Univerzita Karlova

Pedagogická fakulta

Katedra anglického jazyka a literatury

BAKALÁŘSKÁ PRÁCE

Vztah cizojazyčných předpokladů a studijních výsledků v anglickém jazyce u žáků a žákyň středních škol s důrazem na potenciální genderové rozdíly

The relationship between foreign language aptitude and learning outcomes in English in Czech high school students, with an emphasis on potential gender differences

Adam Novák

Vedoucí práce: PhDr. Petra Kacafírková, Ph.D.

Studijní program: Specializace v pedagogice

Studijní obor: Anglický jazyk se zaměřením na vzdělávání – Matematika se zaměřením na vzdělávání

2024

Odevzdáním této bakalářské práce na téma *The relationship between foreign language aptitude and learning outcomes in English in Czech high school students, with an emphasis on potential gender differences* potvrzuji, že jsem ji vypracoval pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále potvrzuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

V Praze dne 9. července 2024

ABSTRAKT

Tato bakalářská práce se zabývá cizojazyčnými předpoklady českých žáků a žákyň středních škol a jejich vztahem ke studijním výsledkům v anglickém jazyce. Zatímco teoretická část seznamuje čtenáře se základními koncepty z oblasti cizojazyčných předpokladů, část empirická vychází z výzkumu provedeného pomocí dvou testových baterií, které byly administrovány 57 studujícím (25 žáků, 32 žákyň) gymnázia. Jedním z hlavních cílů práce bylo prozkoumat souvislosti mezi cizojazyčnými předpoklady studujících a jejich studijními výsledky v anglickém jazyce, operacionalizovanými jako jejich známky ve škole a rozsah jejich písemné receptivní slovní zásoby. Dále byl zkoumán vztah genderu a cizojazyčných předpokladů studujících. Získané výsledky naznačují, že některé z komponent cizojazyčných předpokladů studujících by s rozsahem jejich anglické slovní zásoby mohly souviset. Vztah mezi cizojazyčnými předpoklady a známkami studujících oproti tomu nebyl nalezen. Zatímco mezi studenty a studentkami významný statistický rozdíl v úrovni jejich cizojazyčných předpokladů zjištěn nebyl, v porovnání se svými spolužačkami vykazovali studenti vyšší rozsah písemné receptivní slovní zásoby. Kromě toho se také ukázalo, že všichni studující vnímají znalost anglického jazyka jako důležitou do života. Na závěr studující hodnotili svou úroveň anglického jazyka. Většina z nich uváděla úroveň B1 či B2, přičemž v tomto sebehodnocení nebyly mezi studenty a studentkami nalezeny významné rozdíly.

KLÍČOVÁ SLOVA

ABSTRACT

The bachelor's thesis deals with foreign language aptitude of Czech grammar students and its relationship to their learning outcomes in English. While the theoretical part introduces the basic concepts in the area of foreign language aptitude, the empirical part is based on field research conducted using two test batteries, which were administered to the sample of 57 grammar school students (25 men, 32 women). The primary aim of the research was to examine whether there are any relations between students' foreign language aptitude and their learning outcomes in English, operationalised as their school grades and the size of their written receptive vocabulary. Furthermore, the relationship between students' foreign language aptitude levels and their gender was explored. The results suggest that certain components of students' foreign language aptitude might be related to the size of their vocabulary. By contrast, foreign language aptitude did not turn out to be connected to students' school grades. Moreover, while no significant differences were found between male and female students in terms of their foreign language aptitude levels, male students demonstrated higher written receptive vocabulary knowledge compared to their female classmates. In addition, the results show that all participants perceive the knowledge of English as important in life. Finally, with no gender differences, most students reported their English to be at either level B1 or B2.

KEYWORDS

foreign language aptitude, gender, second language acquisition, learning outcomes

Table of contents

Introduction

With the world becoming a global village, learning a foreign language has become part of almost everyone's life, at least during school years. As early as primary school, Czech children begin to learn English as a second language. Since it is so widespread in today's society, it seems reasonable to address language learning from a scientific point of view. Foreign language acquisition is considered to be a subfield of psycholinguistics, an interdisciplinary scientific discipline, which studies language perception and production, as well as first and foreign language acquisition (Alduais et al., 2022).

For the purposes of this thesis, the broad scope of psycholinguistics has been narrowed down to the role of individual differences in the process of foreign language acquisition. Some learners seem to be more successful at acquiring a foreign language than others and this bachelor's thesis aims to explore some of the potential explanations for these differences. Specifically, the thesis focuses on foreign language aptitude, which could be understood as one's innate readiness for a foreign language learning (Carrol, 1981) or in other words, a certain gift for learning a foreign language.

Firstly, one of the main goals of this paper is to examine potential differences in foreign language aptitude levels between male and female learners of English. Furthermore, the thesis aims to explore the relationships between one's foreign language aptitude and learning outcomes in English, also with an emphasis on gender differences. Learning outcomes will be operationalised as learners' school grades and primarily, the size of their written receptive vocabulary.

As for the structure of the thesis, the theoretical part starts by introducing the discipline of psycholinguistics, primarily the field of second language acquisition. After other factors that influence the acquisition process are presented, existing research into foreign language aptitude is summarized. The main purpose of the empirical part was to examine the above-mentioned relationships between foreign language aptitude and learning outcomes in the Czech environment. Specifically, the sample consisted of Czech grammar school students. After the methodological aspects of the research are introduced, results are presented. The subsequent discussion assesses the potential contributions and limitations of the study, while making certain recommendations for further research.

THEORETICAL PART

1 Second language acquisition

1.1 Broader context of second language acquisition

1.1.1 Psycholinguistics

Scholars have been interested in the human mind and language for hundreds of years. According to Altmann (2006), it is in the ancient Egypt, where the earliest references to language and the brain can be found. The ancient Egyptians were the first to record a case of aphasia. Psycholinguistics has its empirical roots in the latter decades of the 18th century as well as in the 19th century (Levelt, 2012) and its further development is tightly linked to names such as F. Gall, C. Wernicke, P. Broca, W. Wundt, F. de Saussure or N. Chomsky (Alduais et al., 2022). It was not until the year 1936, however, that the term *psycholinguistics* was created by an American psychologist J. Kantor. The term was popularised in 1946 by Kantor's student N. Pronko (Levelt, 2012). Five years later, in 1951, the Interdisciplinary Summer Seminar in Psychology and Linguistics took place at Cornell University, aiming to examine the possible connections between the area of psychology and linguistics. This event is widely believed to mark the beginnings of modern psycholinguistics (Levelt, 2012).

Psycholinguistics is considered to be highly interdisciplinary science, connecting fields such as linguistics, psychology, biology and neuroscience, but also cognitive and computer sciences (Alduais et al., 2022). The scope of today's *psycholinguistics* was well defined by Warren (2013), who states that it is "the study of the mental representations and processes involved in language use, including the production, comprehension and storage of spoken and written language" (p. 4). In addition to the already mentioned, it is also the field of first and second language acquisition that is studied by psycholinguistics (Alduais et al., 2022).

1.1.2 Second language acquisition

Second language acquisition (*SLA*), a subarea of psycholinguistics, is according to Ortega (2009) "the scholarly field of inquiry that investigates the human capacity to learn languages other than the first, during late childhood, adolescence or adulthood, and once the

first language or languages have been acquired" (p. 2). It is necessary to distinguish the field of second language acquisition from the area of bilingual acquisition, which can be defined as "the process of learning two or more languages relatively simultaneously during early childhood – that is, before the age of four" (p. 4). Taking the above-mentioned definitions into account, the key difference resides in the age at which a person starts learning the language.

Finally, it is also important to mention that throughout this thesis, the terms 'second' and 'foreign' language will be used interchangeably, and they will refer to, as Ortega (2009) suggests, "any language learned after the L1 (or L1s)" (p. 5).

1.2   Factors influencing second language acquisition

Individuals differ greatly in the pace with which they learn a second language as well as in the ultimate level they achieve in that language. This disparity can be attributed to a wide variety of factors, which play diverse roles in the process of second language acquisition. These variables include the age of the learners, their mother tongue and linguistic environment, their foreign language aptitude and motivation as well as their personality traits such as cognitive styles and learning strategies, or the degree of extraversion and foreign language anxiety (Dörnyei, 2006; Ortega, 2009; Skehan, 1991). The following section of the bachelor thesis will explore some of these factors in greater detail.

1.2.1   Age

First language acquisition

One of the frequently debated individual differences in the process of learning a second language is the age of the learner (Singleton, 2001). There is considerable evidence supporting the existence of sensitive periods for first language acquisition. It is of crucial importance that children are exposed to language until they are six or seven years old, otherwise they might not learn to speak at all (Goldin-Meadow, 1982, as cited in Nolen-Hoeksema et al., 2012). Šulová (2019) explores the importance of the presence of parents until the fourth year of a child. Their absence may have an irreversible impact on the child's

language development. These children show poorer level of vocabulary, worse understanding of written text and lower ability to create rhymes.

These and other findings strongly support the existence of sensitive periods for first language acquisition. A *sensitive period* can be defined as an optimal period for the development in a certain area. If this area is not stimulated on time, i.e., in the sensitive period, the specific skill or behaviour might not ever develop to its full potential (Nolen-Hoeksema et al., 2012).[1]

Second language acquisition

Similarly, one might expect that there could also be a certain sensitive period for the acquisition of a second language. One of the areas researchers mostly focus on is the relationship between age and the acquisition of morphosyntax, i.e., the combination of morphological and syntactic aspects of language. These studies focus on people who started living in a foreign country at a different age and therefore, the age at which they began to learn a second language varied. Ortega (2009) states that "the accumulated findings suggest that, by and large, learners who began acquiring the L2 before a certain age, which these studies locate to be around puberty, will tend to exhibit intuitions that are very close to those of native speakers of that language" (p. 19). By contrast, the intuitions of those who arrived in the foreign country after puberty and therefore, started learning the L2 later in their life, will most probably not reach the native-like levels (Ortega, 2009). Johnson and Newport (1989) conducted a study of 46 Chinese and Korean L1 speakers, who moved to US at different age (ranging from 3 to 39). A high correlation (r = -.77) was found between age of arrival and test score, which supports the hypothesis that older learners acquire morphosyntactic rules with more difficulty.

It is also the learners who begin to learn the second language before puberty that generally demonstrate higher ultimate level of phonological competences. By contrast, those who start learning a new language later in their life are more likely to develop a foreign

---

[1] The notion of sensitive periods is not restricted only to the field of language and speech, in the first year of life, for example, it is essential that a very close relationship is built between children and their caregivers (Rutter et al., 1990, as cited in Nolen-Hoeksema et al., 2012).

sounding accent (Ortega, 2009; Scovel, 2000). While Scovel (2000) discusses the existence of a potential sensitive period tightly associated with brain plasticity, according to Flege (1999), the phonetic performance of older learners tends to be worse since they perceive the sounds of the new language through a prism of their already stabilized sound system of their first language. Their experience with their first language is considerably higher than the experience of younger learners, and therefore it is harder for them to perceive and produce the unknown sounds.

Although the above-mentioned arguments strongly support the idea that age works in favour of younger learners, infrequent exceptions can be observed. Even people who start learning a second language later in their life can get very close to the performance of native speakers (Bongaerts et al., 1997; Johnson & Newport, 1989). On the other hand, not all early starters reach native-like levels in their second language (Hyltenstam & Abrahamsson, 2000). When considering these exceptions, it is important to realise that language acquisition is a highly complex process influenced by a variety of other factors such as learners' first language, their motivation, cognitive styles, and other personality traits. The age of a learner, therefore, is apparently not the only variable, but rather one piece of a mosaic, which as a whole determines one's ultimate performance in a second language.

1.2.2 Personality

One of the factors which have a substantial impact on the process of learning a second language is personality, which according to American Psychological Association (2018a) refers to "the enduring configuration of characteristics and behaviour that comprises an individual's unique adjustment to life, including major traits interests, drives, values, self-concept, abilities, and emotional patterns". By its definition, personality structure to a large extent influences behaviour and therefore, a tight connection between one's personality and performance in a second language cannot be denied.

Motivation

American Psychological Association (2018b) defines motivation as "the impetus that gives purpose or direction to behaviour". Gardner and Masgoret (2003) conducted a meta-analysis exploring 75 independent samples with 10 489 individuals in total. Across the studies included in the meta-analysis, three different measures of second language

achievement were used – self-ratings, objective measures, and grades. The instrument used to assess the learners' motivation was the Attitude/Motivation test Battery (Gardner, 1985). The AMTB is concerned with 3 dimensions, namely "the amount of effort the individual expends in learning the language; the extent to which the individual wants to achieve a high level of competence in the language; affect experienced while learning the language" (Gardner & Masgoret, 2003, p. 128). Although the three different measures of second language achievement correlated similarly with the construct of motivation, a slight variance was found. The highest correlation was measured between motivational variables and self-ratings, with the mean corrected correlation r = .39. It was slightly lower in the relationship between motivation and grades (r = .37), and the correlation with objective measures was the lowest (r = .29), yet still considerably similar (Masgoret and Gardner, 2003). Taking these findings into account, it is evident that the role of motivation is also significant in the context of second language learning.

Foreign language anxiety

Another highly influential factor in the process of second language acquisition is the level of one's anxiety. It is important to differentiate generally anxious people from those with specific anxiety reactions, i.e., for example, in the context of second language acquisition (Horwitz et al., 1986). Gardner and MacIntyre (1994) define foreign language anxiety as "the feeling of tension and apprehension specifically associated with second language contexts, including speaking, listening, and learning" (p. 284).

All the above-mentioned, Horwitz, Gardner and MacIntyre, contributed to the development of widely used instruments for measuring anxiety in the context of second language acquisition. Horwitz et al. (1986) presented the Foreign Language Classroom Anxiety Scale, whereas MacIntyre and Gardner are the authors of another instrument, the Input, Processing and Output Anxiety Scales (1994). The name of the later instrument indicates that not only does anxiety affect the output stage, i.e., the production of the earlier acquired material, but it is also involved in the input and processing of the second language (MacIntyre & Gardner, 1994).

Horwitz et al. (1986) mention a variety of symptoms which can be connected to the concept of foreign language anxiety. These students find it hard to speak in class and they

can have trouble concentrating. They might as well postpone doing homework or even avoid classes. Some of them experience fear of negative evaluation, while others become forgetful. They can also experience palpitation or excessive sweating. Moreover, the authors mention test anxiety as another frequently occurring phenomenon. As a result of all of these symptoms, the process of learning a second language can be substantially exacerbated.

On the other hand, interestingly, a certain degree of concern can also have a positive influence on the learning process. For example, Suter (1976) found that the level of learners' concern about their pronunciation of English significantly correlates (r = .46) with pronunciation accuracy.

Self-efficacy

Apart from the previously mentioned personality characteristics, namely motivation and foreign language anxiety, there is a plethora of other factors that influence the second language acquisition process. Albert Bandura's concept of self-efficacy, which can be defined as "an individual's subjective perception of their capability to perform in a given setting or to attain desired results" (American Psychological Association, 2023a), is one of these factors. In their meta-analysis of 109 studies, Robbins et al. (2004) found that self-efficacy was an important predictor of academic success, showing moderate correlation (r = .496) with GPA (grade point average). It is also in the context of second language acquisition that self-efficacy plays a significant role. For example, the results of the study conducted by Mills et al. (2006) suggest that there is a positive correlation between reading self-efficacy and reading proficiency.

Cognitive styles

According to Plháková (2023), cognitive style is a unique way in which people process information. American Psychological Association (2023b) defines it as "a person's characteristic mode of perceiving, thinking, remembering, and problem solving". Although there is a wide variety of cognitive styles, the attention will be drawn only to the difference between field dependence and field independence, a cognitive style often examined in the context of second language acquisition (Hoffman, 1997). Plháková (2023) explains that field dependent people are likely to perceive the world around them as a whole, while field

independent ones tend to perceive the surrounding objects separately or in other words, they "look at their world analytically" (Hoffman, 1997, p. 223).

According to Johnson et al. (2000), field dependent learners might be advantaged in some aspects of L2 learning. For example, more field dependent learners of English scored higher when their oral pragmatic abilities were assessed by their ESL teachers. On the other hand, contrary to Johnson et al.'s assumption, field dependence-independence did not prove to be associated with metaphor fluency (2000). Similarly, Hoffman (1997) remarks that field dependent learners show greater social perceptiveness and sensitivity. On the other hand, "field independents tend to outperform field dependents on foreign language proficiency tests in which many of the exercises focus on discrete points of the target language" (p. 224).

Based on the aforementioned observations, some field dependent second language learners could be expected to profit from their potentially better pragmatic and social skills, whereas fields independents might be more likely to approach the second language more analytically, which can be perceived as an advantage, for example, in the context of learning grammar. This summary, however, is rather simplistic and as Hoffman suggest, further research is needed in this field (1997).

Having examined some of the relevant personality variables, it is important to realise that there are also other potentially influential individual personality differences. Dewaele (2012), for instance, discusses the possible role of extraversion and introversion, conscientiousness, or creativity in the process of second language acquisition.

1.2.3 Cognition

In a similar manner that the previous part explored the impact of some of the most important personality characteristics on the process of second language acquisition, this chapter draws attention to the relationship between cognition and language acquisition. American Psychological Association (2018c) defines cognitive processes as "any of the mental functions assumed to be involved in the acquisition, storage, interpretation, manipulation, transformation, and use of knowledge". Cognitive processes include perception, memory, imagination, learning and problem solving (Plháková, 2023).

Memory

Even though all cognitive processes, by definition, influence the process of language acquisition, only the concept of memory will be explored here. It is important to distinguish between long term and working memory. The construct of long-term memory refers to a repository of an immense amount of knowledge (Plháková, 2023). Ortega (2009) focuses on the area of vocabulary, and she presents three important properties of the knowledge stored in long-term memory – its strength, size, and depth. According to her, "strength concerns the relative ability to use a given known word productively or to recognize is passively" (p. 88) and it is connected to the gradual automatization and acquisition of the knowledge. The second quality she describes is the size of the mental lexicon, which can be understood as the total quantity of vocabulary in the long-term memory storage. The last important property is vocabulary depth, and it refers to "how elaborated, well specified and structured the lexical representations are" (p. 88). It is intuitive that all the above-mentioned properties of the stored knowledge influence the way people use and understand vocabulary. For example, Enayat and Derakhshan (2021) showed that the size of receptive vocabulary significantly predicts the second language overall speaking proficiency.

Working memory, on the other hand, functions as a temporary storage, which allows the currently perceived information to be processed or used (Plháková, 2023). Interestingly, the capacity of working memory in the L2 seems to be inferior to the one in the L1 (Harrington & Sawyer, 1992; Ortega, 2009) and this disparity is thought to be decreasing as learners get more proficient in the language, nevertheless, there is not much empirical background behind this assumption (Ortega, 2009).

Although some relevant information about the relationship between memory and language acquisition have been mentioned, the concept of memory will explored in more detail in the subsequent chapter, which addresses foreign language aptitude.

1.2.4 Crosslinguistic influence

Apart from all of the other aforementioned influential factors, learners' first language also plays a significant role in the process of foreign language acquisition. According to Odlin (2012), the term *crosslinguistic influence* "refers to the influence of one language upon another" (p. 1) and it is considered to be roughly synonymous with the phrase *language*

transfer, which can be further divided into two parts - *negative* and *positive transfer* (Odlin, 2012). Therefore, it is important to realise that one's first language can both facilitate and inhibit the second language acquisition. Intuitively, the occurrence of either positive or negative transfer can be expected to be determined by the degree of similarities and differences between the two languages.

Krashen (1981) mentions complex word order and word-for-word translation of phrases as two of the main areas in which the crosslinguistic influence occurs. Dušková (1969, p. 16) illustrated this interference by the case of learners of English whose first language was Czech. According to her, the word order errors included "placing the direct object after an adverbial modifier, e.g. *I met there some Germans (I met some Germans there)*" or "placing a temporal modifier before a local one, e.g. *I returned last month from P. (I returned from P. last month)*". She claims that clear associations can be seen between these incorrect translations and Czech word-order structure. Moreover, she also adds different examples of mistakes caused by word-for-word translations such as "*the journey liked me (cesta se mi líbila – I liked the journey)*", "*another my friend (jiný můj přítel – another friend of mine)*" or "*there are many other cases to solving (případy k řešení – cases to be solved)*" (p. 17-18).

All of the previous examples clearly illustrate that the concept of crosslinguistic influence can by no means be omitted when considering the wide variety of factors influencing the process of foreign language acquisition.

1.2.5 Gender

With the role of gender in the process of foreign language acquisition being one of the research objectives of this thesis, it is important to draw attention to some similarities and potential differences between male and female learners. [2]

According to Hyde's (2005) *gender similarities hypothesis*, "males and females are alike on most – but not all – psychological variables" (p. 590). In this spirit, it could be expected that girls (women) and boys (men) will not significantly differ in language

---

[2] For the purposes of this bachelor thesis, gender has been reduced to two categories – male and female. Nevertheless, by no means does this reduction deny the existence of other gender identities.

acquisition. Nevertheless, there is considerable evidence supporting the existence of certain differences. For example, Heyder et al. (2017) focus on boys' underachievement in language grades. They mention students' ability self-concept, task values, prior achievement and also their socializers' beliefs as possible causes of the potential differences between boys and girls, since all of these factors might be substantially conditioned by the student's gender. It is important to add that this study focused on the acquisition of first language, in particular on the grades of German learners of German language. However, with a certain degree of caution, these findings could perhaps be, based on the similar nature of processes involved in language learning generally, to some extent generalised also to foreign language acquisition.

Moving specifically to the context of foreign language learning, Oga-Baldwin and Fryer (2020) conducted research in Japan exploring the influence of gender on students' motivation for learning a foreign language. They suggest that whereas male students were more likely to use externally driven motives, it was intrinsic motivation that was more frequently demonstrated by girls. The authors interpret these conditions as disadvantageous for male learners.

Another interesting study was conducted by Bernat and Lloyd (2007). They examined the relation between the gender of EFL[3] learners' and their beliefs about language learning, which concerned, for example, areas such as foreign language aptitude, difficulties of language learning, or learning and communication strategies. Their beliefs differed only in two items (out of 34) of the used instrument, with one of them being "*People who speak more than one language are very intelligent.*" (p. 83). Female learners significantly more often than males agreed with this statement. The second of these items focused on the experienced enjoyment when practising English with native speakers. Although with only minor difference, it was found that male students enjoyed it more than female students. Despite these two differences, male and female learners overall showed similar opinions on language learning. Albeit these findings are interesting, Bernat and Lloyd (2007) believe that further evidence is necessary before generalising the obtained results. Additional research is

---

[3] English as a Foreign Language

needed not only to explain the relationship between gender and learners' beliefs, but also to shed more light on the general topic of gender in foreign language acquisition.

2 Foreign language aptitude

Having briefly outlined some of the most prominent factors which affect the process of foreign language acquisition, the reader's attention will be now drawn to the focal subject of this thesis – the concept of foreign language aptitude and its potential influence on the learners of foreign language.

2.1 Introduction to the concept of foreign language aptitude

2.1.1 The concept of foreign language aptitude and its components

Foreign language aptitude is believed to be a "set of abilities particularly useful for the learning of second languages" (Skehan, 2014, p. 367). In other words, it could be understood as a certain gift for L2 learning. Finally, J. B. Carroll (1981), the imaginary father of the concept of foreign language aptitude, defines it as "an individual's initial state of readiness and capacity for learning a foreign language, and probable facility in doing so given the presence of motivation and opportunity" (p. 86). Based on the aforementioned, people with higher aptitude levels might be expected to encounter less difficulties and therefore be possibly faster and more successful in the process of second language acquisition.

Although foreign language aptitude is considered to be connected to cognition, i.e., "distinct from affective and conative variables" (Li, 2015, p. 387), it is clearly distinguishable from general cognitive abilities since it is specifically associated with language learning (Skehan, 2014). Addressing the origin of foreign language aptitude, Skehan (2002) states that "Carroll (1973) was neutral as to whether aptitude is largely fixed at birth, or is the result of early experience" (p. 79). Regardless of when it is formed, aptitude is believed to be relatively stable, it is not subject to major changes. It can, however, be to a certain extent influenced by language learning experience (Li, 2022; Wen, 2011). Lastly, it is essential to realise that foreign language aptitude is perceived as a multidimensional

construct consisting of several subcomponents (Skehan, 2014), all of which play an important role in second language acquisition.

Probably the most influential composition of the aptitude construct was introduced by already mentioned J. B. Carroll. He speaks of 4 components – *phonetic coding ability*, *grammatical sensitivity*, *inductive language learning ability* and *associative memory* (Carroll, 1973; Sparks et al., 2011). *Phonetic coding ability* refers to "the ability to identify, and store in long-term memory, new language sounds or strings of sounds" (Carroll, 1973, p. 7). Both *grammatical sensitivity* and *inductive language learning ability* are tightly connected to the language structure. The former refers not only to one's sensitivity to the syntactic and grammatical functions of words in a sentence but also to the sensitivity to the syntactic patterns of larger linguistic units (Carrol, 1973). People with higher levels of *inductive ability* are more successful at generalising and inductively inferring rules from auditory or printed language material (Carroll, 1973; Sparks et al., 2011). According to Wen (2011), the fourth factor identified by Carroll, namely *associative memory*, refers to "the ability to rote learn vocabulary items paired with their associated translations" (p. 233).

2.1.2 MLAT

The above-mentioned components, except for the inductive ability, have been traditionally measured by the Modern Language Aptitude Test (MLAT), which was introduced by Carroll and his colleague Sapon in 1959 (Sparks et al., 2011). MLAT consists of five subtests. Phonetic coding ability is measured by subtests 1, 2 and 3, i.e., *Number Learning*, *Phonetic Script*, *Spelling Clues*, respectively. *Words in Sentences*, the fourth part of the test, focuses on Grammatical sensitivity. The memory component is measured by the fifth part of the MLAT called *Paired Associates*. (Li, 2015; Ortega, 2009; Sparks et al., 2011) Since the empirical part of this thesis is not based on the MLAT, it is not necessary to further examine the individual parts of the test.

In terms of psychometric properties of the test, Sparks and Patton (2013) refer to Carroll and Sapon, who reported a test-retest reliability of the Long Form of the MLAT of .90 (ninth-grade males) to .91 (ninth-grade females). The MLAT also proved to be an instrument with high predictive validity, mainly in the context of formal instruction setting (Ortega, 2009). Skehan (2002) speaks about its correlations of r = .4 to r = .6 with language

course performance. The success of MLAT can be further confirmed by the fact that many of later developed language aptitude measures, some of which will be mentioned in the following chapters, were influenced by its structure (Singleton, 2017). However, as far as the construct validity, which can be understood as the extent to which an instrument measures the theoretically established construct (American Psychological Association, 2018d), is concerned, the MLAT has been criticised. This critique includes the fact that only three of the four theoretical components of foreign language aptitude are measured in the test, inductive ability was not covered (Ortega, 2009).

2.1.3 More recent concepts of foreign language aptitude

Although Carroll's contribution to this area of psycholinguistic is immense and undeniable, it was not only the structure of the MLAT that has been questioned over the years. His concept of foreign language aptitude itself has not escaped criticism either and to a certain extent, it has been challenged repeatedly.

Different perspective was taken, for example, by Peter Skehan (2002), who relates the concept of foreign language aptitude to the processes which are involved in second language acquisition. These processes can be understood as cognitively various stages of language learning and different aptitude components are important for each of them. According to Skehan (2002), language learning might include the following processes – *noticing* (1), *pattern identification* (2), *extending* (3), *complexifying* (4), *integrating* (5), *becoming accurate* and *avoiding error* (6), *creating a repertoire* and *achieving salience* (7), *automatising rule-based language* and *achieving fluency* (8), and *lexicalising* and *dual-coding* (9).

After the introduction of these elementary stages, he divides them into four larger clusters, namely *noticing* (1), *patterning* (2-5), *controlling* (6-8) and *lexicalising* (8, 9). In Skehan's words, noticing is "the first insight (within any particular domain) that some aspect of form is worth attention" (p. 91). He adds that besides phonemic coding ability, aptitudinal components of working memory and attentional management are relevant in this phase.

The stage of patterning is nicely summarized by Ortega (2009):

During patterning hypotheses about what has been noticed are implicitly or explicitly made, tried out, revised, extended to relevant cases and eventually a final resulting generalization (that is, a new 'rule') is integrated into the existing knowledge of the L2, which thereby undergoes restructuring. (p. 163)

The foreign language aptitude components influencing the process of patterning undoubtedly include Carroll's grammatical sensitivity and inductive learning ability. However, according to Skehan (2002), these two aptitude components cannot explain the restructuring of language, i.e., extending or modifying the already existing knowledge, which also takes place in the phase of patterning. Therefore, he admits that further work might be needed for the purposes of coping with this problem.

During the following learning process, controlling, learners are able to avoid majority of errors, their use of language gradually becomes faster and more accurate. They reach considerable automaticity in the language. According to Skehan, aptitude abilities in this phase are connected to proceduralisation and retrieval processes.

Lexicalisation, the last of the mentioned stages, could be understood as a certain continuation of the preceding stage. Learners demonstrate even more fluency and automaticity, their access to a wide variety of linguistic expressions is faster and smoother. Using foreign language is overall substantially cognitively easier than it was before, and the learners achieve the highest levels of proficiency. One of the crucial aptitude components for the stage of lexicalisation is memory, specifically storage and most importantly retrieval processes, neither of which is sufficiently focused on in the MLAT (Skehan, 2002). In the context of this stage, Skehan mentions chunking as another aptitude strength.

It was not only Peter Skehan but also Peter Robinson who offered an alternative to the Carroll's model of foreign language aptitude. Similarly as Skehan, Robinson takes into consideration different stages and contexts of learning when referring to the concept of foreign language aptitude (2005). Moreover, he also emphasises the importance of combination of various factors and abilities in the process of foreign language acquisition. Wen (2011) concludes that the perspectives of both Skehan and Robinson "demonstrate the potential to go beyond traditional FLA research (which relied heavily on the predictive

power of aptitude scores)" (p. 234), while they both offer a more dynamic view on the foreign language aptitude concept.

2.1.4 Instruments used for measuring aptitude

Along with new theoretical concepts, which emerged even decades after Carroll's publication of his insights into foreign language aptitude, new instruments measuring learner's aptitude were developed. In this chapter, some of them will be briefly introduced.

One of the most used language aptitude measures is Meara's LLAMA (Meara & Rogers, 2020h). Since the empirical part of this thesis is largely based on the LLAMA battery, this instrument will be in more detail described later.

Another aptitude measurement battery is Pimsleur's Language Aptitude Battery (PLAB), designed by Pimsleur in 1960s. Skehan (2002) states that the instrument was created to be administered to high-school students in order to assess their auditory skills as well as their inductive language learning ability. Besides these factors, PLAB also collects the GPA, motivation, and vocabulary of the learner (Li, 2022).

Some of the newly developed instruments are based on the findings of cognitive psychology, as in the case of the CANAL-F test and the HiLAB battery (Skehan, 2014). The theory behind the former, i.e., the Cognitive Ability for Novelty in Acquisition of Language (Foreign) (Grigornko et al., 2000), is based on the individual differences in the knowledge acquisition processes (selective encoding, accidental encoding, selective comparison, selective transfer and selective combination), levels of processing (the lexical, morphological, semantic and syntactic level) and also in the modes of input and output (visual and oral). Finally, the authors refer to the processes of encoding, storage, and retrieval of information (Grigornko et al., 2000). The latter mentioned, i.e., the High-Level Language Aptitude Battery (Linck et al., 2013), also substantially draws from cognitive psychology. According to the authors, the Hi-LAB is designed to be administered to very advanced learners. Apparently, that makes the instrument rather unique. The suggested components of high-level aptitude measured by the Hi-LAB include working memory (phonological short-term memory and executive functioning – namely updating, inhibitory control and task switching), associative memory, long-term memory retrieval, implicit learning, processing speed and auditory perceptual acuity (Linck et al., 2013).

Based on the demonstrated development, it can be expected that the field of foreign language aptitude and its measurement will continue to be tightly connected to cognitive sciences even in future, particularly to cognitive psychology.

2.2 Previous research of foreign language aptitude

As new concepts have emerged over the years and the whole area of foreign language aptitude has developed, a wide range of questions have arisen in the field.

2.2.1 Memory

Several discussions have concentrated on the area of memory (Skehan, 2002). As mentioned above, its importance was emphasised already by J. B. Carroll and to these days, it is considered to play a substantial part in the process of second language acquisition (Sparks et al., 2011). Despite the consensus on the importance of memory, different authors stress different kinds or subsystems of memory. Whereas Carroll spoke specifically about associative memory, sometimes also referred to as rote memory (Li, 2022), more recent authors accent the role of working memory. This switch is regarded to be rather reasonable, for example, by Sparks et al. (2011), who believe that working memory might be more important in predicting foreign language proficiency in comparison to rote (associative) memory, traditionally measured by the fifth part of the MLAT. As it was defined earlier in this thesis, working memory functions as a temporary storage, which allows the currently perceived information to be processed or used (Plháková, 2023). Taking Baddeley's perspective (Baddeley & Hitch, 1974), working memory has four subsystems – the phonological loop, the visuospatial sketchpad, the episodic buffer, and the central executive (Plháková, 2023). The Hi-LAB, for example, measures executive functioning and phonological short-term memory, both of which the authors regard to be potential aptitude components (Linck et al., 2013). Although further research is certainly needed in this field, it is believed that the capacity of working memory significantly influences, for example, lexical, as well as grammar acquisition (Singleton, 2017). According to Singleton, working memory plays a crucial role in the area of language aptitude (2017).

Finally, as can be illustrated on the above-mentioned Hi-LAB, the qualities of long-term memory are also often believed to have an impact on foreign learning acquisition.

2.2.2 Distinction between intelligence and foreign language aptitude

Apart from numerous discussions regarding memory, the potential overlap between the concept of foreign language aptitude and intelligence has also been the subject of many research efforts. It is the area of analytic components of aptitude, which has received considerable attention. Skehan (2002) suggests that "the abilities implicated in searching for, and internalising patterns in language are strongly connected with general cognitive abilities" (p. 82). Similar beliefs were implied by Carroll (1990). Moreover, Skehan (2002) summarises the findings of Sasaki (1991, 1996) and states that "it was the language analytic aspects of aptitude which were closest in their relationship to intelligence, while the phonemic coding ability and memory components still showed separation" (p. 77). Although it cannot be denied that there is a certain relation between foreign language aptitude and intelligence, it is important to realise that they play separate roles in the process of foreign language acquisition (Sparks et al., 2011).

2.2.3 Explicit and implicit learning conditions

Apart from the already mentioned, there have also been speculations about the influence of foreign language aptitude under explicit and implicit learning conditions. In other words, researchers have been interested in whether "aptitude matters across all learning conditions, particularly learning with or without intention, attention, awareness and rules" (Ortega, 2009, p. 159). Findings to date suggest that foreign language aptitude is important under both circumstances. It is, for example, Skehan (2002) who claims that one's aptitude characteristics are important in both explicit and implicit contexts of learning. Similar opinion is held Sparks et al. (2011) or de Graaff (1997). Based on his meta-analytic review, in which he focused on the role of foreign language aptitude on grammar acquisition, Li (2015) also implies that aptitude can be influential in both contexts. However, he found stronger correlations for explicit conditions of learning.

2.2.4 Age

Finally, the uncertain relation between aptitude and age should not be forgotten. Whether the importance of aptitude differs based on the learner's age remains to be unclear. For example, DeKeyser (2000) found that language aptitude was predictive of the ultimate success in foreign language acquisition only for adult immigrants, not for childhood arrivals.

These findings might be in favour of the notion that language aptitude is important predominantly after a certain potential critical period. In a similar manner as DeKeyser (2000), Abrahamsson and Hyltenstam (2008) suggest that "a high degree of language aptitude is required if adult learners are to reach a L2 proficiency that is indistinguishable from that of native speakers" (p. 481). However, unlike DeKeyser, they found that aptitude can be also important for child language learners, not only for adults.

Interestingly, based on his meta-analytic review, Li (2015) suggests that foreign language aptitude might be a greater advantage for high school students compared to university students. According to Li, this phenomenon could be potentially explained by the fact that traditionally measured aptitude is more likely to be influential in earlier stages of language learning and since university language learners have more experience with the target language, aptitude might not that important in their case (2015).

2.3 Foreign language aptitude and foreign language acquisition

As it has been outlined above, foreign language aptitude plays an important role in the process of foreign language acquisition. In this chapter, the relationship between the aptitude and outcomes in the foreign language will be explored in more detail.

Although in a rather simplified way, Carroll (1990) summarises the subject matter well when saying that "different levels of ability on aptitude tests correspond to increasing difficulties and longer times to learn whatever has to be learned in acquiring a foreign language" (p. 26). Taking a more statistical perspective, Li (2016) found a correlation of r = .49 between the global construct of aptitude and overall L2 proficiency, which was measured by course grades or standardised tests scores. He adds that the acquired correlation suggests that language aptitude explains around 25 % of the variance of L2 proficiency. Li (2022) asserts that aptitude is, therefore, more predictive of L2 proficiency than other individual differences such as motivation or anxiety.

It is important to realise, however, that the predictive power of aptitude differs for specific foreign language knowledge and skills. Li (2022) summarises the findings of Li (2016) and states that overall aptitude is "a stronger predictor of listening, reading, speaking, and grammar than vocabulary, and a non-significant predictor of writing" (p. 40).

Specifically, there were moderate associations between overall aptitude and reading, speaking, grammar and listening, with r values .39, .37, .32, and .30, respectively. With vocabulary, aptitude correlated with the r value of .15 (Li, 2016). As for the area of grammar, for example, Li's (2016) results are in accordance with his earlier conducted meta-analytic review, where the association between aptitude and L2 grammar learning reached the r value of .31 (Li, 2015).

After a general summary, Li (2016) examines the data in more detail and he presents other interesting findings. Whereas phonetic coding correlated the most with the variable of vocabulary and the least with listening skills, language analytic ability showed, quite intuitively, highest correlations with grammar learning (r = .39) and reading comprehension (r = .35). Based on the correlation with grammar, it makes sense to believe that language analytic ability plays a significant role in the acquisition of morphosyntax. The construct of rote memory showed the lowest predictivity of L2 success among the three foreign language aptitude components, i.e., phonetic coding, language analytic ability and rote memory. This finding could be considered to be in certain accordance with the opinions of Sparks et al. (2011), who question the role of rote memory, while they suggest that more attention should be directed to the concept of working memory. However, in Li's (2016) meta-analysis, although least predictive, rote memory was still a significant predictor of L2 achievement. It is worth noticing that the above-mentioned Hi-LAB (Linck et al., 2013), an instrument measuring aptitude, includes both working and associative (rote) memory.

2.3.1 Foreign language aptitude in vocabulary learning

For the purposes of the empirical part of this thesis, foreign language outcomes will be operationalised as one's grades and primarily, the size of one's vocabulary. Therefore, it is relevant to explore some of the already existing research into the associations between foreign language aptitude and the knowledge of vocabulary in the target language.

Based on their study, Sáfár and Kormos (2008) suggest that foreign language aptitude might play a certain role in vocabulary acquisition. Dahlen and Caldwell–Harris (2013) examined the relationship between foreign language aptitude measured by the MLAT and "the level of success in initial learning of foreign vocabulary" (p. 902). In the study,

participants with higher aptitude scores demonstrated better ability to both recall and recognise the target words in foreign language.

Clearly, both aforementioned studies imply that there is a certain association between foreign language aptitude and vocabulary learning. However, before any conclusions are drawn, it is advisable to explore the existing evidence more comprehensively, which will be done by means of Li's (2016) meta-analysis. As mentioned above, he found that overall aptitude correlated with learners' vocabulary less than with their scores in reading, speaking, grammar and listening, which suggests that aptitude might not be as important for vocabulary learning. Nevertheless, although aptitude as a whole was as well as rote memory weakly predictive of learners' vocabulary scores, the remaining measured aspects of aptitude correlated moderately with the variable of vocabulary. Namely, these aspects comprise phonetic coding, language analytic ability, Number Learning and Spelling Clues[4]. Furthermore, Li (2016) adds that learning vocabulary might potentially be more conditioned by, for example, PSTM[5] or other abilities, which are typically not assessed by conventional aptitude tests. Indeed, the research into working memory showed significant correlations between phonological memory and vocabulary learning (Hummel, 2009). O'Brien et al. (2006), for example, imply that phonological memory might be connected to vocabulary use. Li (2016) suggests that further research is needed to shed more light on the speculations regarding the association between aptitude and vocabulary. Hopefully, the empirical part of this thesis will be one of the many necessary pieces of this mosaic.

2.4 The role of gender in measuring foreign language aptitude

Finally, as far as the question of gender is concerned, it is advisable to examine its influence on foreign language aptitude measures, specifically on Meara's LLAMA (Meara & Rogers, 2020h), which was used in the empirical part of this theses. Rogers et al. (2016) explored all of the 4 subtests of LLAMA and in none of them did female and male participants differ significantly. A year later, another examination of the role of gender in

---

[4] Number Learning and Spelling Clues are subtests of MLAT. According to Li (2016), "theoretically there is no transparent matchup between these two subtests and the three aptitude components" (p. 822). Therefore, they both have their own category in the meta-analysis.

[5] PSTM stands for phonological short-term memory.

the LLAMA tests was conducted by Rogers et al. (2017). Similar findings were reached, and the authors suggest that gender, along with some other mentioned individual variables, does not "consistently affect the overall variance in LLAMA scores" (p. 56). Therefore, the instrument of LLAMA could be considered to be gender neutral. It other words, it can be administered to both male and female learners with no difference. Other characteristics of the LLAMA test will be discussed in more detail in the following chapters.

EMPIRICAL PART

3 Methodology

3.1 Research goals and hypotheses

After the introduction to the field of second language acquisition, the theoretical part primarily focused on the concept of foreign language aptitude. At first, the concept was presented in the context of its historical development. Subsequently, existing research was summarised, with an emphasis on the relationships between foreign language aptitude and learning outcomes.

The empirical part of this thesis follows the theoretical one, with the purpose to explore the above-mentioned relationships in the context of Czech environment. Specifically, the research sample consisted of grammar school students. The main goal of the conducted research was to ascertain, whether and how foreign language aptitude relates to the language performance of Czech learners of English. Based on the findings presented in the theoretical part, a certain positive relationship was expected to exist between the two variables. Therefore, the first hypothesis was formulated as follows.

H1: *There is a positive correlation between foreign language aptitude of Czech high school students and their learning outcomes in English.*

Moreover, the research also aims to explore the potential relationship between students' gender and their foreign language aptitude. Due to paucity of evidence in this area, the following hypothesis has a rather exploratory character.

H2: *Male and female high school students do not differ in their levels of foreign language aptitude.*

Finally, the research focuses on whether the variable of gender moderates the relationship between students' foreign language aptitude and their learning outcomes in English.

3.2 Data collection

3.2.1 Research sample

A total of 57 participants took part in the research, specifically 25 men and 32 women. This sample consisted of either four-year or eight-year grammar school students from Central Bohemian Region. The school was not chosen randomly, the selection was based on a personal acquaintance. To preserve anonymity, the school will not be further specified.

In terms of age, the sample was considerably homogenous. The youngest participant was 16 years old, whereas the oldest one was 18 years old. The mean age was 16.7 (SD = 0.605). Two participants did not record their age.

Majority of the participants (37) were second year students of the four-year programme (3 classes in total). All the remaining 20 students studied the eight-year program. Eight of them were sexta[6] students (1 class), whereas twelve of them were septima[7] students (1 class). The participants came from a total of 5 classes.

3.2.2 Methods

In the conducted research, various measures were used for the assessment of the data.

Foreign language aptitude

The instrument used to measure language aptitude was Paul Meara's LLAMA (Meara & Rogers, 2020h), or more precisely, three of its subparts. The battery was created, with the exception of the LLAMA_D subtest, based on the above mentioned MLAT (Suzuki, 2021). LLAMA is available online[8] and it is free of charge, both of which aspects were of high importance when choosing the measuring instruments for this thesis.

---

[6] Sexta corresponds to the second year of the four-year programme.

[7] Septima corresponds to the third year of the four-year programme.

[8] https://www.lognostics.co.uk/tools/LLAMA_3/index.htm

Learning new words

The development of the LLAMA_B test (Meara & Rogers, 2020a) was based on the MLAT paired associates task and therefore, it is learning vocabulary which is measured in LLAMA_B. This test consists of two parts – learning and testing phase. In the former, the test-takers have two minutes to remember as many as possible names of the twenty objects presented to them simultaneously on screen. The names come from Central American language and they were "arbitrarily assigned to the target images" (Rogers et al., 2016, p. 185). In the testing phase, the test-takers can gain up to twenty points, depending on how many times they correctly connect the name to its corresponding image, i.e., the maximum score is, as in all other LLAMA subtests, 20 points.

Listening for new words

LLAMA_D (Meara & Rogers, 2020b) focuses on sound recognition, and it does not correspond to any of the MLAT subtests. According to the authors, good recognition of words is important in the process of second language acquisition since "if you can recognise that a sequence of sounds is something that you have heard before, then you are going to find it easier to remember, and it will be easier for you to attach a meaning to it" (Meara & Rogers, 2020c, p. 1). In this subtest, the test-takers successively examine 50 short sound items of unfamiliar language. For each of the items, the test-takers have to decide whether they have previously heard it or not. More specifically, if a sound item appears in the test for the first time, the test-takers should click the left button ("new word"), whereas if it is not the first time they have heard the item in the test, they should click the right button ("repeated word").

Grammar rules

LLAMA_F (Meara & Rogers, 2020d) was also used as an aptitude measure in the conducted research. This subtest is believed to assess test-takers' ability to infer grammatical rules (Rogers et al., 2016). In accordance with the concepts defined in the theoretical part, LLAMA_F purports to measure certain analytic components of foreign language aptitude. This subtest consists, similarly as LLAMA_B, of two parts – learning and testing phase. In the first of them, test-takers have five minutes to explore twenty pictures with their description in an artificial language called Patsi (Meara & Rogers, 2020e; Rogers et al.,

2017). The pictures differ, for example, in the number or colour of the depicted objects in them. Test-takers should "work out how the descriptions relate to the pictures", based on which they "should be able to intuit some of the grammatical and morphological features of the language" (Rogers et al., 2017, p. 51). In the second part of the test, the test-takers are presented with ten new pictures and their task is to describe them using the rules they learned in the previous phase.

Sounds and symbols

The last of the subtests is LLAMA_E (Meara & Rogers, 2020f). It focuses on sound-symbol correspondence, as one of the tasks in the MLAT does (Rogers et al., 2016). In other words, the LAMMA_E is believed to assesses the extent to which the test-taker is able to create a link between familiar sounds and the foreign language writing system (Meara & Rogers, 2020g). For reasons of time, this subtest was not included in the conducted research as an instrument for measuring students' foreign language aptitude.

Psychometric properties of the LLAMA battery

In terms of psychometric properties of the LLAMA battery, it is important to mention that the tests have not been extensively standardised and Meara states that they "should not be used in high stakes situations" (2022). However, as indicated, for example, by Suzuki (2021), the tests have been frequently used in aptitude research. Moreover, although there is not a large body of evidence so far, a few studies addressing the psychometric properties of LLAMA have been conducted. Importantly for the purposes of this study, as has been mentioned above, the LLAMA tests are gender neutral (Rogers et al., 2017). What is more, these authors state that the tests seem to be independent of the test-taker's age, first language and level of education. The test has also been explored by Granena (2013, as cited in Artieda & Muñoz, 2016), according to whom LLAMA showed test-retest reliability and furthermore, its internal consistency was approaching an acceptable level of 0.80. As far as validity is concerned, Rogers et al. (2017) believe that the tests "still need to be validated in terms of their ability to predict language learning" (p. 56). Some steps in this respect were made by Bokander and Bylund (2019). With the potential exception of the LLAMA_B subtest, their results show that the battery as a whole might lack sufficient internal validity. As a result, they suggest that researchers should be careful when interpreting data acquired by the

LLAMA tests. Nevertheless, they still see potential in LLAMA, and they emphasise the need of further research into the battery.

Learning outcomes

Learning outcomes have been operationalised as students' final grades as well as their English written receptive vocabulary size.

Grades

Students' grades were collected through a record sheet, in which they were asked to write down their final grades from the previous school year[9]. Two grades were acquired from each student – from first and second school term. For the purposes of the analysis, an average was made of the two grades.

Vocabulary size

To measure students' vocabulary size, the Vocabulary Levels Test (VLT) (Nation, 1990) was used. Similarly as the LLAMA battery, the VLT is available online[10] and it is free of charge. It is only written receptive vocabulary that this instrument measures. Students' productive vocabulary was not focused on in this research. The test items of the VLT are divided into five sections with increasing difficulty - words at the 2,000-, 3,000-, 5,000-, university and 10,000-word levels. Each of these sections includes six sets of six words, with each of the sets accompanied by three definitions. The test-takers must match three out of the six words in each set with their corresponding definition.

After the test-takers complete the task, they are presented with their scores in the form of percentages. The scores are also divided into the five above mentioned levels. In other words, test-takers are shown how successful they were at each of the word levels separately. Therefore, the score of 100, for example, at the 3,000- word level means that the test-taker matched all the words and the definitions at this level correctly. By contrast, the score of 0 indicates that none of the connections between words and their definitions were

---

[9] There is a 5-point grading system in the Czech Republic, with 1 being the best grade and 5 being the worst one.

[10] https://www.lextutor.ca/tests/vlt/?mode=test

made correctly. The test is believed to "estimate the learners' basic knowledge of common word meanings, specifically, the extent to which learners know the common meanings of words at the 2,000, 3,000, 5,000, 10,000 and University word levels" (Beglar & Hunt, 1999, p. 132).

Other collected information

Apart from aptitude and learning outcomes, students' self-evaluation of their competences in English was examined along with their attitudes towards English. Students were presented with four statements, which they were asked to evaluate using a 7-point Likert scale from 1 (=strongly agree) to 7 (=strongly disagree).[11] The statements were as follows:

- I see the knowledge of English as important in life.
- Overall, I enjoy/like English (in and out of school).
- I am good at English.
- I am satisfied with my level of English.

At the end of this part, students were asked to evaluate their level of English using the labels A1, A2, B1, B2, C1 and C2, which were further explained to them.

The last data collected by the record sheet included basic information about the students - their age, gender and the year they are currently in.

3.2.3 Research execution

In the initial stages of the research, the instruments for measuring students' foreign language aptitude and learning outcomes were chosen. All of these were described above. As mentioned before, the selection of school was based on my personal acquaintance with an English teacher who works at the target school. She put me in contact with the director of the school, who agreed with the execution of the research. Before entering the school, signed informed consent forms were collected from all participating students.

---

[11] The whole Likert scale looked as follows: strongly agree – agree – somewhat agree – neither agree or disagree – somewhat disagree – disagree – strongly disagree

The research itself was conducted in December 2023 (3 classes) and January 2024 (2 classes). With each of the classes, I spent 45 minutes (1 English lesson), during which the whole research was completed. At the beginning of the testing, the research goals were introduced to the students. Afterwards, the students were provided with record sheets, in which they were instructed to record all their answers. First, the online LLAMA battery was administered. Subsequently, the students were asked to record basic information about themselves and their grades from last year. Moreover, as mentioned above, self-evaluation of their own competences in English was also collected. Finally, the students took the online Vocabulary Levels Test. At the end of the lesson, the record sheets were collected.

3.2.4   Ethical aspects of the research

Prior to the realisation of the research itself, informed consent forms were distributed to all students by their English teacher. There were two versions of the consent – one for adult students and the other for the parents of underage students. Submission of a signed informed consent form was a condition for the participation in the research.

The informed consent familiarised the students and their parents with the goals of the research and the type of collected data. Emphasis was also placed on voluntary participation in the research, with the possibility to withdraw even during the testing itself. Finally, complete data anonymisation was guaranteed.

3.3   Statistical analysis

The collected data was analysed using a computer program *jamovi*. Firstly, descriptive statistics was conducted in order to explore and describe the research sample. The main focus was drawn to calculating measures of central tendency and determining the normality of various distributions. Moreover, frequencies of certain variables were calculated.

Subsequently, hypotheses were tested using both parametric and non-parametric tests. To compare scores of two groups, Welch's t-test, Mann-Whitney U test and X^2 test of association were used. Furthermore, in order to explore the relationships between two continuous variables, correlation analyses were carried out.

4 Results

The analysis of results is divided into two parts. In the first of them, descriptive statistics of the collected data will be conducted. The focus of the second part will be on hypothesis testing.

4.1 Descriptive statistics

As mentioned above, the research sample consisted of 57 participants (25 men and 32 women), all of whom were grammar school students. The mean age of the sample was 16.7 (SD = 0.605).

4.1.1 Foreign language aptitude

Students' foreign language aptitude was measured by three subtests of the LLAMA battery (Meara & Rogers, 2020h) – LLAMA_B, LLAMA_D and LLAMA_F (LLAMA_E was not included in the testing). At first, the participants' results in the individual subtests will be examined and subsequently, their total scores in the LLAMA battery will be explored.
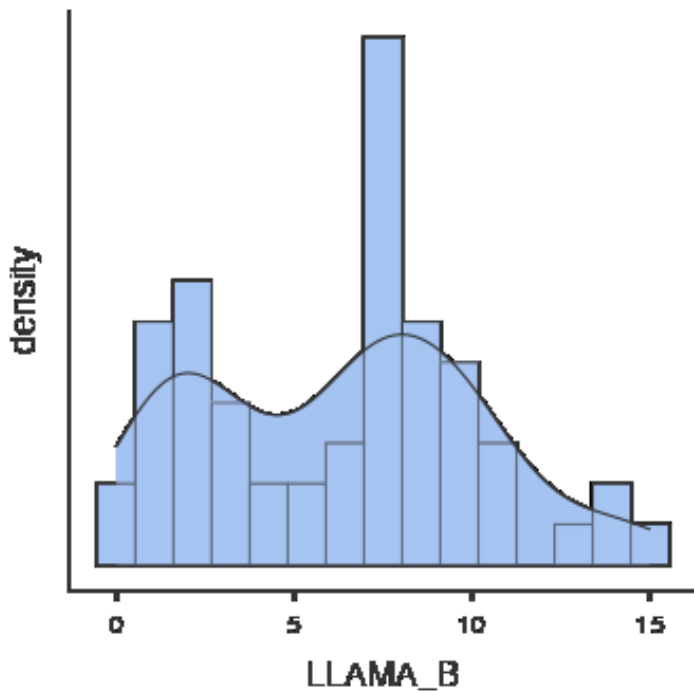
LLAMA_B

Participants could score between 0 and 20 points in the LLAMA_B subtest (the same applies for LLAMA_D and LLAMA_F). As can be seen in Table 1, while the minimum score was 0, the maximum score was 15, with the mean value of 6.26 (SD = 3.94).

For the purposes of statistical testing, it is important to know whether the distribution is normal (Gaussian). Firstly, measures of central tendency should be examined. Table 1 shows that mean, median and mode are relatively similar, which is one of the requirements for a normal distribution. In a normal distribution, skewness should fall within the interval (-1; 1), while the interval of (-3; 3) is acceptable for kurtosis. The distribution of students' results in the LLAMA_B subtests fulfils both of these requirements. Moreover, a normal distribution should be unimodal or in other words, it should have a single peak. As can be seen in Figure 1, the distribution of students' LLAMA_B scores is bimodal (it has two peaks). Therefore, the values are not normally distributed. This assertion can be further

confirmed by the result of Shapiro-Wilk test (p = .019)[12]. Therefore, the null hypothesis of the distribution being normal can be rejected.

Figure 1

LLAMA_B Score Distribution

LLAMA_D

The overall scores were somewhat higher in LLAMA_D subtest compared to the LLAMA_B subtest, with the minimum of 1 point, the maximum of 16 point and the mean value of 8.65 (SD = 3.11). Unlike the distribution in LLAMA_B, students' results in the LLAMA_D subtest are distributed normally. Measures of central tendency are strongly similar and the values of both skewness and kurtosis are very close to the values of a normal distribution. Moreover, when density plot is generated, it is clearly unimodal. Together with Shapiro-Wilk p-value of .589 (Table 1), students' scores in the LLAMA_D subtest can be considered to be distributed normally.

---

[12] Shapiro-Wilk test is a test of normality. When its p-value is lower than 0.05, the null hypothesis of the data set being normally distributed can be rejected. For p-value > 0.05, the rejection of null hypothesis cannot be done.

LLLAMA_F

For LLAMA_F, the scores were even higher than for LLAMA_D. The mean value was 11.2 (SD = 3.43). The lowest score had the value of 2, the highest 18. Although being slightly left skewed (-0.503), as Figure 2 shows, distribution of the LLAMA_F scores is also Gaussian. This assertion is supported not only by the examining the histogram and density plot, but also by the result of Shapiro-Wilk test (p = .181).

Figure 2

LLAMA_F Score Distribution
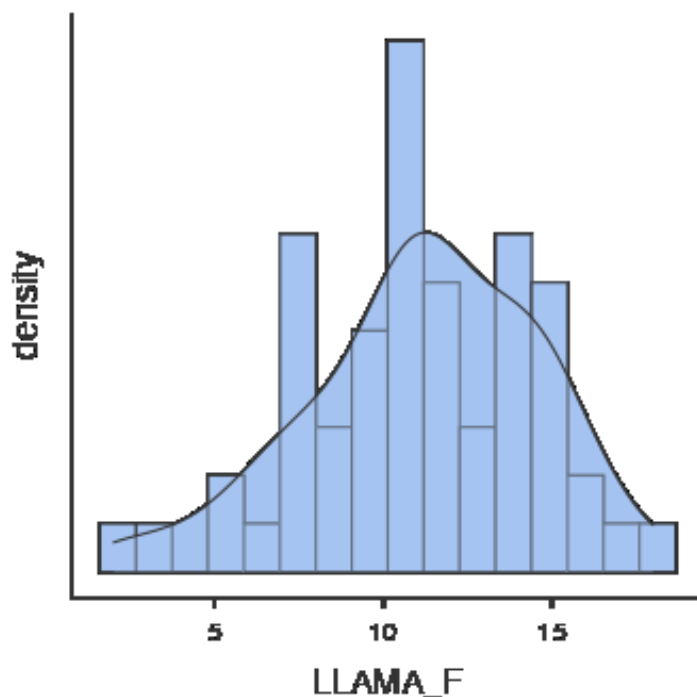
Total scores in the LLAMA battery

When the scores of the three individual LLAMA subtests were totalled for each student, they also showed normal distribution. Out of a maximum of 60 points, the highest result was 40 points and the lowest result 7 points, with the mean value of 26.1 (SD = 7.03). Skewness and kurtosis took on the values of -0.143 and 0.136 respectively. Shapiro-Wilk test did not reject the null hypothesis of the distribution being normal (p = .583).

Table 1

Used LLAMA Subtests Descriptives

	LLAMA_B	LLAMA_D	LLAMA_F
N	57	57	57
Missing	0	0	0
Mean	6.26	8.65	11.2
Median	7	9	11
Mode	7.00	9.00	11.0
Standard deviation	3.94	3.11	3.43
Minimum	0	1	2
Maximum	15	16	18
Skewness	0.168	-0.00612	-0.503
Std. error skewness	0.316	0.316	0.316
Kurtosis	-0.832	-0.196	0.114
Std. error kurtosis	0.623	0.623	0.623
Shapiro-Wilk W	0.950	0.983	0.971
Shapiro-Wilk p	0.019	0.589	0.181

4.1.2   Learning outcomes

The research focused on two types of learning outcomes – students' grades and their written receptive vocabulary size, as measured by the VLT. The recorded values of both learning outcomes will be described below.

Grades

As explained above, two grades were acquired from each student – from first and second school term. For the purposes of the analysis, an average was made of the two grades. Except for one student, everyone recorded their grades. Only one student recorded grade 3 in both school terms. By contrast, most students recorded grade 1 in at least one (but very often in both) term. Whereas mean took on the value of 1.35, both median and mode equalled to the value of 1.00. Both histogram and Shapiro-Wilk test clearly showed that the distribution is not normal.

Vocabulary

Students' vocabulary size was measured by the Vocabulary Levels Test (Nation, 1990), which focuses on written receptive vocabulary. The VLT measures test-takers' vocabulary size on five different levels with increasing difficulty – words at 2,000-, 3,000-, 5,000-, university and 10,000-word levels. Since there was limited time for the testing, several students failed to complete the entire test. Therefore, only the scores from 2,000-, 3,000- and 5,000-word level subtests will be included in the statistical analysis. All participants, except for one student who finished only the first (2,000) level, managed to complete all these three subtests. The missing values of this student (for 3,000- and 5,000-word levels) were replaced with the mean of the whole data set for these levels.

As can be seen in Table 2, students' scores at 2,000- and 3,000-word levels are strongly similar. Most frequently (mode), participants achieved 100 % in these subparts of the VLT test. While mean values were almost identical for these subtests 88.9 (2,000) and 88.8 (3,000), standard deviations slightly differed – 10.7 and 11.6 respectively. Both distributions are substantially left skewed. Neither of them can be considered normally distributed, as Shapiro-Wilk test shows.

Similarly, scores at 5,000-word level also do not have a normal distribution (Shapiro-Wilk $p < .001$). The mean value, however, is significantly lower – 71.6 (SD = 16.3). The range of the data is considerably higher in this subpart compared to the previous two. It takes on the value of 77, with the maximum score of the data set being 94 and the minimum 17.

The previous descriptive statistics of the VLT included all the participating students. However, 36 out of a total of 57 students managed to finish even the University word level subtest. Therefore, additional analysis will be carried out for these 36 students, with the rest of the students (n = 21) left out. Unlike in the previous cases, this distribution is Gaussian (normal). Not only can it be seen in Figure 3, but it is also confirmed by Shapiro-Wilk test ($p > 0.05$). The mean value for this data set is 70.4 (SD = 15.8), with minimum of 28 and maximum of 100.

Table 2

VLT Descriptives

	2,000	3,000	5,000
N	57	57	57
Missing	0	0	0
Mean	88.9	88.8	71.6
Median	89	94	72
Mode	100	100	72.0
Standard deviation	10.7	11.6	16.3
Minimum	56	56	17
Maximum	100	100	94
Skewness	-1.09	-0.989	-1.21
Std. error skewness	0.316	0.316	0.316
Kurtosis	1.03	0.132	1.78
Std. error kurtosis	0.623	0.623	0.623
Shapiro-Wilk W	0.866	0.854	0.907
Shapiro-Wilk p	< .001	< .001	< .001

Figure 3

Score Distribution at University Word Level
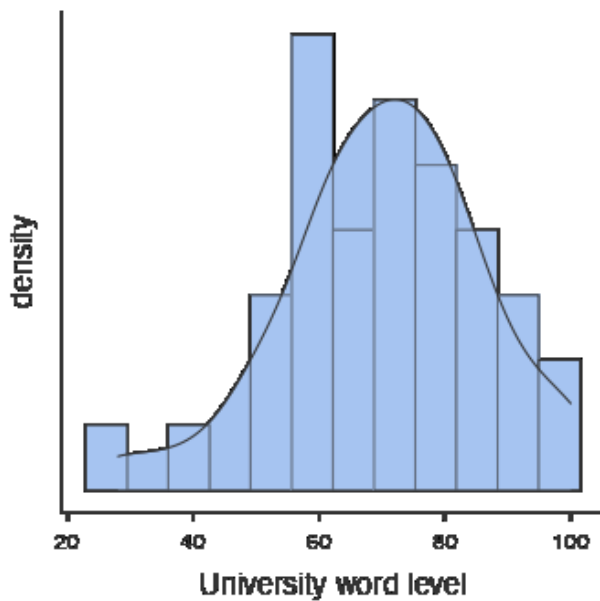
4.1.3 Self-evaluation

Finally, a 7-point Likert scale was used to explore students' self-evaluation of their competences in English and also their attitudes towards English. Since these topics are not included in the research goals and hypotheses, their analysis is regarded to be rather peripheral and exploratory.

In terms of perceived importance of English in life, 48 out of a total of 56 students strongly agreed and 8 students agreed with the statement *I see the knowledge of English as important in life*. The variance of students' answers was substantially higher in the other items. Out of a total of 56 students, 15 strongly agreed, 29 agreed, 9 somewhat agreed, 2 neither agreed or disagreed and 1 somewhat disagreed with the statement *Overall, I enjoy English (in and out of school).* 8 students strongly agreed with the statement *I am good at English*, whereas 19 students agreed, 21 somewhat agreed, 5 neither agreed or disagreed and 3 disagreed with the same statement. For the last statement (*I am satisfied with my level of English*), the whole Likert scale was covered. Strong agreement with the statement was recorded by 4 students, while 12 students agreed, 19 somewhat agreed, 8 neither agreed or disagreed, 6 somewhat disagreed, 4 disagreed and 2 strongly disagreed with the statement.

The last item focused on students' self-evaluation of their level of English in terms of the labels A1, A2, B1, B2, C1 and C2. None of the students saw their English on A1 or C2 level. Majority of them selected B1 or B2 level. Figure 4 shows the results in more detail.

Figure 4

Students' Level of English – Self-evaluation
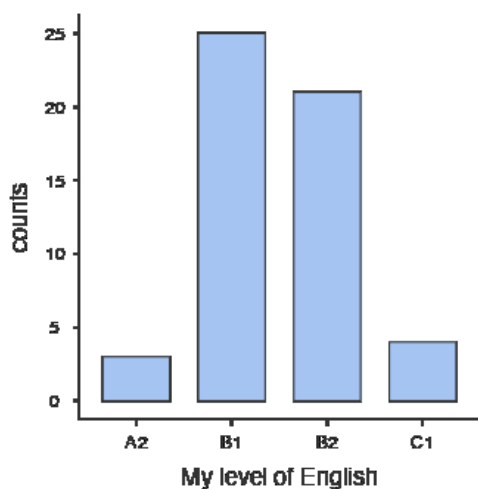
4.2   Hypothesis testing

Having scrutinised and described the data, it is possible to proceed to hypothesis testing. Firstly, the potential differences in foreign language aptitude between male and female students will be examined. Subsequently, the relationship between students' foreign language aptitude and their learning outcomes will be explored. Finally, attention will be drawn to gender differences in students' self-reported level of English.

4.2.1   Hypothesis 2

Hypothesis 2 was chosen to be tested first. It was formulated as follows:

H2: Male and female high school students do not differ in their levels of foreign language aptitude.

To find out whether there are gender differences in foreign language aptitude as measured by the LLAMA battery, the independent samples tests were used. Before conducting the tests, it was necessary to check the distributions of students' scores in the LLAMA subtests. Where the distribution was normal (LLAMA_D, LLAMA_F and the LLAMA battery in total), Welch's t-test was used. Otherwise (LLAMA_B), Mann-Whitney U test was used. This distinction was done in the previous chapter.

It was found that there was no significant difference in foreign language aptitude between male and female students. Importantly, this was true for all the included aptitude measures. For LLAMA_D, LLAMA_F and the LLAMA battery in total, concrete results of the conducted t-tests were as follows: $t(54.9) = -0.194$, $p = .847$; $t(49.2) = 0.582$, $p = .563$; $t(46.5) = 0.732$, $p = .468$ respectively. For LLAMA_B, conducted Mann-Whitney U test showed $U = 365$, $p = .572$. Based on these results, there do not seem to be gender differences in foreign language aptitude in high school students.

4.2.2   Hypothesis 1

According to hypothesis 1, there is expected to be a positive correlation between foreign language aptitude of Czech high school students and their learning outcomes in English. The potential association between the two variables was analysed by means of correlation.

Grades

Firstly, potential relationship between FLA and grades was tested. As students' grades were not normally distributed, nonparametric tests were used. No statistically significant correlations were found between students' grades and their scores in the LLAMA tests. For all four explored relationships (between students' grades and their scores in LLAMA_B, LLAMA_D, LLAMA_F and the LLAMA battery in total), Spearman's rho was very close to zero. More importantly, based on p-value, none of the correlations were statistically significant.

The Vocabulary Levels Test

Subsequently, the relationship between students' foreign language aptitude and their written receptive vocabulary size as measured by the VLT was examined. As with grades, nonparametric tests were used since students' scores in all three VLT subparts did not have a normal distribution. In total, 12 relationships were explored[13]. Only two of them, however, were statistically significant. A positive correlation (r = .273) was found between students' scores in LLAMA_B and 5,000-word level test (p = .04). Similarly positive relationship (r = .270) was found between students' scores in LLAMA_F and 5.000-word level test (p = .042). The rest of the examined relationships did not show statistical significance. Even among these insignificant results, however, certain differences can be found. For example, whereas there is a certain positive correlation (r = .247) between LLAMA_B and 3,000-word level (p = .064), LLAMA_D generally shows zero correlations with the VLT with p-values substantially higher than the threshold value of 0.05.

---

[13] 3 subparts of the VLT (words at 2,000-, 3,000- and 5,000-word levels) combined with 4 LLAMA measures (LLAMA_B, LLAMA_D, LLAMA_F and the LLAMA battery in total)

Table 3

Correlations between LLAMA and the VLT

		LLAMA_B	LLAMA_D	LLAMA_F	LLAMA
2,000	Spearman's rho	0.114	-0.057	0.088	0.104
	df	55	55	55	55
	p-value	0.396	0.673	0.517	0.443
3,000	Spearman's rho	0.247	0.006	0.204	0.235
	df	55	55	55	55
	p-value	0.064	0.966	0.128	0.078
5,000	Spearman's rho	0.273	-0.070	0.270	0.232
	df	55	55	55	55
	p-value	0.040 *[14]	0.604	0.042 *	0.082

Furthermore, the entire sample (N = 57) was divided into two groups by gender – male (n = 25) and female (n = 32) students, and similar analyses as described above were performed separately for the two groups. Interestingly, in neither of the groups was there any significant relationship between foreign language aptitude as measured by the LLAMA battery (or its subtests) and the VLT (2,000-, 3,000- and 5,000-word levels).

Finally, the relationship between the LLAMA tests and University word level scores was examined. This correlation analysis was conducted only for the sample of the 36 students (male and female students together) who managed to finish the University word level subtest. No significant correlations were found between foreign language aptitude and students' vocabulary size.

4.2.3 Additional testing

Since the collected data offers other possible analyses, some of these will be carried out in this section. However, it should be noted that these analyses will not directly relate to the hypotheses set in this thesis and therefore, their character is regarded to be rather peripheral and exploratory.

As students' written receptive vocabulary size was collected for the purposes of exploring its relationship with FLA, I decided to compare the VLT scores of male and female

---

[14] * indicates statistically significant correlation, with $p < .05$

students. Due to missing data, one female student was not included in this analysis. Significant differences between male and female students were found on all three VLT word levels (2,000, 3,000 and 5,000) in favour of male students. For all of them, Mann-Whitney U test computed medium effect size of 0.388 (p = .011), 0.333 (p = .029) and 0.457 (p = .003) respectively. Table 4 shows more details regarding this comparison.

Table 4

The VLT Men and Women Descriptives

	Group	N	Mean	Median	SD
2,000	man	25	93.5	94.0	6.18
	woman	31	85.4	89.0	12.3
3,000	man	25	92.7	94.0	9.25
	woman	31	85.7	89.0	12.6
5,000	man	25	78.8	78.0	10.04
	woman	31	65.8	67.0	18.4

Potential differences between male and female students were explored not only in the areas of foreign language aptitude and vocabulary size, but also in students' self-evaluation of their competences in English. For the purposes of this analysis, the item in which students were asked to evaluate their level of English using the labels A1, A2, B1, B2. C1 or C2 was chosen. A chi-square test of association was performed to examine the relation between gender and self-reported level of English. The relation between these variables was not statistically significant, X^2 (3; N = 53) = 6.19, p = .103. In other words, based on the research sample, there is not a statistical association between gender and self-reported level of English. The contingency tables (Table 5) show both observed and expected values.

Table 5

Contigency Tables – Gender and Self-reported Level of English

gender		My level of English				Total
		A2	B1	B2	C1	
woman	Observed	2	16	10	0	28
	Expected	1.58	13.2	11.09	2.11	28.0
man	Observed	1	9	11	4	25
	Expected	1.42	11.8	9.91	1.89	25.0
Total	Observed	3	25	21	4	53
	Expected	3.00	25.0	21.00	4.00	53.0

5   Discussion

Hypothesis 1

As mentioned above, one of the goals of the conducted research was to ascertain, whether and how foreign language aptitude relates to the language performance of Czech learners of English. It was hypothesised that there will be a positive correlation between FLA and learning outcomes in English. In other words, students with higher aptitude levels were expected to perform better in English than students with lower aptitude levels. This performance was operationalised as one's grades and primarily, the size of one's written receptive vocabulary.

Based on the analysis of the collected data, students' foreign language aptitude was not found to be correlated with their grades. This finding is in conflict with Li (2016), who reported a correlation of r = .49 between the global construct of aptitude and overall L2 proficiency measured by course grades or standardised tests scores. This contrast could be possibly explained by the nature of the data obtained in this research. Majority of the students recorded grade 1 in at least one (but very often in both) term. Some of the students reported grade 2, whereas grade 3 was recorded only by 2 students. None of the students reported grade 4 or 5. Evidently, the data was highly homogeneous, as a result of which it might not have been possible to distinguish grades of students with higher and lower foreign language aptitude.

Similarly, students' foreign language aptitude did not turn out to be largely associated with their written receptive vocabulary size. Students' scores in the LLAMA battery in total as well as in the LLAMA_D subtest showed no significant correlations with their VLT scores. Nevertheless, certain relationships have been found between the two variables. Specifically, a statistically significant correlation of r = .270[15] was found between students' scores in LLAMA_F and 5,000-word level test. Moreover, it was also scores in LLAMA_B, which significantly correlated (r = .273) with 5,000-word level tests. In other words, students' scores in the two subtests of the LLAMA battery weakly correlated with their

---

[15] Although statistically significant, the correlation of r = .270 is regarded to be relatively weak. The same applies for the correlation of r = .273.

receptive vocabulary size at 5,000-word level. It therefore seems that students' written receptive vocabulary size is related differently to the individual subcomponents of their foreign language aptitude, as measured by the LLAMA subtests. This assumption is in accordance with the notion of foreign language aptitude being a multidimensional construct, with each of its subcomponents influencing different aspects of second language acquisition. LAMMA_B is rather directly connected to the aspect of memory, and it is believed to focus on one's ability to learn vocabulary. Therefore, its positive relationship to vocabulary size is relatively intuitive. The insignificant correlations between students' results in LLAMA_D and their vocabulary size might be explained by the focus of LLAMA_D – the ability to recognise sounds is perhaps not as important in vocabulary learning. The existence of a positive relationship between LLAMA_F, which purports to measure one's language analytic ability, and students' vocabulary size is somewhat more surprising. Possibly, the disposition to learn vocabulary might be influenced, apart from other factors, by learners' general cognitive abilities, which to some extent overlap with language analytic ability (Carroll, 1990; Skehan, 2002). In other words, the relationship between LLAMA_F and students' vocabulary size could be mediated by another variable – students' intelligence. It should be recognised, however, that this explanation is based on my personal interpretation and it is not supported by further research.

What can also be regarded as somewhat surprising, is that no correlations have been found between students' foreign language aptitude and their scores in 2,000- and 3,000-word level tests. Possibly, in a similar manner as grades, these two levels might have been too easy for most students and therefore, no differences could be found between high and low aptitude students. Perhaps, more difficult test might have better discriminating power. Contrary to this assumption, however, not even in the relationship between students' foreign language aptitude and their scores in the university VLT subtest, which is considerably more difficult and students' scores in it thus varied, was there found a significant correlation.

As far as the first hypothesis is concerned, I was also interested whether the above-mentioned relationship differs for male and female students. In neither of the groups was there found a significant association between foreign language aptitude and the VLT subtests.

The overall relatively low correlations between FLA and vocabulary size might be also potentially explained by the fact, as Li (2022) concludes, that overall aptitude is "a stronger predictor of listening, reading, speaking, and grammar than vocabulary, and a non-significant predictor of writing" (p. 40). Therefore, vocabulary does not seem to be as strongly associated with foreign language aptitude as other aspects of language. Li (2016) adds that aptitude correlated with vocabulary with r value of .15, which is to some extent comparable with the results of this research. Finally, it is important to emphasise once again that although weakly, two LLAMA subtests correlated significantly with students' vocabulary size.

Hypothesis 2

The second hypothesis focused on the relationship between gender and foreign language aptitude, and it assumed that male and female high school students do not differ in their levels of foreign language aptitude. The conducted tests confirmed this assumption for all the included aptitude measures (LLAMA_B, LLAMA_D, LLAMA_F and the LLAMA battery in total). These findings are in accordance with those of Rogers et al. (2016) who concluded that there are no gender differences in the LLAMA tests. Findings of this research therefore support the claims that the LLAMA battery is not susceptible to external individual differences (Rogers et al., 2016). However, as these authors add, "the LLAMA tests still need to be validated in terms of their ability to predict language learning" (p. 56).

Other findings

In addition to the research hypotheses, some other relationships were explored. Having compared the VLT scores of male and female students, it was found that there are significant differences of medium size effect in favour of male students on all three VLT word levels (2,000, 3,000 and 5,000). In other words, male students showed larger written receptive vocabulary size than female students. These findings might be somewhat counterintuitive considering the above-mentioned Heyder et al.'s (2017) thoughts about boys' underachievement in first language grades or Oga-Baldwin and Fryer's (2020) beliefs about boys being disadvantaged in terms of their motivation for learning a foreign language. Based on these claims, one might expect girls to outperform boys also in the current study. However, it is important to realise that neither of these studies focused specifically on the

area of vocabulary size. In fact, studies exploring the relationship between gender and foreign language vocabulary size are scarce. Therefore, findings of this thesis should be considered rather exploratory since they are among the first of their kind. Although this gender difference might sound interesting, it was inferred from a relatively small non-random sample, and I would therefore not venture to generalise it to a broader population.

Furthermore, students' self-evaluation of their competences in English and their attitudes towards English were collected. Interestingly, while almost everyone perceives the knowledge of English as important in life, not all students seem to be satisfied with their level of English. Whereas some reported their English to be on level A2 or C1, vast majority of students selected either level B1 or B2. None of them reported level A1 or C2. Given that these students were only in their second or third year (out of a total of four), their self-reported level of English seems to somewhat exceed what Czech school system expects of them - the state English maturita exam is at B1 level (MŠMT, 2014). However, it is quite common for grammar school students to have a higher level of English than B1 and these findings are therefore not too surprising. Finally, based on the conducted analysis, significant differences were not found between male and female students in their self-reported level of English.

Implications for teachers of English

Finally, it is important to realise that some of the above presented findings might be transferable to educational practice. If further research confirms that foreign language aptitude, as measured by the LLAMA battery or other aptitude tests, is indeed related to the size of Czech students' vocabulary, but also, for example, to their reading, speaking, or listening skills, teachers of English could then use these instruments to map their students' potential strengths and weaknesses, and provide them with support according to their individual needs.

Moreover, the findings of the current study which indicate that there are no differences between male and female learners of English in terms of their foreign language aptitude might help teachers to resist some of the widespread unsubstantiated stereotypes regarding gender differences in foreign language acquisition.

Lastly, the information that most students find the knowledge of English as important in life can be considerably motivating for teachers of English. Knowing that what they teach is perceived as useful by their students is a great sign of the meaningfulness of their work.

5.1 Limitations

The conducted research has certain limitations. One of them concerns the research sample. As has been mentioned above, the selection of school was based on my personal acquaintance with an English teacher who works at the school. The sampling method could therefore be described as convenience sampling rather than random, which would be preferable. As far as sampling is concerned, choosing grammar school students as research participants might have also brought some limitations. On average, knowledge of English can be expected to be rather high in grammar school students and primarily, quite homogeneous. This assumption was well illustrated on participants' grades (most of them recorded grade 1 or 2) and similarly, on their knowledge of words at 2,000- and 3,000-word levels (majority of students reached very high scores). The fact that the variable of grades (and vocabulary size) took on very similar values for most students might have subsequently complicated hypothesis testing. What is more, due to the nature of the sample (relatively small, non-random, grammar school) the findings might not be fully representative of a broader population. Future research might therefore focus, for example, on students from other types of high schools. Only then will we know whether these findings can be further generalised.

Another possible limitation is connected to the operationalisation of the variable of learning outcomes. In the current research, it was operationalised as one's grades and as one's written receptive vocabulary size. However, as Li (2022) states, overall aptitude is "a stronger predictor of listening, reading, speaking, and grammar than vocabulary, and a non-significant predictor of writing" (p. 40). Therefore, in future research, it might be preferable to operationalise learning outcomes as, for example, certain reading or listening skills. On the other hand, the choice to measure students' vocabulary size seems rather reasonable for the purposes of this study since assessing their reading or listening skills would probably require to spend considerably more time with the students.

As far as the measuring instruments used in this research are concerned, it is important to mention that one subtest of the LLAMA battery, namely LLAMA_E, was not included in the testing. The reasons for its exclusion were purely time-related. Although necessary, this exclusion might have influenced the results of the conducted analyses and if the LLAMA_E subtest had been incorporated as a measuring instrument of foreign language aptitude, the findings of this thesis might be somewhat different. Therefore, it would be advisable to use the whole LLAMA battery in the future research, given there is time for that.

As discussed earlier, the LLAMA battery has not been sufficiently validated so far (Rogers et al., 2017), as a result of which interpreting its relationship to learning outcomes might be hindered. In spite of this weakness of the battery, it was rather sensible to choose it as a measuring instrument of FLA since it is available online and importantly, it is free of charge, both of which aspects were of high importance. Moreover, certain quality of the LLAMA tests is supported by the fact that they are used frequently in aptitude research (Suzuki, 2021).

Lastly, the employment of the VLT in the research is also connected to some limitations. As mentioned above, 2,000- and 3,000-word levels turned out to be too easy for most students. Therefore, follow-up research might, based on the type of its target sample, consider omitting these VLT subtests and thus save some time for different tests. Furthermore, it is also important to add that I was not able to find many validation studies of Nation's version of the VLT (1990). Therefore, it might seem more appropriate to use, for example, Schmitt et al.'s (2001) version of the VLT. Although their revision of the VLT may be superior to Nation's version in terms of its psychometric properties, its administration is more time-consuming and from that perspective, it would have been less suitable for this study, in which the testing in school was strictly time-limited.

Conclusion

The main goal of this bachelor's thesis was to examine the relationship between foreign language aptitude and learning outcomes, with an emphasis on gender differences. The theoretical part focused primarily on the introduction of the concept of foreign language aptitude, to a large extent in the context of its historical development. For the purposes of the empirical part, test batteries assessing foreign language aptitude and written receptive vocabulary size were administered to the sample of 57 grammar school students. Based on the conducted analyses of the data obtained, research hypotheses were addressed.

The research findings support the assumption that there could be a certain relationship between foreign language aptitude and learning outcomes. Specifically, statistically significant weak correlations were found between students' written receptive vocabulary size and two subcomponents of their foreign language aptitude, as measured by the LLAMA battery. Nevertheless, one of its subtests, as well as the total score in the LLAMA battery, showed no relation to students' vocabulary size. Similarly, no significant relationship was found between students' foreign language aptitude and their grades. Furthermore, the hypothesis that male and female students do not differ in their levels of foreign language aptitude was confirmed. Not only did they show equal levels of foreign language aptitude, male and female students were also similar in terms of their self-reported level of English. Vast majority of them reported to be at B1 or B2 level. By contrast, certain gender differences were found to exist in students' size of written receptive vocabulary. Compared to female students, male students demonstrated somewhat better results.

One of the main advantages of the conducted research is certainly its sample comprising Czech participants. The topic of foreign language aptitude has not been explored extensively in the Czech environment and therefore, the current study sheds some light on at least a small part of a large unexplored area. For example, the findings confirm the assertion of foreign studies that foreign language aptitude does not differ between men and women. This thesis therefore contributes to dispelling the persisting myths about gender differences in the area of foreign language acquisition. What is more, although on a relatively small sample, the research shows how Czech grammar school students self-evaluate their competences in English, which might be found interesting by Czech teachers of English.

Moreover, teachers could also appreciate the information that students perceive the knowledge of English as important in life. Finally, if the results of the current study are further confirmed and extended, teachers of English might then possibly use aptitude tests to identify their students' potential individual needs.

Overall, this bachelor's thesis contributes to the extension of knowledge in the field of psycholinguistics, specifically in the area of foreign language acquisition, and I believe that the presented findings might serve as an inspiration or a starting point for further research.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native Second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. https://doi.org/10.1017/s027226310808073x

Alduais, A., Alfadda, H., Baraja'a, D., & Allegretta, S. (2022). Psycholinguistics: Analysis of Knowledge Domains on children's language acquisition, production, comprehension, and dissolution. *Children*, *9*(10), 1471. https://doi.org/10.3390/children9101471

Altmann, G. (2006). Psycholinguistics: History. *Encyclopedia of Language & Linguistics*, 257–265. https://doi.org/10.1016/b0-08-044854-2/04157-2

American Psychological Association. (2018a, April 19). Personality. In *APA Dictionary of Psychology*. Retrieved June 2, 2024, from https://dictionary.apa.org/personality

American Psychological Association. (2018b, April 19). Motivation. In *APA Dictionary of Psychology*. Retrieved June 2, 2024, from https://dictionary.apa.org/motivation

American Psychological Association. (2018c, April 19). Cognitive process. In *APA Dictionary of Psychology*. Retrieved June 2, 2024, from https://dictionary.apa.org/cognitive-process

American Psychological Association. (2018d, April 19). Construct validity. In *APA Dictionary of Psychology*. Retrieved June 2, 2024, from https://dictionary.apa.org/construct-validity

American Psychological Association. (2023a, November 15). Self-efficacy. In *APA Dictionary of Psychology*. Retrieved June 2, 2024, from https://dictionary.apa.org/self-efficacy

American Psychological Association. (2023b, November 15). Cognitive style. In *APA Dictionary of Psychology*. Retrieved June 2, 2024, from https://dictionary.apa.org/cognitive-style

Artieda, G., & Muñoz, C. (2016). The LLAMA tests and the underlying structure of language aptitude at two levels of foreign language proficiency. *Learning and Individual Differences*, *50*, 42–48. https://doi.org/10.1016/j.lindif.2016.06.023

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 47–89. https://doi.org/10.1016/s0079-7421(08)60452-1

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295x.84.2.191

Beglar, D., & Hunt, A. (1999). Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing*, *16*(2), 131–162. https://doi.org/10.1177/026553229901600202

Bernat, E., & Lloyd, R. (2007). Exploring the gender effect on EFL learners' beliefs about language learning. *Australian Journal of Educational and Developmental Psychology*, *7*, 79-91.

Bokander, L., & Bylund, E. (2019). Probing the internal validity of the LLAMA language aptitude tests. *Language Learning*, *70*(1), 11–47. https://doi.org/10.1111/lang.12368

Bongaerts, T., van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(4), 447–465. https://doi.org/10.1017/s0272263197004026

Carrol, J. B. (1990). Cognitive Abilities in Foreign Language Aptitutude: Then and Now. In Parry, T. S., & Stansfield, C. W. (Eds). *Language aptitude reconsidered. language in education theory and practice, no. 74* (pp. 11–29). Distributed by ERIC Clearinghouse.

Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign-language teaching. *Linguistics*, *11*(112). https://doi.org/10.1515/ling.1973.11.112.5

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, MA: Newbury House

Dahlen, K., & Caldwell–Harris, C. (2013). Rehearsal and aptitude in foreign vocabulary learning. *The Modern Language Journal*, *97*(4), 902–916. https://doi.org/10.1111/j.1540-4781.2013.12045.x

De Graaff, R. (1997). THE EXPERANTO EXPERIMENT: Effects of Explicit Instruction on Second Language Acquisition. *Studies in Second Language Acquisition*, *19*(2), 249–276. https://doi.org/10.1017/s0272263197002064

DeKeyser, R. M. (2000). The Robustnesss of Critical Period Effects in Second Language Acquisition. *Studies in Second Language Acquisition*, *22*(4), 499–533. https://doi.org/10.1017/s0272263100004022

Dewaele, J. (2012). Personality in Second language acquisition. *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0904

Dörnyei, Z. (2006). Individual differences in Second language acquisition. *AILA Review*, *19*, 42–68. https://doi.org/10.1075/aila.19.05dor

Dušková, L. (1969). On sources of errors in foreign language learning. *IRAL - International Review of Applied Linguistics in Language Teaching*, *7*(1), 11–36. https://doi.org/10.1515/iral.1969.7.1.11

Flege, J. E. (1999). Age of learning and second language speech. In D. Birdsong (Ed.), *Second language acquisition and the Critical Period Hypothesis* (pp. 101–131). Lawrence Erlbaum Associates Publishers.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Edward Arnold.

Grigornko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of Foreign Language Learning Ability: The canal-f theory and test. *The Modern Language Journal*, *84*(3), 390–405. https://doi.org/10.1111/0026-7902.00076

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, *14*(1), 25–38. https://doi.org/10.1017/s0272263100010457

Heyder, A., Kessels, U., & Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: Not a lack of aptitude but students' motivational beliefs and parents' perceptions? *British Journal of Educational Psychology*, *87*(2), 205–223. https://doi.org/10.1111/bjep.12145

Hoffman, S. Q. (1997). Field dependence/independence in second language ACQUISITION1 and implications for educators and instructional designers. *Foreign Language Annals*, *30*(2), 222–234. https://doi.org/10.1111/j.1944-9720.1997.tb02344.x

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, *70*(2), 125–132. https://doi.org/10.2307/327317

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in Nonnovice Adult Learners. *Applied Psycholinguistics*, *30*(2), 225–249. https://doi.org/10.1017/s0142716409090109

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581–592. https://doi.org/10.1037/0003-066x.60.6.581

Hyltenstam, K., & Abrahamsson, N. (2000). Who can become native-like in a Second language? all, some, or none? *Studia Linguistica*, *54*(2), 150–166. https://doi.org/10.1111/1467-9582.00056

Janebi Enayat, M., & Derakhshan, A. (2021). Vocabulary size and depth as predictors of second language speaking ability. *System*, *99*, 102521. https://doi.org/10.1016/j.system.2021.102521

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in Second language learning: The influence of maturational state on the acquisition of English as a Second language. *Cognitive Psychology*, *21*(1), 60–99. https://doi.org/10.1016/0010-0285(89)90003-0

Johnson, J., Prior, S., & Artuso, M. (2000). Field dependence as a factor in second language communicative production. *Language Learning*, *50*(3), 529–567. https://doi.org/10.1111/0023-8333.00137

Krashen, S. D. (1981). *Second language acquisition and Second language learning*. Pergamon Press.

Levelt, W. (2012). *A History of Psycholinguistics*. https://doi.org/10.1093/acprof:oso/9780199653669.001.0001

Li, S. (2015). The Associations Between Language Aptitude and Second Language Grammar Acquisition: A Meta-Analytic Review of Five Decades of Research. *Applied Linguistics*, *36*(3), 385–408. https://doi.org/10.1093/applin/amu054

Li, S. (2016). The construct validity of Language Aptitude. *Studies in Second Language Acquisition*, *38*(4), 801–842. https://doi.org/10.1017/s027226311500042x

Li, S. (2022). Explicit and Implicit Language Aptitudes. In S. Li, P. Hiver & M. Papi (Eds), *The Routledge handbook of second language acquisition and individual differences* (pp. 37-53). Routledge.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R., Smith, B. K., Bunting, M. F., & Doughty, C. J. (2013). Hi-LAB: A New Measure of

Aptitude for High-Level Language Proficiency. *Language Learning*, *63*(3), 530–566. https://doi.org/10.1111/lang.12011

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the Second language. *Language Learning*, *44*(2), 283–305. https://doi.org/10.1111/j.1467-1770.1994.tb01103.x

Masgoret, A. –M., & Gardner, R. C. (2003). Attitudes, motivation, and Second language learning: A meta–analysis of studies conducted by Gardner and Associates. *Language Learning*, *53*(1), 123–163. https://doi.org/10.1111/1467-9922.00212

Meara, P. (2022). *The LLAMA tests*. Lognostics. https://www.lognostics.co.uk/tools/LLAMA_3/index.htm

Meara, P. & Rogers, V. (2020a). The LLAMA Tests v3. LLAMA_B. Cardiff: Lognostics. https://www.lognostics.co.uk/beta/LLAMA_B3/LLAMA_B3.htm

Meara, P. & Rogers, V. (2020b). The LLAMA Tests v3. LLAMA_D. Cardiff: Lognostics. https://www.lognostics.co.uk/beta/LLAMA_D3/LLAMA_D3.htm

Meara, P. & Rogers, V. (2020c). *LLAMA_D v3.0 Listening for new words: The Manual*. https://www.lognostics.co.uk/tools/LLAMA_3/LLAMA_Manuals/LLAMA_D3_Manual.pdf

Meara, P. & Rogers, V. (2020d). The LLAMA Tests v3. LLAMA_F. Cardiff: Lognostics. https://www.lognostics.co.uk/beta/LLAMA_F3/LLAMA_F3.htm

Meara, P. & Rogers, V. (2020e). *LLAMA_F v3.0 Grammar Rules: The Manual*. https://www.lognostics.co.uk/tools/LLAMA_3/LLAMA_Manuals/LLAMA_F3_M anual.pdf

Meara, P. & Rogers, V. (2020f). The LLAMA Tests v3. LLAMA_E. Cardiff: Lognostics. https://www.lognostics.co.uk/beta/LLAMA_E3/LLAMA_E3.htm

Meara, P. & Rogers, V. (2020g). *LLAMA_E v3.0 Sounds and Symbols: The Manual*. https://www.lognostics.co.uk/tools/LLAMA_3/LLAMA_Manuals/LLAMA_E3_M anual.pdf

Meara, P. & Rogers, V. (2020h). The LAMMA Tests v3. Cardiff: Lognostics. https://www.lognostics.co.uk/tools/LLAMA_3/index.htm

Mills, N., Pajares, F., & Herron, C. (2006). A reevaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals*, *39*(2), 276–295. https://doi.org/10.1111/j.1944-9720.2006.tb02266.x

MŠMT. (2014). *Katalog požadavků zkoušek společné části maturitní zkoušky, Anglický jazyk*. https://maturita.cermat.cz/files/files/AJ_Katalog_pozadavku_MZ-17-18_opr_2.pdf

Nation, P. (1990). *Teaching and learning vocabulary*. New York: Newbury House.

Nolen-Hoeksema, S., Fredrickson, B., Loftus, G. R., & Wagenaar, W. A. (2012). *Psychologie Atkinsonové a Hilgarda* (Přepracované vydání, přeložila Hana ANTONÍNOVÁ). Portál.

O'Brien, I., Segalowitz, N., Collentine, J., & Freed, B. (2006). Phonological memory and lexical, narrative, and grammatical skills in second language oral production by Adult Learners. *Applied Psycholinguistics*, *27*(3), 377–402. https://doi.org/10.1017/s0142716406060322

Odlin, T. (2012). Crosslinguistic Influence in Second language acquisition. *The Encyclopedia of Applied Linguistics*. https://doi.org/10.1002/9781405198431.wbeal0292

Oga-Baldwin, W. L. Q., & Fryer, L. K. (2020). Girls show better quality motivation to learn languages than boys: Latent Profiles and their gender differences. *Heliyon*, *6*(5). https://doi.org/10.1016/j.heliyon.2020.e04054

Ortega, L. (2009). *Second language acquisition*. Hodder Education.

Plháková, A. (2023). *Učebnice Obecné Psychologie*. Academia.

Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, *130*(2), 261–288. https://doi.org/10.1037/0033-2909.130.2.261

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics*, *25*, 46–73. https://doi.org/10.1017/s0267190505000036

Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude. *EUROSLA Yearbook*, *16*, 179–210. https://doi.org/10.1075/eurosla.16.07rog

Rogers, V. E., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude. *EUROSLA Yearbook*, *16*, 179–210. https://doi.org/10.1075/eurosla.16.07rog

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the Llama Aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49–60. https://doi.org/10.22599/jesla.24

Rogers, V., Meara, P., Barnett-Legh, T., Curry, C., & Davie, E. (2017). Examining the Llama Aptitude tests. *Journal of the European Second Language Association*, *1*(1), 49–60. https://doi.org/10.22599/jesla.24

Sáfár, A., & Kormos, J. (2008). Revisiting problems with foreign language aptitude. *IRAL - International Review of Applied Linguistics in Language Teaching*, *46*(2). https://doi.org/10.1515/iral.2008.005

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213–223. https://doi.org/10.1017/s0267190500200135

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

Singleton, D. (2001). Age and Second Language Acquisition. *Annual Review of Applied Linguistics*, *21*, 77–89. https://doi.org/10.1017/s0267190501000058

Singleton, D. (2017). Language aptitude: Desirable trait or acquirable attribute? *Studies in Second Language Learning and Teaching*, *7*(1), 89–103. https://doi.org/10.14746/ssllt.2017.7.1.5

Skehan, P. (1991). Individual differences in Second language learning. *Studies in Second Language Acquisition*, *13*(2), 275–298. https://doi.org/10.1017/s0272263100009979

Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69–93). Amsterdam: Benjamins.

Skehan, P. (2014). Foreign language aptitude and its relationship with grammar: A critical overview. *Applied Linguistics*, *36*(3), 367–384. https://doi.org/10.1093/applin/amu072

Sparks, R. L., & Patton, J. (2013). Relationship of L1 skills and L2 aptitude to L2 anxiety on the Foreign Language Classroom Anxiety Scale. *Language Learning*, *63*(4), 870–895. https://doi.org/10.1111/lang.12025

Sparks, R. L., Humbach, N., Patton, J., & Ganschow, L. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal*, *95*(2), 253–273. https://doi.org/10.1111/j.1540-4781.2011.01176.x

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning1. *Language Learning*, *26*(2), 233–253. https://doi.org/10.1111/j.1467-1770.1976.tb00275.x

Suzuki, Y. (2021). Probing the construct validity of LLAMA_D as a measure of implicit learning aptitude. *Studies in Second Language Acquisition*, *43*(3), 663–676. https://doi.org/10.1017/s0272263120000704

Šulová, L. (2019). *Raný Psychický Vývoj Dítěte*. Univerzita Karlova, nakladatelství Karolinum.

Vocabulary Levels Test (n.d.). Compleat Lexical Tutor. https://www.lextutor.ca/tests/vlt/?mode=test

Warren, P. (2013). *Introducing psycholinguistics*. Cambridge University Press.

Wen, Z. (2011). Foreign language aptitude. *ELT Journal*, *66*(2), 233–235. https://doi.org/10.1093/elt/ccr068