout a writing task can have on students' vocabulary learning. Ultimately, this thesis aims at providing insights into collaborative writing and vocabulary learning that can help language teachers to make informed choices about their approach to language teaching. To achieve this aim, not only were numerous research findings consulted and compiled in the literature review of this thesis, but also an empirical study was conducted to obtain first-hand data. In the following an overview of this thesis' structure is given:

In the first part of this thesis (chapters 2 to 4) relevant literature will be reviewed in order to establish a common understanding of the constructs under investigation and to provide the theoretical grounding for the conducted research project. The second part (chapters 5 and 6) is devoted to a thorough presentation of the empirical study. These two parts are then

brought together in the discussion of chapter 7. The conclusion and implications of the findings are to be found in chapter 8.

In more detail: The subsections of chapter 2 “Development of vocabulary knowledge” provide information on the meaning of the terms “word” and “word knowledge”, the cognitive process of language use, Nation’s (2013) framework for supporting vocabulary knowledge, and the productive vocabulary use in writing. Chapter 3 “Learning through interaction” then elaborates on Vygotsky’s *sociocultural theory of mind* (1978) and its application to language learning through Swain’s concept of *languaging* (2006). After having clarified the term “collaborative writing”, the subsections of chapter 4 engage with the effects collaborative writing can have on language learning and the factors that influence the effectiveness of this method. Chapter 5 presents the research project step by step. It starts with an overview of the research questions and hypotheses. Then it continues with the contextualisation of the project and a description of the research design, its participants and the materials for data collection. The chapter ends with information on data preparation and scoring. Chapter 6 offers a detailed analysis of the test results of each hypothesis after having clarified some basics of statistics. This chapter is followed by the discussion of the findings linking the results to previous research as well as to students’ feedback. Finally, chapter 8 summarizes the main findings and finishes off with practical implications for language teachers.

2. Development of vocabulary knowledge

This chapter aims at clarifying the term *vocabulary knowledge* by firstly explaining what is linguistically understood when referring to a *word*, secondly, how knowledge thereof can be characterized, thirdly, how the process of language production cognitively works, and lastly, how the development of vocabulary knowledge can be supported in an EFL classroom with additional focus on written productive vocabulary use.

2.1. Definition of the term "word"

Depending on the research area within linguistics, scholars investigate and focus on different aspects of what we understand as *word*. Starting broadly a *word* can be defined as "a basic lexical unit" (Schmitt 2010: 8), yet a *word's* meaning can be specified according to its main characteristics: its orthography (i.e. written form), semantics (i.e. meanings), phonology (i.e. pronunciation) and grammar (e.g. its position in phrases). From an orthographic stance one might conclude that a word is a bit of language that has a space on either side of it. The semantic definition highlights that words represent single, complete concepts. When we view a *word* as "a unit of language that is subject to the language's particular word-based phonotactic constraints (i.e. which sounds can go next to which other ones) and phonological processes" (Murphy 2010: 12), then we see it from a phonological perspective. Lastly, grammatically speaking, a *word* is an expression that has a part of speech which can be detected on the basis of its morphological inflections and its place in phrases (Murphy 2010: 12-13).

The term *word* can also be further specified in the study of language by using the terms *tokens*, *types, lemmas* and *word families*. A sentence, paragraph or page of a book consists of words. When we count each of these words, no matter whether they are of the same form or not, we count *tokens* – also called *running words*. For example, the sentence 'It is not easy to say it correctly' contains eight *tokens*. If we only count the words of the same form once, we count word *types.* Therefore, the example sentence only includes seven *types,* since 'it' occurs twice. From this viewpoint, the words 'book' and 'books' would be counted separately; yet, it would appear odd to say that these words are learned independently of each other. Therefore, scholars investigating words from a language acquisition perspective, count *lemmas* – headwords and their inflected and reduced forms. A *lemma* includes the different forms of the headword that are of the same part of speech. If we take the noun 'walk' as a *lemma* then

it includes its plural form 'walks' but not the inflected forms of the verb 'to walk' ('walks', 'walked', 'walking') since those are a different part of speech. *Word families* is a more comprehensive term and, unlike *lemmas,* includes closely related derived forms of the headword in addition to its inflected forms of the same part of speech. Thus, the verb 'walks' like in 'She walks to school every day' would not be counted as a separate entity as it belongs to the *word family* 'walk' (Nation 2013: 9-11).

This distinction is useful for further discussion on the development of vocabulary knowledge, especially the term *lemma,* which will be revisited when exemplifying the process of language production. But first, the various aspects of *word knowledge* will be elaborated.

2.2. Word knowledge

Despite the ambiguous meaning of the term *word* and the even more complex notion of *knowledge*, this section will try to pinpoint what it means when we say somebody knows a word.

Generally, word knowledge is multifaceted and can be described with regards to its *breadth, depth* and *fluency* (Nation 2020: 15). *Breadth* is located on the quantitative side of vocabulary knowledge and refers to the number of words of which one has established a form-meaning pairing. *Depth* and *fluency,* on the other hand, are located in the qualitative dimension of lexical knowledge. The former is characterized by knowing different aspects of words, such as collocations, derivations and register (Webb 2020: 6-7). Fluency describes the ability to use words with ease or speed and accuracy when producing the language (Daller, Milton & Treffers-Daller 2007: 8).

Furthermore, it is common to divide word knowledge into *receptive* and *productive* knowledge. Whereas *receptive* knowledge allows for recalling meaning when a word form is met (for example, in a reading exercise), *productive* knowledge allows for recalling a word form in order to communicate a certain meaning when speaking or writing (Nation 2020: 16). This distinction can be drawn across the three main spheres of vocabulary, namely *form*, *meaning*, and *use*, which are further divided into nine aspects that vocabulary knowledge consists of: *spoken form of words, written form of words, word parts* (belonging to *form*)*, form and meaning, concept and referents, associations* (belonging to *meaning*)*, grammatical functions, collocations,* and *constraints of use* (belonging to *use*) (Nation 2020: 16).

Table 1, provided by Nation (2020: 16), illustrates these aspects.

Table 1 Aspects of word knowledge

Form	Spoken	R*	What does the word sound like?
		P**	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Word parts	R	What parts are recognizable in this word?
		P	What word parts are needed to express the meaning?
Meaning	Form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints of use (register, frequency, …)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

*R= receptive knowledge; **P= productive knowledge

For instance, knowing the *spoken form* involves the receptive level of knowing what a word sounds like as well as the productive level of knowing how to pronounce it. Knowing the *written form* involves the ability to read and write the words. Being able to recognize and use *word parts* (i.e. affixes and word stems) meaningfully is another part of knowledge of word form. Such a knowledge, however, would be hollow if no meaning to the form was attached. Therefore, one of the most important aspects of vocabulary knowledge is the connection between *form and meaning*. That is being able to infer the meaning of a word from a form as well as using the correct word form for the corresponding meaning. Knowledge of meaning of

a word further involves knowing about related *concepts and referents* (e.g. homonyms, homographs and homophones). Furthermore, *associations* of word meanings, such as synonyms, antonyms and hyponyms are also related to vocabulary knowledge. When it comes to using the words in context it is important to be aware of their *grammar*, *collocations* and *constraints* on their use (e.g. which expressions are formal/informal) (Nation 2020: 17-22).

These different aspects can be perceived as different stages of a continuum of word knowledge that begins with a superficial familiarity with the word and ends with an ability to use the vocabulary correctly in free production (Quian & Lin 2020: 68). As various studies have found out (see Zarva et al. 2005; Crossely, Salsbury & McNamara 2014), vocabulary depth is closely associated to learners' productive language skills and can be regarded as one of the strongest indicators for language proficiency. For example, Crossley et al.'s (2014: 583-584) study showed that the variance in assessment of lexical proficiency of written samples was mainly explained by the collocation accuracy. As it is EFL teachers' objective to lead EFL learners towards certain proficiency levels in the course of their schooling, considerable attention should be drawn to the progression from receptive and quantitative word knowledge (breadth) to productive, qualitative word knowledge (depth and fluency).

Before we answer the question of how teachers can help their students to achieve this progress, we shall have a closer look at what is involved in the process of producing language.

2.3. The cognitive process of language use

As a detailed description of all factors and steps involved in the production of speech would go beyond the scope of this thesis, only the parts relevant in understanding the connections between the various aspects of vocabulary knowledge – which can further inform teaching and learning vocabulary – will be described.

Nation refers to Levelt's (1989; 1992) model of language use and adaptations thereof by, for instance, Bierwisch and Schreuder (1992). This model visualizes the partly conscious and partly unconscious steps of a speaker when producing a spoken message. An illustration of this model is provided in Figure 1 below. This figure deliberately depicts two differently shaped boxes, square boxes and round components, in order to indicate the two different types of knowledge that are involved. The square boxes illustrate cognitive processing that involves procedural knowledge, that is knowledge, which is not accessible through introspection. The

round boxes represent processing that draws on declarative knowledge, thus knowledge that can be examined through conscious thought (Nation 2013: 62).

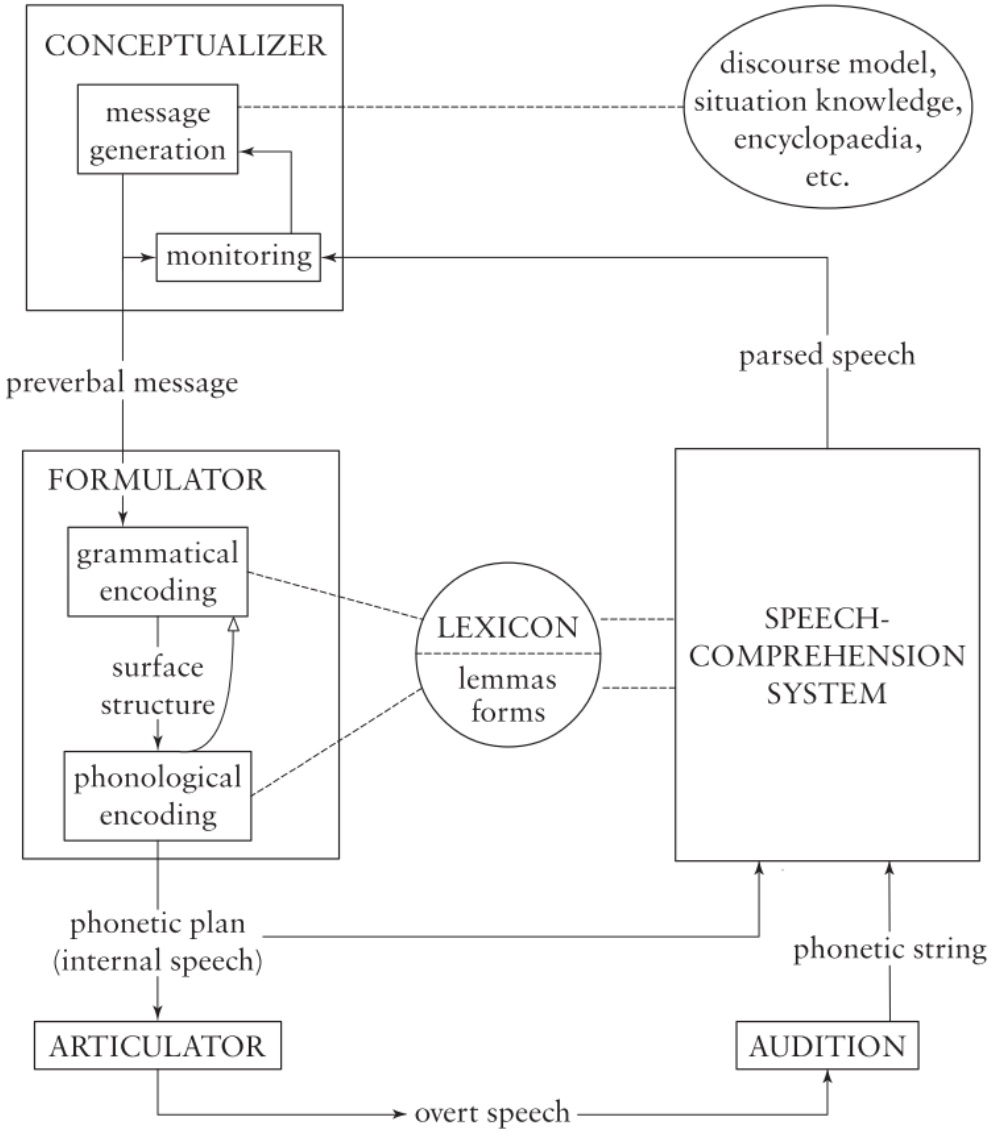

Figure 1. Model of speech production according to Levelt (1989: 9)

In general, the process starts with the *conceptualizer*, which involves firstly the intention to say something, secondly, choosing the relevant information, thirdly, sequencing the information and lastly, checking that the message fits with what has been communicated before. The fragments that have been chosen at this stage are not yet verbalized; that only happens in the next step, which is called *formulator.* There, the fragments from the *conceptualizer* are changed into a phonetic plan, which means that the message is grammatically and then phonologically encoded. The grammatical and phonological encoding are "lexically driven" (Levelt 1989: 181); that is, the words from our mental *lexicon* – or more

specifically, information in the *lemmas* – determine the grammar, morphology and phonology of the message. This assumption that the *lexicon* serves as mediator between conceptualization of a message and its grammatical and phonological encoding is what Levelt (1989: 181) calls the "lexical hypothesis". Next, this phonetic plan in turn is changed into actual speaking by the *articulator.* The circle of this model is closed by the step of self-monitoring, that is, listening to what has been articulated, comprehending the articulation, and using it to adjust further language use (Nation 2013: 61-62). This model also shows that most parts of language production draw on procedural knowledge, except for the information for the conceptualizer (discourse model, situational knowledge, etc.) and the lexicon.

Regarding the mental lexicon, two aspects are important to clarify. First, in contrast to the other parts of speech production, the mental lexicon involves declarative knowledge, which means that the knowledge is accessible through our consciousness and therefore can be broadened both by formal study as well as incidental learning. Second, the *lexicon* – according to Levelt – consists of two parts: *lemmas* and *forms*. The former comprises semantic and grammatical knowledge and possibly also information about appropriateness and register that help choose the fitting lemma for the given context. The latter is the morpho-phonological form of the word and is linked to the *lemma* by a so-called *pointer*. Following his lexical hypothesis, it is the lexicon that triggers the other aspects of language knowledge (Levelt 1989: 183-188). We will take the word "painter" as an example to help illustrate this process. When accessing the *lexicon* the *formulator* finds the lemma with the meaning 'a person who paints'. This meaning is connected to the grammatical information that it is a singular countable noun "expressing the agentive of the action expressed by the verb stem" (Levelt 1989: 183). Next these two components are connected to the morphological form *paint* + *er* as the suffix *-er* adds the agentive function to the word stem and thereby produces a noun. Then the phonological form and stress pattern matching the morphological form, namely /ˈpeɪntər/, are chosen and finally uttered (Levelt 1989: 182-183; Nation 2013: 63-64).

This view highlights the essential role of vocabulary knowledge for developing communication skills in another language. As grammatical and phonological aspects of words are so closely related to the lexicon and even determined by it, it is vital for language learners to meet vocabulary that is used in context in order to develop their vocabulary knowledge (Nation 2013: 62-63).

2.4. Framework for supporting vocabulary learning

This leads us to the question of how teachers can best support their learners in their development of vocabulary knowledge. While it appears rather easy for teachers to help students enlarge their vocabulary size by inviting them to study wordlists for example, supporting them to use those words correctly in context poses a much bigger challenge for them. To overcome this challenge, it is helpful firstly to visualize the cornerstones of a vocabulary learning supportive language course and in a next step to examine concrete methods in language teaching that are conducive to deepen students' vocabulary knowledge.

Paul Nation provides a useful framework for establishing a language course that accounts for supporting vocabulary learning. According to him (Nation 2013: 579) vocabulary learning should occur throughout the four strands of *meaning-focused input, language-focused learning, meaning-focused output* and *fluency development*. Each of these strands should be implemented for an equal amount of time throughout the language course in order to unfold its full potential for vocabulary learning.

A task that involves *meaning-focused input* aims at developing an understanding of the main messages conveyed by the language that are met in a reading or listening task. Here, 95% of the vocabulary should already be known in order to draw learners' attention to unfamiliar vocabulary. Thereby, incidental vocabulary learning can occur as learners try to guess the meaning from the context.

A *language-focused learning* task involves direct teaching and learning of selected vocabulary as well as training strategies for learning vocabulary (Nation 2013: 579). This strand can comprise, for example, a *word-focused activity* that "requires the understanding of the words on which the exercise focuses, with or without producing them". Due to the learners' attention on establishing a form-meaning link such decontextualized learning can quickly increase their vocabulary size (Laufer 2020: 352). Here, Nation emphasizes once more that no more than 25% of class time should be occupied by explicit teaching of language form and system (2013: 85).

When turning to *meaning-focused output* activities and *fluency* development tasks, the focus shifts towards conveying messages again. While the former requires 95% coverage of known words in order to encourage students to use unfamiliar items (which can involve communication activities with written input for instance), fluency tasks should include even

fewer unfamiliar words as their aim is to enhance repetition and performance at a faster than usual speed. Repeated speaking on familiar topics or graded reading are just two examples thereof (Nation 2013: 579)

Implementing Nation's framework bears two advantages: First, repetition – which is an essential part of learning vocabulary since it was found that it takes between 5 and 16 repetitions of an item in the input to establish a rich and accurate lexical representation (Webb 2007: 60-63) – is ensured, and second, opportunities for different conditions of learning can be provided. This way varied contexts for encountering and using vocabulary are guaranteed, which not only Ellis & Ferreira (2009) found as a crucial factor for helping learners develop depth of vocabulary knowledge.

2.5. Productive vocabulary use in writing

Yet, it should also be mentioned that more time and repeated effort is required for the development of vocabulary for productive use, i.e. speaking and writing (Nation 2013: 56). Research on productive use of vocabulary in writing supports this view.

The study by Laufer and Paribakht (1998) revealed that productive vocabulary developed slower and in less predictable manners than receptive vocabulary. Furthermore, Horst and Collins (2006) showed that increased fluency does not necessarily lead to increased richness of vocabulary use, as learners might be more likely to make more accurate use of words they already know well instead of taking risks with vocabulary they are not that familiar with yet. This finds support in Coxhead (2007), who found that learners were reluctant to use words in a written task that they had just encountered in a reading text, and led Nation (2013: 270) to the conclusion that "productive knowledge needs to be built up and is usually not immediately achieved". Lee and Muncie (2006) found that building up productive knowledge by employing a multitude of pre-writing tasks (such as watching a movie, doing cloze exercises, reading a text, writing a first draft of the text, etc.) on the same topic was effective, as a large amount of the target words were produced in the final writing task. Therefore, a written assignment at the end of a learning unit can support vocabulary to become productive.

When assigning a writing task that aims at the development of productive vocabulary knowledge, teachers should bear three aspects in mind: learners' level of language proficiency, learning burden of words, and a motivational task design.

First, the level of students' proficiency should set the objectives for which aspects of vocabulary should be learnt. For example, referring to Nation's nine aspects of vocabulary knowledge (see Table 1), beginners should only be required to know the spoken and written word form, form-meaning connection and some very basic grammatical rules. With increasing level of proficiency more aspects can be taught and practised (Nation 2020: 25).

Second, it is advisable to consider the learning burden of particular words – that is "the amount of effort needed to learn [them]" (Nation 2020: 24). The learning burden of a word can be light or heavy depending on the relationship to learners' language repertoire – their first, second and other language knowledge – as well as on the word's regularities with respect to how far it fits the systems of form, meaning and use within the second language. For example, if the spelling of a words differs vastly from its pronunciation, (e.g. *yacht*), if the core meaning cannot be inferred from the first language (e.g. *fork*), if the grammar is irregular (e.g. *bring – brought*), or if the collocations are different to the first language, the learning burden can be heavier. In these cases, teachers can help reduce the learning burden by comparing second language features to features of the first language or by emphasizing systematic patterns and analogies within the second language (Nation 2013: 45). Such awareness-raising exercises can be done during a pre-writing activity (Nation 2013: 270).

Third, motivation to use particular vocabulary items plays a vital role as well. Motivation in this context should be understood in Corson's (1985) sense, namely as the desire and opportunity to use a word. If the task does not provide the opportunity or does not elicit the wish to use particular words, then these words remain in the learners' *unmotivated* vocabulary (Nation 2013: 270).

In order to bring receptive vocabulary into productive use, Nation (2013: 271-276) suggests the following activities among others:

- Post-reading activities including completions with words from the text ranging from simply copying them from the text to using them with a different grammatical form.
- Paraphrasing a given sentence with a target word that is provided.
- Translating short texts from the first language into English.
- Training learners in dictionary use, especially highlighting the usefulness of checking example sentences of dictionary entries to help them for writing sentences.

- Dicto-comp and dicto-gloss activities in which learners listen to a text and then reconstruct it from memory, either individually or in groups.
- Email interaction with a more proficient user, e.g. a native speaker, which can lead to learners noticing and copying new words.
- Linked-skills tasks that include speaking that encourages learners to use particular vocabulary, then they orally report their results back to class and finally prepare a written report on the same topic.

One type of such a linked-skills task – namely collaborative writing – will be explained in detail in chapter 4 as it is also the main method under investigation in this thesis. However, before the discussion of this method is reached, the underlying learning principles should be outlined. Therefore, the next section will present the *sociocultural learning theory of mind* (SCT) by Vygotsky (1978), and how researchers have applied this theory to language learning by taking the example of Swain's concept of *collaborative dialogue* (1997).

3. Learning through interaction

3.1. Vygotsky's sociocultural theory of mind

The following outline of Vygotsky's SCT is not intended to be exhaustive, rather it should provide an insight into its core principles that have been taken up by many scholars in particular of the fields of second language teaching and learning.

The SCT is a sociocognitive theory that assigns verbal interaction a leading role in the cognitive development of individuals. It postulates that cognitive development, especially the development of higher order cognitive functions – that is, among others, voluntary attention, intentional memory and language learning – happens within a social context. More precisely, it happens during interaction between humans of different degrees of knowledge, i.e. between an expert and a novice (such as an adult and a child), where the expert offers assistance to the novice (Vygotsky 1981, referred to in Storch 2013: 13).

In essence, Vygotsky's claims that:

> [W]hat is the *zone of proximal development* today will be *the actual developmental level* tomorrow – that is, what a child can do *with assistance today* she will be able to *do by herself tomorrow* (Vygotsky 1978: 87 [my emphasis]).

I will take this quote as a basis for briefly explaining the key concepts of this proposition that I have highlighted in italics.

Vygotsky distinguishes between two developmental levels: the *actual developmental level* and the *zone of proximal development* (ZPD). Whereas the former grants insight into the already completed developmental cycles, thus things that children are capable of doing on their own already (= mental development retrospectively), the latter characterizes the individual's prospective mental development that can be achieved through problem solving in collaboration with more capable peers. Therefore, functions that are in the process of maturation, i.e. have not yet matured but will mature in the near future, can be considered to be in the ZPD (Vygotsky 1978: 85-87).

With this distinction Vygotsky further suggests that a child's development happens on two planes: the *social* and the *psychological* plane. First, the child's cognitive development occurs while they are among other people as an "intermental category" and afterwards inside the child as an "intramental category" (Vygotsky 1966: 40). The transformation process from an *intermental* to an *intramental* – or what he later calls 'interpersonal' and 'intrapersonal' – category is called 'internalization'. Thereby an "external activity is reconstructed and begins to occur internally" (Vygotsky 1978: 56-57). Thus, the proposition is that cognitive functions originating in dialogues between individuals – interpersonally – can undergo the process of internalization and thereby turn maturing functions gradually into one's own independent – intrapersonal – resources (Lantolf 2005: 342).

It is important to note that Vygotsky also differentiates between *development* and *learning:* "[L]earning is not development; however, properly organized learning […] sets in motion a variety of developmental processes" (1978: 90). From this perspective, *development* presupposes *learning* and according to him, *learning* best takes place in the ZPD, where reaching a new stage of development is the objective. If learning was only focused on developmental levels that have been acquired already, then this learning would be ineffective for the child's overall development (Vygotsky 1978: 89).

In the ZPD, spatially visualized as "the distance between the actual developmental level […] and the level of potential development […]" (Vygotsky 1978: 86), the nature of *assistance* offered by the more knowledgeable peer plays a key role. For the interaction to have a positive effect on the learning of the novice, the expert needs to be sensitive towards the

current state of the novice's knowledge and the next possible state that could be achieved with assistance (Vygotsky 1978: 89). Consequently, careful assessment of the novice's developmental state by the expert and a subsequent adjustment of their assistance is necessary. Clues for such can be the novice's utterances, for example. To support the learning of the novice, the expert provides finely tuned assistance, which they then reduce gradually, inviting the novice to take increasing responsibility for the activity (Storch 2013: 14). This kind of assistance is also commonly known as *scaffolding* (which is not a term Vygotsky coined but was introduced by Wood et al. who based it on his SCT) (Wood et al. 1976 referred to in Storch 2013: 14).

While Vygotsky refers to the necessary participants in the interaction to create a ZPD as the *more knowledgeable* and *less knowledgeable*, it is also important to mention that the former term is not restricted to refer to adults/teachers and the latter to children/students only; meaning that, for example at school, it is not only teachers who can provide their students with the interaction necessary for learning to take place. As confirmed in research in Second Language Teaching peers can also provide each other with scaffolding (e.g. Kowal & Swain 1994). Donato (1988, 1994) found that with some groups of learners *collective scaffolding* emerged. These are scaffolding episodes in which learners pool their partial knowledge of their L2 to solve various linguistic tasks. According to his observations, those learners performed beyond their existing level of linguistic expertise (which can be interpreted as the *actual developmental level* in Vygotskian terms). Furthermore, Donato found partial evidence that the knowledge the learners have co-constructed as a group was internalised and used independently by them in a follow-up activity. Considering these findings, Lantolf (2000: 17) suggests understanding the ZPD as the "collaborative construction of opportunities for individuals to develop their mental abilities".

3.2. Swain's theory of "languaging"

One scholar that applied Vygotsky's sociocultural theory to second language acquisition (SLA) and paved the way for research in this area is Merrill Swain. Her body of work provides insights into the vital role of interaction for language learning. Concepts which she has coined and are briefly outlined in this subsection are *output hypothesis*, *languaging*, *collaborative dialogue* and *language-related episodes* (LREs).

Swain's *output hypothesis* (1985) can be viewed as a reaction to Krashen's *comprehensible input hypothesis* (1981, 1982, 1985) and Long's *interaction hypothesis* (1983, 1985). Krashen's main claim was that for SLA to be facilitated, comprehensible input is demanded, that is linguistic input that is slightly more advanced than the learner's current L2 knowledge level. According to Krashen, the reasons for understanding this comprehensible input and further developing one's language skills lie within the learner, namely in their existing L2 knowledge and extralinguistic knowledge. Long agrees with the importance of comprehensible input but highlights the relevance of verbal interactions in the form of *negotiations for meaning* for learning to take place. Negotiations for meaning during a conversation can assume the shape of clarification requests, confirmation and comprehension checks and can lead to modification of input in order to avoid communicational breakdowns. To him, this modification allows for comprehensible input and thus for language learning (Storch 2013: 8).

Swain acknowledged these SLA theories but since research at that time also suggested that comprehensible input is not sufficient for L2 acquisition, she decided to investigate the role of output. In 1985 she proposed the *comprehensible output hypothesis* which states that in order to successfully learn an L2, learners need to be required to also produce language orally or in written form in an understandable way. The reasons being that input only stimulates comprehension which is achieved by processing language for gist "by relying on understanding key content words or on general knowledge" (Storch 2013: 9) whereas production initiates processing language syntactically (Storch 2013: 9).

Almost ten years later she reconsidered her hypothesis and modified it to *pushed output* instead of *comprehensible output* (Swain, 1993)*. Pushed output* then also took into account the importance of learning to produce language accurately and not only to the extent that it can be understood. Therefore, her claim was that teachers need to push their learners to produce language with grammatical accuracy which should help them to notice gaps in their current language knowledge and consequently invite them to explore and reflect on their linguistic resources. For example, when learners are trying to communicate a certain meaning, they might not know yet how to express it properly, they then need to look for other ways to overcome this gap and can test hypothesis. Thereby deeper cognitive processes during L2 learning are initiated and greater learning gains can be achieved. Swain and Lapkin (1995) could find evidence for these deeper cognitive processes by analysing think-aloud protocols

of L2 learners while they were producing a text. Izuzmi (2002) and Izumi and Bigelow (2000) also confirm Swain's claims as they showed that learners who were asked to produce output significantly outperformed in their attainment levels learners who were only asked to comprehend input (Storch 2013: 9).

Up to this point, Swain was mainly focussing on L2 learning from an information processing angle. SCT, however, affected her research interest which is why it shifted towards the role of language in interaction and how it mediates L2 development. Along with this shift, she refined the concept of *output* to *languaging* (Swain 2006, 2010; Swain et al. 2011). Swain (2006: 98) defines *languaging* as follows:

> Languaging […] refers to the process of making meaning and shaping knowledge and experience through language. It is part of what constitutes learning. Languaging about language is one of the ways we learn language. This means that the languaging (the dialogue or private speech) about language that learners engage in takes on new significance. In it, we can observe learners operating on linguistic data and coming to an understanding of previously less well understood material. In languaging, we see learning taking place.

Therefore, *languaging* is not only the process of conveying a message (i.e. *output*) but also a cognitive tool for learning (Swain & Lapkin 2002: 285) – a tool in the shape of articulated thinking about how to overcome a linguistic problem. This articulation can happen in two ways: as *private speech* or *collaborative dialogue*. Private speech can be understood as self-talk or, as referred to in psychology, self-explanation (Chi et al. 1989, 1994). It is the process of how learners engage in a dialogue with themselves during a solitary task. They evaluate their options for a given problem and draw on their own knowledge, and thereby might be able to talk themselves through to the solution (Lantolf 2005). Collaborative dialogue, however, is dialogue taking place between learners that work together to solve linguistic problems or to co-construct knowledge about language (Swain et al. 2002: 172). It becomes apparent that *collaborative dialogue* is interwoven with Donato's theory of *collective scaffolding* as mentioned in section 3.1.

According to Swain and Lapkin (2002: 286) "language learning occurs in dialogue, not as a result of it" undergoing the following steps. When collaborating, learners need to speak to each other. Speaking allows thoughts to be externalized resulting in an utterance which is an object that can be analysed, questioned, agreed or disagreed with, changed, reflected upon,

or disregarded. Applied to SLA settings that means that during their dialogue, learners will deliberate about language form and lexical choice in order to make their intended meaning as clear and precise as possible. The vocalized deliberation about language, referred to as *languaging*, then mediates L2 learning since the talk also reinforces the process of internalization. As already described in the previous section, it is the claim that linguistic forms and meanings that have been jointly worked on – in an intermental activity – can move inside the individual learner to become an intramental/psychological activity (Swain & Lapkin 2002: 286). Thereby collaborative dialogue allows for a co-construction of knowledge which can instantiate individual L2 learning.

Research on collaborative dialogue operationalized *languaging* in order to make it a unit of analysis through the so-called *language-related episodes* (LREs). An LRE is "any part of a dialogue where the students talk about the language they are producing, question their language use, or correct themselves or others" (Swain & Lapkin 1998: 326). These episodes are therefore marked by learners' attention focusing on language use and usually take place while students are jointly completing a given task. LREs comprise deliberations about any aspect of language, i.e. grammatical form, wording, meaning of vocabulary, punctuation, cohesion, etc. (Storch 2013: 28) and can vary in length and depth of engagement (Storch 2013: 35). For instance, the following LRE is an example of a form-based LRE that arose between two students during their completion of a writing task (Storch 2009: 145):

16 Olivia: Figure 3 shows that there were Vietnamese and Laotians participate
17 Shirley: Participating
18 Olivia: Participating
19 Shirley: Is correct? (sounds incredulous)
20 Olivia: (laugh). . . Because there were. We are using there were

Olivia forms a sentence including the simple present form of the verb *participate*. Her colleague then suggests using *participating*. Olivia then agrees with her colleague and even provides the justification for using this form (Storch 2009: 145) by referring to the phrase they used in their text.

This excerpt from a collaborative dialogue also shows that their LRE was correctly resolved, which is the desired outcome. However, not all LREs may be resolved or resolved correctly –

there are also instances in which learners neither solve the linguistic problem nor manage to do so correctly. Nevertheless, multiple studies (e.g. Adams & Ross-Feldman 2008; Alegría de la Colina & García Mayo 2007; Niu 2009; Storch 2001b; Storch & Aldosari 2013) substantiate that the majority of learners were able to resolve the encountered LREs and to resolve them correctly.

In conclusion, this section showed that Vygotsky paved the way for a new understanding of language learning, highlighting the significance of peer interaction for developing one's language competence. Research in *languaging* showed that through collaboration, learners not only discussed language-related problems, but were also able to solve many of them – problems that they possibly could not have solved if they had worked on their own. Therefore, collaboration can create a space in which learners enter the zone of proximal development. The subsequent section will provide more insights into this phenomenon in collaborative writing tasks.

4. Collaborative writing in L2 classes

One method that has learning through peer interaction at its core is *collaborative writing*. Now, the characteristics of collaborative writing will be described, the effects collaborative writing can have on language learning will be discussed and possible factors influencing learning in such a context will be addressed.

4.1. Specification of the term "collaborative writing" and task types

In the broadest sense *collaboration* refers to sharing labour with others. The term originates from the Latin word *collaborare*, meaning *to labour together*. Essentially it can be defined as "to work with another person or group in order to achieve or do something" (Merriam-Webster online dictionary). Therefore, collaborative writing can be understood as the co-authoring of a text by two or more authors (Storch 2013: 2). According to Ede and Lunsford (1990) collaborative writing is characterised by three aspects: considerable interaction throughout all stages of the writing process, shared power and responsibility over decisions made regarding the text produced, and the production of one single text.

Noteworthily, writing – from this point of view – is both a process and a product. As a process, it requires the authors to work together in the writing process, which comprises planning, generating ideas, deliberating about textual features and structure, editing, revising. Thereby

they not only exchange ideas but also negotiate meaning in order to reach a common understanding of the issue at hand (Schrage 1994 referred to in Storch 2013: 2). This characterisation shows that collaborative writing can be a very extensive task; however, it can also be implemented in the L2 classroom on a smaller scale by excluding the editing and revising phase for example. The most striking feature of a collaboratively written product then is that it is a jointly produced text, that cannot simply be divided into parts that have been written by separate individual authors (Stahl 2006). If the latter is the case, or if different tasks involved in the writing process (e.g. drafting or editing) are distributed among the writers, then *cooperative* writing is the more fitting term (Dillenbourg et al. 1996). In collaborative writing there is no such division of roles, but rather a mutual engagement and collective effort. As a result, such a text is also jointly owned by all contributing authors (Storch 2013: 2-3).

In the L2 classroom collaborative writing can be implemented through different task types. In literature on collaborative writing scholars differentiate between meaning-focused and language-focused writing tasks. Examples of the former can be descriptive and argumentative texts, reports, and jigsaws[1]. These tasks prompt students to compose a text with the help of pictures or written input. What makes them meaning-focused is that learners' attention to language form is incidental. They are mainly occupied with writing a meaningful text and pay attention to form only when difficulties with language use arise. Language-focused tasks, on the other hand, have it as their objective to shift learners' focus to certain language forms that have been selected in advance. A common language-focused task for collaborative writing is the dictogloss activity or adaptions thereof. A dictogloss task has different stages. First, the teacher reads out a text passage which learners are simply listening to. Next, the teacher reads the same text again with learners taking notes this time. After that, learners usually work in pairs or small groups to reconstruct the read text with the help of their notes. The read passage usually incorporates certain grammatical forms, and the reconstruction process should push learners to use those forms. Although this task focuses on certain language aspects, it differs from controlled grammar exercises such as editing or open cloze as learners are required to compose a complete text, not just changing or inserting pieces of a text (Storch 2013: 53-54).

[1] A jigsaw activity is an information gap task in which each member of the group or pair receives one part of the information and in order to get access to the whole information and complete the task they need to exchange their bits of information (Storch 2013: 53).

4.2. Effects of collaborative writing in the L2 classroom

As has been outlined in section 3.2 collaborative tasks in SLA settings provide opportunities for *languaging* and are thus regarded as a site for language learning. As *collaborative writing tasks* are one method of implementing collaboration among learners, it is argued that collaborative writing tasks create similar supportive conditions for L2 learning: *languaging*, *collective scaffolding* and *L2 use*.

During the composition of a text learners naturally attend to their own language skills and notice possible gaps in their knowledge. While they are thinking about which words best express their ideas or which verb form is the most suitable for transporting their intended meaning, their attention focuses on language choice and form (Storch 2013: 28). Research (Cumming 1990; Storch & Wigglesworth 2007) suggests that such deliberations account for about 30 percent of the learners' talk during the composition of a text. The amount appears to be the same for both conditions, whether students composed the text individually or collaboratively. As we have seen with Swain's theory of *languaging,* both private speech and collaborative dialogue allow for language learning. However, there is one distinct difference between the two of them: the quantity of linguistic resources learners can draw on. Whereas learners engaging in solitary writing tasks can only rely on their own language competence (drawing on their first and second language skills) to solve linguistic problems, learners engaging in a collaborative task can draw from a larger pool of knowledge. As a result, solitary writing rather prompts students to use existing knowledge than to generate novel language creations (Manchón 2009 referred to in Storch 2013: 37), which in turn makes it more difficult to resolve LREs.

Apart from analysing students' dialogues for LREs, researchers also tried to pin down effects of collaboration on the quality of the texts produced. Studies by Fernández Dobao (2012) and Wigglesworth & Storch (2009), among others, analysed the written products (stories and essays) with respect to their length, linguistic complexity, and accuracy. While Wigglesworth & Storch compared individually written texts with texts written in pairs, Fernández Dobao additionally included texts written by small groups. Their assumption was that texts written in collaboration might also be longer, linguistically more complex and more accurate than text written alone. While neither of them could prove that collaboration leads to longer and more complex texts, they could provide significant findings regarding the accuracy of collaboratively

written texts. Both studies showed that collaboratively written texts indeed contained significantly fewer errors than individually written texts. Moreover, Fernández Dobao (2012: 54) found that individually written texts were the least accurate and that texts written in small groups was significantly more accurate than those texts written in pairs. In combination with the analysis of the participants' LREs, the researchers attribute the greater accuracy of collaboratively produced texts to the greater availability of linguistic knowledge (Storch 2012: 76).

The pooling of linguistic knowledge while collaborating can be seen as a form of *collective scaffolding*. Learners can provide each other with assistance, which can be expressed through various forms of feedback: suggestion of alternatives, corrections, recasts and confirmation. This kind of peer feedback bears a number of advantages. First, it is provided timely and immediately as a response to learners' signs of requesting help. Thereby questions can be answered as soon as they occur. This would not be possible in teacher feedback on completed texts. Here students only get access to their feedback once their work has been returned. At that time, they might not even remember the difficulties they had faced or simply do not care much about the corrections since they might consider their text as 'done and dusted'. Secondly a study conducted by Brooks and Swain (2009) suggests that such immediate peer feedback, compared to teacher feedback, is also more useful for learning as it tends to be more appropriate to learners' developmental level. The researchers' explanation was that the assistance provided by the teacher in this study was involving structures above learners' developmental stage (Brooks & Swain 2009). In Vygotsky's terms one could say that the assistance was beyond learners' zone of proximal development. On a related note, these research findings might also point to the importance of teachers' sensitivity to learners' developmental level and evaluation of what lies within and what lies beyond their ZPD. Therefore, it appears that pair work can be a good solution to offer students scaffolding that is appropriate to their needs.

Lastly, collaborative writing tasks and the dialogue triggered by it quite naturally create the need to use the L2. Not only are learners writing in the L2 but also practicing their speaking skills while discussing their ideas. Thereby the activity becomes a linked skills task, as was already touched on in section 2.5. Furthermore, they are practicing numerous communicative functions by taking on different roles in the writing process. As they are co-authors, sounding

boards, tutors and critical peers they can practice giving explanations, asking for opinions, offering feedback, voicing criticism and suggesting improvements or alternatives (Weissberg 2006; Camps et al. 2000). However, in order to perform these communicative functions in the L2 a certain level of language competence is required. If learners do not feel equipped enough to vocalize their suggestions in their L2, they are likely to turn to their L1. In addition, when students possess the same L1 it often comes more natural to them to converse in their L1 than in their L2. Such observations are not only common among language teachers but there are also studies confirming this experience (Kang 2005). Thus, to encourage students to use as much of their L2 as possible, language teachers could provide language help (such as phrases for expressing the various communicative functions) and remind students of using the L2.

These attributes of collaborative writing show the potential collaborative writing tasks offer for language learning. Nevertheless, to what extent this kind of activity is able to exploit its potential is influenced by various factors. The following section will examine a few of them.

4.3. Factors influencing effectiveness of collaborative writing for L2 learning

When implementing the method of collaborative writing in the language classroom, teachers also need to make decisions regarding the grouping of their learners. As they want to enhance their students' language learning, they need to consider how to best group their learners. Any teacher will agree that assigning two students to a dyad who do not get along will end in little collaboration and most likely in little language learning. So, teachers try to take students' preferences for each other into account. However, there are more aspects that are worth to be considered when deciding on the grouping of learners. The proficiency level of group members, the group size, and the relationships formed between the group members in combination with the roles they have adapted also have an impact on the language-learning opportunities.

Two studies (Leeser 2004; Storch & Aldosari 2013) investigated the impact different proficiency level pairings can have on language learning by means of analysing the dyad's dialogues regarding the types of LREs produced and the resolution of their LREs. Both studies involved three different groups of pairings: high-high proficiency pairs, high-low proficiency pairs and low-low proficiency pairs; they differed however in task type. While Leeser's participants were asked to complete a dictogloss task (2004: 62), Storch & Aldosari's (2013: 35) students completed a composition task. The results of Leeser's study suggest that L2

proficiency pairing has an effect on how many form-focused LREs came up, and to what extent the dyads were able to resolve their LREs. In particular, it was found that high-high proficiency pairs generated the largest number of LREs, of which most of them were correctly resolved. Low-low proficiency groups produced few LREs and only two thirds of the LREs could be solved. Mixed proficiency pairs generated more LREs than low-low pairs, however this difference could not be proven significant. Furthermore, the analysis of the dialogue of the high-low pairs showed that most of the LREs were resolved by the higher-proficiency member and that the lower-proficiency member contributed rather minimally (Leeser 2004: 67-71). Leeser, therefore, calls the benefit of pair work for low-proficiency learners in the context of his study[2] into question (Leeser 2004: 73). With regards to the resolution of LREs in combination with the proficiency level grouping, Storch & Aldosari (2013) obtained similar results. The quantity of incorrectly resolved LREs was also bigger in the low-low proficiency pairs than in the high-low and high-high proficiency pairs. Surprisingly, they found that some mixed-proficiency pairs generated fewer LREs than low-low pairs, which differs to Leeser's findings. Yet, the interaction between the high and low proficiency students seemed similar to what Leeser had observed: the high-proficiency learner dominated during the dialogue while the lower-proficiency learner was rather passive (Storch & Aldosari 2013: 45-47).

These findings show that there seems to be a relation between students' proficiency level and the role they adapt during pair work. By adapting certain roles, learners also establish a particular pattern of interaction, which is likely to influence the quantity and quality of language learning opportunities for the respective students. Storch (2009) proposes that pairs form relationships that vary in their degree of mutuality and equality. Whereas the degree of mutuality describes "the learners' level of engagement with each other's contributions" (Storch 2013: 61), the degree of equality describes their level of contribution to the task and distribution of decisions. Depending on the level of each of these criteria, students' predominant interaction pattern can be characterized as *collaborative* (high mutuality & high equality), *dominant/dominant* (low mutuality & high equality), *dominant/passive* (low mutuality & low equality), *expert/novice* (high mutuality & low equality). The most preferable relationship, of course, is the *collaborative*, in which both members of the dyad contribute to

[2] He also discusses possible reasons for these findings, among others he questions the appropriateness of the dictogloss task and the developmental readiness of low-proficiency learners to learn grammatical information through such a collaborative task (Leser 2004: 74-75).

all parts of the task and engage with their partner's suggestions. *Dominant/dominant* relationships – or also referred to as cooperative patterns – are rather prone to conflict. Although they contribute to the task equally, they do so rather side by side instead of together as they are not willing to engage with each other's contributions. In a *dominant/passive* pattern, one of the students adopts an authoritarian role taking over the decision making for the most part leaving the other student to contribute minimally. However, if the dominant learner acts as a tutor who encourages the other learner to participate in the task, the interaction pattern is described as *expert/novice* (Storch 2013: 61-62).

If pair work already poses challenges for students to contribute to the task, the question arises whether it is useful to group students in teams of more than two people. When in small groups, some learners might not feel the responsibility to participate, since the task can also be completed by their other group members. As a consequence, group work can be an invitation for some learners to lean back, contribute little, and in turn learn little. Teachers who made such experiences with their group of learners might resort to pair work as it minimizes the possibility of such behaviour (Storch 2013: 60). Yet, there are also studies (such as Fernández Dobao 2012) that suggest greater learning success – measured by the amount of resolved LREs – when students worked together in groups of four than when they worked in pairs. Fernández Dobao's later study (2014) then zoomed in on the amount of contribution of each of the four team members and how this influences their learning of vocabulary. There it was found that only 25 percent of LREs were interactively resolved by all four members and the majority of LREs (75 percent) was only resolved by three or two students, raising the question how beneficial the LREs were for those learners who contributed less (Fernández Dobao 2014: 510). This is why she analysed students' subsequent vocabulary learning by means of a pre- and post-test vocabulary task as well as by evaluating their performance on an individual writing task. Her findings show that small group interaction leads to significantly more instances of learning new vocabulary than pair work does (Fernández Dobao 2014: 510-511). Furthermore, individual participants who did not contribute verbally to an LRE, were also found to have improved their vocabulary knowledge from observing their group members' dialogue – for example, one student provided the wrong translation for the word *boat* in the pre-test, did not contribute to their group's lexical LRE about this word, but then was able to give the correct translation in the post-test as well as use it in their written text (Fernández

Dobao 2014: 512). These findings accentuate the potential group work can have for language learning, even for those who seem rather passive during the collaborative task.

In this section, it became apparent that assigning students to group work does not necessarily result in genuine collaboration between them. The proficiency level of the team members and the resulting roles and interaction patterns also play a part in creating learning opportunities. In order to support collaborative interaction among group members, teachers can assign tasks which predetermine students' roles during the activity (e.g. giving a tendentially passive or lower proficient student information that the other student needed [Yule & Mcdonald 1990]) or require students to distribute roles with their respective responsibilities in order to keep each of the team members engaged.

This chapter attempted to provide some insights into the nature of collaborative writing tasks, the potential they bear for the second language learning and which factors can influence the success of language learning throughout a collaborative writing task. The compilation of these research findings can help language teachers to make more informed choices about how to implement collaborative writing tasks effectively.

5. Study

5.1. Overview

Against the backdrop of the aforementioned theories such as Vygotsky's *sociocultural theory of mind* (1978) and Swain's application to second language learning through the concept of *collaborative dialogue* (1997) this study intends to investigate the effects the method of collaborative writing (CW) can have on students' vocabulary knowledge. More precisely, this study tries to answer the following three research questions by testing the corresponding hypotheses:

R1: What effect does collaborative writing have on (less proficient) students' development of vocabulary knowledge?

- H1a: CW is more beneficial to learners' development of word knowledge than individual writing (IW).
- H1b: Less proficient learners improve their vocabulary knowledge through CW with a more proficient peer.

R2: How do texts written collaboratively differ from individually written texts with regards to fluency, lexical diversity, accuracy and target word fulfilment?

- H2a: Collaboratively written texts show greater fluency, higher lexical diversity and accuracy than individually written texts.
- H2b: Collaboratively written texts include more target words than individually written texts.

R3: How do the use of target words and the level of accuracy correlate with the development of vocabulary knowledge?

- H3a: Using target words in written texts has a positive effect on the development of vocabulary knowledge.
- H3b: Producing a more accurate text has a positive effect on the development of vocabulary knowledge.

In order to confirm or reject these hypotheses a quasi-experiment with 40 students was conducted. The empirical study took place in the researcher's own teaching setting, namely throughout a new learning unit with her students of two third grade lower secondary classes in Lower Austria (i.e. school year 7). To measure the development in vocabulary knowledge, students completed a vocabulary knowledge test twice: before and after a writing task. These tests included the same vocabulary items which they had encountered throughout their new learning unit. During the writing task, the classes were divided into two groups by the teacher: individuals and pairs. Therefore, half of the students completed the writing task individually and the other half collaboratively in pairs. The grouping took the students' proficiency level and the relationship between the students into consideration. The task prompted students to use the newly learnt vocabulary items in context. The vocabulary post-test then helps to reveal in how far students' knowledge of the given vocabulary items improved and whether there are differences between those having worked individually and collaboratively. The post-test also allows to zoom in on the learning gains of less-proficient learners who worked with a partner or alone. Furthermore, the written products are the basis for analysing the texts' fluency, lexical diversity, accuracy and target word fulfilment. To answer the last research question, the results of the texts' target word fulfilment and accuracy are linked to the post-test results.

5.2. Contextualisation and steps of project

The project was conducted in two third grade classes at the beginning of summer term 2022 (February/March). Their main course material are the books “MORE! 3 Student’s book - enriched course” Gerngross et al. 2021a) and “MORE! 3 Workbook - enriched course” (Gerngross et al. 2021b). From these books Unit 7 “You’ve got a friend!” had been chosen as the new learning unit within which students’ vocabulary knowledge was tested. This learning unit comprised six 50-minute English lessons which had the following objectives:

- Learners can read and understand texts about friendship.
- Learners can listen to and understand recordings about friendship.
- Learners can talk and write about friendships. In particular:
    - They can say how long they have been friends with someone (using the present perfect tense).
    - They can describe features of a good friend (e.g. to be honest with each other, to be kind, …).
    - They can describe what happened in a friendship/between friends (to have an argument, to fall out, to storm out, to make up,…).

During these six lessons, the teacher tried to adhere to Nation’s principle of teaching & learning vocabulary throughout the four strands of *meaning-focused input, language-focused learning, meaning-focused output* and *fluency development* (Nation 2013: 579; see section 2.4).

5.2.1. Preparational lessons for experiment

The unit started off by presenting students five different quotes about what a good friend is like (for example, *A good friend is always honest with you*). Learners could choose one quote they agreed with the most and then work in their groups to find an example to illustrate the importance of this characteristic of a good friend. In the next step they shared their ideas with their classmates. This activity already included some new words & phrases which were also packed into a short follow-up gap-fill activity (see Figure 2).

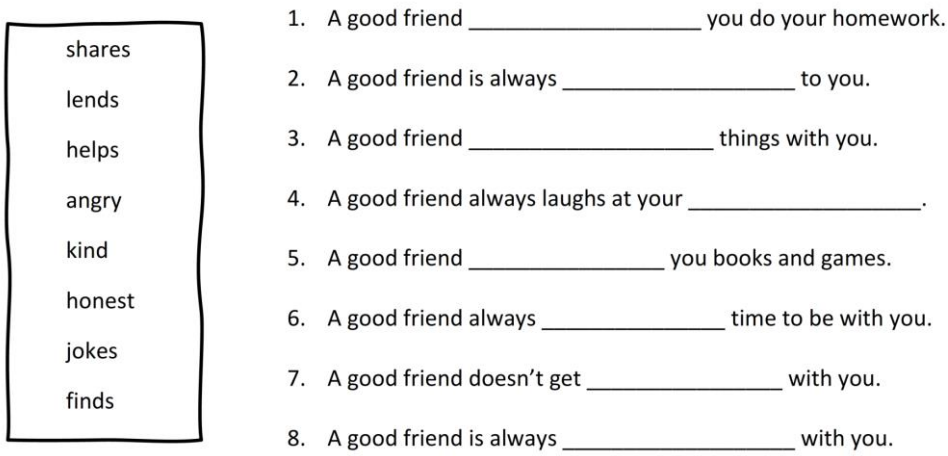

Figure 2. Gap-fill activity with vocabulary about friendship

In the next lesson, students read a text about the friendship of George and Alessia (Gerngross et al. 2021a: 58), where they encountered some more new vocabulary items on the topic of friendship. After a short reading comprehension task, the teacher drew learners' attention to the grammar structure of the present perfect tense in this text (*I've been friends with Alessia for two years*; *I've had those books for years*). Students were asked to make assumptions on what the implied meaning of this tense was as opposed to the past simple tense (*I was friends with Alessia*). In the next step, a revision of the formation and use of the present perfect tense with the help of an informative grammar sheet took place. To practice the formation of this grammatical structure, students were asked to complete a worksheet as homework.

The third lesson continued with students checking their answers with the key and the teacher answering questions or clarifying any concerns. Next, the difference between using past simple tense and present perfect tense as well as important key words (*for, since, last week, ago*,…) were discussed and practiced. Then the focus shifted to a listening comprehension task about objects people possess involving the present perfect tense (*I've had my MP3 player for 3 years*). By this time students had encountered the present perfect tense several times and completed grammar exercises, which is why students were asked to use the new grammar structure in a follow-up speaking activity. During a mingle-activity learners needed to complete a grid (see Figure 3) with names of their colleagues by asking each other questions and answering them (A: "Have you known your best friend for more than 5 years?" B: "Yes, I have". → A writes down name of colleague).

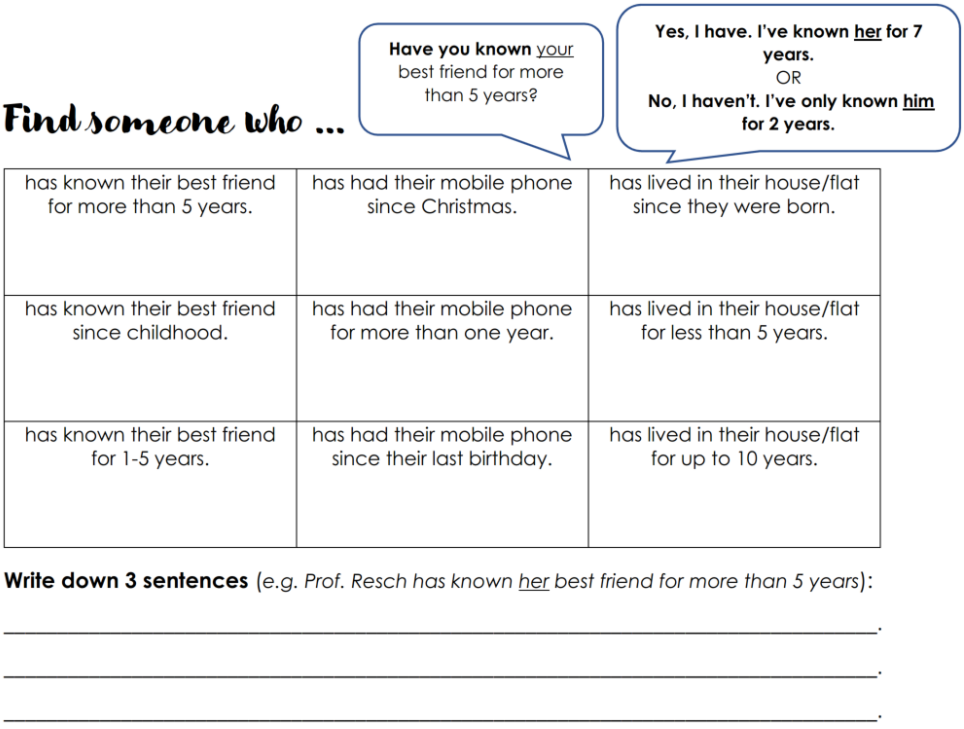

Figure 3 Grid for mingle activity to practice the present perfect tense

The next lesson was a crucial lesson for this project as it included many new vocabulary items that learners should acquire. Together with their teacher, students read the questionnaire "What kind of friend are you?" (Gerngross et al. 2021a: 60) and thereby highlighted and discussed new words/phrases. Students were asked to note down translations of those words/phrases they were unfamiliar with, for example *to have an argument with someone = mit jemanden streiten* or *to introduce someone = jemanden vorstellen*. After having completed this questionnaire, learners reviewed the meaning of the newly encountered vocabulary by matching the phrases with their meanings, e.g. *1 break up with = to finish a romantic relationship* (Gerngross et al. 2021a: 61) and then practiced the correct preposition for the phrasal verbs during a single-choice gap-fill activity, e.g. *James has fallen _____ a) up b) out c) through with Brenda* (Gerngroß 2021b: 61).

5.2.2. Lessons involving the experiment

In the fifth lesson a short revision of the new vocabulary items took place by completing another vocabulary task in the MORE! 3 Workbook. In order to expose learners to these items once more but targeting a different skill and sense, a listening comprehension task about a radio show talking about friendships was implemented. Thereby they heard and saw some of the words and phrases for a last time. During the last 20 minutes of this lesson, students

completed the vocabulary knowledge test (pre-test) in hand-written form, individually, and without any help from the teacher. In order to ensure learners did not face any difficulties with filling in their answers properly, students were made familiar with this type of test a few times before.

In the final lesson of this learning unit, learners were invited to make use of their freshly learnt words and phrases and the revised grammatical structure throughout the composition of a text. More specifically, students were prompted to write an email to their cousin Claire and tell her about two of their friends, Alice and Dan. Students were given a number of content points that they should include in their text. These content points were provided in German in order to elicit the target English vocabulary. The texts were composed digitally in Microsoft Word documents.[3] The crucial point for the project here was that two different interaction formats were applied for completing this task, constituting the treatment of the quasi-experiment. Half of the students of each class wrote the text individually and the other half wrote it collaboratively with a partner. All of the students had 30 minutes to complete the task and were not allowed to use dictionaries or other means of help. Those completing the task on their own were also provided with a desk divider, so that they would not be tempted to copy from their seatmate. As the nature of the collaborative writing task requires students to talk to each other, but those who work on their own should not hear what the groups are discussing, a solution had to be found to avoid such unwanted interference. Therefore, the dyads were placed in one half of the classroom and the individuals in the other half. Moreover, pairs were asked to whisper to each other to keep the volume at a minimum. Once students had finished their text, they handed it in via Microsoft Teams. The final step of this learning unit and experiment was that all students completed the same vocabulary knowledge test a second time (= post-test), just with the items being in a different order[4].

5.2.3. Feedback lesson

While the experiment ended at this stage, the teacher found it worthwhile to devote a part of her follow-up English lesson to collecting feedback from her learners. Students filled in an

[3] Using laptops for text production is common at this school since it has a focus on STEM (Science, Technology, Engineering, Maths). Therefore, students were used to working with this word processing program and no difficulties appeared during this phase. Which advantages and disadvantages this choice of writing mode had on the experiment will be discussed in section 6.2.

[4] Note that the students did not know that the same vocabulary items would be tested in the post-test.

online survey about how they experienced the writing task[5]. On a four-point Likert-scale, they could report how *useful* they found the writing task for learning vocabulary – ranging from *not useful, a little, rather,* to *very useful*. In the next question, they could indicate in how far they agreed with the following statements by ticking *I agree not at all, rather not, rather,* or *fully*: *I have the feeling that I have extended my vocabulary through completing the writing task*; *I have the feeling that I can use the vocabulary better/more accurately now.* Then, they were asked in how far they agree – using the same interval as before – with the statement: *Because of Microsoft WORD's suggestions for correction ("Korrekturvorschläge") or rather because of my correction of the words I learn something new ("lerne ich dazu")*. Lastly, learners were encouraged to reflect on aspects of the writing task they liked or did not like and which of them they found or did not find helpful. These reflections were collected in the comment area at the end of the survey.

There were three reasons for collecting this particular feedback from students and for including it in this Master thesis. Firstly, the teacher could get insights into learners' perception of vocabulary learning in combination with writing texts from a different angle. Secondly, the results of the feedback can be compared to the quantitative findings of the study to show in how far their perception is in line with their observed performance. Lastly, since the use of *Microsoft Word* for composing the text can be regarded as a major confounding variable during the experiment, the answers to the respective question could give some indication of the program's influence on students' written performance. Therefore, some of the feedback will be mentioned in the discussion of the results as well.

5.3. Research design

For the present study a mixed methods research approach was used that combines quantitative and qualitative data collection throughout the course of a quasi-experiment. Two different types of data were collected and analysed respectively. On the one hand, a vocabulary knowledge test was used, the answers to which were examined both quantitatively by calculating a score and qualitatively by analysing the written example sentences. On the other hand, students' written texts were analysed quantitatively by, for

[5] The survey was conducted in German in order to ensure students understand each question and could answer accordingly. In order to avoid unnecessary repetition, the items of the survey and their answers will be presented in English in this thesis. Moreover, only a subset of the most relevant items of this survey that correspond to the concepts under investigation will be included. The complete survey in German with its results can be found in the appendix.

example, calculating the type/token ratio and qualitatively by identifying mistakes in the texts. As the mixed methods are applied simultaneously, they can be referred to as 'parallel mixed designs' (Cohen et al. 2011a:25). The use of different methods bears the great advantage that the weakness of one method can be overcome by the strength of another method and thus can add more depth to the research at hand (Dörnyei 2007: 45).

As already mentioned, the core of this study is a quasi-experiment that involved a "pre-test-post-test two treatment design" (Cohen et al. 2011: 319) with one within-subjects factor, namely vocabulary knowledge, and two experimental groups. The experiment is classified as a *quasi-experiment* because of its lack of randomization of the test groups (Cohen et al. 2011a: 322). Real experiments use randomization in order to minimize the confounding effects of extraneous variables (Cohen et al. 2011a: 313), however, this was not possible in this study due to two reasons: the setting (school) and the nature of one of the hypotheses. The participants of this study attend two different classes and assigning them randomly to new groups proves highly impractical. Due to such practical constraints, it has become an accepted methodology in field studies (such as in the educational context) to work with non-equivalent groups (Dörnyei 2007: 117). Furthermore, as hypothesis 2 reveals, this study is also interested in the development of vocabulary knowledge of a specific group of learners (less proficient learners) through the interaction of another group of learners (more proficient learners). For that reason, the groups were assigned deliberately by the researcher according to students' proficiency level.

Another characteristic of this quasi-experiment is that there are two experimental groups and no control group. Due to the small sample size (*N*=40) adding a control group would minimize the number of observations per group considerably and thereby complicate statistical undertakings. Therefore, treatments were administered for both groups. One group was asked to complete a writing task individually, which will be referred to as GR_{IW} (group individual writing), and the other group was asked to complete the writing task collaboratively with a partner, referred to as GR_{CW} (group collaborative writing). This will allow for comparison between the two treatments; however, no comparison can be made to students who have not received any treatment. Therefore, conclusions regarding the effect of the treatments need to be drawn carefully since there are no results of a control group that help attribute the

outcomes solely to the intervention (and eliminate confounding effects) (Cohen et al. 2011: 313).

The design of this quasi-experiment can be summarized and depicted as follows:

Experimental GR_{IW}	O_1	X_1	O_2
Experimental GR_{CW}	O_3	X_2	O_4

Both experimental groups completed a vocabulary quiz as pre-test, which represents observation 1 (O_1) and observation 3 (O_3). Then they were asked to complete a writing task either individually or collaboratively – which was the treatment (X_1 and X_2) – and finally completed the same vocabulary quiz a second time to serve as post-test (O_2 and O_4). Thereby the difference between the post-test and pre-test can be calculated as well as the difference between the groups.

5.4. Participants and grouping

42 students of two third grade classes from a lower secondary school with a focus on sciences (*Realgymnasium Unterstufe*) in Lower Austria participated in this study. These students were chosen because they are the researcher's own English students and her teaching setting appeared to be the most relevant and suitable field for conducting her research. Since the phenomenon under investigation is a teaching & learning method, which means that the quasi-experiment resembled a rather regular English lesson, and no consent from students and their parents was obligatory. However, the headmaster, parents and students of course had been informed about the empirical undertaking prior to its implementation.

Out of these 42 students two students had to be excluded since they missed one of the two lessons in which data collection took place. As a result, the sample size of this study comprises 40 students. A short questionnaire was filled in by the students to collect demographic information as well as some information on their language background.

The group of participants included 24 male and 16 female pupils who were between 12 and 14 years old. The majority (*n*= 33) of students has German as their first language. For one student Croatian is their first language and for another student it is English. Four pupils grew up bilingually. Each of those four learnt German in combination with either Hungarian, English, Chinese or Italian. Lastly, one student was raised plurilingually, learning German, English and

Polish. Most of the participants (*n*= 31) indicated that the next language they have learnt (after their first language/s) was English. One student acquired German as their second language. Seven students reported that besides learning English they can also speak (to varying degrees) one of the following languages: German, Ukrainian, Italian, Hungarian, Spanish, and "a few other languages" (as one of them said). Two participants did not indicate that they can speak further languages as they have two or three first languages.

Furthermore, students were asked whether they had relatives or friends who were English native speakers, and if so, how often they communicated with each other. Nine students reported that they had English native speakers as friends or relatives. Five of them indicated that they spoke to them once to twice a week, one of them almost every day, another one once a day and one said that they did not speak to each other very often. One student did not provide an answer to this question. Next, they were asked if they had been to an English-speaking country and, if so, for how long. Out of the 40 participants almost half of them (*n*= 19) reported that they had been to an English-speaking country before. While the majority of these students (*n*= 16) went abroad for holidays (1-4 weeks), one of them lived in New Zealand for 2 years during infancy, one lived in the US for half a year and another one said that they travelled to Boston and New York twice a year.

Lastly, information on their extramural English engagement was collected. Students could indicate how many hours per week they listen to, read, speak, and write English outside of school. The bar chart below shows the number of students who reported zero, one, two, three, four, five or more than five hours of English engagement in each category. Overall, the most hours are spent on listening to English with 13 students spending five hours or more on it, and only twelve students spending one hour or less on it. Least hours are spent on speaking English. Eleven pupils stated that they do not speak English outside of school at all and 16 only one hour per week. There are only three students who speak English for 5 hours or more. A few more hours are spent on writing English; however, most students (*n*= 26) only write in English for one hour or less per week. It appears that reading in English is more common among students than speaking or writing, since more than half of them (*n*= 22) read English for 2 hours or more per week. In the comment section, they could add the activities in which they are in contact with the English language. Half of the participants reported that they engage with English while they are playing computer games. Ten of them stated that they

watch videos on YouTube or social media. It was mentioned 14 times that they watch movies or TV series in their free time. Less often (four times) they read books or online articles in English. Equally often (also four times) it was mentioned that they speak or write to someone in English. Only 3 students included listening to English music and two students stated that they are in contact with English when they are programming.

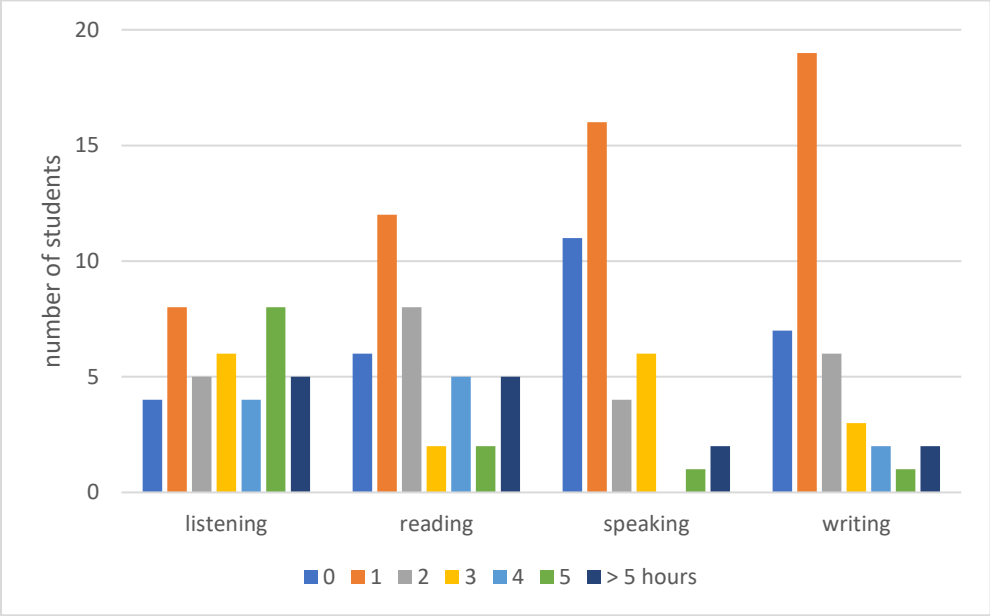

Figure 4 Students' weekly extramural English engagement across the four skills (listening, reading, speaking, writing)

In brief, this language background data shows that the majority of participants grew up monolingually with German as their first language and started learning English at school. Apart from a few students, the cohort does not display a high diversity in students' language repertoire. With regards to their English engagement, all of them indicated that they are exposed to English outside of school to some extent, with more hours spent on English input than with English output. Therefore, it has to be assumed that students' learning of the English language also happens in extramural settings.

At the time of the research project, the students had formal English language instruction at school for at least two and a half years[6] (for some it might be even more, depending on how English lessons had been implemented throughout their elementary school years). The first two years they had four English lessons per week and continued in the third grade with three lessons per week. The curriculum for English in the *AHS Unterstufe* provides that students

[6] Although it should be noted that those two and a half years have been marked by the Corona Virus pandemic and a great number of English lessons were conducted online. The school year 2021/22 was their first school year of continuous contact teaching.

acquire a communicative competence that corresponds to the competence levels A1, A2 and partly B1 of the Common European Framework of Reference for Languages (Bundesministerium für Bildung 2018: 58). For this particular group of third grade students, it therefore can be assumed that their English language competence for spoken interaction and spoken production is A1-A2, and for listening, reading and writing A2 (Bundesministerium für Bildung 2018: 60). The Common European Framework of Reference (Council of Europe 2001) defines this level of English competence as being able to

> understand sentences and frequently used expressions related to areas of most immediate relevance […] [and to 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 *[sic]* background, immediate environment and matters in areas of immediate need (Council of Europe 2001: 24).

English learners at this stage of English competence are referred to as "Basic Users" (Council of Europe 2001: 24) or pre-intermediate language users.

Naturally, however, there are differences in the language competence between the participants of this study. These differences in the level of knowledge play a major role in this thesis since it argues – following Vygotsky's theory of the *ZPD* and Dalton's concept of *collective scaffolding* (see section 3.1) – that more knowledgeable peers can support less knowledgeable peers in their learning process. Therefore, students' proficiency level was a key indicator for the grouping of the participants. For this reason, a brief outline of the criteria for assessment and how the students' proficiency level for this study was determined are provided next.

The Austrian school system requires teachers at secondary schools to assign grades to students' performances that adhere to the "Beurteilungsstufen" (rating scale) of the Austrian "Leistungsbeurteilungsverordnung" (Bundesministerium für Bildung 2021). The grade should represent a student's level of knowledge and competence in a subject. Depending on how far students meet the requirements of the curriculum, how well they complete the assigned tasks, and how well they can apply their knowledge to novel tasks independently, they will receive either *Sehr gut* (1), *Gut* (2), *Befriedigend* (3), *Genügend* (4) or *Nicht genügend* (5). For a *Sehr gut*, they need to meet the requirements and fulfil the tasks to an extent that is way above the essentials and should be clearly able to apply their knowledge independently. For

a *Gut*, task fulfilment should be above the essentials, but independent application also needs to be clearly visible. For a *Befriedigend,* tasks need to be fully completed in their essential areas ("in den wesentlichen Bereichen zu Gänze erfüllt", Bundesministerium für Bildung 2021: 10) and lacks in performance can be compensated by showing a noticeable tendency towards independency ("Ansätze zur Eigenständigkeit", Bundesministerium für Bildung 2021: 10). A *Genügend* is assigned if students' performances predominantly fulfil the requirements in their essential areas. If students do not fulfil the requirements of a *Genügend,* then a *Nicht genügend* is assigned. For English these grades can be indicative to how far students have mastered the desired language level with its set learning objectives. This particular group of students received grades on their two *Schularbeiten* (i.e. written tests) during winter semester and a grade for their overall performance[7] during the winter semester in their *Schulnachricht* (report on grades of winter semester).

For the grouping of the students the teacher researcher decided to consider students' grades on their first and second *Schularbeit* as key indicators since it is a more accurate representation of their English competence/knowledge level (as it does not include other factors such as homework or participation in class). The following pie chart (Figure 5) shows the distribution of students' average grade on their *Schularbeiten* (resulting from their first and second English *Schularbeit*). As can be seen from the graph out of 40 students seven achieved the grade *Sehr gut*, eight a *Gut*, 17 a *Befriedigend*, seven a *Genügend* and only one of them a *Nicht genügend*.

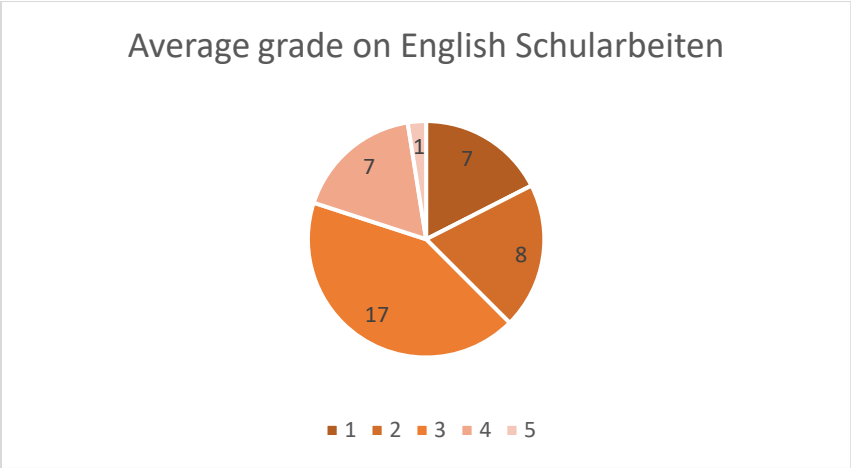

Figure 5 Distribution of students' average grades on English *Schularbeiten* from the first semester

[7] Including their grades on the Schularbeiten, but also their performance on vocabulary & grammar quizzes, fulfilment of homework, work during projects etc.

In order to follow Vygotsky's theory of the ZPD and Dalton's collective scaffolding and to test hypothesis 1b, learners' English language competence had to be classified as *more proficient* (MP) or *less proficient* (LP). To do that, two aspects were closely considered: a) their average grade from their first and second *Schularbeit* and b) their score on the vocabulary knowledge pre-test. Students with a *Sehr gut* or *Gut* were automatically considered as MP and students with a *Genügend* or *Nicht genügend* as LP. For those with a *Befriedigend* the vocabulary knowledge pre-test score was consulted. If their score was above the mean score of 53.35 points (out of 75) then they were classified as MP; if their score was below 53.35 then they were classified as LP. After the proficiency level had been determined, thought was given to the social skills of the learners and their relationship to each other in order to decide who could work with whom most effectively. I observed these patterns already in some of my previous projects with my students and chose the allocation accordingly. The final grouping looks as follows:

Table 2 List of participants

GROUP	PARTICIPANT	GENDER	GRADE	VK-PRETEST SCORE	PROFICIENCY LEVEL
GR_{IW1-20}	IW1	m	2	66	MP
	IW2	f	2	56	MP
	IW3	m	1	55	MP
	IW4	m	2	63	MP
	IW5	m	3	42	LP
	IW6	m	3	50	LP
	IW7	f	3	51	LP
	IW8	m	3	47	LP
	IW9	m	2	55	MP
	IW10	m	1	58	MP
	IW11	f	3	45	LP
	IW12	m	3	65	MP
	IW13	f	3	65	MP
	IW14	f	3	51	LP
	IW15	m	3	43	LP
	IW16	f	2	58	MP
	IW17	m	4	42	LP
	IW18	f	3	53	LP
	IW19	m	4	50	LP
	IW20	m	3	52	LP
GR_{CW1}	CW1	m	4	57	LP
	CW8	m	2	72	MP
GR_{CW2}	CW2	m	1	65	MP
	CW3	m	3	40	LP
GR_{CW3}	CW4	f	1	71	MP
	CW7	f	4	54	LP

GR_{CW4}	CW5	f	4	48	LP
	CW9	f	1	72	MP
GR_{CW5}	CW6	m	4	42	LP
	CW10	m	3	62	MP
GR_{CW6}	CW11	f	4	53	LP
	CW14	f	1	73	MP
GR_{CW7}	CW12	m	2	68	MP
	CW16	m	3	46	LP
GR_{CW8}	CW13	m	1	62	MP
	CW17	m	3	40	LP
GR_{CW9}	CW15	f	2	72	MP
	CW18	f	3	42	LP
GR_{CW10}	CW19	m	5	42	LP
	CW20	f	3	67	MP

The GR_{IW} (*n*= 20) comprises nine *more proficient* students and eleven *less proficient* students. With regards to the GR_{CW} (*n*= 20) it was ensured that each pair consisted of one *more proficient* (*n*= 10) and one *less proficient* student (*n*= 10).

5.5. Materials for data collection

For this research project two different materials were used for data collection. First, the vocabulary pre- and post-test, for which an adaptation of the Vocabulary Knowledge Scale by Paribakth and Wesche (1997) was used. Second, a self-designed writing prompt for students' individual and collaborative text production was provided.

5.5.1. Vocabulary Knowledge Scale

As this study wants to investigate the development of vocabulary knowledge, and as such knowledge is highly multifaceted, an instrument that is able to measure different levels of word knowledge was chosen. Paribakht and Wesche's Vocabulary Knowledge Scale (1997) was considered as a fitting test as it was created with the purpose of "distinguish[ing] stages in learner's developing knowledge of particular words" (Paribakht & Wesche 1997: 179). As figure 6 below shows, this test involves a five-point scale that includes not only self-report items but also performance items with the aim of eliciting self-perceived as well as demonstrated knowledge of particular written words. The scale ranges from "I= I don't remember having seen this word before", i.e. total unfamiliarity of the vocabulary item, to "V= I can use this word in a sentence", implying that the word can be used with grammatical and semantical accuracy. Category II indicates learners' recognition of the word without the

ability to recall the meaning of the word, III indicates the recognition of the word plus having an idea of the word's meaning, and IV shows the self-report that the meaning of the word is known. If learners choose category III, IV, or V they also need to provide evidence for their self-reported vocabulary knowledge. Concretely, III and IV ask them to provide a synonym or translation of the word, V additionally asks them to provide a sentence with the given word.

With regards to Nation's categorization of word knowledge (see section 2.2), it can be said that with categories I to IV learners' receptive knowledge of the meaning of words is assessed. Category V also takes their productive word knowledge into account, namely their knowledge of grammatical functions and collocations.

Self-report categories	
I	I don't remember having seen this word before.
II	I have seen this word before, but I don't know what it means.
III	I have seen this word before, and I think it means _____. (synonym or translation)
IV	I know this word. It means _____. (synonym or translation)
V	I can use this word in a sentence: _____. (Write a sentence.) *(If you do this section, please also do Section IV.)*

Figure 6 Vocabulary Knowledge Scale (Paribakht & Wesche 1997: 180)

At this point the instrument's quality criteria and limitations should be mentioned. In 1996 Wesche and Paribakht (qtd in Milton 2009: 160) conducted a pilot study with 95 participants to determine the validity and reliability of their VKS. In this study they found out that the self-report scores correlated highly with actual evidence given by the participants ($r > 0.9$). Two weeks later a test-re-test reliability check resulted in a reliability of above 0.8 and lastly the analysis of concurrent validity with the Eurocentres Vocabulary Size test revealed high correlations of 0.5 as well (Wesche & Paribakht 1996, qtd in Milton 2009: 160).

Despite the instrument's good validity and reliability, it bears some obvious limitations as well. While it aims at measuring different levels of word knowledge, it still is not able to measure certain aspects of word knowledge, such as multiple word meaning, which was a major

critique voiced by Read (2000: 136) and Milton (2009: 160). Furthermore, productive knowledge of vocabulary cannot be measured fully as learners are asked to provide one sentence with the given word only. From such an isolated sentence it might be difficult to identify if the learner has actually understood the item's meaning or not (Read 2000: 137). Lastly it was criticised that the level of measurement is rather coarse and only takes the basic stages of the process of vocabulary learning into account (Meara 1996: 6). Yet, Paribakht and Wesche (1997: 179) themselves stated that it was their intention that the VKS should measure the skill to use words in initial contextualized production and not those abilities that go beyond it, e.g. multiple word meanings. In order to overcome this weakness of the VKS, students' composed texts were analysed as well in order to see whether students were able to use the words in a broader context as well. Nevertheless, it should be born in mind that the scores of this project's VKS only represent partial aspects of receptive and productive vocabulary knowledge.

For this particular research project two small adaptions to the VKS were made. First, learners were only asked to provide a translation and not given the option to provide a synonym. This adaption was made because this group of learners have been trained to study vocabulary by means of translating them from English into German and vice versa since their first year of English at this Gymnasium. Furthermore, the selected vocabulary items were discussed and translated into German during the English lessons (as described in section 5.2). Thus, students were familiar with this format of testing vocabulary. Secondly, the layout of the VKS was changed in order to facilitate learners' rating of the items. Additionally, category 3 and 4 were highlighted in yellow as well as the column "translation" in order to remind students to give the German word if they have ticked 3 or 4. The wording of category 5 was also changed to "I know that this word means ____ (= translation) AND I can use this word in a sentence" instead of "If you do this section, please also do Section IV", to avoid any confusion. The colour green should remind them to write a sentence in the column "sentence". Two example answers (00 and 0) were also given to illustrate the procedure. The final version of the VKS is presented in figure 7 below. Since this instrument was implemented in previous vocabulary quizzes, students were already familiar with this kind of test, and a smooth completion of the test was ensured.

		I don't remember having seen this word before.	I have seen this word before but I don't know what it means.	I have seen this word before and I **think** it means ____ (= translation).	I know this word. It means ____ (= translation).	I know that this word means ____ (=translation) **AND** I can use this word in a sentence.		**Vocabulary Knowledge Quiz I** **Name:** _____ **Class:** _____ **Date:** _____	**points**
No	**Word/Phrase**	**1**	**2**	**3**	**4**	**5**	**translation**	**sentence**	
00	dog					X	Hund	My dog is called Rocky.	
0	to stroll		X						
1	childish								
2	get on well with someone								
3	be honest with someone								
4	break up with someone								
5	sulk								
6	have an argument with someone								
7	to be kind								
8	jealous								
9	solve one's problems								
10	disappointed								
11	fall out with someone								
12	introduce								
13	make fun of someone								
14	childhood								
15	make up with someone								

Figure 7 Final version of VKS used in this study

This adapted version of the VKS included 15 vocabulary items. Most of these items were selected based on the tasks that were covered during the English lessons of the learning unit "You've got a friend" in the MORE! 3 books (see section 5.2). The only item that was not explicitly dealt with is number 8 "jealous". In this case the teacher researcher wanted to know how many of her students know this word anyhow. Like in the original VKS learners were asked to rate their knowledge about the English words/phrases and to give a German translation and English example sentence if applicable. For instance, with item 1 *childish*, one student ticked "5" in the self-report category and then filled in the German translation *kindisch* and provided the sentence *My brother is childish*. Such an answer for item 1 would obtain 5 out of 5 points in the scoring which is explained in more detail in the next section.

5.5.2. Writing Task

The writing prompt, which is shown in figure 8, aimed at eliciting most of the vocabulary items from the VKS. However, this time students' productive English knowledge should be tested, i.e. their ability to recall the English form for the meaning they want to express and their ability to use these words accurately in context. Therefore, they were provided with a prompt that consisted of German bullet points which students should include in their composition. Bullet point *verstehen sich sehr gut*, for example, should elicit the VKS item 2 *get on well with someone* and should be preferably used in this context by saying *They get on well with each other*.[8]

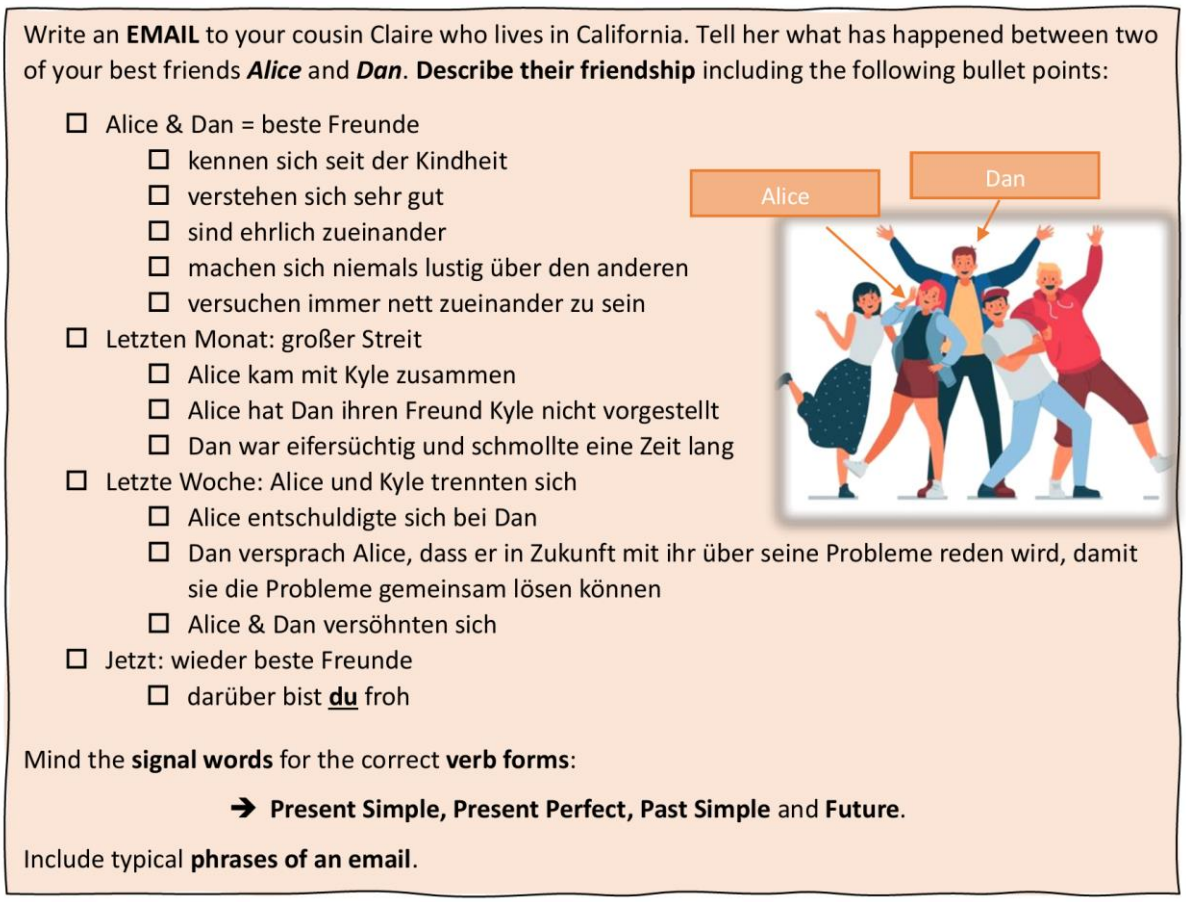

Figure 8 Writing prompt

Previous studies in this field used various writing tasks for data collection, such as Jigsaws, dictogloss, text reconstructions (Alegría de la Colina & García Mayo 2007), picture stories (Fernández Dobao 2012), data commentaries and argumentative compositions (Storch & Wigglesworth 2007, 2009), and essays (Watanabe & Swain 2007, 2008). Most of these studies

[8] This example illustrates that by using the phrase, *someone* needs to be substituted by the reciprocal pronoun *each other*. During the correction process, the teacher found out that many students struggled with this grammatical structure, which is why she reviewed reciprocal pronouns in one of the follow up lessons.

also used a pre-post-test research design in combination with analysing students' LREs and thereby obtained mixed results. One reason for those mixed results can be that it could not be ensured that learners really focused on the target items during their collaborative dialogue or individual think-aloud protocols. As a consequence, what they have covered during their deliberations might not correspond to the post-test items and thus might have not improved in post-test scores (Storch 2013: 80). Therefore, Storch argues that a "pre-post test research design is problematic […] unless we consider data of pair talk and individual think-aloud protocols" (2013: 80).

Yet, this study does not include an investigation of LREs since such an additional analysis (including recording students, transcribing the recording, coding the dialogues, etc.) would have gone far beyond the scope of this master thesis. Therefore, to tackle the aforementioned problem, the I decided to design a writing task that is tailored to the target vocabulary from the pre- and post-test. I found that the most direct way to elicit the English phrases was to provide the German translations. The prompt asked students to use the bullet points in their text – adhering to Corson's (1985) principle of creating a desire and opportunity to use words (see section 2.5). Therefore, if students did not use the target phrase, it could be assumed that they most likely did not know the target vocabulary. Therefore, the criterion *target word fulfilment* was created to investigate students' productive knowledge of the target words.

Furthermore, inspired by previous research (Storch 2005; Storch & Wigglesworth 2007; Fernández Dobao 2012) the texts were also analysed with respect to the criteria of fluency, accuracy and lexical diversity.

5.6. Data preparation and scoring

This section aims to clarify the scoring procedure of the vocabulary knowledge tests as well as of the texts produced by the students. All the students of this study who were present at the two dates of data collection participated responsibly and fulfilled all the requirements, which is why all tests and texts could be used for data analysis and none had to be excluded.

5.6.1. Scoring of VKS

First, the students' answers of the VKS-tests were assessed according to Paribakht and Wesche's scoring system (1997: 180-181), which is illustrated in Figure 9 below.

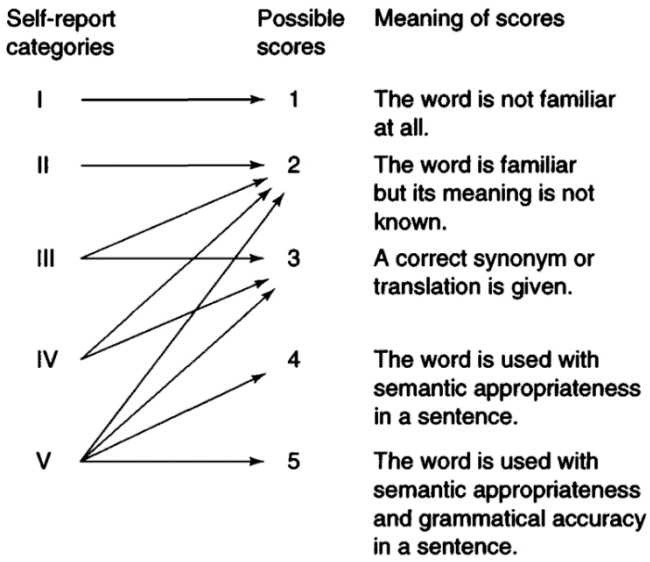

Figure 9 Paribakht & Wesche's VKS scoring system (1997: 181)

If students crossed category I their score was 1, indicating that the word is not familiar at all. A cross in category II resulted in a score of 2 which means that although the word is familiar, its meaning is not. With category III, IV and V multiple scores were possible according to the correctness of their demonstrated knowledge. For example, if the translation learners gave was incorrect, they only received a score of 2. The following excerpt from a VKS demonstrates this: For item 7 "to be kind" the student crossed category V, however, he provided an incorrect translation, which is why he only obtained 2 points for this item. On the other hand, another student translated this phrase correctly and therefore received 3 points.

Figure 10 Excerpt from VKS – incorrect translation of item 7

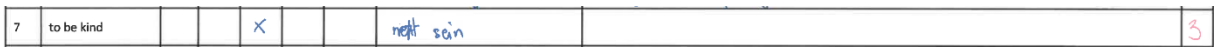

Figure 11 Excerpt from VKS – correct translation of item 7

If, for category III and IV, they were able to provide a correct translation then they received a score of 3 (e.g. to be kind = freundlich/ nett sein). A score of 4 was assigned if students indicated category V and the word was used meaningfully in a sentence – thus the given sentence reflected the learner's knowledge of the word's meaning – but mistakes on the level of grammatical accuracy or collocations happened. If learners managed to use the word with

semantic and grammatical accuracy in the sentence, then a score of 5 was assigned. Mistakes in other parts of the sentences not directly related to the target word had no bearing on the score. For each of these possibilities an example is provided below:

a. Item 11 *fall out with someone*: I falled out with Lisa. → 3
b. Item 3 *be honest with someone*: My friend is honest to me. → 3
c. Item 6 *have an argument with someone*: I have a argument with my mom all the time. → 4
d. Item 5 *sulk* → Lusi sulks every time we argument. → 5

Example sentence a. only reached a score of 3 since the wrong past tense form of the verb *fall* was used, which is grammatically incorrect. Example sentence b. got a score of 3 because an incorrect preposition (*to* instead of *with*) for the phrasal verb *be honest with* was used, which was categorized as a mistake on the level of collocation. The mistake that happened in example c. (*a argument* instead of *an argument*) is decisive for the score as the grammatical aspect of using the indefinite article *an* before a noun starting with a vowel sound was disregarded. The error that appeared in example d., however, has no influence on the score since the sentence part *we argument* has no direct relation to the correct use of the verb *to sulk*.

Once scores were assigned to each item of the VKS-test on paper, the teacher researcher manually transferred them into an excel file, where the total score of the test was calculated. This was an important preparational step for subsequent statistical undertakings to test hypotheses 1a and 1b:

- H1a: CW is more beneficial to learners' development of word knowledge than individual writing (IW).
- H1b: Less proficient learners improve their vocabulary knowledge through CW with a more proficient peer.

5.6.2. Scoring of written texts

As this set of data involved multiple scoring procedures, each of them will be explained and exemplified by means of extracts from a student's data set (CW5 belonging to GR_{CW4}). The complete analysis will be provided only in the subsequent section.

As already described in section 5.5.2, students were asked to compose a text following the given prompt (see figure 8). In total 30 texts (20 text of GR_{IW} and 10 texts of GR_{CW}) had to be prepared for analysis. The analyses of the texts focused on four criteria: fluency, lexical diversity, accuracy and target word fulfilment, which correspond to hypotheses 2a and 2b.

- H2a: Collaboratively written texts show greater fluency, higher lexical diversity and accuracy than individually written texts.
- H2b: Collaboratively written texts include more target words than individually written texts.

The following table 3 allows for an overview of how each criterion has been operationalized through which variable:

Table 3 Criteria for text analyses

Criterion	Fluency	Lexical diversity	Accuracy	Target word fulfilment
variables	No. of words per text	Type/token ratio	No. of mistakes per 100 words	No. of target vocabulary used correctly
	No. of sentences per text	vocd-D		No. of target vocabulary used incorrectly
	No. of words per sentence	MTLD		Total No. of target vocabulary used

While previous research (Storch 2005; Storch & Wigglesworth 2007; Fernández Dobao 2012) that investigated written texts against the backdrop of collaborative writing mainly measured the texts' fluency (length of a text), accuracy (mistakes or errors of a text) and complexity[9] (T-units and clauses of a text), this study omits complexity due to the great amount of time it would have consumed to manually code 30 texts with regards to T-units and clauses. Instead, this study includes the dimension of lexical diversity, that is "the range of different words used in a text, with a greater range indicating a higher diversity" (McCarthy & Jarvis 2010: 381).

Before going into detail with data preparation of the hypothesis at hand, a short excursion into the field of lexical diversity is provided.

Lexical diversity measures have been used in a wide range of research fields, such as stylistics, data mining, forensics and of course language acquisition. In applied linguistics it was found

[9] Complexity was measured by identifying clauses and T-units and subsequently calculating the number of words per clause and T-unit as well as number of clauses per T-unit (Fernández Dobao 2012: 47).

that lexical diversity indices can be indicative of writing quality, speaker competence and vocabulary knowledge. Yet, identifying a robust index to represent lexical diversity has posed a challenge due to its measure being sensitive to variations in text length. The best-known lexical diversity index is the type-token ratio; however, it is also the index that has a particularly strong sensitivity to text length (McCarthy & Jarvis 2010: 381). When interpreting a type-token ratio, one needs to be aware of its underlying composition. Namely, while the number of tokens (words) increases when a text becomes longer, the number of types (different words) gradually slows down. As a result, texts that are longer automatically decrease in the calculated value of lexical diversity and increase in the rate of lexical repetition. However, these rates do not necessarily reflect the real lexical diversity of a longer text and therefore researchers have tried to improve the underlying mathematical model for lexical diversity, resulting in *vocd-D* and *MTLD* (McCarthy & Jarvis 2010: 382).

Vocd-D is a measuring device that calculates the value D three times in a series of random text samplings of one text and thereby provides an average value as output. Its output values usually are located between 10 and 100, with higher values suggesting larger lexical diversity (McCarthy & Jarvis 2010: 381-383). Additionally, Duran et al. (2004) applied vocd-D to various cohorts in Bristol and provided a summary of the means and sub-ranges for selected ages and adult writing. Their graph (see Figure 12) can serve as reference point to give meaning to the value of *D (*Duran et al. 2004: 237-238).

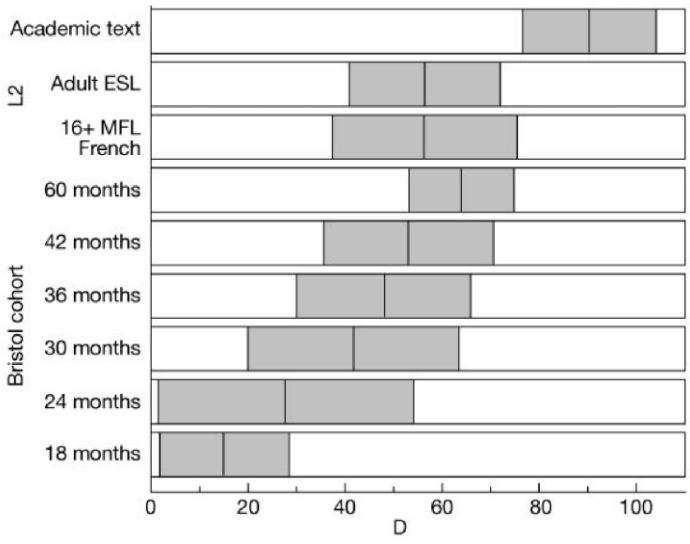

Figure 12 Summary of means and sub-ranges of D for various cohorts (Duran et al. 2004: 238)

McCarthy and Jarvis (2007) found that the vocd-D, however, also has two deficits, namely that it varies as a function of text length and that vocd-D replicates the hypergeometric distribution function[10]. Therefore, they have investigated the validity of another model for the calculation of lexical diversity: MTLD (Measure of textual lexical diversity) (McCarthy & Jarvis 2010). Their study showed that MTLD does not correlate with flawed lexical diversity approaches such as the type-token ratio. Internal validity was confirmed by providing consistent results regardless of text length. Despite implications of MTLD being a powerful index, the authors advise to consider more than just one single index when assessing lexical diversity (McCarthy & Jarvis 2010: 390-391). For this reason, this study includes values from type-token ratio, vocd-D and MTLD for the analysis of lexical diversity.

Data preparation for hypothesis 2a took place in two ways: First, each digitally written text was entered into the online text analysing tool *Text Inspector* to obtain necessary figures to measure fluency and lexical diversity. Second, each text was corrected manually by the teacher researcher in Microsoft Word, which served as basis for the scoring of accuracy.

After entering the text into *Text Inspector* a multitude of figures were provided as output. Those relevant for fluency and lexical diversity – namely word count, sentence count, average number of words per sentence, types, tokens, type-token ratio, vocd-D and MTLD were collected in an excel spreadsheet. Attention had to be drawn to the generated number of sentences since occasionally two sentences were mistaken as one. Therefore, the number of sentences was double-checked and manually amended if necessary.

After the word documents had been corrected, their mistakes were counted. However, not all mistakes were counted towards the scoring. The mistakes that were not counted include punctuation, lexical choice if it did not hinder understanding, and repetition of the exact same mistake. Some of the previous studies did not include spelling mistakes either, however, for this study orthography was regarded as an important indication for productive vocabulary knowledge, which is why spelling mistakes were counted. To make the number of mistakes comparable, the number of mistakes per 100 words was calculated. A smaller number of mistakes represents a greater accuracy of the text. The figure below is a screenshot of the correction of GR_{CW4}'s text. The corrections which are framed are the ones that were counted.

[10] which is a discrete probability distribution in statistics/ probability theory. For a discussion of hypergeometric distribution see Wu 1993.

Hi Claire,

how are you? We want to tell you about the friendship of our best friends Alice and Dan.

They have been friends since their childhood. In school Alice and Dan got to know each other very well. Both made a promise. They won't lie to each ~~other, when~~ other when there's a problem. Alice and Dan promised~~,~~ that they will never make fun of each other. Dan is always kind to Alice and appreciates her very much.

But last month they had an argument. The reason was that Alice had a relationship with Kyle. Dan wasn't happy about it, because Alice didn't introduce Kyle to Dan. Dan was so jealous and sulked for a long time. Dan fell out with Alice, so they haven't heard from each other for a long time.

Last week Alice and Kyle broke up. She decided to go to Dan and apologize to him. Dan said to Alice~~:~~ "I wasn't a good friend and I'm so sorry about it, but in the future, I will talk to you ~~speak~~ about my problems ~~with you~~, so we can solve them together." They saved their friendship and made up.

Now they are best friends again and spend a lot of ~~very much~~ time together.

That was the long story about how our ~~my~~ friends saved their friendship. We're so glad ~~and couldn't believe~~ that everything is alright. Theresa and I are really looking forward to meeting you in California.

We hope you're well.

Lots of love,

Figure 13 Correction of GR_{CW4}'s text

In conclusion, the text at hand achieved the following scores:

Table 4 Scores of GRCW4's text

Text Inspector								Manually in Word	
Fluency			Lexical diversity					Accuracy	
words	sentences	Words/ sentence	Tokens	Types	Type-token-ratio	vocd-D	MTLD	Mistakes	Mistakes/ 100 words
244	*23*	*10.6*	*244*	*122*	*0.50*	*76.21*	*75.16*	*10*	*4.10*

To test Hypothesis 2b – collaboratively written texts include more target words than individually written texts – a tailored writing prompt was devised to elicit the newly learnt target vocabulary that was also quizzed in the vocabulary knowledge test. As has been briefly explained in section 5.5.2, the written products were analysed with regards to the criterion *target word fulfilment*, meaning counting how many of the target words were used correctly, incorrectly, and used overall. The difference between these three categories is described in Table 5 below.

Table 5 Target word fulfilment

Distinction of target vocabulary use	Definition
Correctly used	Used meaningfully with lexical and grammatical accuracy
Incorrectly used	Used with grammatical or lexical inaccuracy (e.g. wrong tense of verb or wrong preposition used with phrasal verb)
Not used	Target word/phrase not used
Used overall	Total number of correctly and incorrectly used vocabulary

The text above, for example, used nine target phrases *correctly*, one *incorrectly,* and two *not* at all, resulting in ten target phrases *used overall* out of twelve. For the first bullet point, they used "They have been friends since their childhood" (GR_{CW4}). Thereby they used the target vocabulary *childhood* in a meaningful and accurate way, which is why it was counted as correctly used target word. The bullet point *schmollte eine Zeit lang* elicited the target vocabulary, but it was not used grammatically correctly as the past-tense suffix *-ed* was missing: "Dan […] sulk__ for a long time" (GR_{CW4}). Therefore, this item was counted as incorrectly used. Lastly, instead of using *get on well with each other* for bullet point *verstehen sich sehr gut*, the pair used "got to know each other very well" (GR_{CW4}), which differs in meaning to the target vocabulary and therefore was considered as not used target vocabulary. Table 6 illustrates the target word fulfillment for each bullet point.

Table 6 Target word fulfilment

Bullet point eliciting target vocabulary	Correctly used	Incorrectly used	Not used
seit der Kindheit	x		
verstehen sich sehr gut			x
sind ehrlich zu einander			x
machen sich niemals lustig über den anderen	x		
versuchen immer nett zueinander zu sein	x		
großer Streit	x		
Alice hat Dan ihren Freund Kyle nicht vorgestellt	x		
Dan war eifersüchtig	x		
und schmollte eine Zeit lang		x	
Alice und Kyle trennten sich	x		
damit sie die Probleme gemeinsam lösen können	x		
Alice & Dan versöhnten sich	x		
No of target words used (in-)correctly	**9**	**1**	**2**
Total no of target words used overall	**10**		

In order to prove hypotheses 3a and 3b – using target words in written texts has a positive effect on the development of vocabulary knowledge, and producing a more accurate text has a positive effect on the development of vocabulary knowledge – no additional data had to be prepared. The scores of the vocabulary *post-test* as well as the scores of the categories *target word fulfilment* and *accuracy* were used for the relevant statistical computations.

6. Data analysis and results

This section will present the quantitative analyses conducted with respect to the six hypotheses 1a, 1b, 2a, 2b, 3a and 3b. For each hypothesis, the descriptive and inferential statistics will be provided together with a brief explanation for the chosen statistical computations. For all statistical tests the open-source program JASP has been used. But first, a short introduction into statistical models and terms shall be given.

6.1. Basics of statistics

Descriptive statistics involve for example frequency counts, measures of central tendency (e.g. mean values) and deviance (e.g. standard deviation). These statistics however only provide information for the participants of that exact sample group (Field 2018: 22-30). In order to test hypothesis and to confirm or reject predictions for a wider population, inferential statistics need to be applied. The most commonly used approach to test research questions is the *null hypothesis significance testing*. The null hypothesis states that an effect is absent, and it represents a baseline against which scientists can evaluate how plausible their alternative hypothesis is (Field 2018: 73). In this study, the H0 of my alternative hypothesis 1a would be that CW has no effect on learners' development of word knowledge and that no difference between CW and IW can be observed. In order to gain confidence that the alternative hypothesis is true (e.g. CW is more beneficial to learners' development of word knowledge than IW), the statistical probability that the effects are random should fall below the threshold of 5%, which is represented through the value $p < 0.05$ (Field 2018: 72-74). Thus, if the probability is smaller than 0.05, it is to be said that "the model explains a sufficient amount of variation to reflect a genuine effect in the real world" (Field 2018: 79). As a result, the test statistic is regarded as statistically significant, allowing scientists to reject their null hypothesis and to assume that their alternative hypothesis is very likely to be true (Field 2018: 79). It is

important to note that the sample size has a big effect on whether differences between groups turn out to be significant or not. Generally, large sample sizes have greater power to detect effects, which is why in large samples even tiny differences can be significant. On the other hand, this also means that in rather small samples even apparently large differences turn out to be non-significant (Field 2018: 89-90). This should be borne in mind when interpreting the results of this study.

6.2. First hypotheses

As explained in detail in the previous section, hypothesis 1a and 1b draw on data from the pre- and post-test of the Vocabulary Knowledge Scale (VKS). As there were two measurement time points (pre-test and post-test) and two groups (GR_{IW} and GR_{CW}) the Repeated Measures ANOVA (analysis of variance) was selected as statistical computation. The following distribution plots in figure 14 were created to prove whether normal distribution was given.

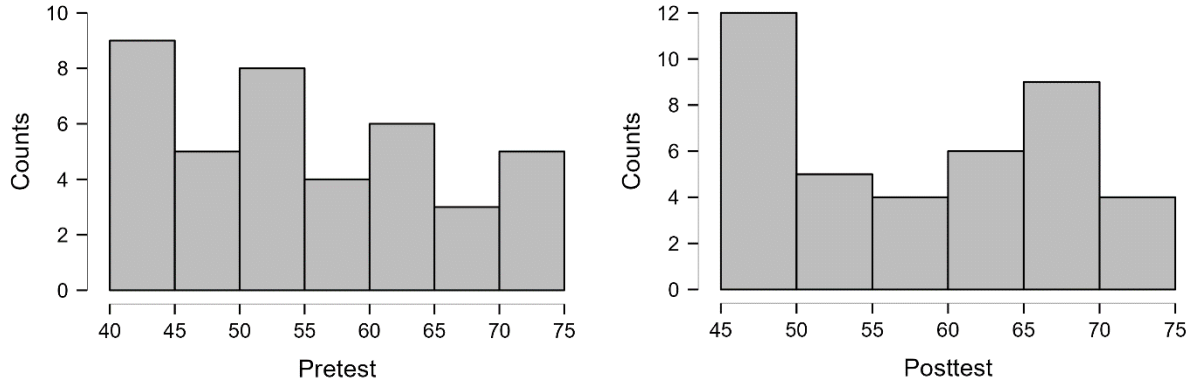

Figure 14 Distribution plots of VKS pre-test and post-test

The vertical axis shows how often each value from the horizontal axis occurred in the data set. In the case of normal distribution, the bars should indicate a bell-shaped curve. The bell shape is treated as an indication that most scores lie around the centre of distribution (Field 2018: 22). As neither the data from the pre-test nor from the post-test suggest a bell-shape curve, no normal distribution can be assumed. Normally, the lack of normality would imply that scientists should resort to a non-parametric test instead of a parametric test, for example using a Mann-Whitney-U test instead of a t-test. In the case of an ANOVA, however, there is no comparable, established non-parametric test. Fortunately, studies such as Blanca et al.'s (2017: 556) found that the ANOVA was a robust test even if non-normal distributions were given. Therefore, the repeated measures ANOVA was used for testing hypotheses 1a and 1b.

6.2.1. Hypothesis 1a

On a descriptive level, it can be observed that both groups' VKS scores have increased in the post-test (see Table 7). The mean score of GR_{IW} in the pre-test was 53.35 (*SD*= 7.58) and in the post-test 56.05 (*SD*= 8.04), which results in a learning gain of 2.7. The mean score of GR_{CW} in the pre-test equalled 57.40 (*SD*= 12.44) and in the post-test 61.05 (*SD*= 10.13), resulting in a learning gain of 3.65.

Table 7 Descriptives of pre- and post-test scores of both groups

RM Factor 1	Group	Mean	SD	N
Pre-test	CW	57.400	12.437	20
	IW	53.350	7.576	20
Post-test	CW	61.050	10.133	20
	IW	56.050	8.036	20

Thus, there has been a slightly greater increase in the post-test score of GR_{CW} than of GR_{IW}, as the following line graph in figure 15 depicts.

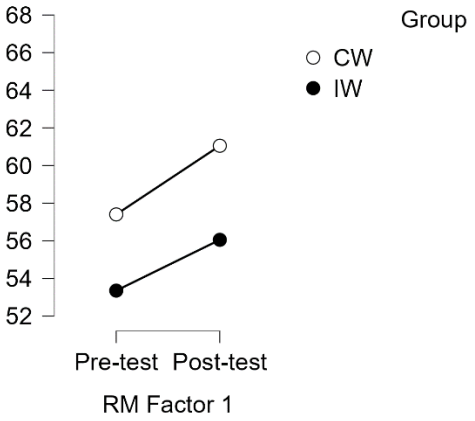

Figure 15 Pre-test and post-test mean scores of GRIW and GRCW

The Repeated Measures ANOVA helped to find out whether the effect of time (pre-test and post-test) is significant and whether the effect of interaction between time and groups (GR_{IW} and GR_{CW}) was significant. The test revealed that there is a significant main effect of time, $F(1, 38)= 22.44$, $p < .001$. Thus, both groups could significantly improve their vocabulary after having written the texts. However, the test did not show any significant effect on the interaction between time and groups $F(1, 38)= 0.50$, $p= .483$, which means that the difference in vocabulary learning gain between GR_{IW} and GR_{CW} is not statistically significant. Therefore,

hypothesis 1a, CW is more beneficial to learners' development of word knowledge than individual writing, cannot be confirmed with the given number of participants (*N*= 40).

6.2.2. Hypothesis 1b

For the analysis of hypothesis 1b, a subset of the pre- and post-test scores were included, namely the scores of the less proficient learners in GR_{IW} (*n*= 11) and of the less proficient learners in the CW group (*n*= 10) as is shown in table 8.

Table 8 Descriptives of pre- and post-test scores of less proficient learners of both groups

RM Factor 1	Group	Mean	SD	N
Pre-test	CW	46.400	6.293	10
	IW	47.818	4.167	11
Post-test	CW	52.500	6.884	10
	IW	50.545	4.719	11

As the graph in figure 16 illustrates the mean scores of both groups have increased in the post-test, however, the LP learners of GR_{CW} achieved a higher score (*M*=52.50, *SD*= 6.88) than the LP learners of GR_{IW} (*M*= 50.55, *SD*= 4.72). Thereby LP learners of GR_{CW} were able to improve by 6.10 whereas LP learners of GR_{IW} could only improve by 2.73. As a result, the learning gain in GR_{CW} was 3.37 greater than in GR_{IW}.

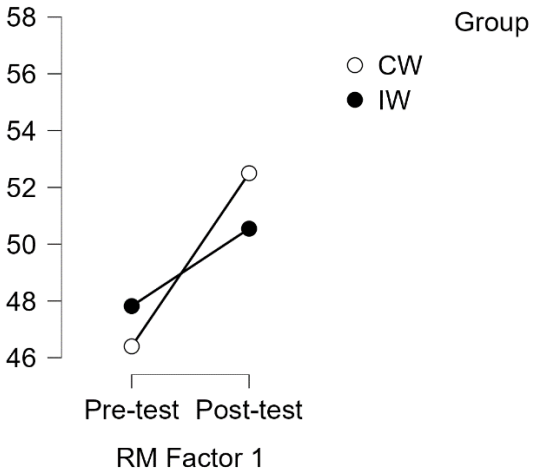

Figure 16 Pre- and post-test scores of LP learners of GRIW and GRCW

The repeated measures ANOVA showed that the LP learners of both groups could significantly improve their vocabulary knowledge from the first measurement time point to the second, $F(1, 19)= 29.26$, $p < .001$. The difference between the groups was slightly not significant, which

is why no significant effect in the interaction between measurement time points and groups could be found, $F(1, 19)= 4.27$, $p= .053$. This is likely due to the even smaller sample size ($n= 21$). With regards to hypothesis 1b this means that it can be confirmed that less proficient learners improve their vocabulary when writing a text with a more proficient peer. Yet, less proficient learners who wrote the text on their own improved significantly as well. While there is a tendency for collaborative writing being more beneficial for less proficient learners, no statistical evidence thereof could be found.

6.3. Second hypotheses

In order to investigate hypotheses 2a and 2b, data extracted from the texts of GR_{IW} ($n= 20$) and GR_{CW} ($n= 10$) were analysed by means of independent samples t-tests. For a t-test to be applied, two assumptions must not be violated: the data sets need to be non-significant in the test of normal distribution (Shapiro-Wilk test) as well as in the test of homogeneity of variances (Levene's test). If the data appears to be significant ($p < 0.05$) in these tests, normality and equality of variance cannot be assumed, and it is thus advised to use the non-parametric equivalent, namely the Mann-Whitney-U-test (Field 2018: 248). As clarified in section 5.6.2, the data set for H2a comprised the categories *fluency* (expressed through the variables *number of words*, *number of sentences* and *words per sentence*), *lexical diversity* (i.e. the variables *type-token ratio, vocd-D* and *MTLD*), and *accuracy* (operationalised through the variable of *number of mistakes per 100 words*). For H2b the data set included the category *target words fulfilment* with the four dependent *variables correctly used, incorrectly used,* and *used* target words. For each of these dependent variables a Shapiro-Wilk and Levene's test was conducted. All but two variables proved to be non-significant. The only variables that deviated from normality was the *number of words* ($p_{CW}= .022$; $p_{IW}< .001$) and *incorrectly used* target vocabulary ($p_{IW}= .002$). Therefore, for these variables the Mann-Whitney-U-test was applied additionally to the t-test. Generally, the following presentation of test results should be treated with caution due to the considerable difference in sample size of the two groups (20 versus 10 texts).

6.3.1. Hypothesis 2a

The descriptive statistics in table 9 shows that collaboratively written texts include more words ($M_{CW}= 162.90$, $SD_{CW}= 34.57$; $M_{IW}= 152.60$, $SD_{IW}= 40.38$), and more sentences ($M_{CW}= 15.90$, $SD_{CW}=3.32$; $M_{IW}= 13.85$, $SD_{IW}= 2.54$) than individually written texts. Interestingly,

individually written texts tend to have longer sentences, that is a larger number of words per sentence (M_{CW}= 10.30, SD_{CW}= 0.88; M_{IW}= 11.04, SD_{IW}= 1.82). With respect to lexical diversity, texts composed by GR_{CW} have slightly higher type-token ratio scores (M_{CW}= 0.58, SD_{CW}= 0.05; M_{IW}= 0.57, SD_{IW}= 0.05), vocd-D scores (M_{CW}= 78.96, SD_{CW}= 18.85; M_{IW}= 71.32, SD_{IW}= 11.70) and MTLD scores (M_{CW}= 68.13, SD_{CW}= 16.04; M_{IW}= 65.07, SD_{IW}= 15.68) than texts composed by GR_{IW}. The most striking difference between the groups was to be found in the category *accuracy*. IW texts included double the number of mistakes per 100 words than CW texts (M_{CW}= 4.43, SD_{CW}= 2.50; M_{IW}= 8.21, SD_{IW}= 3.28).

Table 9 Descriptives of variables of GRCW and GRIW's texts

	Group	N	Mean	SD	SE
Words	CW	10	162.900	34.565	10.931
	IW	20	152.600	40.376	9.028
Sentences	CW	10	15.900	3.315	1.048
	IW	20	13.850	2.540	0.568
Words per sentence	CW	10	10.300	0.882	0.279
	IW	20	11.043	1.816	0.406
Type-token ratio	CW	10	0.584	0.052	0.016
	IW	20	0.579	0.051	0.011
VOCD	CW	10	78.955	18.749	5.929
	IW	20	71.319	11.704	2.617
MTLD	CW	10	68.126	16.044	5.074
	IW	20	65.070	15.680	3.506
Mistakes per 100 words	CW	10	4.426*	2.498	0.790
	IW	20	8.213*	3.277	0.733

Note: * $p < .007$

To test whether these differences are statistically significant, independent samples t-tests were conducted as well as an independent-samples Mann-Whitney U test for the variable *number of words.* Since the same group of participants was used for each of these seven t-tests the multiple comparisons problem occurred. To counteract this problem, the Bonferroni correction method was applied. Thereby the level of significance was adjusted from the typical $p < 0.05$ to $p < 0.007$. Results with a p value lower than 0.007 were considered to be significant. The computation revealed that none of the variables from the table above include a significant difference between the two groups, except for one variable – the variable *mistakes*. The raincloud plot in figure 17 below illustrates the distribution of the data.

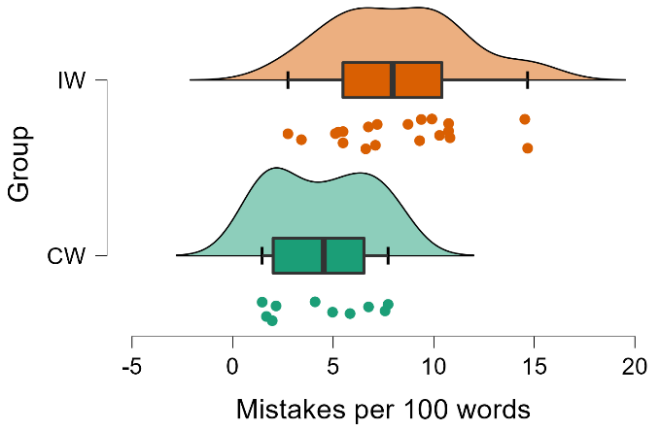

Figure 17 Raincloud plots of number of mistakes per 100 words in GRIW and GRCW

Whereas the number of mistakes per 100 words in IW texts range from around 3 to 15 mistakes with the majority ranging from 5 to 11 mistakes, the number of mistakes in CW texts only range from around 2 to 8. This difference in the range of number of mistakes is also visible in the standard deviation of the mean values M_{CW}= 4.43, SD_{CW}= 2.50, M_{IW}= 8.21, SD_{IW}= 3.28. Statistically it could be proven that collaboratively written texts include significantly fewer mistakes than individually written texts, $t(28)$= -3.21, p= .003. Hence, it can be assumed that CW leads to a greater accuracy in written products. Yet, as the chart clearly shows, fewer data was obtained from GR_{CW} than from GR_{IW} which can be a reason for the smaller range in number of mistakes of GR_{CW}'s texts. To obtain more solid findings, future research should aim at a greater sample size and an equal amount of data in both groups. In conclusion, most parts of hypothesis 2a cannot be confirmed since there is no statistical evidence for the method of CW to lead to a higher *fluency* and *lexical diversity*. However, there is a tendency for CW to lead to higher *accuracy* of written texts.

6.3.2. Hypothesis 2b

The following table 10 shows the mean values of the variables from the criterion of *the target word fulfilment*: target words *correctly used, incorrectly used,* and *used overall* in texts produced by GR_{CW} and GR_{IW}. As multiple independent samples t-tests were conducted, the Bonferroni correction method was applied once more, resulting in a significance level of $p < 0.017$. Significant differences could be found for the variables *target words correctly used* and *used overall,* however the difference in the variable *target vocabulary used incorrectly* did not prove to be significant. With a mean of 7.20 (SD= 1.87) GR_{CW} used 3.25 more target words correctly than GR_{IW} (M= 3.95, SD 1.91), which represents a significant difference between the

groups, $t(28)= 4.43$, $p < .001$. For the variable *incorrectly used target words* a t-test as well as a Mann-Whitney U test was conducted (due to the lack of normality). Neither the t-test nor the U-test could provide statistical significance for the difference in the number of incorrectly used target words between the two groups, $t(28)= 0.09$, $p = .926$, $U(n_{CW}= 10, n_{IW}= 20)= 95$, $p= .0839$. With regards to the number of *target words used overall*[11], GR_{CW} was able to use approximately 3.3 more target words overall than GR_{IW} ($M_{CW}= 8.90$, $SD_{CW}= 1.60$, $M_{IW}= 5.60$, $SD_{IW}= 1.90$), proving to be a significant difference, $t(28)= 4.71$, $p < .001$.

Table 10 Descriptives of target vocabulary correctly used, incorrectly used and used overall of both groups

	Group	N	Mean	SD	SE
Correctly used	CW	10	7.200*	1.874	0.593
	IW	20	3.950*	1.905	0.426
Incorrectly used	CW	10	1.700	1.636	0.517
	IW	20	1.650	1.226	0.274
Used overall	CW	10	8.900*	1.595	0.504
	IW	20	5.600*	1.903	0.426

* $p < .017$

Consequently, hypothesis 2b can be confirmed: Collaboratively written texts include more target words overall than individually written texts.

6.4. Third Hypotheses

Hypotheses 3a and 3b investigated the correlation between target word fulfilment as well as accuracy and students' scores on their vocabulary knowledge post-test. For interpreting the correlation coefficient *r*, three different effect sizes are conventionally distinguished: small, medium and large effect sizes. An *r* equalling .10 is regarded as a small effect, .30 as a medium effect and .50 as a large effect (Cohen 2013: 79-80). Since the data of the post-test violated the assumptions of normality, a non-parametric test had to be chosen. In the case of calculating a correlation coefficient, this means that Spearman's correlation coefficient – also called Spearman's rho – should be used (Field 2018: 351).

[11] which is the sum of correctly and incorrectly used target words

6.4.1. Hypothesis 3a

The results of Spearman's correlations between the variables *post-test scores*, *correctly used* and *target words used overall* are presented in table 11 below. The test revealed a significantly positive, large correlation (r= 0.553, p < .01) between the variables *correctly used* target vocabulary and *post-test* scores. The correlation between *target words used overall* and *post-test* turned out to be highly significantly positive and large (r= 0.583, p < .001).

Table 11 Spearman's correlations between post-test scores and target word fulfilment

Variable		Post-test	Correctly used	Used
1. Post-test	Spearman's rho	—		
2. Correctly used	Spearman's rho	0.553 **	—	
3. Used overall	Spearman's rho	0.583 ***	0.873 ***	—

* p < .05, ** p < .01, *** p < .001

These results suggest that the more target vocabulary was used, the higher the post-test results of the vocabulary knowledge scale turned out, implying that the use of vocabulary is linked to an improvement in vocabulary knowledge.

6.4.2. Hypothesis 3b

Table 12 shows that the Spearman's correlation calculated a significant medium negative correlation between the post-test scores and the number of mistakes per 100 words. This means that the higher the number of mistakes is, the lower the scores of the vocabulary post-test are. The converse argument then says that the production of a more accurate text is linked to a higher learning gain in vocabulary knowledge.

Table 12 Spearman's correlation between post-test scores and accuracy

Variable		Post-test	Mistakes per 100 words
1. Post-test	Spearman's rho	—	
2. Mistakes per 100 words	Spearman's rho	-0.383 *	—

* p < .05, ** p < .01, *** p < .001

In conclusion, the results of Spearman's correlations confirm both hypotheses: Both variables – using target words in written texts and producing a more accurate text – have a positive effect on the development of vocabulary knowledge.

7. General discussion

After the presentation of the individual findings of the respective hypotheses in the previous chapter, the results will now be linked to previous research that was mentioned in the literature review and to some of the students' feedback collected during a follow-up English lesson (see section 5.2.3). Pointing to some of the shortcomings of this study, I attempt a careful interpretation of the results.

This study intended to explore the effects collaborative writing can have on pre-intermediate learners' vocabulary knowledge and text production in order to help foreign language teachers to make informed decisions about whether or how to apply collaborative writing tasks in their classroom.

7.1. First research question

The first research question is concerned with the effect the overarching theme of this thesis – collaborative writing – has on the dependent variable *vocabulary knowledge*: what effect does collaborative writing have on (less proficient) students' development of vocabulary knowledge? With this question it should be investigated whether collaborative writing bears benefits over individual writing for learners' vocabulary learning, additionally focusing on those learners' who can be considered as less proficient (for the definition of less proficient in this thesis' context see section 5.4). These sub questions are converted into hypotheses 1a and 1b, which will be discussed in the following.

The directional research hypothesis 1a states that collaborative writing is more beneficial to learners' development of word knowledge than individual writing. This assumption derived from the fundamental learning theory outlined by Vygotsky (1978) that cognitive development is initiated and facilitated by interaction with others, and was further strengthened by research into *collaborative dialogue* by Swain (1997) that positions the dialogue as the site for language learning. During collaborative dialogues learners can pool their linguistic resources to solve problems and together co-construct knowledge about language (Swain et al. 2002: 172). To encourage collaborative dialogue among students, the method of collaborative writing was selected. Due to the larger pool of resources in pair work as opposed to individual work, it was assumed that learners composing the text in dyads are more likely to reach larger learning gains in their vocabulary knowledge than learners composing it on their own. This assumption was supported by research conducted by Kim

(2008) and Fernández Dobao (2012). Kim (2008: 123), who also compared group work with individual work with regards to their gains in vocabulary knowledge, found that the collaborative group significantly outperformed the individual group in the vocabulary post-test. Fernández Dobao (2012), who compared small group work with pair work, found that a greater number of group members (four instead of two) also resulted in more instances of learning vocabulary. The results of the repeated measures ANOVA of this thesis' study, however, do not coincide with these previous findings, despite the descriptive indications for GR_{CW} to have reached a higher amount of vocabulary knowledge. As shown in section 6.2 the difference in vocabulary learning gain between GR_{IW} and GR_{CW} was not statistically significant.

While H1a took all learners into consideration, H1b focused on the impact CW has on less proficient learners (LP) and whether a difference in vocabulary learning between CW and IW can be observed. H1b was formulated as a directional hypothesis stating that LP learners improve their vocabulary knowledge through CW with a more proficient (MP) peer. This hypothesis resulted again from Vygotsky's concept of the zone of proximal development in which less knowledgeable individuals can learn from interacting with more knowledgeable peers (1978). By co-constructing knowledge, learners can serve as a collective scaffold, allowing them to perform beyond their individual momentary state of expertise. It is argued in Vygotsky's theory as well that what these learners were able to do with each other's assistance, they will be able to do by themselves in the future. In essence, it is this proposition that hypothesis 1b set out to test: whether the LP students working in pairs were able to transfer the knowledge they have constructed with their more proficient partner on the intermental plane to the intramental plane (Vygotsky 1966: 40), i.e. whether they were able to internalize the co-constructed vocabulary knowledge during the collaborative writing task to make use of it independently in the post-test. As an indication for students having internalized new knowledge, the teacher researcher considered the increase in the vocabulary knowledge scale scores from the pre-test to the post-test. As was pointed out in the section before, the GR_{CW} LP students' increase in vocabulary knowledge from the first measurement point to the second was significant, indicating that new knowledge was internalized. In order to trace the increase back to the peer interaction, the results were compared to the results of GR_{IW} LP students, who have improved significantly too. Although GR_{CW} LP students' learning gain scores were more than twice as high as the scores of GR_{IW} LP students, the repeated measures ANOVA did not reveal any significant difference between the groups.

There are hardly any studies that investigate vocabulary learning of students of different proficiency levels during a collaborative writing task the way this experiment designed it, since they mostly analysed LREs without focusing on vocabulary explicitly. Therefore, only cautious comparisons to the study by Leeser (2004) can be made (see section 3.3.). Leeser's findings indicate that in high-low proficiency pairs, LREs were mainly solved by the high proficient learner resulting in little learning gain for the rather passive low proficient counterpart. This thesis' experiment cannot determine which member of the pair resolved the language problems and which roles they took on, but its finding do show that the collaborative writing activity yielded significant learning gains for less proficient learners and thereby contradict Leeser's findings. From the feedback students gave after the experiment a tendency to experience collaborative writing as more conducive to learning vocabulary instead of individual writing could be observed. One student from GR_{IW} learner stated that they neither find writing on their own enjoyable nor helpful ("Alleine nicht so lustig [,] [a]lleine hilft nicht so", IWb[12]). From GR_{CW} three learners voiced that working with a partner was helpful, among whom one explicitly stated that their partner helped them to express themselves ("Er hat mir geholfen mich auszudrücken", CWd).

The significant increase in vocabulary knowledge of GR_{IW} LP learners, raises the question of which factors contributed to their learning gain since the collaborative dialogue with a peer can be ruled out. Intuitively, I decided to look into the most apparent confounding variable: the medium used for text production, namely *Microsoft Word*. Therefore, as explained in section 5.2, I included this variable in the feedback form by asking students whether they feel they learn something new from noticing *Word*'s suggestions for correction and correcting the underlined parts. The feedback from the participants revealed that on average they rather agree with this statement (*M*= 3.0) with 75 percent *rather agreeing* or *fully agreeing* to it. In the final comment section of the feedback form, one student of GR_{IW} highlighted the advantage of writing in *Word* over writing on paper by stating that *Word* points to some mistakes ("Und ich finde, dass es auf Word auf jeden Fall besser ist als auf einen Zettel, weil es auch auf ein paar Fehler hinweist […]", IWa). From this statement it can be concluded that some students experience *Word* as helpful for noticing mistakes and learning from them. It

[12] Unfortunately, these feedback answers could not be traced back to individual participants but only to the group they were allocated to. Therefore, the anonymised abbreviation for the students' feedback uses letters instead of numbers to distinguish them from the previously established abbreviations (IW1-20 and CW1-20).

appears that *Word* can function as an important corrective feedback-giver, drawing learners' attention to mistakes and sometimes also providing suggestions for correction.

Before continuing with the discussion of research question two, two major limitations of this part of the study are worth mentioning.

Firstly, the sample size (*N*= 40 for hypothesis 1a and *n*= 21 for hypothesis 1b) was relatively small which minimized the possibility for obtaining significant results. As explained in section 6.1 the larger the sample sizes are the greater is their power to detect significant differences between groups. Thus, for the difference to turn out to be significant a larger sample size most likely would have been needed.

Secondly, a major shortcoming of the chosen study design should not be overlooked: the systemic variation due to *practice effects* and *boredom effects* resulting from the repeated measures (Field 2018: 20). While students were naïve during the first vocabulary quiz, they completed the second quiz with prior experience on how to perform the task. This means that an improvement in vocabulary post-test results could also be attributed to the participants' *practice effects*, that is their familiarity with the vocabulary items. The teacher-researcher even observed that some students checked the translations of some vocabulary items right after the first vocabulary quiz. On the other hand, the decline or stagnation in post-test scores of ten participants could be explained by the *boredom effect* among others. It could be that those students experienced fatigue during the second round of the vocabulary test due to their familiarity with the test items.

In brief, although neither of the two statistical tests yield significant findings to support that CW is more beneficial to vocabulary learning for all learners and less proficient learners likewise, a tendency for collaborative writing being more helpful for vocabulary learning can be observed from previous research in this area as well as from students' feedback. Nevertheless, more research – especially including qualitative analyses – is needed to prove this tendency.

7.2. Second research question

The second research question focuses on the written product and how certain properties thereof differ between GR_{IW} and GR_{CW}: How do texts written collaboratively differ from

individually written texts with regards to fluency, lexical diversity, accuracy, and target word fulfilment? This question was divided into two hypotheses – 2a and 2b.

Hypothesis 2a proposes that collaboratively written texts show greater fluency, higher lexical diversity and accuracy than individually written texts. This direction of the hypothesis mainly derived from the previously mentioned assumption that pairs have more linguistic resources than individuals, which could result in a longer text, more diverse lexis, and fewer mistakes. The independent t-tests, however, could only confirm this proposition, with respect to one variable, namely accuracy. The differences in the scores of fluency (number of words and sentences) as well as in the scores of lexical diversity (type-token ratio, VOCD and MTLD) between the groups was descriptively observable – GR_{CW} outperforming GR_{IW} – but statistically not significant.

Regarding the category of fluency previous studies present mixed results. Storch & Wigglesworth (2007: 163) could not detect any significant differences between individually and jointly written texts either. In Fernández Dobao's study (2012), those learners who worked in small groups or pairs overall produced shorter texts than those learners who worked alone. She ties this result to the rather lengthy discussions learners working collaboratively had, emerging from the need to agree on the content of their stories and to the language they wanted to use (Fernández Dobao 2012: 54). While Fernández Dobao used a picture story as writing impulse, which allows for more freedom regarding lexical choice, the present writing prompt should guide students to the use of particular vocabulary. Consequently, it is likely that there were fewer or shorter discussions among the group members, which could have led to longer texts.

As briefly mentioned in section 5.6.2, studies investigating collaborative writing and its effect on written products were concerned with complexity of the obtained texts instead of lexical diversity. Therefore, the findings of lexical diversity cannot be compared. However, what should be pointed out is the similarity of results regarding the texts' accuracy. Both researcher teams – Storch and Wigglesworth as well as Fernández Dobao – also reported significant differences between the accuracy of individually written and collaboratively written texts. Measuring error-free T-units and clauses (resulting from the analysis of complexity) Storch and Wigglesworth (2007: 165-166) found that texts written by dyads include significantly more accurate error-free T-units and more error-free clauses than texts by individuals. Similar

findings were presented by Fernández Dobao (2012: 54): collaboratively written texts were more accurate than individually written texts on average. While this thesis' study cannot provide insights into the reasons for the greater accuracy of collaboratively written texts, the analysis of LREs in the studies mentioned before could shed a light on it. Fernandez Dobao (2012) additionally investigated the LREs produced among group members during the composition. There she found that small groups produced more LREs and could correctly resolve a higher number of them than pairs and individuals. These solutions to the encountered language-related problems were subsequently transferred into the texts, resulting in more accurate texts (Fernández Dobao 2012: 54). Finally, she attributes the greater success of groups to resolve LREs to the greater number of linguistic resources that could be pooled for solving linguistic problems.

While these are promising findings on the level of accuracy that support collaborative writing, a major limitation of this study shall be highlighted. Although the teacher researcher attempted to maintain a consistent, objective assessment throughout the correction of all texts – by following her set guidelines for which mistakes are counted and which are not – subjectivity in this process still cannot be ruled out. It appears that accuracy is a relatively complex and subjective criterion that makes it hard to provide fully reliable results. To increase objectivity and thus the reliability of this criterion, a second and third rater could have been consulted. Thereby interrater reliability could have been ensured. Although being aware of the importance of this quality criterion, I decided not to consult my two only English teacher colleagues at school. This was mainly due to the fact that I did not want to burden them with even more work on top of their already challenging job[13]. Nevertheless, for any future research it is highly advisable to have the texts assessed by more than one rater.

Hypothesis 2b changes the focus of text analysis to *target word fulfilment* by suggesting that collaboratively written texts include more target words than individually written texts. As outlined in section 5.5 and 5.6 the target words in the text correspond to the majority of vocabulary items quizzed in the vocabulary pre- and post-test. The direction of this hypothesis was influenced by the above presented findings that groups were able to solve more linguistic problems. In the present study, translating the German bullet points into English and using

[13] Note: As the school is very young and just setting up, the teaching staff is very small. This is why I only had two English colleagues, who moreover were also just in their first years of teaching, and I could see that they were filled to the brim with their own teaching responsibilities.

the phrases correctly can pose such linguistic problems. In turn, having used the target word correctly can be interpreted as evidence for students having solved the linguistic problem. From that stance, the results of this study can be cautiously compared to previous research on LREs. As the independent t-test revealed, there was a significant difference in the overall number of target words used by GR_{CW} and GR_{IW}, whereby GR_{CW} managed to use 3.3 more target words on average than GR_{IW}. Thus, the constructed hypothesis could be confirmed. Linking this result to LRE research, one can interpret this higher number in target words used as a higher number of linguistic problems solved, supporting Fernández Dobao's findings that during CW more linguistic problems could be resolved. Yet, whether the success of using more target words resembles an increase in vocabulary learning was only tested in hypothesis 3a, which will be discussed below.

Before, it is worth paying attention to what it means if students have used or not used target words. Having not used the target phrases does not necessarily mean that they omitted this bullet point of the writing prompt completely. It could also mean that they used an alternative formulation. For example, some students used *nice* instead of *kind* for expressing the meaning of *versuchen immer <u>nett</u> zueinander zu sein*; or *Alice and Dan are friends again* instead of *Alice and Dan made up* to express the meaning of *Alice and Dan* versöhnten sich. While a similar meaning is conveyed by these alternative formulations contributing to successful communication, the target words were not used, whereby learners are missing an opportunity to practice the newly learnt vocabulary. This phenomenon of learners resorting to words they already know or words that are more frequent (e.g. the adjective *nice* is more frequently used than *kind*) could be explained by research into productive vocabulary use as it was outlined in section 2.5 (productive vocabulary use in writing). The participants in Horst and Collins' study (2006) showed a tendency to use words they already know correctly, avoiding the risk of using newly encountered vocabulary incorrectly. A similar reluctance was observed among Rott et al.'s participants, who – when they were asked to do written recall of a previously presented text – did not use the target words but words of higher frequency, for instance they used *tree* instead of *oak* and *animals* instead of *livestock* (Rott et al. 2002 referred to in Nation 2013: 270). From the feedback comments of the present study's participants, one serves as an indication that some students in fact did not know the target vocabulary well enough to use them. This student reported, "ich habe den Schreibauftrag als nicht so cool gefunden, da ich die meisten Vokabeln zu den Zeitpunkt nicht so gut konnte" (CWa).

In conclusion, the results showed that with regards to accuracy and target word fulfilment collaboratively written texts outperformed individually written texts giving rise to the assumption that collaboration leads to more success in solving linguistic problems. Whether the solution of linguistic problems, manifested in a higher number of target words used and a higher accuracy of the written products, is linked to a development in vocabulary knowledge was the concern of research question 3.

7.3. Third research question

Essentially, research question 3 – how do the use of target words and the level of accuracy correlate with the development of vocabulary knowledge – investigates the usefulness of writing for learning vocabulary (leaving the collaborative aspect aside). The hypotheses focus on two aspects of writing, namely using target words and producing a more accurate text. Formulated directionally, they suppose that each of those aspects has a positive effect on vocabulary learning, thus: using more target words and writing a more accurate text correlates with more vocabulary learning. The direction derived from research into productive vocabulary use in writing (section 2.5) and the conclusion that a writing task at the end of a learning unit – in which target vocabulary was met several times – can help receptive vocabulary knowledge become productive vocabulary knowledge. The Spearman's correlation for both variables in combination with the vocabulary post-test results revealed a significant correlation, confirming both hypotheses. These results therefore can be interpreted as – broadly speaking – writing being a useful means to learn vocabulary.

"Broadly speaking" was intentionally inserted in this interpretation as the quantification of the VKS scores resulted in a reduction of the complexity of the phenomenon *vocabulary knowledge*. Hence, we can only refer to general learning gains in vocabulary knowledge without having any insights into the quality of vocabular learning. Whether the writing task was more conducive to enlarging vocabulary size or to increasing vocabulary depth cannot be answered. For example, it could be that students improved their score from 0 (not being familiar with the word at all) to 3 (knowing the translation of the word), which would represent an increase in vocabulary size on a receptive level since a new form-meaning pairing was acquired. On the other hand, it could be that students at first received a score of 3 by showing that they had already established a form-meaning pairing, as they provided a correct translation, and then received a score of 5 in the post-test because they were able to use the

word correctly in a sentence. This increase of 2 points then needs to be viewed as an increase in vocabulary depth on the productive level. Hence, it should be borne in mind that these findings do not provide nuanced information on which aspects of vocabulary knowledge had been developed.

For more nuanced conclusions, the data would need more in-depth qualitative analyses. Since the analysis of learners' LREs is not an option for the data obtained during this study, another possibility could be to link each learner's response of each item of the vocabulary pre-test to the use of the corresponding target word in the text and then to the response of the item in the post-test. Such an approach was attempted at first, but soon considered as too laborious for this master thesis. Yet, an example of this attempt is provided in the following, since it could be useful for future research.

Table 13 Quantitative analysis of development of vocabulary knowledge

Item	**Parti-cipant**	**Pre-test response & *score***	**Text extract & *target word fulfilment***	**Post-test response & *score***	**Result:**
No. 2: get on well with someone	IW8	**Transl.:** Sich gut verstehen **Sentence:** I get on well with Eric. *5*	… they get very well with each other. *Incorrectly used*	**Transl.:** Sich gut verstehen **Sentence:** I get on well with my dad. *5*	Vocabulary item used correctly in pre-test and post-test, yet used incorrectly in the text by leaving out the preposition "on". **Vocabulary knowledge remained the same although incorrectly used in the text.**
No. 13: make fun of someone	CW5	**Transl.:** Spaß haben mit jemanden **Sentence:** -- *2*	Alice and Dan promised, that they will never make fun of each other. *Correctly used*	**Transl.:** sich über jemanden lustig machen **Sentence:** Sara make fun of Peter. *4*	Vocabulary item incorrectly translated in the pre-test and no sentence provided. During CW item was correctly used. In post-test item was correctly translated and used in a sentence however with grammatical inaccuracy leaving out the third person 's'. **After having used the vocabulary item in the text, the student established a form-meaning pairing – increasing size on a receptive level – and was able to use it in a sentence syntactically correctly but grammatically incorrectly – improving one aspect of productive vocabulary knowledge.**

To complement this shortcoming, the development of vocabulary knowledge through writing shall be viewed from a different angle, namely from the perspective of students. The teacher researcher collected students' feedback on their perception of how useful the writing task was for increasing a) vocabulary size, b) vocabulary depth, and c) for learning vocabulary overall. They could voice their (dis-)agreement with the statements: *I have the feeling that I have extended my vocabulary through completing the writing task* (referring to vocabulary size); and *I have the feeling that I can use the vocabulary better/more accurately now* (referring to vocabulary depth). Moreover, they could report their opinion on how useful they found the writing task for learning vocabulary in general.

With regards to the usefulness of the writing task for extending one's vocabulary, students had mixed feelings. While the majority (*n*= 19, equalling 48 percent) of the students rather agreed to this statement, there were also many students (*n*= 13 equalling 28 percent) not agreeing at all. The mean value of their responses equalled 2.6 (on a scale from 1 to 4). One reason for almost a quarter of students disagreeing with the idea that writing the text is useful for enlarging one's vocabulary size could be that they are more used to the strategy of studying vocabulary from word lists. One student points to this assumption by stating that they did not find writing a helpful strategy for learning vocabulary, they are rather used to reading vocabulary a few times and then being quizzed by their parents ("Mir persönlich bringt Schreiben nicht so viel, ich lese sie mir öfters durch und dann fragen meine Eltern mich ab", IWc). A higher mean value of agreement was achieved in the second statement (*M*= 2.8) and the third (*M*= 3.1). Regarding the second statement, the majority of students (*n*= 26, equalling 65 percent) either *agreed rather* or *agreed fully* that writing helped them to make use of the vocabulary more accurately. Lastly, overall, they experienced the writing task as *rather useful*, with the vast majority of students (86 percent) reporting that the writing task was *rather* (*n*= 25) or *fully* (*n*= 9) useful for vocabulary learning.

In brief, students reported that they found the writing task more useful for increasing the productive aspect of vocabulary knowledge to use words more correctly and less useful for increasing their vocabulary size. But they mostly agreed that writing was useful for vocabulary learning after all.

Before the conclusion and implications for teaching are presented, one last weakness of the research project should be mentioned: the short period of data collection. The experiment

took place within four days and the post-test was conducted immediately after the writing task. Therefore, it is important to be aware that what has been referred to as *vocabulary learning* in the context of this study, only captures immediate learning gains of a fairly limited duration. Since no delayed vocabulary test was conducted, no evidence for long-term vocabulary learning can be provided. In order to find out whether gains in vocabulary endured, future research involving a longitudinal study design is needed.

8. Conclusion and implications for language teaching

The purpose of this study was to explore the effects collaborative writing tasks can have on students' vocabulary learning. Thereby, this thesis scrutinized the construct of *vocabulary knowledge*, finding that it is an essential but highly complex part of linguistic knowledge. Not only should learners acquire a considerable amount of vocabulary, but also should they be able to make use of it meaningfully and accurately. This involves that students of a foreign language need to acquire both receptive and productive knowledge of words. Therefore, teachers need to create a vocabulary learning supportive environment that takes the different aspects of vocabulary knowledge into account. Linguist Paul Nation (2013: 579) among others, therefore, argues that vocabulary learning needs to happen throughout the four strands of *meaning-focused input*, *language-focused learning*, *meaning-focused output* and *fluency development*. A meaningful output, such as a writing task, requires productive vocabulary use. As productive vocabulary knowledge takes more time to build up, it is suggested that students are assigned writing tasks at the end of a learning unit (Nation 2013: 270). By designing a writing task that motivates students to use the newly learnt vocabulary, teachers create the opportunity for receptive word knowledge to become productive.

Yet, we know that, in each foreign language class, there are students that have difficulties with writing tasks and using vocabulary accurately. Assigning solitary writing tasks that students are asked to do as home exercise might put the less proficient students at a disadvantage for their vocabulary development. Still, it appears to be common practice among many English teachers in Austria. With the compilation of research findings and the results of the present study, I hope that I could call attention to an alternative method that can be more supportive to language learning: collaborative writing. When applying a collaborative writing task, teachers can turn the communicative language classroom into a space where students support each other's language learning process. They can pool their linguistic resources and are not

left alone with their own limited linguistic knowledge. They are given the chance to enter the *zone of proximal development* by interacting with a peer. Thereby they can serve as *collective scaffolds* for each other and discuss linguistic problems amongst themselves in a way that might be more accessible to them than when explained by a teacher.

While this thesis' study could not provide definite support for collaborative writing being more beneficial for vocabulary learning than solitary writing, due to insignificant findings in hypotheses 1a and 1b, a tendency is still observable. On the one hand, the descriptive statistics of H1a and H1b suggest that pairs improved their vocabulary knowledge to a greater extent than individuals. On the other hand, when combining the research findings of the second hypotheses – pairs included more target vocabulary and less mistakes – and the third hypotheses – using more target words and producing a more accurate text correlate with more vocabulary learning – a relatively clear picture emerges: Through collaborative writing, learners were given a greater opportunity to learn vocabulary than those having worked alone since they used more target words and made less mistakes.

The degree to which learners benefit from a collaborative writing task, however, is also influenced by a number of factors. The degree of language proficiency of team members, the roles they take on, and the resulting relationship that they form in their group are some of them. It can be that students of lower proficiency take on a passive role and engage little in the activity, while the higher proficient student dominates the pair work activity. Thereby they establish a dominant/passive relationship that fails to meet the criteria of genuine collaboration and is likely to end in little learning for the less proficient student. Thus, assigned group work does not always result in true group work, which can be frustrating for teachers and learners alike.

Which brings me to the initially posed question: Can teamwork make the dream work? My answer is: Yes, it can. However, to unfold the full potential collaborative writing tasks offer for vocabulary learning, it is pivotal for teachers to put enough thought into the learner allocation and task creation. With clear instructions added on top, collaborative writing can be turned into a successful communicative linked skills activity, during which learners not only produce language but also collectively deliberate about language. And it is this deliberation about language, this *languaging*, that constitutes a crucial part of learning.

In languaging, we see learning taking place (Swain 2006:98).

9. Bibliography

Adams, Rebecca; Ross-Feldman, Lauren. 2008. “Does writing influence learner attention to form?” In Belcher, Diane; Hirvela, Alan (eds.). *The Oral-literate Connection*. Ann Arbor MI: The University of Michigan Press, 243–266.

Alegría de la Colina, Ana.; García Mayo, María P. 2007. “Attention to form across collaborative tasks by low-proficiency learners in an EFL setting”. In García Mayo, M.P. (ed.). *Investigating Tasks in Foreign Language Learning*. Clevedon: Multilingual Matters, 91–116.

Bierwisch, Manfred ; Schreuder, Robert. 1992. “From concepts to lexical items”. *Cognition*, 41, 23–60.

Blanca, Maria. J.; Alarcón, Rafael; Arnau, Jaume; Bono, Roser; Bendayan, Rebecca. 2017. “Non-normal data: Is ANOVA still a valid option?” *Psicothema 29*(4), 552-557. doi:10.7334/psicothema2016.383

Brooks, Lindsay; Swain, Merill. 2009. “Languaging in collaborative writing: Creation and response to expertise”. In Mackey Alison and Polio Charlene. (eds.). *Multiple perspectives on interaction in SLA*. Mahwah, NJ: Lawrence Erlbaum, 58–89.

Bundesministerium für Bildung. 2018. *Lehrplan der AHS-Unterstufe*. https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10008568 (25 Feb. 2022)

Bundesministerium für Bildung. 2021. *Leistungsbeurteilungsverordnung.* https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009375 (1 March 2022).

Chi, Michelene T.H., Bassok, Miriam, Lewis, Matthew W., Reimann, Peter; Glaser, Robert. 1989. “Self explanations: how students study and use examples in learning to solve problems”. *Cognitive Science* 13(2), 145–182.

Cohen, Louis, Lawrence Manion & Keith Morrison. 2011. *Research Methods in Education* (7th ed.). London: Routledge.

Corson, David J. 1985. *The Lexical Bar*. Oxford: Pergamon Press.

Council of Europe. 2001. Common European framework of references for languages: learning, teaching, assessment (CEFR). https://www.coe.int/en/web/common-european-framework-reference-languages (26 Feb. 2022)

Coxhead, Averil. 2007. “Factors and aspects of knowledge affecting L2 word use in writing”. In Davidson, P., Coombe, C., Lloyd, D. and Palfreyman, D. (eds.). Teaching and Learning Vocabulary in Another Language. Dubai: TESOL Arabia, 331–42.

Crossley, Scott A.; Salsbury, Tom; McNamara, Danielle. S. 2014. “Assessing lexical proficiency using analytic ratings: A case for collocation accuracy”. *Applied Linguistics* 36(5), 570–590.

Cumming, Alister. 1990. “Metalinguistic and ideational thinking in second language composing”. *Written Communication* 7(4), 482–511.

Daller, Helmut; Milton, James; Treffers-Daller, Jeanine. 2007. “Editor’s introduction”. In Daller, Helmut; Milton, James; Treffers-Daller, Janine (eds.). *Modelling and assessing vocabulary knowledge*. Cambridge, UK: Cambridge University Press, 133–149.

Dillenbourg, Pierre; Baker, Michael; Blaye, Aagnès; O’Malley, Claire. 1996. “The evolution of research on collaborative learning”. In Reimann Peter.; Spada, Hans. (eds) *Learning in Humans and Machine: Towards an Interdisciplinary Learning Science*. Oxford, Elsevier, 189–211.

Donato, Richard. 1988. *Beyond group: A psycholinguistic rationale for collective activity in second-language learning*. Doctoral dissertation: University of Delaware, Newark.

Donato, Richard. 1994. “Collective scaffolding in second language learning”. In Lantolf, James P (ed.). *Vygotskian Approaches to Second Language Research.* Norwood, NJ: Ablex, 33–56.

Dörnyei, Zoltán. 2007. *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Durán, Pilar; Malvern, David; Richards, Brian; Chipere, Ngoni. 2004. “Developmental trends in lexical diversity”. *Applied Linguistics* 25(2), 220-242.

Ede, Lisa; Lunsford, Andrea. 1990. *Singular texts/plural authors*. Carbondale: Southern Illonois University Press.

Ellis, Nick. C.; Ferreira-Junior, Fernando. 2009. “Construction learning as a function of frequency, frequency distribution, and function”. *The Modern Language Journal* 93(3), 370–385.

Fernández Dobao, Ana. 2012. “Collaborative writing tasks in the L2 classroom: Comparing group, pair, and individual work”. *Journal of Second Language Writing* 21(1), 40–58.

Fernández Dobao, Ana. 2014. “Vocabulary learning in collaborative tasks: A comparison of pair and small group work”. *Language Teaching Research* 18(4), 497-520.

Field, Andy. 2018. *Discovering statistics using IBM SPSS statistics*. Los Angeles: SAGE.

Gerngross, Günter; Puchta, Herbert; Holzmann, Christian; Lewis-Jones, Peter; Stranks, Jeff. 2021. *MORE! 3 Student’s book – enriched course*. (4th edition). Innsbruck: Helbling Languages.

Gerngross, Günter; Puchta, Herbert; Holzmann, Christian; Lewis-Jones, Peter; Stranks, Jeff. 2021. *MORE! 3 Workbook – enriched course*. (4th edition). Innsbruck: Helbling Languages.

Horst, Marlise; Collins, Laura. 2006. From faible to strong: How does their vocabulary grow? *Canadian Modern Language Review*, 63(1), 83-106.

Izumi, Shinichi. 2002. “Output, input enhancement, and the noticing hypothesis: An experimental study in ESL relativisation”. *Studies in Second Language Acquisition* 24(4), 541–577.

Izumi, Shinichi; Bigelow, Martha. 2000. “Does output promote noticing and second language acquisition?” *TESOL Quarterly* 34(2), 239–278.

Kang, Su-Ja. 2005. “Dynamic emergence of situational willingness to communicate in a second language”. *System* 33(2), 277–292.

Kim, Youjin. 2008. “The contribution of collaborative and individual tasks to the acquisition of L2 vocabulary”. *The Modern Language Journal* 92(1), 114–130

Kowal, Maria; Swain, Merill. 1994. “Using collaborative language production tasks to promote students’ language awareness”. *Language Awareness,* 3(2), 73–93.

Krashen, Stephen D. 1985. *The input hypothesis: Issues and implications*. London: Longman.

Lantolf, James P. 2000. “Introducing sociocultural theory”. In Lantolf, James P. (ed.). *Sociocultural theory and second language learning*. Oxford: Oxford University Press, 1–26.

Lantolf, James P. 2005. “Sociocultural theory and L2 learning: An exegesis”. In Hinkel, Eli (ed.) *Handbook of research in second language teaching and learning*. Mahwah, NJ: Erlbaum, 335–354.

Laufer, Batia. 2020. “Evaluating exercises for learning vocabulary. In Webb, Stuart (ed.). *The Routledge handbook of vocabulary studies*. Milton: Routledge, 351-368.

Laufer, Batia; Paribakht, Sima T. 1998. The relationship between passive and active vocabularies: Effects of language learning context. *Language Learning*, 48(3), 365–391.

Lee, Siok H.; Muncie, James. 2006. From receptive to productive: Improving ESL learners’ use of vocabulary in a postreading composition task. *TESOL Quarterly*, 40(2), 295–320.

Leeser, Michael J. 2004. “Learner proficiency and focus on form during collaborative dialogue”. *Language Teaching Research* 8(1), 55–81.

Levelt, Willem J. M. 1989. *Speaking: From intention to articulation.* Massachusetts: MIT Press.

Levelt, Willem J. M. 1992. *Accessing words in speech production: Stages, processes and representations*. Cognition, 42, 1–22.

Long, Michael H. 1983. “Native speaker/non-native speaker conversation and the negotiation of comprehensible input”. *Applied Linguistics* 4(2), 126–141.

Manchón, Rosa, M. 2009. Introduction: Broadening the perspective of L2 writing scholarship: The contribution of research on foreign language writing. In Manchón, Rosa (ed.). *Writing in foreign language contexts: Learning, teaching, and research*. Clevedon, UK: Multilingual Matters, 1–22.

McCarthy, Philip; Scott, Jarvis. 2010. “MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment”. *Behavior Research Methods* 42(2), 981-392.

Meara, Paul. 1996. "The vocabulary knowledge framework". http://www.lognostics.co.uk/vlibrary/meara1996c.pdf (23 May 2021).

Merriam-Webster online dictionary. https://www.merriam-webster.com/dictionary/collaboration (5 Jan. 2022).

Milton, James. 2009. *Measuring second language vocabulary acquisition.* Bristol: Multilingual Matters.

Murphy, Lynne M. 2010. *Lexical meaning.* Cambridge: Cambridge University Press. doi:10.1017/CBO9780511780684

Nation, Paul I. S. 2013. *Learning Vocabulary in Another Language*. Cambridge: University Press. https://doi.org/10.1017/CBO9781139858656

Nation, Paul. 2020. “The different aspects of vocabulary knowledge”. In Webb, Stuart (ed.). *The Routledge handbook of vocabulary studies.* Milton: Routledge, 15-29.

Paribakht, Sima T. & Marjorie Wesche. 1997. "Vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition". In James Coady & Thomas N. Huckin (eds.). *Second language vocabulary acquisition: A rationale for pedagogy.* Cambridge: Cambridge University Press, 174–200.

Qian, David D.; Lin, Linda H. F. 2020. “The relationship between vocabulary knowledge and language proficiency”. In Webb, Stuart (ed.). *The Routledge handbook of vocabulary studies*. Milton: Routledge, 66-80.

Read, John. 2000. *Assessing vocabulary.* Cambridge: Cambridge University Press.

Schrage, M. (1994) Writing to collaborate; collaborating to write. In Leonard, J.S.; Wharton, C.E.; Davis, R.M. ; Harris J. (eds.). *Authority and textuality. Current views of collaborative writing*. West Cornwall, CT: Locus Hill Press, 17–24.

Stahl, Gerry. 2006. *Group cognition: Computer support for building collaborative knowledge.* Michigan: MIT Press.

Storch, N. and Wigglesworth, Gillian. 2007. “Writing tasks: Comparing individual and collaborative writing”. In García Mayo, M.P. (ed.). *Investigating tasks in formal language learning*. London: Multilingual Matters, 157–177.

Storch, Neomy. 2009. *The nature of pair interaction. Learners’ interaction in an ESL class: its nature and impact on grammatical development*. Saarbrücken, Germany: VDM Verlag.

Storch, Neomy. 2013. *Collaborative writing in L2 classrooms*. Bristol: Multilingual Matters.

Storch, Neomy; Aldosari, Ali. 2013. “Pairing learners in pair work activity”. *Language Teaching Research* 17(1), 31–48.

Swain, Merill. 1985. “Communicative competence: some roles of comprehensible input and comprehensible output in its development”. In Gass, Susan M. and Madden, Carolyn G. (eds.). *Input in second language acquisition*. Rowley, MA: Newbury House, 235-256.

Swain, Merill. 1993. “The output hypothesis: Just speaking and writing aren’t enough”. *The Canadian Modern Language Review* 50(1), 158–164.

Swain, Merill. 1997. “Collaboratie dialogue: Its contribution to second language learning”. *Revista Canaria de Estudios Ingleses* 34, 115-132.

Swain, Merill. 2006. “Languaging, agency and collaboration in advanced second language learning”. In Byrnes Heidi. (ed.) *Advanced language learning: The contributions of Halliday and Vygotsky*. London, UK: Continuum, 95–108.

Swain, Merill; Lapkin, Sharon. 1995. “Problems in output and the cognitive processes they generate: A step towards second language learning”. *Applied Linguistics* 16(3), 371–391.

Swain, Merill; Lapkin, Sharon. 1998. “Interaction and second language learning: two adolescent French immersion students working together”. *The Modern Language Journal* 82(3), 320–337.

Swain, Merill; Lapkin, Sharon. 2002. “Talking it through: Two French immersion students’ response to reformulation”. *International Journal of Educational Research* 37(3-4), 285–304.

Vygotsky, Lev S. 1978. Mind in society: Development of higher psychological processes. Cambridge, Mass: Harvard University Press.

Vygotsky, Lev S. 1981. “The genesis of higher mental functions”. In Wertsch, James V. (ed.). *The concept of activity in Soviet Psychology*. Armonk, NY: M.E. Sharpe, 144-188.

Watanabe, Yuko; Swain, Merill. 2007. “Effects of proficiency differences and patterns of pair interaction on second language learning: Collaborative dialogue between adult ESL learners”. *Language Teaching Research* 11(2), 121–142.

Webb, Stuart A. 2007. “The effect of repetition on vocabulary knowledge”. *Applied Linguistics* 28(1), 46–65.

Webb, Stuart. 2020. “Introduction”. In Webb, Stuart (ed.). *The Routledge handbook of vocabulary studies*. Milton: Routledge, 1-12.

Weissberg, Robert. (2006) *Connecting speaking and writing*. Ann Arbor: University of Michigan Press.

Wigglesworth, Gillian; Storch, Neomy. 2009. “Pairs versus individual writing: Effects on fluency, complexity and accuracy”. *Language Testing* 26(3), 445–466.

Williams, Jessica. 1999. “Learner-generated attention to form”. *Language Learning* 49(4), 583–625.

Wood, David; Bruner, Jerome S.; Ross, Gail. 1976. “The role of tutoring in problem-solving”. *Journal of Child Psychology and Psychiatry,* 17(2), 89–100.

Wu, Trong. 1993. “An accurate computation of the hypergeometric distribution function”. *ACM Transactions on Mathematical Software*, 19, 33-43.

Yule, George; Macdonald, Doris. 1990. “Resolving referential conflicts in L2 interaction: The effects of proficiency and interactive role”. *Language Learning* 40(4), 539–556.

Zareva, Alla; Schwanenflugel, Paula J.; Nikolova, Yordanka. 2005. „Relationship between lexical competence and language proficiency: Variable sensitivity”. *Studies in Second Language Acquisition* 27(4), 567–595

Appendices

Appendix 1: Abstract (English)

In the English as a Foreign Language (EFL) classroom, writing is one of the four skills that learners need to develop throughout their schooling. This skill is inevitably linked to the development of vocabulary knowledge, especially the ability to use words correctly in context. In order to support their learners in this process, teachers can assign writing tasks in various formats. Following Vygotsky's sociocultural theory of mind (1978) and Swain's application to second language learning through the concept of collaborative dialogue (1997), there is support for collaborative writing to be more beneficial to students' vocabulary learning than individual writing. This thesis' study therefore set out to investigate the effects the method of collaborative writing can have on students' vocabulary learning and their written performance. 40 students of two third grade classes of a Gymnasium in Lower Austria participated in this study's quasi experiment. The quasi-experiment included two groups: 10 pairs who wrote a text collaboratively (GR_{CW}) and 20 students who wrote the text individually (GR_{IW}). Data was collected and analysed in two ways: on the one hand, the learners' scores of their vocabulary pre-and post-test were evaluated and on the other hand, their scores of their written product in the categories *fluency, lexical diversity, accuracy* and *target word fulfilment* were compared. The data obtained from the vocabulary pre- and post-test showed that, on a descriptive level, GR_{CW} improved their vocabulary knowledge to a greater extent than GR_{IW}. Especially less proficient learners appear to have benefited from the collaboration with a more proficient peer, since their vocabulary knowledge scores increased significantly. Furthermore, GR_{CW}'s texts achieved higher scores in the categories *accuracy* and *target word fulfilment*, which gives rise to the assumption that collaboration leads to greater success in solving linguistic problems. Lastly, the overall usefulness of writing texts for learning vocabulary was tested. It was found that there is a positive correlation between vocabulary learning and using target words & producing a more accurate text. Feedback from the participants indicated that writing texts might be more useful for increasing their productive aspects of vocabulary knowledge (e.g. using words more correctly) and less useful for enlarging their vocabulary size. These findings not only highlight the importance of writing texts in the EFL classroom but should also encourage teachers to implement collaborative writing in their lessons.

Appendix 2: Abstract (Deutsch)

„Writing“ ist eine der vier Fertigkeiten, welche Lernende im Fach Englisch während ihrer schulischen Laufbahn entwickeln müssen. Diese Fähigkeit ist unweigerlich mit der Entwicklung des Wortschatzes – insbesondere mit der Fähigkeit, Wörter im Kontext richtig zu verwenden – verbunden. Um die Lernenden dabei zu unterstützen, können Lehrende Schreibaufgaben in verschiedenen Formaten aufgeben. In Anlehnung an Vygotskys *sociocultural theory of mind* (1978) und Swains Anwendung dieser Theorie auf das Erlernen einer Zweitsprache durch das Konzept des *collaborative dialogue* (1997), gibt es Grund für die Annahme, dass kollaboratives Schreiben für das Vokabellernen vorteilhafter ist als individuelles Schreiben. Die Studie dieser Masterarbeit machte es sich daher zum Ziel, die Auswirkungen der Methode des kollaborativen Schreibens auf das Lernen von Vokabeln zu untersuchen, wie auch die Auswirkungen auf die verfassten Texte der Lernenden. 40 Schüler*innen aus zwei dritten Klassen eines niederösterreichischen Gymnasiums nahmen an dem Quasi-Experiment dieser Studie teil. Die Teilnehmer*innen wurden in zwei Gruppen geteilt. GR_{CW} umfasste zehn Paare, die gemeinsam kollaborativ einen Text verfassten. GR_{IW} umfasste 20 Lernende, die den Text individuell verfassten. Die Daten wurden auf zwei Arten erhoben und analysiert. Einerseits wurden die Ergebnisse des Vokabel Pre- und Post-Tests herangezogen und andererseits die Ergebnisse, die sie mit ihrem Text in den *Kategorien Flüssigkeit, lexikalische Vielfalt, Richtigkeit* und *Zielworterfüllung*, erreichten. Die aus den Vokabeltests gewonnenen Daten zeigten, dass GR_{CW} ihre Wortschatzkenntnisse auf deskriptiver Ebene stärker verbesserten als GR_{IW}. Insbesondere weniger geübte Lernende schienen von der Zusammenarbeit mit geübteren Mitschüler*innen profitiert zu haben, da ihre Vokabelkenntnisse signifikant gestiegen sind. Darüber hinaus erreichten die Texte der GR_{CW} höhere Punktzahlen in den Kategorien *Richtigkeit* und *Zielworterfüllung*, was die Vermutung nahelegt, dass Kollaboration zu mehr Erfolg bei der Lösung sprachlicher Probleme führt. Abschließend wurde die allgemeine Nützlichkeit des Schreibens von Texten für das Vokabellernen getestet. Die statistische Rechnung ergab, dass es eine positive Korrelation zwischen dem Lernen von Vokabeln und der Verwendung von Zielwörtern sowie des Verfassens eines Textes mit weniger Fehlern gibt. Das Feedback der Teilnehmer*innen deutete außerdem darauf hin, dass das Schreiben von Texten für die Entwicklung produktiver Aspekte des Wortschatzes (Wörter richtig verwenden zu können) nützlicher sein kann als für die quantitative Erweiterung des Wortschatzes. Diese Ergebnisse unterstreichen nicht nur die Bedeutung des Schreibens von Texten im Englischunterricht, sondern sollen Fremdsprachenlehrpersonen auch dazu ermutigen, kollaboratives Schreiben in ihrem Unterricht zu implementieren.

Appendix 3: Information for parents and students

Sehr geehrte Erziehungsberechtigte!

Hiermit möchte ich Sie über die baldige Durchführung meiner **Forschung im Englischunterricht** in den dritten Klassen informieren.

Ich befinde mich gerade am Ende meines Master-Lehramtsstudiums, welches das Schreiben einer Masterarbeit inklusiver empirischer Forschung umfasst. Ich verfasse diese Arbeit im Fach Englisch und lege dabei mein Augenmerk auf die Entwicklung der englischen Sprachkenntnisse der Schüler*innen. Genauer gesagt, möchte ich untersuchen, **welche Auswirkungen die Methode des kollaborativen Schreibens – das Verfassen eines englischen Textes in Partnerarbeit – auf die Schreibkompetenz der einzelnen Schüler*innen hat**.

Um diese Forschungsfrage beantworten zu können, werde ich **zwei Methoden vergleichen**: das individuelle Verfassen eines Textes und das kollaborative Verfassen eines Textes. Dazu werden die Schüler*innen von mir in zwei Gruppen geteilt: jene, die einen Schreibauftrag alleine ausführen und jene die ihn in Partnerarbeit ausführen. Zusätzlich werden die Kinder je ein Vokabelquiz vor und nach dem Schreibauftrag vervollständigen. Der geschriebene Text sowie die Ergebnisse der Quizzes werden zur Datenerhebung und -auswertung herangezogen und sollen Vergleiche zwischen den Lerneffekten der zwei unterschiedlichen Methoden ermöglichen.

Sie sehen also, **die Schüler*innen der 3A und 3B dürfen einen wichtigen Beitrag zur Forschung im Feld der Fremdsprachendidaktik leisten**. Die Ergebnisse können dann nicht nur meine Unterrichtstätigkeit, sondern auch jene vieler anderer Fremdsprachenlehrkräfte beeinflussen. Deshalb ist es mir auch ein Anliegen, dass die Schüler*innen meiner Forschung mit **Ernsthaftigkeit** und **Bemühen** entgegentreten.

Bezüglich der **Bewertung** der Texte und Quizzes: wie gut oder schlecht diese ausfallen, hat keinen Einfluss auf die Note (hier sei auch angemerkt, dass die Quizzes ohne Vorbereitung/ „Auswendig Lernen“ der Vokabeln stattfinden werden). Die **Mitarbeit** (also das Befolgen der Anweisungen, das sorgfältige und gewissenhafte Bearbeiten der Arbeitsaufträge, sowie das Zusammenarbeiten während der Partnerarbeit) wird aber natürlich gewertet.

Die **Datenerhebung** wird voraussichtlich in drei aufeinanderfolgenden Englischstunden im **Februar/März** erfolgen. Selbstverständlich werden alle Daten vertraulich behandelt und **anonymisiert**.

Ich freue mich sehr darüber mit Ihren Kindern diese Studie durchführen zu können und hoffe, aufschlussreiche Ergebnisse sammeln zu können.

Sollten Sie Fragen oder Anregungen haben, können Sie mich gerne per E-Mail kontaktieren.

Mit freundlichen Grüßen

Katja Resch, BEd

Appendix 4: Vocabulary Knowledge Quiz I & II - blank

No	Word/Phrase	I don't remember having seen this word before. 1	I have seen this word before but I don't know what it means. 2	I have seen this word before and I think it means ___ (= translation). 3	I know this word. It means ___ (= translation). 4	I know that this word means ___ (=translation) AND I can use this word in a sentence. 5	translation	Vocabulary Knowledge Quiz I Name: _____ Class: _____ Date: _____ sentence	points
00	dog					X	Hund	My dog is called Rocky.	
0	to stroll		X						
1	childish								
2	get on well with someone								
3	be honest with someone								
4	break up with someone								
5	sulk								
6	have an argument with someone								
7	to be kind								
8	jealous								
9	solve one's problems								
10	disappointed								
11	fall out with someone								
12	introduce								
13	make fun of someone								
14	childhood								
15	make up with someone								

		I don't remember having seen this word before.	I have seen this word before but I don't know what it means.	I have seen this word before and I **think** it means ___ (= translation).	I know this word. It means ___ (= translation).	I know that this word means ___ (=translation) **AND** I can use this word in a sentence.		**Vocabulary Knowledge Quiz II** **Name:** _____ **Class:** _____ **Date:** _____	**points**
No	**Word/Phrase**	**1**	**2**	**3**	**4**	**5**	**translation**	**sentence**	
00	dog					X	*Hund*	*My dog is called Rocky.*	
0	to stroll		X						
15	make up with someone								
14	childhood								
13	make fun of someone								
12	introduce								
11	fall out with someone								
10	disappointed								
9	solve one's problems								
8	jealous								
7	to be kind								
6	have an argument with someone								
5	sulk								
4	break up with someone								
3	be honest with someone								
2	get on well with someone								
1	childish								

Appendix 5: Writing prompt

Name(s): _____ Class: _____ Date: _____

WRITING TASK:

Write an **EMAIL** to your cousin Claire who lives in California. Tell her what has happened between two of your best friends ***Alice*** and ***Dan***. **Describe their friendship** including the following bullet points:

- ☐ Alice & Dan = beste Freunde
  - ☐ kennen sich seit der Kindheit
  - ☐ verstehen sich sehr gut
  - ☐ sind ehrlich zueinander
  - ☐ machen sich niemals lustig über den anderen
  - ☐ versuchen immer nett zueinander zu sein
- ☐ Letzten Monat: großer Streit
  - ☐ Alice kam mit Kyle zusammen
  - ☐ Alice hat Dan ihren Freund Kyle nicht vorgestellt
  - ☐ Dan war eifersüchtig und schmollte eine Zeit lang
- ☐ Letzte Woche: Alice und Kyle trennten sich
  - ☐ Alice entschuldigte sich bei Dan
  - ☐ Dan versprach Alice, dass er in Zukunft mit ihr über seine Probleme reden wird, damit sie die Probleme gemeinsam lösen können
  - ☐ Alice & Dan versöhnten sich
- ☐ Jetzt: wieder beste Freunde
  - ☐ darüber bist **du** froh

Mind the **signal words** for the correct **verb forms**:

➔ **Present Simple, Present Perfect, Past Simple** and **Future**.

Include typical **phrases of an email**.

- ➔ You **MUST NOT use a dictionary**, **the internet** or any other help.
- ➔ The whole text **MUST be written in English**.
- ➔ If you don't know the exact words/phrases in English:
  - try to use synonyms (= andere Wörter mit einer ähnlichen Bedeutung) or
  - paraphrase the words (= umschreibe die Wörter).
- ➔ You've got **30 minutes** to complete the task.

Appendix 6: Feedback form

IQES online

Feedback & Reflexion Schreibaufgabe

1 - Ich habe die Schreibaufgabe

- ○ alleine gelöst.
- ○ mit einem Partner/ einer Partnerin gelöst.

2 - Für das Lernen von Vokabeln empfand ich die Schreibaufgabe als

	nicht	wenig	eher	sehr
2.1 - nützlich	○	○	○	○
2.2 - motivierend	○	○	○	○
2.3 - schwierig	○	○	○	○
2.4 - vergnüglich	○	○	○	○

3 - Ich habe das Gefühl, dass ich durch die Schreibaufgabe

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft zu
3.1 - meinen englischen Wortschatz erweitert habe.	○	○	○	○
3.2 - die Vokabeln nun besser/korrekter anwenden kann.	○	○	○	○

4 - Texte verfassen mithilfe von WORD

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft zu
4.1 - Ich bemerke die Korrekturvorschläge (z.B. wenn Wörter rot/blau unterwellt sind) meist.	○	○	○	○
4.2 - Ich verbessere die rot/blau unterwellten Wörter meist.	○	○	○	○
4.3 - Ich verstehe meist, warum die unterwellten Wörter korrigiert werden müssen.	○	○	○	○
4.4 - Durch die Korrekturvorschläge bzw. durch mein Verbessern der Wörter lerne ich dazu.	○	○	○	○

5 - Welche Aspekte haben dir am Schreibauftrag (nicht) gefallen oder hast du als (nicht) hilfreich empfunden? Z.B. in Hinblick auf Angabe, Medium (Word Datei), Austausch mit Partner/Partnerin... Hast du Verbesserungsvorschläge?

Schreibe eine kurze Antwort - Stichwörter reichen.

Appendix 7: Language background survey

IQES online

Background Info English

1 - Name*

2 - Gender*

- ○ Male
- ○ Female
- ○ Diverse

3 - Age*

4 - What is/are your first language/s (=Erstsprache(n)/Muttersprache(n))?*

- ☐ German
- ☐ English
- ☐ Polish
- ☐ Hungarian
- ☐ Italian
- ☐ Other

If "other": which? _____

5 - Which further languages do you speak? (Welche weiteren Sprachen sprichst du noch neben deiner Erstsprache?)*

- ☐ German
- ☐ English
- ☐ Polish
- ☐ Hungarian
- ☐ Italian
- ☐ Other

How well do you speak it?

_____

6 - Have you got relatives (=Verwandte) or friends who are English native speakers/ who only speak English with you?*

- ○ Yes
- ○ No

If "yes": how often do you speak with each other? (once a day/ week/ …)

_____

7 - Have you been to an English speaking country?*

- ○ Yes
- ○ No

If "yes", which country, when and for how long? (e.g. Canada in 2016 for 3 months)

_____

8 - Outside of school: How many hours per week do you*

	0	1	2	3	4	5	more than 5
8.1 - listen to English?	○	○	○	○	○	○	○
8.2 - read English?	○	○	○	○	○	○	○
8.3 - speak English?	○	○	○	○	○	○	○
8.4 - write in English?	○	○	○	○	○	○	○

9 - Outside of school, I'm mostly in contact with English when...*

e.g. ... I play an online computer game.

10 - In general I enjoy working with the English language...*

- ○ Not at all
- ○ A little bit
- ○ Quite a lot
- ○ Very much

11 - I would enjoy English more if... *