

University of Pardubice
Faculty of Arts and Philosophy

Evaluating Teacher-made Language Tests for Assessing Upper-Secondary
Students' English Proficiency

Bachelor's Thesis

2024

Daniel Hofman

Univerzita Pardubice
Fakulta filozofická
Akademický rok: 2022/2023

ZADÁNÍ BAKALÁŘSKÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení: **Daniel Hofman**
Osobní číslo: **H21198**
Studijní program: **B0231A090018 Anglický jazyk**
Specializace: **Anglický jazyk pro vzdělávání**
Téma práce: **Hodnocení testů dosažené jazykové úrovně angličtiny připravených učitelů na středních školách**
Téma práce anglicky: **Evaluating Teacher-made Language Tests for Assessing Upper Secondary Students English Proficiency**
Zadávací katedra: **Katedra anglistiky a amerikanistiky**

Zásady pro vypracování

Student se v této bakalářské práci zabývá problematikou struktury testů vytvořených učitelem a zaměřených na testování gramatiky. V úvodu teoretické části vysvětlí problematiku spojenou s testováním. Po představení základních otázek, spojených s testováním bude následovat analýza zásad konstrukce testů. Dále se student zaměří na konstrukci testu z pohledu struktury, tedy výběru typu testu, jeho realizaci a následné bodování. V praktické části student zrealizuje výzkum kombinací ústních pohovorů s učiteli a analýzou jejich testů na gramatiku a slovní zásobu. Hlavním cílem praktické části bude zjistit, jak učitelé připravují své testy s ohledem na cíl, který testováním chtějí dosáhnout.

Rozsah pracovní zprávy:
Rozsah grafických prací:
Forma zpracování bakalářské práce: **tištěná/elektronická**
Jazyk zpracování: **Angličtina**

Seznam doporučené literatury:

Brown, Henry Douglas. 2004. Language assessment: principles and classroom practices. San Francisco: Longman
Burke, Kay. 2010. Balanced Assessment: From Formative to Summative. Bloomington: Solution Tree Press
Dolin, Jens, Paul Black, Wynne Harlen, and Andrée Tiberghien. 2018. "Exploring Relations Between Formative and Summative Assessment" In Transforming Assessment, edited by Jens Dolin and Robert Evans, 53-80. Cham: Springer International Publishing
Gottlieb, Margo. 2016. Assessing English Language Learners: Bridges to Educational Equity: Connecting Academic Language Proficiency to Student Achievement. Thousands Oak: Corwin Publishers
Kolář, Zdeněk a Renata Šikulová. 2009. Hodnocení žáků 2., dopl. vyd. Praha: Grada.
Nikolov, Marianne. 2016. "Assessing Young Learners of English: Global and local Perspective" in Educational Linguistics, edited by Francis M. Hult, volume 25. Cham: Springer International Publishing

Vedoucí bakalářské práce: **Mgr. Kateřina Keplová**
Katedra anglistiky a amerikanistiky

Datum zadání bakalářské práce: **2. dubna 2023**
Termín odevzdání bakalářské práce: **30. března 2024**

doc. Mgr. Jiří Kubeš, Ph.D. v.r.
děkan

Mgr. Olga Roebuck, Ph.D. v.r.
vedoucí katedry

V Pardubicích dne 30. listopadu 2023

Prohlašuji:

Práci s názvem Hodnocení testů dosažené jazykové úrovně angličtiny připravených učiteli na středních školách jsem vypracoval samostatně. Veškeré literární prameny a informace, které jsem v práci využil, jsou uvedeny v seznamu použité literatury.

Byl jsem seznámen s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon), ve znění pozdějších předpisů, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Beru na vědomí, že v souladu s § 47b zákona č. 111/1998 Sb., o vysokých školách a o změně a doplnění dalších zákonů (zákon o vysokých školách), ve znění pozdějších předpisů, a směrnici Univerzity Pardubice č. 7/2019 Pravidla pro odevzdávání, zveřejňování a formální úpravu závěrečných prací, ve znění pozdějších dodatků, bude práce zveřejněna prostřednictvím Digitální knihovny Univerzity Pardubice.

V Pardubicích dne 14.6. 2024

Daniel Hofman v. r.

ACKNOWLEDGEMENT

I am grateful to my supervisor, Mgr. Kateřina Keplová, for guiding me during the creation of this thesis. I would also like to thank my friends and family for their support and patience.

ANNOTATION

This thesis focuses on the creation process of teacher-made tests assessing upper-secondary students' English proficiency with a focus on grammar and vocabulary. The theoretical part researches the testing of these aspects of language, principles applied to tests, and planning and designing a test in detail to establish the foundation for the practical part. The practical part describes the way teachers prepare their tests and analyse their methods based on the insights derived from the theoretical part.

KEYWORDS

teacher-made tests, grammar, vocabulary, test qualities, upper-secondary education

NÁZEV

Hodnocení testů dosažené jazykové úrovně angličtiny připravených učiteli na středních školách

ANOTACE

Tato práce se zabývá tím, jak učitelé na středních školách vytváří své testy zaměřující se na gramatiku a slovní zásobu. Teoretická část podrobně zkoumá testování těchto aspektů jazyka, principy, které by měly být na testy aplikovány a plánování a designování testu, navíc slouží jako základ pro praktickou část. Praktická část popisuje, jak učitelé vytvářejí své testy a analyzuje jejich metody srovnáváním s poznatky vyvozené z teoretické části.

KLÍČOVÁ SLOVA

testy vytvořené učiteli, gramatika, slovní zásoba, kvality testu, střední vzdělání

Table of contents

Introduction	9
Theoretical part	10
1. Fundamentals of Testing	10
1.1. Understanding the term test	10
1.2. Assessing Grammar and Vocabulary	11
2. Principles Applied on Tests	14
2.1. Reliability	14
2.2. Validity	19
2.3. Practicality	22
2.4. Authenticity	23
3. Planning and Designing Language Tests	24
3.1. Test Types	24
3.2. Test Specifications	25
3.2.1. Item Types and Tasks	28
3.3. Scoring and Grading	30
Practical part	32
4. Teachers' method of creating language tests	32
4.1. Question 1 - Can you describe your method of creating a unit test in bullet points?	36
4.2. Question 2 - How long, on average, does it take to create this type of test?	37
4.3. Question 3 - With what goal in mind do you create your tests?	38
4.4. Question 4 - If you use the help of AI for the creation of your tests, how exactly do you use it and why?	40
4.5. Question 5 - If you have your class divided into two groups, do you give both groups the same test, or do you give each group a different test, and why?	42
4.6. Question 6 - If a test uses a point system, what scale do you use in order to convert the points to a grade? ..	43
4.7. Question 7 - What challenges do you face when creating your tests?	44
4.8. Question 8 - How often do you use these types of grammar and vocabulary tasks?	46
4.8.1. Analysis of the Acquired Tests	48
Conclusion	50
Resumé	51
Bibliography	54
Appendix	56

LIST OF FIGURES

Figure 1 Pie Chart – Methods of creating a unit test	36
Figure 2 Pie Chart - Time it takes teachers to create a unit test.....	37
Figure 3 Pie Chart - The goal teachers have in mind when creating a test.....	39
Figure 4 Pie Chart - The usage of AI by teachers	40
Figure 5 Pie Chart - The usage of scholarly literature from the field of testing	41
Figure 6 Pie Chart - The usage of the same or a different test for two groups	42
Figure 7 Pie Chart - Points to grade conversation	43
Figure 8 Table - Unique points to grade conversation systems	44
Figure 9 Pie Chart - Challenges that teachers face when creating their tests	45
Figure 10 Pie Chart - Task type preferences	47

Introduction

The sphere of education goes hand in hand with testing; therefore, it is essential to construct tests well to obtain the information needed about the students to make further decisions regarding them. This thesis's theoretical part comprises a short introduction that answers fundamentals regarding the reasons for language testing in general and when applied to testing grammar and vocabulary. Following the introduction is the chapter that delves deep into principles that should be applied to the test: reliability, validity, practicality, and authenticity, each described in their sub-chapter. The last chapter in the theoretical part focuses on planning and designing a language test; this chapter will introduce the test types that might be used within a classroom and the stages of test creation, which involve mentioning task types and scoring. The practical part of the thesis strives to learn how teachers testing upper-secondary students create their tests with regard to the goal they have in mind while making the test. This is accomplished by analysing answers obtained from teachers through an online questionnaire, with the consequent verification of the results of a question tied to the usage of task types through analysis of their grammar and vocabulary tests.

Theoretical part

1. Fundamentals of Testing

1.1. Understanding the term test

The term "test" is crucial to establish before further delving into the problem of testing because understanding what a test is can help us understand the purpose of testing as a whole. This introductory segment explores the similarities and differences in definitions of tests found in scholarly literature, intending to offer a thorough comprehension of the term.

To fully grasp the meaning of a test, it is paramount to analyse different definitions by professionals in the field and find parallels between them. Brown (2004, 3) describes a test as "A method of measuring a person's ability, knowledge, or performance in a given domain." Similar views are held by Douglas (2014, 2), who says, "When it comes right down to it, a test is a measuring device, no different in principle from a ruler, a weighing scale, or a thermometer." Carr (2018, 5) also assumes that the function of a test is that of a tool and emphasises that it needs to be well crafted for it to be useful, just as it is with tests. Therefore, a test is a tool or a method used to measure something; that particular something is the information the test creator is looking for, for example, a test taker's ability, knowledge, or performance in a given domain.

Furthermore, it is essential to mention that the purpose of a test is closely intertwined with what a test is since a tool without a purpose would have no use. Harris and McCann (1994, 26-27) state that the reason for a test is to make decisions concerning the students and if there are no decisions to be made as a result of a test, then the test is meaningless. In like manner, Carr (2018, 6) expresses that the purpose of a test is to help the test maker make decisions. Douglas (2014, 17) also says that tests are used as input when making decisions about the test takers. Thus, the reason for a test or, in other words, its purpose is to help the test creator make certain decisions about the test taker.

Additionally, decisions made as the result of the test help the test maker choose the right test type; the test types are categorised by the type of decision they help to make. The decisions the test helps to make, Carr (2018, 6-9) divides into admission, placement, diagnostic, progress, achievement, proficiency, and screening. Brown (2004, 43-48) similarly lists the following test types: language aptitude, proficiency, placement, achievement, and diagnosis.

The decision the test maker needs to make is reflected in the type of the test. Even though some of these test types may be similar, each one is suited for slightly different decisions.

In conclusion, by comparing and contrasting definitions and insights of various professionals in the field, it is possible to define a test as a tool used by a test maker to measure the test taker's ability, knowledge, or performance in a given domain to make further decisions which involve the test taker. It is essential to state that in the case of classroom tests, the test maker is also in the role of test administrator and assessor.

1.2. Assessing Grammar and Vocabulary

The previous sub-chapter established what a test is and why teachers use it as a tool. The terms grammar and vocabulary are closely tied to this. This sub-chapter will explain what those terms mean and their reflection on language assessment.

Firstly, it is essential to stress the link between teaching and testing regarding certain aspects of language. Purpura (2004, 4) states that during different points in time, grammar in relation to assessment stood for different requirements because of what was taught under this term; for example, being able to memorise grammatical rules, being able to provide accurate translation or being able to select the correct answer on a multiple-choice test. Read (2000, 16-17) emphasises similar parallels in vocabulary, to some teachers, vocabulary is connected to memorising a list of words and finding out if the learner can match the words with their synonyms and definitions or be able to translate the word; while others state that the focus lies in learners effectively using the vocabulary knowledge when they perform the language tasks. Therefore, teaching trends at different points in time and understanding what a particular part of language stands for greatly influences the assessment and, consequently, the requirements of the learner.

The next point to consider is what grammar currently means in pedagogical space. Purpura (2004, 22) describes pedagogical grammar as a "principled description of the target-language forms, created for the express purpose of helping teachers understand the linguistic resources of communication." Similarly, Davies (2007, 21) views pedagogical grammar as a grammatical description of the language that is intended for pedagogical use, in other words, to help with the teaching process. Livia (2006, 7) states that pedagogical grammar, in terms of its position, stands between the linguistic description of language and the use of the language in the classroom. Hence, pedagogical grammar focuses mainly on teaching the language rather than teaching about the language, which would be the focus of linguistic grammar.

Consequently, it is also crucial to analyse grammar for assessment purposes. Regarding grammar, two types of grammatical knowledge are being assessed: explicit grammatical knowledge and implicit grammatical knowledge (Ma 2008, 114). In terms of assessment, it is advisable to test both implicit grammatical knowledge and explicit grammatical knowledge (Purpura 2004, 45). Ellis et al. (2009, 40) state that if the learning process involves both implicit and explicit knowledge, then assessing both is paramount. So, in terms of assessment, two types of grammatical knowledge should be tested: implicit and explicit.

Understanding the difference between these two types of grammatical knowledge is very important for designing the test, starting with the explicit grammatical knowledge. Purpura (2004, 42) describes explicit grammatical knowledge as conscious knowledge related to grammatical forms and their meaning. On top of that, Ma (2008, 115) states that it is a type of knowledge that is accessed slowly. In addition, Ellis (2014, 11) claims that explicit grammar rules are teachable. Explicit grammatical knowledge can, therefore, be understood as conscious knowledge related to grammatical forms and their meaning, which is accessed slowly and can be taught to the students.

On the other hand, implicit grammatical knowledge significantly differs in its meaning in relation to explicit grammatical knowledge. Implicit grammatical knowledge is the type of knowledge used in language behaviour, which is considered natural (Purpura 2004, 42). Similarly, Ma (2008, 115) mentions that implicit grammatical knowledge occurs in natural language behaviour, refers to unconscious knowledge, and, on top of that, can be accessed quickly. Ellis (2014, 11) highlights the fact that in comparison to explicit knowledge, implicit grammatical knowledge cannot be taught because it is built upon a complex neural network of associations. Implicit grammatical knowledge can be understood as a type of knowledge used in naturally occurring language unconsciously that is accessed quickly and cannot be taught.

Furthermore, there is a direct link between the type of knowledge being tested and the type of tasks in a test. Ma (2008, 115-16) links the assessment of explicit knowledge to selected-response, e.g., multiple choice, true-false and matching activities, and limited-production tasks, e.g., gap-filling, short-answer and information transfer activities; meanwhile, extended-production tasks, e.g., essays, dialogues, reports, are associated with testing implicit knowledge. Similarly, Purpura (2004, 45) suggests that when it comes to testing explicit knowledge, it might be done through multiple-choice or short-answer tasks, while implicit knowledge should be assessed through tasks that make learners use grammar in spontaneous

speech – full production in real-time. Thus, tasks focused on testing explicit grammar knowledge take the form of selected-response and limited-production tasks, and for testing implicit grammar knowledge, they focus on extended-production tasks.

Proceeding to the vocabulary section, before delving into its assessment, a short definition of what is seen as vocabulary is required. Cambridge Dictionary defines *vocabulary* as "all the words that exist in a particular language or subject", and by going further and trying to look for the definition of a *word*, one can find "a single unit of language that has meaning and can be spoken or written" (Cambridge University Press, n.d.). Pearson et al. (2007, 284) state that definitions found in dictionaries are lacking in terms of specificity when used for research purposes of vocabulary acquisition and understanding. It is not clear what precisely the term *word* refers to; it can refer to a variety of lexical units, so it is vital for that relevant lexical unit to be defined for the purposes of specific assessments (Read 2000, 20). When defining vocabulary while keeping assessment in mind, the tester must be specific and transparent in what he wishes to assess in terms of vocabulary.

If focusing purely on vocabulary in regard to assessment, it is possible to find more specific answers to the meaning of vocabulary. Pearson, Hiebert and Kamil (2007, 284) call vocabulary assessment "the assessment of students' knowledge of word meaning." However, this creates a question of what it means to know a word, which many researchers in this field discuss. For example, Stahl and Bravo (2010, 567) state that knowing a word is a phenomenon that has multiple parts, and it is not an all-or-nothing phenomenon. For example, Dale (1965; quoted in Read 2000) created four stages of knowing a word, those ranging from stage 1 – "I never saw it before", stage 2 – "I have heard of it, but I don't know what it means, stage 3 – "I recognise it in context, it has something to do with..." to stage 4 - "I know it." Similarly, Paribakht and Wesche (1993, 15) developed a vocabulary knowledge scale for research related to vocabulary learning that has five levels of word knowledge that begins with level 1 – "I have never seen this word" all the way to level 5 – "I can use this word in a sentence." Thus, vocabulary assessment concerns the knowledge of the word's meaning, but knowing a word is a phenomenon that consists of multiple parts.

In conclusion, both grammar and vocabulary have a unique approach to their teaching and, consequently, their assessment methods, which change based on the research done during the time. When testing grammatical knowledge, the tester should keep in mind if their goal is to test explicit or implicit knowledge and, based on this information, design the test format and

the tasks. Similarly, testers creating vocabulary tests must be specific and transparent in what they wish to test.

2. Principles Applied on Tests

This chapter will concern principles that should be followed while creating a test. If the information obtained from administering the test is to be useful, the following principles must be applied to the test: reliability, validity, practicality, authenticity, and impact (Brown 2004, 19). Similarly, Bachman and Palmer (1996, 17) created a model of usefulness which provides a metric for evaluation of the usefulness of the test, which similarly consists of reliability, construct validity, authenticity, impact, practicality and additionally interactiveness. On the other hand, other authors such as Arthur Hughes (1992, 42) or Dan Douglas (2014, 10) are mainly concerned with validity and reliability. Therefore, if the test is deemed to be useful, it needs to follow the qualities mentioned, mainly validity and reliability, but even the other ones play their role; for the purposes of this thesis, this chapter will discuss reliability, validity, practicality, and authenticity.

2.1. Reliability

When it comes to reliability in testing, it is important to explain what it means for the test to be reliable. Brown (2004, 20-1) states that if the test is to be reliable, it has to be consistent and dependable; e.g., if the test is taken on multiple occasions by the same students, the results should be similar. Douglas (2014, 10) describes reliability as the accuracy of the measurement that the test gives, and if the test is to be reliable, it should give as accurate a measurement as possible. Bachman and Palmer (1996, 19-20) view reliability in the same manner; they describe it as consistency of measurement. If the scores of the test takers are similar after taking the test on multiple occasions, the test is perceived to be more reliable (Hughes 1992, 29). Kubiszyn and Borich (2003, 311) state for a test to be considered reliable, it must give the same or similar results over repeated administrations. Thus, reliability can be understood as consistency of measurement; consequently, it should be accurate and dependable.

Additionally, there are a number of factors that can influence the consistency of measurement which has to be taken into consideration when talking about reliability. According to Brown (2004, 21-2), those factors may be fluctuations in students, scoring, test administration, and the test itself. There are many reasons for inconsistencies in the measurement of a test,

stemming from the side of the student or the test itself (Douglas 2014, 10). Bachman and Palmer (1996, 20) also state that it is not possible to remove all inconsistencies, but the job of the tester is to minimise them through the test design. So, the inconsistency comes partly from the side of the student, the test itself and its scoring and administration; consequently, the tester's job is to minimise the inconsistencies through the design of the test.

Firstly, delving into the inconsistencies can be caused from the side of the student. Deviations from the true score can be caused by the student's illness, fatigue, anxiety, or other factors related to their well-being (Brown 2004, 21). Similar causes are also mentioned by Douglas (2014, 29), who highlights students' energy levels and mental state, e.g., fatigue, daydreaming, misreading instructions, or forgetfulness to fully finish certain tasks. Hughes (1992, 33) elaborates on the idea of a distinction between the true score and the actual score and states that it is probable that the actual score does not represent their best score (true score). Inconsistencies on the student's side are mainly caused by their mental or physical state, which deviates them from their true score.

In contrast, there are also some inconsistencies that can be caused by the rater. Brown (2004, 21) divides these inconsistencies into two groups: inter-rater reliability and intra-rater reliability; the difference between these two groups is that inter-rater reliability is when two or more scorers' results differ on the same test, e.g., the causes such as biases, inattention, and inexperience, while intra-rater reliability usually occurs to the classroom teachers due to reasons such as fatigue, unclear scoring criteria or bias towards certain students. While Douglas (2014, 27) does not divide rater's inconsistencies into these groups, he still mentions that inconsistencies can arise due to a lack of experience in rating performance and scoring errors. The concept of inter-rater reliability could be connected to Hughes (1992, 35-6), who introduces the idea of the scorer reliability coefficient, which is an agreement between different raters of the test in their results and is represented by a numeral value. Inconsistencies that could be caused by raters are divided into two groups, intra-rater and inter-rater reliability, both stemming from different problems. Inter-rater reliability can be measured using the scorer reliability coefficient.

Furthermore, reliability may be affected by the administration of the test and the test itself. Administration inconsistency stems from the physical elements during testing, e.g., noise, temperature, condition of equipment in the classroom or even readability of the test; similarly, even the test itself can cause inconsistencies, e.g., the length of the test, poorly designed test

tasks and the time limit (Brown 2004, 21-2). Douglas (2014, 10) also states that the test itself can cause inconsistencies because of the design of the tasks; bad design can stem from instructions, which could be unclear, or the difficulty, which may exceed the targeted level of the student, and quality can also be influenced due to tasks having multiple correct answers. Thus, administration inconsistencies are tied to the physical state of the classroom and the equipment used, while the inconsistencies caused by the test stem from its design.

All these inconsistencies play into the reliability coefficient of the test, which can help interpret the quality of the test in terms of reliability. The reliability coefficient is a numeral value between 0 – results that are unconnected with each other -and 1 – results that are the same every time - that represents the reliability of the test (Hughes 1992, 31-2). Carr (2018, 108) states that "reliability values range from .00 – all errors – to 1.00 – perfect measurement." Kubiszyn and Borich (2003, 311) claim that the most used methods to measure reliability are test-retest, alternate form, and internal consistency. Therefore, the reliability coefficient is a numeral value between 0 and 1, which represents the reliability of the test. There are three methods to obtain this value: test-retest, alternate form, and internal consistency.

It is essential to mention that each test type has slightly different expectations for the reliability coefficient. Lado (1961; quoted in Hughes 1992) claims that, for example, vocabulary and reading tests should be in the .90 to .99 range, while listening tests are usually in the .80 to .89 range, meaning that reliability in testing different abilities varies. Hughes (1992, 32) additionally mentions that based on the decision made regarding the test, the reliability coefficient must be adequate to the decision, e.g., entrance exam tests. Carr (2018, 109) states that when it comes to high-stakes testing, the minimally accepted reliability is .80. Depending on the test type and decision made in relation to the test, the reliability coefficient differs, but it should be at least .80 or around this value.

As mentioned previously, there are multiple methods to arrive at the reliability coefficient, starting with the test-retest method (sometimes called stability). As the name implies, the test-retest method makes use of two sets of scores from two same tests given at different points in time (Kubiszyn and Borich, 311-12). For the test-retest method, one needs to have two sets of scores from a group of subjects that take the same test two times (Hughes 1992, 34). Carr (2018, 109) describes test-retest as an approach that makes the student take the same test twice, and based on two sets of scores the reliability correlation is calculated. Hence, the test-

retest method makes use of participants taking the same test twice to obtain two sets of scores, which are used to calculate the reliability coefficient.

However, there are some drawbacks to using this method as a means to obtain a reliability coefficient. Hughes (1992, 32) highlights the problem of this method in regard to the time gap between the two times needed for the administration of the test; if the second test is too soon, then the participants may recall the answers, but on the other hand, if the second test is too late then participants may forget the subject matter. Similarly, Carr (2018, 109) mentions if the two tests are administered too close to each other, the participants may achieve better results on the second attempt thanks to the familiarity, but if they are too far apart, then the participants may learn something new resulting in better results. Kubiszyn and Borich (2003, 312) mention that the results of the participants may differ not because of the unreliability of the test but because the participants themselves have changed. When it comes to using the test-retest method, the biggest drawback comes from the reliability of the second score due to the change in the participants between those two tests.

Another method that can be used in order to obtain the reliability coefficient is called alternate forms (sometimes called parallel forms or equivalence). This approach, similar to the test-retest method, makes use of two test administrations, but the main difference is that instead of the same test being administered twice, the tests, in this case, are parallel using the same task formats, assessing the same construct but containing different tasks (Carr 2018, 110). An estimate of reliability can be obtained through the usage of two equivalent forms of a test, both being administered to a particular group of participants and correlation is determined between the two sets of scores (Kubiszyn and Borich 2003, 313). The alternate forms method makes use of two equivalent forms of a test with the same task formats, assessing the same construct but with different tasks.

By using alternate forms method, it is possible to eliminate certain problems that stem from using test-retest. Effects of the time on the participants, e.g., forgetting the subject matter, familiarity with the test, or learning something new, are reduced when using the alternate test method (Hughes 1992, 32). Kubiszyn and Borich (2003, 313) state that "This estimate eliminates the problems of memory and practice involved in test-retest estimates." Therefore, with the usage of the alternate forms method, the tester can eliminate problems tied to the time gap between the administrations of the test-retest method.

However, the alternate forms method still has certain drawbacks, keeping it from being the ideal method to obtain reliable data. Carr (2018, 110) states that in practice, parallel forms are never truly parallel, which means that the score is never exactly identical. This is tied to the fact that it is not a good idea to create two tests just to check reliability. Similarly, Kubiszyn and Borich (2003, 313) emphasise the fact that it is pretty challenging to create two good tests in order to measure reliability, mentioning that this method is mainly used by test publishers. The alternate forms method mainly suffers from the practicality and the amount of effort required to create two high-quality tests that could be considered truly parallel for the purpose of reliability measurement.

The last approach that is used to measure consistency is called internal consistency. Carr (2018, 110) describes the method in the following way: "As the name suggests, look at the extent to which different parts of the test are measuring the same thing; or, to look at it differently, the degree to which they are measuring it in the same way." Kubiszyn and Borich (2003, 313) state that if the test is internally consistency, then items in the test are correlated to each other. Hughes (1992, 32) adds that this is the most common way to obtain reliability data because it requires only one administration of one test. The internal consistency method requires only one test administration, and the data are obtained through the correlation of test items, the degree to which they are measuring the same thing.

The internal consistency approach makes use of the split-half reliability method in order to measure consistency just from one administration. This method makes use of a test which is split into two equivalent halves, and then the correlation between them is determined (Kubiszyn and Borich 2003, 313). In this method, participants get two scores, similar to test-retest or alternate-forms, but instead of multiple administrations or tests, the scores are given for each half of the same test (Hughes 1992, 32-33). The internal consistency method derives the consistency coefficient from one administration of one test, which is divided into two halves.

For this method to work, it is crucial to divide test items equally. The best approach to divide the test items – if they are not randomly spread in terms of difficulty - in the test appropriately is to place all even-numbered items into one half and odd-numbered items into the other half, making the split-half reliability method, sometimes called odd-even reliability (Kubiszyn and Borich 2003, 313). To ensure that the split-reliability method works, both halves of the tests must be equivalent; if the test is ordered in terms of difficulty, then it is advised to split items

into odd-numbered and even-numbered (Hughes 1992, 33). One of the safest ways for the test to be divided is to put even-numbered questions into one half and odd-numbered questions into the other (Carr 2018, 110). If the test items are ordered in terms of their difficulty, then the best way to divide them is by using odd-even reliability, i.e., putting even-numbered items into one half and the odd-numbered items into the other half.

However, this approach requires a special formula so the correct reliability coefficient can be obtained. The Spearman-Brown prophecy formula is used to adjust the correlation upwards because, in the split-half method, each half is treated as its separate test, even though two tests would be longer (Carr 2018, 110). The formula used to adjust the coefficient upwards is due to the fact that the split-half method uses a test that is significantly shorter; the formula is called the Spearman-Brown prophecy formula (Kubiszyn and Borich 2003, 314). The formula used to obtain the adjusted reliability coefficient for the split-half method is called the Spearman-Brown prophecy formula.

In summary, reliability is one of the key qualities of the test and it is possible to define it as consistency of measurement. The inconsistencies stem from the student, tester, administration process, and the test itself. The inconsistencies deviate students from their true scores. There are many methods used to obtain the reliability coefficient of the test; the one that is most used is called the split-half method; however, the coefficient must be adjusted using the Spearman-Brown prophecy formula. It is essential to state that it is not possible to remove all inconsistencies, but the tester's job is to minimise them through the design of the test.

2.2. Validity

Along with reliability, another important principle that should be applied to a test is validity. Hughes (1992, 22) states that for a test to be valid, it needs to accurately measure what it is intended to measure. If the test measures what it is supposed to measure, then it can be considered valid (Kubiszyn and Borich 2003, 299). Carr (2018, 151) states that validity often refers to whether the test assesses what it is meant to assess, but he remarks that it is an oversimplification. A slightly more complex description is provided by Gronlund (1998; quoted in Brown 2004), who describes validity as "the extent to which inferences made from assessment results are appropriate, meaningful, and useful in terms of the purpose of the assessment." On a similar note, Douglas (2010, 10) says that "validity concerns the appropriacy of the inferences made on the basis of the test performance." Thus, many

interpretations slightly differ or oversimplify the principle of validity, but for the basic understanding, any of the mentioned descriptions can be used.

Validity is established based on the evidence that may support it. First of all, there is content validity (or content-related evidence), which is described by Mousavi (2002, quoted in Brown 2004) as a type of evidence that revolves around the test taker performing behaviour, which is the subject matter of the test from which conclusions are drawn. Similarly, Kubiszyn and Borich (2003, 300) state that content validity evidence is concerned with the question of whether the test measures, matches and fits the instructional objectives of the test. Hughes (1992, 22) says that if the test is valid in terms of its content, then it is made of the language skill, structure, etc., with which it is concerned. Carr (2018, 315) defined content validity as "The adequacy of with which the test items sample the universe to which the test developer wants scores to generalise." Therefore, content validity nowadays is not viewed as a separate type of validity but just as evidence that focuses on the content which supports the validity on the whole level.

There are certain issues that may influence the content validity of the test. Sometimes, the test may look valid in terms of content but measure something different than intended, for example, the ability to guess, due to the test items being too difficult or poorly constructed; content validity does not guarantee that a test will be good, but it is a minimum requirement for it to be useful (Kubiszyn and Borich 2003, 300). If the areas the test is supposed to measure are under-presented or not present at all, it is not likely to be accurate in terms of its validity. Additionally, tests of these types can have harmful backwash effects – the areas that are not present in tests will be less touched upon in teaching (Hughes 1992, 21). Difficulty and construction of test items, together with their selection, may significantly influence the validity of the test and even the teaching process as a whole due to the backwash effect.

Another piece of evidence related to the validity of the test is criterion-related validity. To establish criterion-related validity, the test score is correlated with an external criterion (Kubiszyn and Borich 2003, 300). Likewise, Brown (2004, 24) states that "Criterion-related validity refers to the extent to which the "criterion" of the test has actually been reached." Moreover, there are two kinds of criterion-related validity: concurrent validity and predictive validity (Hughes 1992, 22). Criterion-related validity makes use of a correlation between the score and certain criteria; moreover, there are two types of criterion-related validity: concurrent and predictive validity.

The first type of criterion-related validity that will be discussed is concurrent validity. Kubiszyn and Borich (2003, 300-301) remark that concurrent validity deals with measures that are to be administered at the same time with the goal of validating certain measures; furthermore, this method makes use of a correlation coefficient, and it is determined by the correlation of two sets of scores – the new test and the already established test. On a similar note, Brown (2004, 24-25) states that for a test to have concurrent validity, it has to have results that are supported by other performance, e.g., a high score on a language test is supported by actual proficiency in the language. Whether the level of agreement – validity coefficient – is acceptable depends on the importance of the decision made as a consequence of the test and the purpose of the test (Hughes 1992, 24). Concurrent validity uses a correlation between specific scores and criteria that are administered at about the same time, and based on the coefficient, it is possible to tell the particular level of validity.

The second type of criterion-related validity is predictive validity, which differs in its usage. Hughes (1992, 25) states that this type of validity is concerned with the degree to which the test can predict the future performance of the candidate. Brown (2004, 25) gives examples of types of tests for which this type of evidence is essential, e.g., placement tests, language aptitude tests and admissions assessments batteries because the criterion to assess in these types of tests is whether the test taker will be successful in the future. Similarly to concurrent validity, even predictive validity makes use of correlation to obtain a validity coefficient; the predictive validity evidence of the test is obtained by measuring the subject of the test after a period of time has elapsed (Kubiszyn and Borich 2003, 301). Predictive validity is concerned with predicting future criteria and makes use of the validity coefficient obtained through the correlation of the scores from the test that is supposed to predict the future criterion and the test given to the subjects on the topic after a certain period of time.

The third type of evidence that reinforces validity is construct-related evidence (often referred to as construct validity). Carr (2018, 315) defines construct validity as the "Degree to which it is appropriate to interpret a test score as an indicator of the construct of interest." For this definition to work, one has to understand what a construct is; Fulcher and Davidson (2007, 7) state that constructs are concepts that are measurable and defined in relation to other different constructs. Brown (2004, 25) explains a construct as a model, hypothesis or theory that tries to make sense of phenomena in what it is possible to perceive. For example, Hughes (1992, 26) focuses more on the language aspect and states that construct refers to any underlying trait or ability which is presumed in a theory of language learning. Construct validity revolves

around the measurement of the construct, i.e., the operational concept, which can be measured and defined.

The last evidence of validity is face validity. Mousavi (2002; quoted in Brown 2004) claim that face validity is the degree to which the test appears to look right and measure the ability or knowledge it claims to measure. Similarly, Hughes (1992, 27) states that for the test to have face validity, it means that the test appears to measure what it is supposed to measure, e.g., a test measuring pronunciation ability without requiring the test taker to speak is missing the face validity. Carr (2018, 165) claims that face validity refers to whether or not the test appears to be valid based on the looks alone, which does not work well and claiming for the test to be valid based on face validity alone is perceived to be highly unprofessional. Therefore, face validity is a highly subjective judgment regarding the validity of a test based on looks alone.

To sum up, validity can be understood as the accuracy of measurement of what it is supposed to be measuring. There are many types of evidence that can support the validity of the test, the main being content validity, criterion-related (validity which is divided into concurrent and predictive validity), construct validity and face validity. It is paramount to state that they are not types of validity that stand on their own but merely evidence that supports the validity of the test on the whole level.

2.3. Practicality

Another aspect that a well-designed test should follow is practicality, differing in nature from other qualities mentioned in this chapter. Authors such as Carr (2018, 20) and Fulcher and Davidson (2007, 269) describe the practicality of a test as the number of resources one has to invest in its design and usage. Similarly, Harris and McCann (1994, 92) define practicality as all aspects concerning tests which affect time and resources. Bachman and Palmer (1996, 36) state that practicality is the relation between the resources required for the creation and administration of the test and the resources that are available. Thus, practicality can be understood as the number of resources one has to invest in the test development, administration, and evaluation in contrast to the available resources.

Furthermore, it is crucial to elaborate on what is meant by resources. Bachman and Palmer (1996, 36-37) divide resources into three categories: human resources (test writer, scorer, administrators – in the case of the classroom tests, it is the same person), material resources (space, equipment, materials) and time (development time). Brown (2004, 19-20) states that if

the test is practical, then it follows certain principles, but similarly to Bachman and Palmer, Brown mainly focuses on time constraints (development, administration, evaluation) and material and human resources (materials, equipment, examiners, evaluators). Therefore, resources are, in terms of practicality, understood as time, material, and human resources.

In closing, even though practicality is the shortest quality described in this chapter, this does not imply that it is any less important than others. Practicality revolves around the number of resources – time, material, human - one can allot to the test. If the number of resources exceeds a certain threshold, then the test can be seen as impractical. In other words, practicality could be understood as a balance between all other qualities the test should follow.

2.4. Authenticity

Another quality which is essential to follow is authenticity. Fulcher (2007, 15) defines authenticity as "The relationship between test task characteristics and the characteristics of tasks in the real world." Bachman and Palmer (1996, 9-10) state that authenticity can be understood as the extent to which the tasks on the test are similar to real-life language use, though it is difficult to define authenticity due to many terms being used to describe the same, e.g., pragmatic, functional, communicative, performance, and how should one distinguish real-life language from a non-real one. In other words, if a test task is authentic, then it can be put into practice in the real world, yet many test tasks fail to do this for reasons such as artificially targeting certain lexical items or grammatical forms (Brown 2004, 28). Authenticity can be understood as a connection between a task and usage in the real world; moreover, it can be hard to achieve due to the need to target particular aspects of language in assessment.

Another aspect of authenticity that should be discussed is its implementation in language testing. Brown (2004, 28) states that authenticity may be present in a test in the following ways: relevancy of topics for the learner, contextualisation of test items, thematic organisation of items, the language used in the test being as natural as possible, and task representing real-world tasks. On the other hand, Wang (2008, 34) remarks tests are created in artificial contexts, which affects the authenticity; moreover, the tasks are simply trying to imitate real life and are not real communicative tasks, e.g., job-relevant interviews. The implementation of authenticity is a difficult task due to the understanding of the term itself and the limitation of the tests as a tool, but there are particular strides to be made.

Authenticity represents the connection between the language used in a task and the usage in the real world, while the term itself and the comprehension of authenticity may be, to a certain degree, unclear in the field of language assessment due to many terms being used for the same concept and the concept itself being affected by the artificiality of testing. However, for the test maker, it is crucial to keep the language as natural as possible, together with contextualising the test items.

3. Planning and Designing Language Tests

The previous chapter was concerned with principles that should be applied to the test and their consequent measurements, while the focus of this chapter is on steps taken during the planning and designing process; therefore, the selection of the test type, the construction process of the test, selection of tasks and scoring method. Knowing the variety of options the test maker has at hand while developing a test can positively affect the principles that are applied to the test and, as a result, create a better test.

3.1. Test Types

As mentioned in the first chapter, tests are tools that help the test maker make certain decisions about the test taker. There are many decisions that can be made, e.g., placement, diagnosis, progress, and entrance, etc., but this thesis is concerned with teacher-made tests; therefore, decisions made as a consequence of the tests must correspond to that, which means that this chapter will focus mainly on test types that occur in a classroom with a goal of explaining the specifications behind the test type.

Firstly, it is paramount to state the kind of test types that can appear in a classroom. Brown (2004, 43) states that classroom teachers will most probably find themselves creating achievement, diagnostic, and placement tests, while aptitude and proficiency tests are highly unlikely to be created by a classroom teacher. Similarly, Carr (2018, 7-9) divides tests based on whether they are related to teaching or learning curriculum, called curriculum-related decisions, and if they are not, then they are classified into groups called other decision types. Into curriculum-related decisions, Carr includes placement, diagnostic, progress, and achievement tests. Alderson, Clapham, and Wall (1995, 11-12) also divide tests into placement, progress, achievement, and diagnostic tests; Alderson, Clapham, and Wall do not per se state that these test types are classroom-based, but it can be understood from the way the test types are described. Thus, the tests that are related to the teaching and learning

process that are likely to appear in a classroom are achievement, diagnostic, placement, and progress tests.

Most importantly, this thesis is concerned with achievement tests; therefore, they need to be introduced appropriately. Carr (2018, 7) says that achievement tests are concerned with whether the students have learned what they were meant to learn and subsequently achieved the course objectives. Since the main objective of the achievement test is to see if the goals of the course objectives have been met, they are usually administered at the end of the term of study or a unit, which makes them summative in nature (Brown 2004, 47-48). Additionally, Alderson, Clapham, and Wall (1995, 12) remark that the content of achievement tests is based on the course textbook or course syllabus. Achievement tests are mainly concerned with the result of learning and are used at the end of a unit or a course.

However, there are certain parallels between achievement tests and progress tests that should be mentioned. Carr (2018, 7-9) emphasises that a test that is, for example, administered in the middle of the course can measure how well the materials were learned up to this point in time but also on what to focus on in the future if they were not mastered; hence the test can function both as an achievement test and a progress test. While the achievement tests are summative in nature, they can also take on a formative role through the washback effect in regard to students' performance (Brown 2004, 48). Therefore, a test can take on the role of both achievement and progress test at the same time; these roles are not mutually exclusive.

When it comes to types of decisions that are made in the context of a classroom, teachers usually focus on measuring and diagnosing specific aspects of language in order to place students into certain levels or to see whether the results of learning were achieved and if not what to focus on in the future. The test types, as mentioned before, are simply decisions that the test maker makes in consequence of the test, and it is possible to make multiple decisions after a test, which means that we can have a test functioning both as an achievement test and a progress test; the test types mentioned in the literature are not to limit the test maker when creating a test but to give options on what to focus on.

3.2. Test Specifications

A test creation process consists of many moving parts which the test maker needs to take into consideration because if the test maker wants to accomplish the goal that they have in mind by administering the test, the understanding of these moving parts and their planning becomes of the utmost importance. This section of the chapter will focus on introducing these moving

parts, i.e. test specifications, and their place in the construction planning phase of the test creation.

Firstly, it would be beneficial to briefly introduce what is understood as a test specification. Test specifications are a document which functions as a blueprint for test creation, and it gives information regarding what and how the test is testing (Alderson, Clapham, and Wall 1995, 9). Brown (2004, 50) similarly states that test specifications are used as an outline for a test; he adds that for large-scale testing, these specifications are distributed while being more formal and detailed than those intended for classroom use. Carr (2018, 47) describes test specifications as "Plans for a test, which are used as an outline when the test is written, and as directions for the people writing the test." Therefore, test specifications are plans for the test that serves as an outline, and they may differ slightly in form and usage if they are created for classroom testing or large-scale standardised testing.

Before delving into the kinds of test specifications the test maker needs to consider, it is worthwhile asking a question regarding the importance of the planning process as a whole. Carr (2018, 47) states that the planning process of the test should be equal to the test's importance and length; moreover, for certain test types, multiple people can be involved in the planning process. Furthermore, Alderson, Clapham, and Wall (1995, 10) emphasise the importance of test specifications for different users regarding the scope and use of the test, e. g., test writers, people who review the validity of the test, test users, teachers preparing learners for public exams or even admission officers making decisions on the basis of the test. It is possible to see that test specifications play a more significant role depending on the importance of the test; moreover, test specifications are helpful not only for the person creating the test but also for people who come in contact with the test.

The categorisation and types of test specifications may slightly differ based on the author's writing about the topic; illustrating some of the views on test specification can prove to be helpful for the test maker in terms of a better understanding of what to focus on. Brown (2004, 50-51) divides test specifications into three categories: 1) a broad outline of the test, 2) skills tested, and 3) what the items will look like. Carr (2018, 48) also uses three categories for the test specifications: 1) specifications regarding test context and purpose, 2) specifications of the overall test structure, and 3) specifications of the individual test tasks. Hughes (1992, 49) states that test specifications should consist of information regarding scoring procedures, criterial levels of performance, format, timing and content. Some authors such as Bachman

and Palmer (1996, 88) and Alderson, Clapham, and Wall (1995, 11-12) do not divide test specifications into these broader categories but, right away, state what should be included in them. The division into broader test specification categories is not done by all authors, but it is possible to say that test specifications could be divided based on the test structure, its purpose and context, and the selection of test items.

An outline of the test and skills tested would be the first step to take when writing test specifications. Brown (2004, 50-51) uses the example of a midterm test; for the outline of the test, the test maker should have in mind the allocation of time for the test and its passage and what skills will be tested. Similarly, Hughes (1992, 49-50) states that the test maker should have a specific test structure in mind, which includes time allocation to components of the test and test specification contents (entire potential content of any possible version of a test) should comprise of the language skill tested, though the description of the contents varies depending on its nature. Thus, the outline of the tests should cover the allocation of time for the test and its passages, and the contents should cover the language skills tested, which are described based on their nature.

The thesis focuses on grammar and vocabulary achievement tests; for that reason, commenting on the test specification contents of such tests is paramount. Hughes (1992, 142-147) states that for grammar tests, the content test specifications should contain all the grammatical structures being taught; similarly, the vocabulary tests should include all the relevant vocabulary being taught. The primary considerations that apply to specifications for grammar and vocabulary tests are the distinction between the ability to use a structure or a word and comprehend that structure or recognise a word correctly; other considerations are regarding the range of grammatical structures or vocabulary along with the degree of mastery of the structures and words produced (Carr 2018, 75). Specifications that consider purely grammar and vocabulary part of the test are considering its contents in terms of the relevant grammar structures and vocabulary while also focusing on the distinction between the ability to recognise and use structures and words correctly.

Therefore, the test specifications vary depending on their usage, be it in the context of classroom testing or standardised testing. Furthermore, the importance of test specifications may differ depending on the importance of the test, but it is possible to state that their usage should not be limited only to standardised testing because they can prove to be beyond helpful, even for classroom testing. In terms of classroom testing, the main test specifications the

teacher should have in mind when creating a test are the outline of the test, skills that will be tested and items and tasks.

3.2.1. Item Types and Tasks

As previously mentioned, test specifications should also contain information regarding test items and tasks, but the decisions made in regard to test items and tasks are more complex and vital; therefore, they will be described in a sub-chapter of their own. This sub-chapter will focus on eliciting responses, responding and types of tasks that can be included in grammar and vocabulary tests.

Before delving into specific task types, it is pivotal to list the elicitation modes and response modes the test makers can use when writing their tests. Brown (2004, 51-52) divides elicitation and response modes into two categories, those being oral and written; the written elicitation modes can be a word, set of words, and question directions, and the response modes can be a multiple-choice option, fill in the blank, spell a word, define a term or a short answer. Similarly, Carr (2018, 26) groups these response types into two main groups based on the examinee's production; those groups are selected response tasks in which the examinee selects the correct answer out of multiple choices and constructed response tasks in which the examinee writes something in response; furthermore, the latter group can be subdivided into extended production and limited production based on the length of the response. Ultimately, elicitation modes can range from a simple word up to question directions and response modes can be either a selected response or a constructed response.

Now that two main categories of response modes have been established, it is possible to focus on the task formats of each, starting with the selected response task formats. Carr (2018, 29-31) lists four different task formats, which include multiple-choice, true-false questions, matching, and ordering tasks, the multiple-choice task format being the most recognised among the selection. Despite the fact that the multiple-task format may be the most recognised, both Brown (2004, 55-56) and Carr (2018, 29) state this task format is exceptionally difficult to design correctly. Moreover, Hughes (1992, 60) states that the multiple-choice task format suffers from certain weaknesses that were recognised by the professional at the time, but before that, it seemed like the best way to test. Thus, the selected response tasks range from multi-choice to ordering tasks, with the former being the most recognised but challenging to design while also suffering from certain weaknesses that were overlooked in the past.

Given the popularity of the multiple-choice task format, it is essential to mention the weaknesses it suffers from, together with the reasons for its popularity. Hughes (1992, 60-62) lists the following weaknesses: 1) task format tests only a recognition knowledge, 2) guessing plays a considerable role in test scores, 3) restricts what can be tested, 4) difficult to write successful items 5) harmful backwash effect and finally 6) makes cheating easier. Conversely, there are many upsides to using this task format; both Brown (2004, 55-56) and Carr (2018, 29) emphasise the positive effects this task format has on scoring in terms of the time saved and consistency process. In conclusion, the multiple-choice task format can shine in the large-scale standardised test setting more than in the classroom setting due to the ease of scoring and its consistency coupled with the fact that creating items for this task format is demanding on the test creator.

Moreover, it is pertinent to mention the way some of these weaknesses can be reflected in the context of testing grammar and vocabulary. Hughes (1992, 60-61) states that the multiple-choice task format may provide a distorted view of the examinee's ability since the examinee shows they can identify the correct response but not produce it correctly; furthermore, it can be challenging to come up with distractors¹ because there may not be that many alternatives to the correct structure. In terms of multiple-choice weaknesses reflected in vocabulary testing, Wesche and Paribakht (1996; quoted in Read 2000) mention that the examinee may know a different meaning of the word but not the one sought or that instead of testing the ability to identify a targeted word the items may test examinees' knowledge of distractors. Thus, as can be seen, the multiple-choice task format has its weaknesses when it comes to testing both grammar and vocabulary.

A constructed response is the second response mode category, which is often used when testing grammar and vocabulary; however, for the purposes of this thesis, the focus will be solely on limited production tasks. Carr (2018, 31) states that independent limited production tasks used for testing grammar are most frequently short-answer questions, but other examples are gap-fills and sentence combining or writing tasks. Additionally, Hughes (1992, 143-148) lists specific techniques which can be considered as limited production tasks in nature; techniques for testing grammar encompass paraphrasing, completion and modified cloze types of tasks, while techniques for testing vocabulary consist of definitions and gap filling; though Hughes remarks that all techniques can be adapted to multiple-choice format but finding suitable distractors would be problematic. Therefore, limited production tasks are

¹ The incorrect answers in the multiple-choice task format.

a viable choice for testing both grammar and vocabulary, and all of the limited production techniques can be adapted to multiple-choice task format if needed.

Similar to the multiple-choice task format, using limited production tasks has its advantages and disadvantages. Carr (2018, 31-32) claims that it is less difficult to construct good limited production tasks than selected-response tasks, but the main disadvantages of limited construct tasks revolve around scoring; they take longer to grade than selected-response tasks, and there is a possibility for more than one correct answer. Despite these disadvantages of limited production tasks, Hughes (1992, 143) emphasises their importance for testing grammar over selected response tasks due to the weaknesses of multiple-choice tasks. The problems tied to limited production tasks lie mainly in their scoring, but their usage is recommended for testing grammar by authors such as Arthur Hughes.

In conclusion, the main two response modes used in testing grammar and vocabulary are selected response and constructed response. The most frequently used task format of selected response tasks is multiple-choice, which has its weaknesses when it comes to testing grammar and vocabulary; most notably, it is challenging to construct, but if constructed properly, it dramatically increases the practicality of the test in terms of the scoring process. On the other hand, constructed response tasks are not as challenging to construct, but the disadvantages can become evident in the scoring process.

Once the test specifications have been established and test items and tasks have been selected, it is possible to start writing the test. The process of writing the tests is intertwined with the test specifications; the creation of a test consists of sampling, which focuses on the selection of vocabulary and structures from the test specification contents, item writing and moderation, which involves a selection of task types together with the quality of its construction and finally, it is essential to write a scoring key to be able to award points for the appropriate responses to the selected test tasks (Hughes 1992, 51-52). Therefore, in order to create the best possible test for their students, the teacher should create test specifications, which then work as a blueprint for the test writing process.

3.3. Scoring and Grading

As mentioned previously, the consequent scoring of a test is an important aspect of a test creation process. For this reason, it is essential to dedicate a sub-chapter describing the basics of scoring and the assignment of grades to the performance of the students.

The basis of scoring relies on assigning a numerical value to each test item. Alderson, Clapham, and Wall (1995, 148) state that objective items may be assigned with a mark of 0 if wrong and with a mark of 1 if correct. If the test has multiple sections, the test maker needs to decide the importance of each section and, based on that, assign the numerical value of each item in a set section (Brown 2004, 61). Thus, it can be said that if the task items can be objectively correct or incorrect, it is possible to give 1 point for each correct answer, though this may change if the test comprises multiple sections of different importance.

When it comes to scoring grammar, it is paramount to be precise about what each item is testing. Hughes (1992, 145) emphasises the fact that no points should be deducted for mistakes that are not being tested by the item; moreover, if the test item is testing more elements, each should be assigned its numerical value in terms of points. The aspect that needs to be kept in mind when scoring grammar items is that the element being tested should award or deduct points, not any other unrelated mistake.

Upon the test being scored, it is possible to grade the student's performance, but there are many aspects influencing the points to grade conversation. Brown (2004, 62) lists myriads of reasons that shape the process of grading, for example, the expectations of the institution, the country, culture, and students' expectations caused by the previous tests, to name a few.

To conclude, this sub-chapter only consisted of bare basics that need to be kept in mind when scoring a test and factors that influence the conversion of points to a grade. The main key points are each item should have an assigned numerical value based on its importance; if the answer contains mistakes that are not relevant to the phenomenon tested, no points should be deducted, and finally, the grading performance of students varies based on the institution, country, culture and the expectations set by previous tests.

Practical part

The purpose of the practical part of this thesis is to learn how teachers in upper secondary education create their tests and consequently analyse their tests in terms of the used task types. This is done through the analytical interpretation of the data obtained; at first the thesis was supposed to obtain data through focus group method, but this method was quite time-consuming for the teachers, so it was not possible to organise the meeting, which made this way of obtaining data impossible. Therefore, another method was chosen, which was an online questionnaire that was sent to the selected upper secondary schools for the English teachers to fill in in their free time. Though an online questionnaire is not as flexible as the focus group interview in terms of the amount of information obtained, in this case, the usage of an online questionnaire method was the only way to obtain the data needed for the analysis to be possible.

4. Teachers' method of creating language tests

There are many methods and steps in regard to test creation, and the focus of this part will be to gain insights into the way English teachers in upper secondary education prepare their tests. The main area this section will be concerned with is the test creation process the teachers follow with a focus on the end-of-unit tests, which can be seen as the achievement tests in this regard; tied to this is the amount of time it takes for them to construct a test, the goal of the test they have in mind while creating it, the usage of external resources, the types of tasks the teachers tend to include in their tests, the way they convert the points to the grade and finally the challenges they face when they create their tests.

To obtain the data needed for this part, an online questionnaire was sent to various upper secondary schools for the English teachers to fill in and consequently share their experience in creating achievement tests focused on grammar and vocabulary. Furthermore, to obtain as many responses as possible and therefore have a larger data pool for analysis, the help of the author's university colleagues, who taught or had connections to teachers at upper secondary schools, was used to distribute the online questionnaire.

The questionnaire itself contained seventeen questions, and they could be divided into two types. The first type of questions was focused on the creation process, design choices, and external sources used, which would be primarily open questions or closed questions consisting of yes or no. The second type of question was focused on the frequency of usage of

different types of grammar and vocabulary tasks; in this part, the Likert scale consisted only of four options, ranging from one – "I never use this type of exercise in my tests," to four "I always use this type of exercise in my tests," this was done in order to avoid the centrality bias and force teachers to form an opinion.

Now that the structure and the design choices of the questionnaire have been introduced, it is possible to go over the specific questions and their purpose. However, it is essential to mention that the questions were originally written in Czech and distributed that way.

Therefore, the questionnaire consisted of the following open questions:

1. Can you describe your method of creating a unit test in bullet points?
2. How long, on average, does it take to create this type of test?
3. With what goal in mind do you create your tests?
4. If you use AI to create your tests, how exactly do you use it and why?
5. If you have your class divided into two groups, do you give both groups the same test, or do you give each group a different test, and why?
6. Do you prefer the type of tasks in which students choose between pre-defined answers or tasks in which they need to come up with the answer by themselves, and why?
7. If a test uses a point system, what scale do you use in order to convert the points to a grade?
8. What challenges do you face when creating your tests?

Then, there were closed questions with yes or no answers:

9. While creating tests, do you use literature from professionals in the field of testing?
10. Do you use the help of AI for the creation of your tests?

Lastly, the questionnaire involved questions using the Likert scale:

11. How often do you use the grammar task "Paraphrase"?
12. How often do you use the grammar task "Completion"?
13. How often do you use the grammar task "Modified cloze"?
14. How often do you use the vocabulary task "Synonyms"?
15. How often do you use the vocabulary task "Definitions"?
16. How often do you use the vocabulary task "Gap filling with multiple choice"?
17. How often do you use the vocabulary task "Gap filling"?

The main reason for the inclusion of the listed questions in the questionnaire was to connect and highlight various theoretical parts of the thesis and consequently to gain data on the chosen topics from the English teachers' practices. It is essential to mention that the way the questions were intentionally phrased in a way that could be understood without the knowledge of any literature on the topic of testing, and their amount was selected in order to keep the questionnaire short but simultaneously still able to give the data needed for the analysis to be productive.

As the test creation process is complicated, all of the chosen questions are intertwined, but each one of them focuses on a different aspect that has been described in the theoretical part. The following paragraphs will explain the connections between the questions and the theoretical part of the thesis while giving the thought process behind their inclusion.

The questionnaire starts with a question regarding the method the teachers use while creating their tests, which is the focus of the third chapter in the theoretical part; although the question sounds quite simple, it is used to gain general insight into the steps the teachers take when creating their test which can help in learning the trends that the teachers follow in the current climate of classroom testing. The results will be analysed in contrast to the test specifications to see whether the teachers follow the steps recommended by the experts in the field.

The question that follows is closely tied to the first one and concerns the time it takes for the teachers to create their tests. This question, in terms of the structure of the bachelor's thesis, covers certain parts of the practicality of the test, at least from the point of the time perspective. By analysing the amount of time, a teacher spends creating a test, it is possible to judge whether the test is practical in terms of time and reveal some more data on the teacher's test creation habits.

The third question is focused on the goal the teachers have in mind when creating a test. The answers given to this question can be analysed in contrast to the understanding of what a test is for as a tool, as established in the first chapter.

The world is ever-changing, and new technologies keep rising and influencing the daily lives of many. For this reason, the decision was made to include two questions regarding the usage of AI in the language test creation process. As of right now, there is no literature on the usage of AI in the field of language test creation, which means it is not included in the theoretical part of the thesis, but this question can provide entrancing data on flexibility and the ability to adapt to the newest trends.

The question regarding the usage of the same test or two different tests for a class divided into two groups provides the thesis with data on the validity of the test because it is fair to assume that groups usually differ in their language abilities. As stated in the first chapter, the test may suffer in terms of reliability and validity if the test items are too difficult for the students. The question involves the reason for the teachers' decision, which can furthermore highlight if their decision is influenced by the pursuit of reliability and validity or if there are other reasons why they find their choice beneficial.

There are many types of exercise a teacher can choose to include in their language tests, but at the same time, it should be done with the knowledge of the benefits and drawbacks of each one to ensure that the test results are reliable and the content validity of the test does not suffer. The question which teachers have to answer if they prefer pre-defined answers or having the students come up with their own answers is in the questionnaire in order to acquire an understanding of their experience with these types of tasks and, at the same time, analyse the reliance of their chosen type through their explanation for their usage.

Another critical aspect of a test is the point distribution and its conversion to the grade. This element plays a significant role in the student's understanding of their performance and directly influences their own performance. For this reason, it was included in the questionnaire to analyse the point conversion charts employed by the various English teachers in order to gain an understanding of their point allocation practices.

The last open question in the questionnaire concerned the problems the language teachers face when creating their tests. A test creation process is a complicated one, and there are many moving parts; for this reason, it was decided that including this question may reveal compelling information about the problematic parts of this process from the viewpoint of the language teachers, as well as obtaining data for the analysis.

The following section of closed questions was included to reveal whether or not the teachers use additional external sources in the creation of their tests. While the first question provides specific steps in their step-creation process, it does not specify if they are acquainted with the literature in the field of language testing. As mentioned previously, one of the external sources mentioned is also the use of AI, but in this regard, it is used to simply know if the information is used or not, instead of the specific details on its usage for the test creation.

The thesis is focused on grammar and vocabulary tests; therefore, the last part of the questionnaire is divided into two sections, each focused on different types of tasks. The first

section is focused on the usage of grammar tests; meanwhile, the second section revolves around the usage of vocabulary tasks.

Now that the questions and the reasons for their selection have been explained, it is possible to analyse the results of the questionnaire answers and their implications. Though, before continuing an important aspect to mention is that the questionnaire was filled out by sixteen teachers; therefore, that is the data pool the analysis is working with.

4.1. Question 1 - Can you describe your method of creating a unit test in bullet points?

Based on the answers to the questions, the results can be divided into two categories, which are visible in the pie chart below. The majority, which consists of 56% of answers, create a unit test by adjusting already existing tests, while only 44% of teachers create a new test from grounds up to test their students.

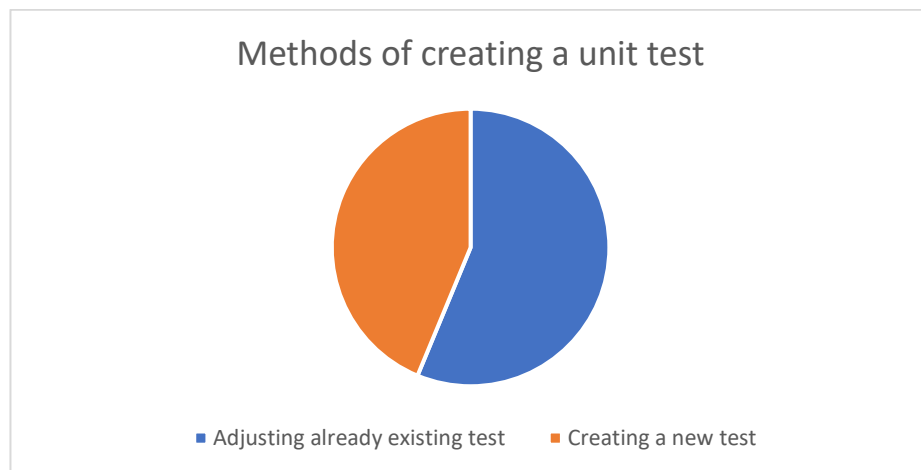


Figure 1 Pie Chart – Methods of creating a unit test

It has been established that before it is possible to start writing a test, the test maker must create test specifications, which he uses as a blueprint during the writing process, which consists of an outline of the test, skills tested and items and tasks. Once the test specifications have been established, the test maker should start creating their test, which consists of selecting appropriate grammar structures and vocabulary, then item writing with emphasis on its construction and creating a scoring key in order to be able to award points for the correct responses to the tasks.

None of the answers to the questions above mentioned writing up test specifications, though they are certainly required for standardised testing, their usage for classroom testing is encouraged by authors such as Henry Douglas Brown or Arthur Hughes. However, there is an

argument to be made for both groups of respondents that even while their answers did not mention the usage of test specifications explicitly, a large majority of respondents who stated they adjusted the already existing tests did so to adjust the contents to the difficulty adequate to the level of the students. Similarly, the teachers create a brand new test for their students to do with the usage of the student's book, which functions as the content for the test and is used for their sampling process when writing the tests.

Therefore, none of the teachers explicitly stated that they wrote plans for the test; however, it can be assumed that they used students' books or other relevant materials used in their lessons from which they selected the contents used for the test, which is later used for sampling of appropriate grammar and vocabulary. Then, they differ in the writing process; while 56% of the respondents simply adjust the contents of already existing tests, the remaining 44% need to create a new test, which involves additional selection and creation of test tasks with appropriately created scoring key.

4.2. Question 2 - How long, on average, does it take to create this type of test?

Surprisingly, the answers to this question were also quite similar, as represented in the pie chart below, where the vast majority of answers were either 30 minutes or between 30 and 60 minutes. The pie chart was used for better visualisation of the answers, as visible that 44% of the respondents stated 30 minutes, 44% stated a time between 30 and 60 minutes, and the last two groups were outliers, though they stated that their answers depended on the length of the test 6% stated time under 30 minutes if the test is short and the last 6% reported that if the test is more significant, it takes above 60 minutes.

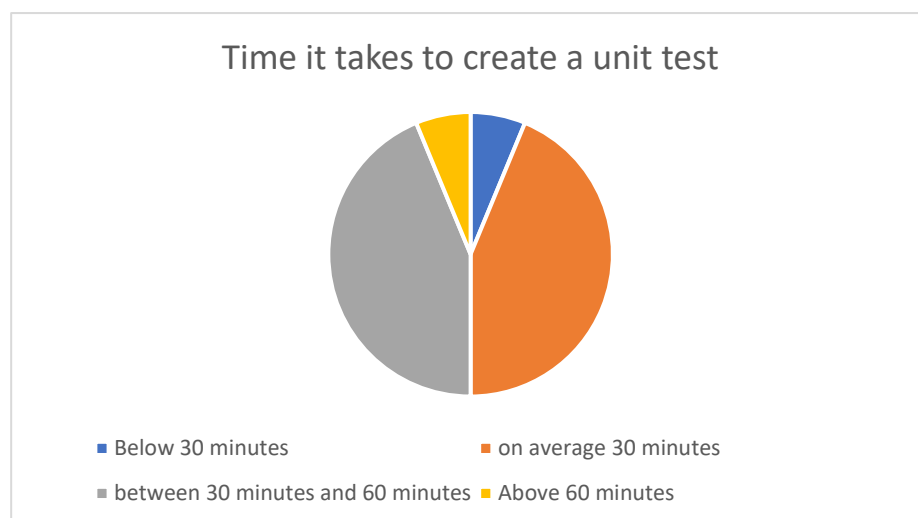


Figure 2 Pie Chart - Time it takes teachers to create a unit test

As stated previously, the information regarding the time it takes to create a unit test can be associated with its practicality. In the sub-chapter on practicality, it was mentioned that practicality can be understood as the number of resources one has to invest in the test development, administration, and evaluation in contrast to the available resources. Moreover, Bachman and Palmer (1996, 36-37) divide resources into three categories: human resources (test writer, scorer, administrators – in the case of the classroom tests, it is the same person), material resources (space, equipment, materials) and time (development time). It is essential to state that in the case of this question, the focus lies only in the development time of the test as that was the information obtained from this question, although it is possible to assume that the other categories of resources are also available since the human resources are the teachers that create, score, and administer the test and material resources such as space, equipment and materials needed for the test should be obtainable from the school.

By following the established understanding of practicality, assuming that the other two categories are available, it is possible to state that all the unit tests the teachers would create could be considered practical if the test creator (the teacher) had at least one hour for the test's development time. However, the group with the outliers could be considered impractical for most of the teachers who filled out this questionnaire. It is vital to point out that establishing practicality is a difficult task because of the data needed; moreover, the data would still be describing only certain times because the resources that the test creator has at hand, such as time, keep changing, while it could be practical to create a test in one hour at one point in time it could be impractical at a different point in time.

4.3. Question 3 - With what goal in mind do you create your tests?

The answers to this question were divided into two main groups, though since this was an open question, some answers were longer and contained more keywords to analyse, while others were shorter and not so in-depth, which is not incorrect, but it should be kept in mind while analysing. To make the analysis of the results clear, the pie chart below will be used to show the main keywords that were used. The additional goals the teachers have in mind while creating a test will be mentioned in the paragraph below.

The main two goals, as represented in the pie chart below, were either "verify knowledge/ability", which was a goal of 63% of teachers, while the other group used "to test knowledge/ability", which would be the remaining 38% of teachers. The results were

unexpected in that all of the teachers used quite similar words to describe the goal in mind while creating a test.

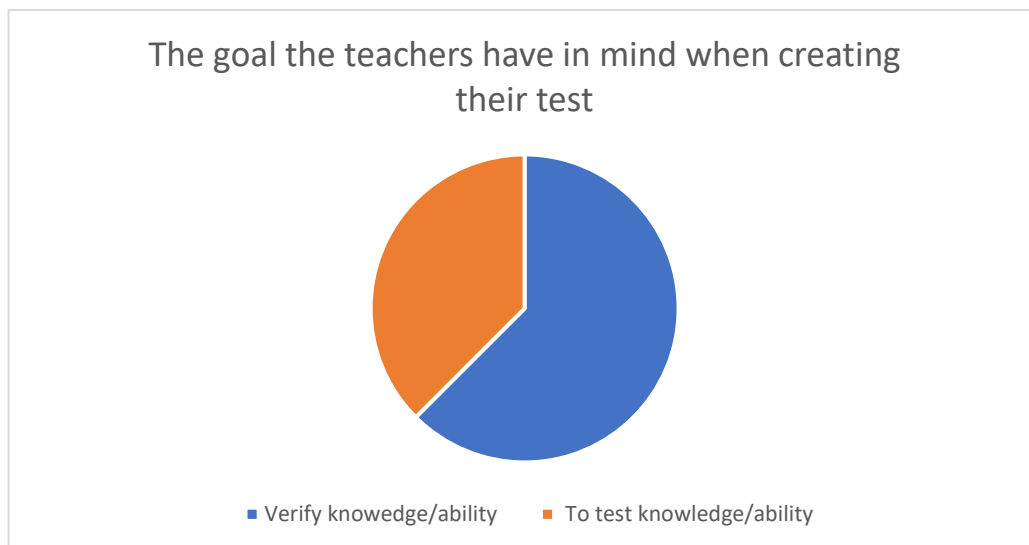


Figure 3 Pie Chart - The goal teachers have in mind when creating a test

Additionally, there were other goals mentioned in the longer answers, but they would be a minority, which is the reason for not including them in the main pie chart. Those goals would consist of the following: 1) to obtain grades for both parents and the students, 2) for students to have a chance to practice, and 3) to motivate the students.

In the first chapter, it was established that a test is a tool used by a test maker to measure the test taker's ability, knowledge, or performance in a given domain in order to make further decisions that involve the test taker. Therefore, by using the test as a tool, its goal would be to measure, as previously stated, the test taker's ability, knowledge, or performance, and in this case, the focus would be on English grammar and vocabulary. The goals that teachers stated will be used in contrast to the goal of a test to highlight possible similarities and differences.

The majority of answers used the word "verify",² which is defined by the Cambridge Dictionary as "to prove something exists or is true, or to make certain that something is correct", and the other word mentioned was "test",³ which defined as "to give someone a set of questions, in order to measure their knowledge or ability" (Cambridge University Press, n.d.). As can be seen from the definitions of these words, they both work with a concept of making judgements based on obtaining information; in the case of language testing, the judgements made are regarding the knowledge or ability, as the teachers and the definition

² The original word used in Czech was "ověřit"

³ The original word used in Czech was "otestovat"

stated. Therefore, it is possible to state that the goal of the test and the goal that the teacher has in mind when creating a test align.

The other goals mentioned, such as the need to obtain grades, students having a chance to practice, or as the way to motivate students, could be considered secondary goals of a test, but standalone, they would not be worth the time to develop, admission and score the test because as the established definition states that the in consequence of the test the teacher is making decisions concerning the students; as a result goals such as grades, practice or motivation cannot be considered as a primary goal of a test.

4.4. Question 4 - If you use the help of AI for the creation of your tests, how exactly do you use it and why?

The topic of AI usage in testing was approached with two questions. The first question focused on whether teachers use the help of AI or not, and the second question, as stated above, asked for more details regarding the usage. The answers to the first question were overwhelmingly negative, with 75% of respondents not using AI for the test creation; while these results do not paint AI usage in a positive light, it is still quite unexpected that the remaining 25% uses the help of AI due to the fact it such a new technology, which makes the results quite unexpected.

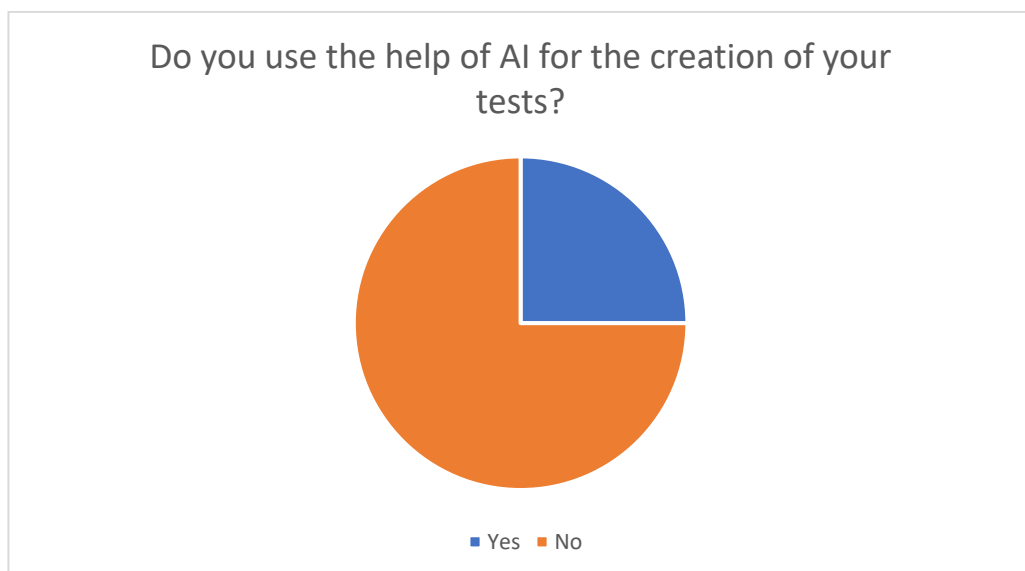


Figure 4 Pie Chart - The usage of AI by teachers

There are still certain parallels to be made between how teaching trends influence what is assessed and put to the test and the technology used in the creation process. As mentioned in the sub-chapter on assessing grammar and vocabulary, teaching trends at different points in

time and understanding what a particular part of language stands for significantly influences the assessment and, consequently, the requirements of the learner. Thus, it is possible that as teaching trends influence the assessment and content of the tests, the technology, in this case, the usage of AI, will have a dominant effect on the test creation in the future.

Now, to examine the ways in which AI is used for the test creation process by 25% of respondents. Responses were as follows: 1) to generate questions for reading comprehension tasks, 2) to generate questions for speaking practice, 3) to generate instructions for test items and finally, the last usage, which was vague and simply stated, 4) to make the test creation process faster. Some of these responses focused on different aspects of language, other than grammar and vocabulary, but an argument can be made that there is a variety of uses for AI, and if used correctly, the usage of AI can make the test creation process more effective.

Another fascinating insight could be gained by comparing the teachers' usage of AI and scholarly literature in the field of testing. Only three out of sixteen respondents stated that they use scholarly literature, and while both scholarly literature and AI usage seem to be used by the same amount of people, it is essential to state that while AI emerged only recently, the literature on the field of testing has been available for at least few decades. This further supports the point that in the future, AI will play a paramount role in making the process of test creation more efficient for teachers.

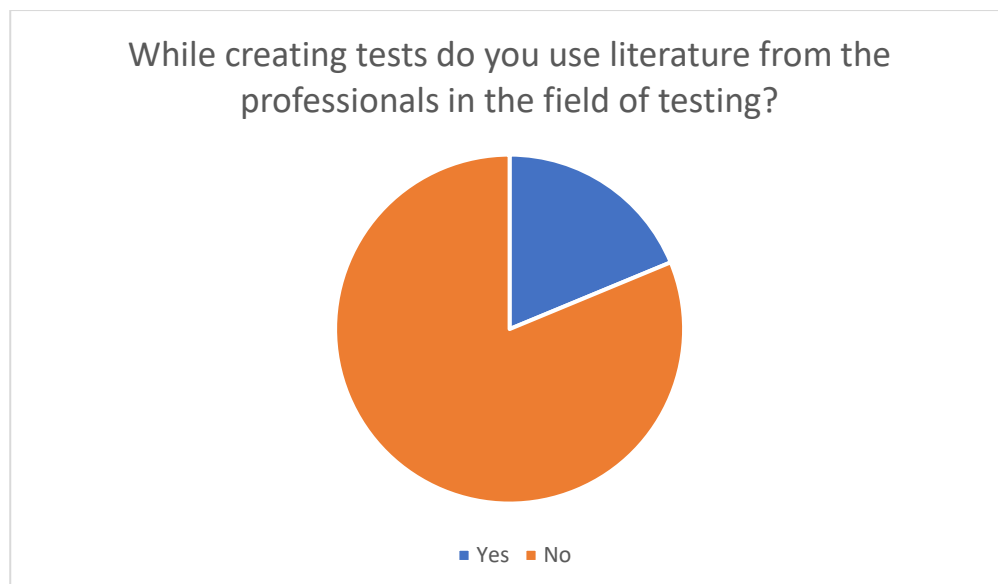


Figure 5 Pie Chart - The usage of scholarly literature from the field of testing

4.5. Question 5 - If you have your class divided into two groups, do you give both groups the same test, or do you give each group a different test, and why?

The results of this question can be divided into three groups, as shown in the pie chart below. The first group of respondents that use the same tests in every situation consists of 44% of the answers; 31% of respondents prefer to create different tests or, to the very least, alter them depending on the group and the minority, consisting of 25% of answers stated that they use different tests depending on the situation.

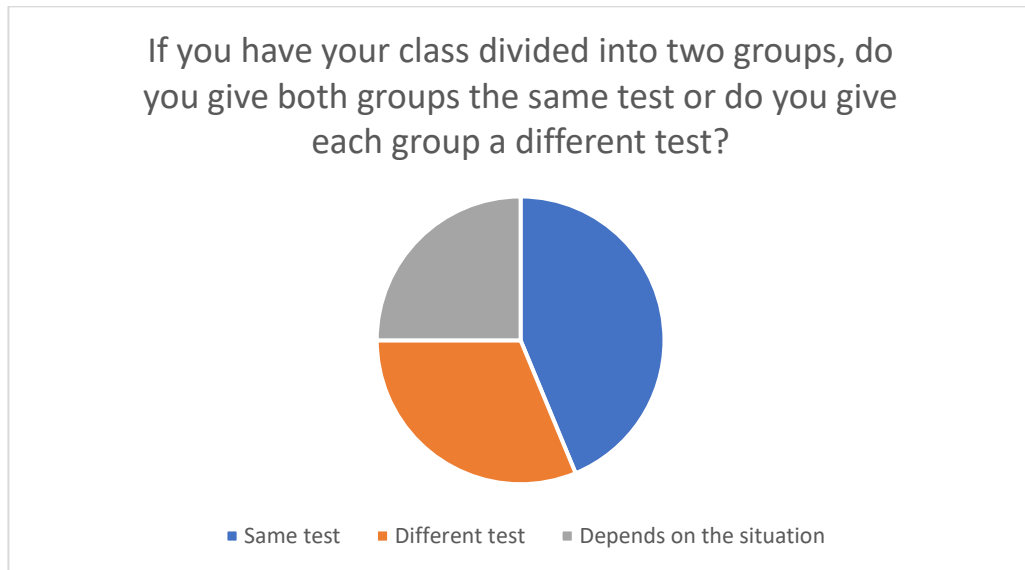


Figure 6 Pie Chart - The usage of the same or a different test for two groups

This question mainly focuses on the test's reliability and validity, which is influenced by the difficulty of the test. As stated in the sub-chapter on reliability, Douglas (2014, 10) also states that the test itself can cause inconsistencies because of the design of the tasks in terms of instructions, which could be unclear, difficulty which may exceed the targeted level of the student and quality because tasks may have multiple correct answers. Furthermore, in regard to the validity, sometimes the test may look valid in terms of content but measure something different than intended, for example, the ability to guess or read leave due to the test items being too difficult or poorly constructed; content validity does not guarantee that a test will be good, but it is a minimum requirement for it to be useful (Kubiszyn and Borich 2003, 300).

Before commenting on the results of this question, it is crucial to list the reasons that respondents mentioned when supporting their answers. Those who create different tests do so mainly out of concerns for cheating and varying levels of the groups. Moreover, certain teachers create different tests only under certain conditions; for example, if the tests are on different days, the teacher creates two different tests to counter potential cheating, the time

needed to create the tests, if the teacher does not have the time, they will create only one test, and if one group is more advanced than the other the teacher creates two tests of varying difficulties. The teachers that only use the same tests for both groups stated that they do so to guarantee the same difficulty for both groups, although one of the respondents using the same tests for both groups mentioned that in the future, they would like to create two different tests due to students cheating.

Therefore, the teachers who use the same test for both groups may see their tests suffering in terms of reliability and validity if the students are not on the same level in terms of English proficiency. Furthermore, another factor contributing to the deterioration of the mentioned test qualities would be sharing the test's content and answers between the students. The remaining two groups of teachers that create a majority in this case (56%) do not face the problems that come with using a single test for both groups, which enhances the validity and reliability of their tests, though this comes at the expense of their time which they need to invest into a test creation process.

4.6. Question 6 - If a test uses a point system, what scale do you use in order to convert the points to a grade?

The answers to this question varied greatly; therefore, for better visualisation, the answers were divided into two groups based on their similarities or differences with each other, which is shown in the pie chart below. The first group of answers to this question were similar; 31% of respondents stated that they convert the points of the test to percentages, and for each 10%, the grade worsens. The second group consists of the majority of answers; 69% of the respondents use unique points to grade conversation methods.

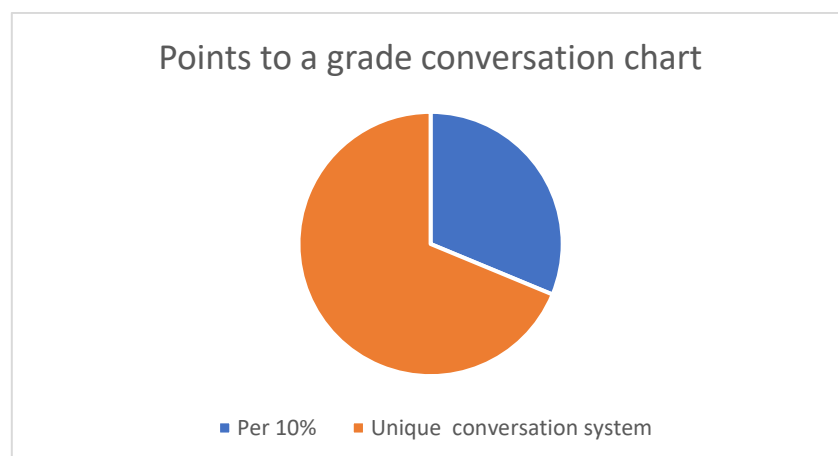


Figure 7 Pie Chart - Points to grade conversation

The difference between the points-to-grade conversation charts highlights unique approaches to grading by teachers and schools. In the table below, the unique conversation systems will be shown so that further comments can be made.

	3x	3x	2x	1x	1x	1x
1	100-90%	100-90%	100-90%	100-90%	100-95%	100-88%
2	89-75%	89-75%	89-75%	89-70%	94-85%	87-75%
3	74-60%	74-65%	74-60%	69-50%	84-70%	74-62%
4	59-45%	64-55%	59-50%	49-20%	69-55%	61-50%

Figure 8 Table - Unique points to grade conversation systems

As can be seen in the table and pie chart above, for the majority of teachers, the threshold to grade a test with 1 is 90%, with deviations being 2-5%, demonstrated by some teachers setting the threshold at 95% and 88%. The most common threshold for grade 2 would be 75%, with deviations being 10-5%, therefore 70%, and the outlier in this situation is 85%. The threshold for grade 3 varies considerably; most used would be 70% and 60%, respectively, with an outlier ranging down to 50%. Lastly, the threshold set for a grade 4 would be 60% for the majority, but there are quite a few unique conversation scales that deviate downwards; most of these thresholds would deviate 10-15%, with one outlier being 20%, differing vastly from the rest.

The questionnaire was anonymous, but it can be stated that the school specialisation can influence the deviations in these conversion scales since it is possible to see certain patterns that may be influenced by the questionnaire being filled out by multiple teachers at the same school. As could have been observed from the conversion scales that teachers use, the differences become increasingly pronounced as the percentage decreases; for example, grade 1 deviates in 2-5%, grade 2 deviates mainly⁴ in 5%, grade 3 deviates mainly in 10% and grade 4 deviates mainly in 15%.

4.7. Question 7 - What challenges do you face when creating your tests?

The answers to this question were analysed and categorised into the following groups. The most mentioned challenge that teachers face in their test creation process is adjusting the difficulty for the students; this was the answer of 44% of the respondents. Another challenge mentioned in 25% of the answers is the time management for both the test creation and the

⁴ Not accounting for the outlier.

consequent correction of the tests. The last two challenges mentioned were tied to creating test tasks, which make up 19% of answers and creating instructions for the test tasks, which is the remaining 13% of answers.

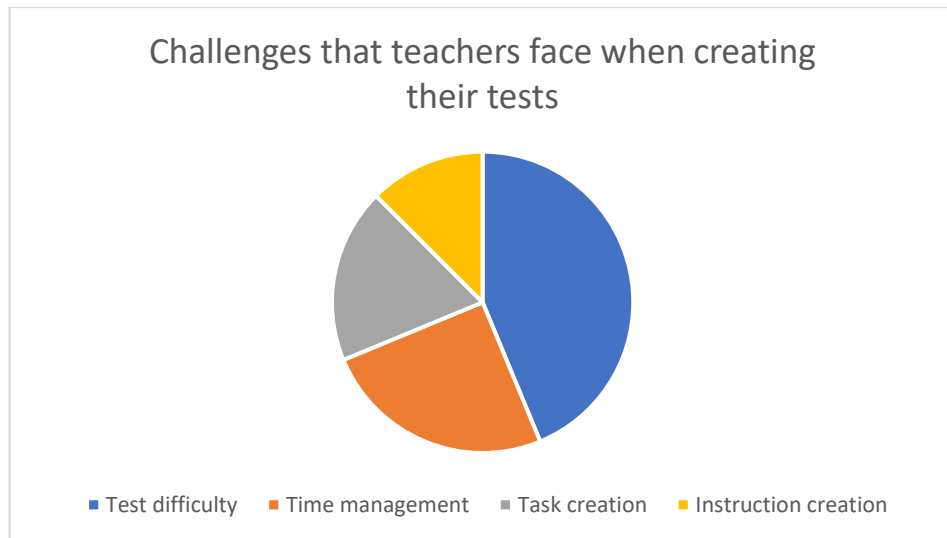


Figure 9 Pie Chart - Challenges that teachers face when creating their tests

The qualities of the tests can be influenced by the challenges that arise during the test creation process. The challenge regarding the difficulty of the test that teachers are facing is closely related to the reliability and validity of the test; therefore, it is important that the difficulty is adequate to the level of the students. The next major challenge is time management, which ties into certain aspects of the practicality of the test. Tasks can influence both reliability and authenticity, depending on the problem. Lastly, the instruction creation challenge, if the instructions are unclear, it may further resolve the test suffering in its reliability.

The point of this question was not to comment on whether or not the teachers overcome these challenges correctly but more so to highlight possible problems the challenges may cause if not resolved correctly. Therefore, by contrasting these challenges that teachers face and the qualities of the tests, it is possible to state that if these challenges are not approached correctly, they can result in the deterioration of mainly reliability and validity. Secondary qualities that can be harmed are the practicality and authenticity of the test.

4.8. Question 8 - How often do you use these types of grammar and vocabulary tasks?

The final question being analysed in this section concerns the usage of various grammar and vocabulary tasks and whether the teachers prefer the usage of limited production tasks or selected response tasks. The charts regarding the usage of the selected tasks will not be included in this section because of their amount; therefore, if visualisation is needed, they can be inspected in the appendix A to C.

Firstly, the usage of the selected grammar tasks from the answers is evident in deducing that certain task types are more popular than others. But before delving into the results, it is essential to state that all of the selected grammar types are limited production tasks, though they slightly differ in the amount of production they involve, which may influence their usage. The most popular task out of the three was modified cloze with seven teachers stating they always include this task type and five usually including this task type. This was followed by the task type paraphrase; ten teachers stated that they usually include this task type, while the remaining three respondents stated that they do not usually use it and the rest never use it. The last task type, which was the least used and the most divided, was completion, which six teachers never use, four teachers do not really use, three teachers usually use, and the remaining three teachers always use. Thus, the most popular task types for testing grammar are modified cloze closely followed by paraphrase, while the usage of completion task type is largely divided based on the data obtained, but it can be stated that it is not nearly as used as the previously stated task types.

Secondly, the usage of the selected vocabulary tasks, similar to grammar tasks, certain vocabulary tasks are much preferred than the others. While the selected grammar tasks consisted of only limited production tasks, three out of four selected vocabulary tasks were selected response task format, with only one being a variation of the limited response task. The task that is by far the most used by the teachers for testing vocabulary is gap filling, designed as a selected response task; nine teachers stated they always include it in their tests, six of them usually include it in their tests, and only one teacher stating they never use this type of task. The second most used task was again gap filling but, in this instance, designed as a limited production task; though still prevalent, the usage significantly decreased now only three teachers stated that they always use this task, eight teachers stated that they usually use this task and the remaining five teachers either do not usually use it or never use it. Following this is the task type synonyms, which is almost evenly divided in usage among the

respondent's usage; seven teachers stated that they usually use it, seven teachers stated that they do not usually use it, and the remaining two teachers never use it. The least used task type for testing vocabulary is definitions; eight teachers stated that they never use it, two teachers do not usually use it, three teachers usually use it, and the remaining three teachers always use it. Therefore, testing vocabulary is mainly dominated by gap-filling tasks, with the selected response being the more popular than limited production, followed by the task synonyms and the least used being definitions.

Before proceeding further, it is imperative to state that all of the mentioned task types are a viable way of testing both grammar and vocabulary; moreover, all of the task types mentioned except for task type synonyms can be transferred either to limited production or selected response tasks. Furthermore, authors such as Nathan Carr, Henry Douglas Brown and Arthur Hughes state that the main advantage of limited production tasks over selected response is the construction part, which is easier, but the main drawbacks are scoring, which is more time-consuming and when the task is not constructed correctly there may be more correct answers, meanwhile selected response tasks may be easier and more efficient when it comes to scoring, but its construction is complicated due to the needed selection of appropriate distractors.

The second question that will be discussed in this section is whether the teachers prefer using selected responses or limited production tasks. The answers are divided into three groups, as shown in the pie chart below. The majority of teachers, which consists of 44% of respondents, prefer to use selected response tasks, 38% of respondents prefer to use a combination of both in their tests, and only 19% of respondents prefer using selected response tasks.

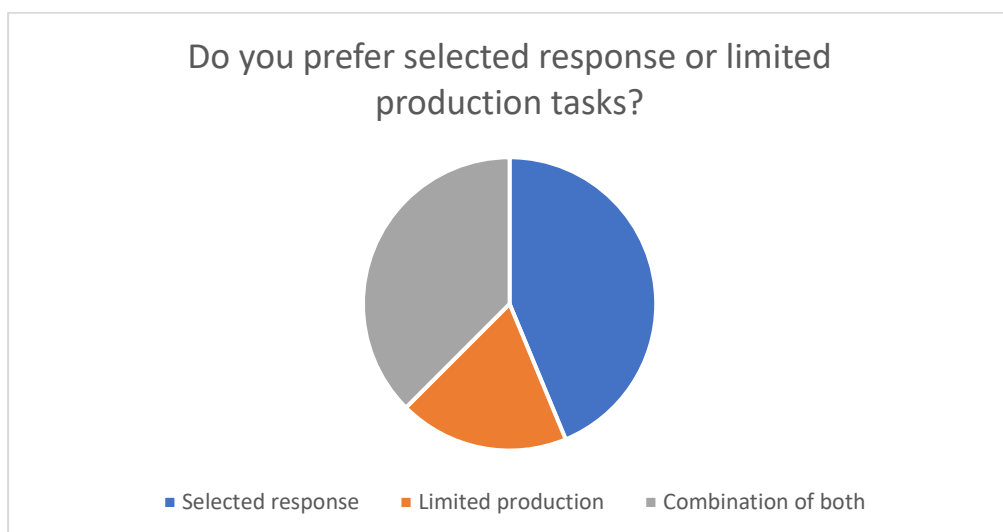


Figure 10 Pie Chart - Task type preferences

Furthermore, it is essential to mention why the teachers prefer to use their preferred task type. The majority of respondents prefer the usage of the selected response task because they believe it is a more efficient way of testing in terms of grading and administration; the other reason for the selection of this task format was that it was easier for students. Respondents who prefer to use limited production tasks do so because, according to them, it is easier to verify whether the students truly learned the material being tested. The last group of respondents who prefer to use the combination of both stated that they choose either limited production or selected response based on what exactly is being tested; based on their perspective, some tasks may be better suited for testing one area of language while not the other, the secondary reason stated was to balance out the difficulty, specifying that they believe selected response tasks are easier while the limited response is more demanding.

In summary, out of the selected task types, the most popular ones for testing grammar are modified cloze and paraphrase, while vocabulary is mainly tested by gap-filling tasks. The reasons that were used to support the teachers' preference in the usage of limited production or selected response tasks were mostly the same reasons as stated in the literature on the topic; selected response being efficient in terms of scoring and grading and limited response verifying the targeted language area to higher degree. On the other hand, some answers were concerning because they were using weaknesses of the tasks as the reasons for their usage, most notably stating that the reason for using limited response tasks is their easier difficulty and, therefore, the higher success rate for the students.

4.8.1. Analysis of the Acquired Tests

The initial goal of this part was to evaluate tests in terms of their qualities, those being validity, reliability, authenticity and practicality, but the amount of data acquired would not allow for such evaluation, and therefore, the analysis will concern the tests' design in terms of the task types used. The analysis will, therefore, be classified under the grammar and vocabulary task usage segment of questions; while in the beginning, the focus was on analysing the answers of the respondents to their method of creating grammar and vocabulary tests, in this part it is also possible to see whether the answers of the respondents correspond to their tests.

All of the acquired tests were analysed based on the response modes used, those being either constructed responses or selected responses. Moreover, this was done while paying attention to whether the tasks were grammar or vocabulary focused for a more in-depth view of the

matter. The data for this analysis consists of seventeen tests acquired from teachers, and all of the analysed tests will be visible in the appendix of the thesis.

The results of the analysis differ from the responses in the questionnaire, though this could be due to the limited amount of data acquired due to not every respondent providing a test. While the majority of the teachers preferred the usage of selected response tasks in the questionnaire, the tests yielded a different answer. Most of the tasks were using a limited production response mode, both for vocabulary and grammar. In terms of grammar tasks, the disparity between the selected response and the constructed response would be more significant; on average, if the test were to include four tasks, only one would be the selected response. While for vocabulary tasks, the usage of the selected response task format is more frequent than in the case of grammar tasks; it is still lower than expected; for every two vocabulary tasks, there would be only one selected response task type. There is one aspect of the test creation that should be mentioned here, the tests that were acquired and analysed from the same teacher in most cases followed the same pattern, which heavily influences this analysis but at the same time gives further insight into the degree the test creation is affected by the pattern the teacher gets used to.

Conclusion

The aim of this thesis was to research the way teachers in upper-secondary education create their tests, with the focus being on grammar and vocabulary. The thesis consists of two main parts: the theoretical and practical. The theoretical part aimed to research the aspects that influence the usefulness of the test while also covering the creation process of the test together with the design choices and their possible impact on the test. This was accomplished by covering the two main topics: principles applied to the test and planning and designing a language test.

The practical part used two methods of obtaining data: firstly, an online questionnaire was used to gain insight into how teachers in upper-secondary education create their language tests through a set of questions which were derived from the topics of the theoretical part; secondly, a test analysis focusing on the selection of the response modes to determine whether the results of the questionnaire are reflected in the created tests. Thus, the aim of the practical part was to research how teachers create their tests with the aim they have in mind. The insights into the test creation were met, though there were certain limitations due to the amount of data acquired.

The findings of the thesis consist of the following: teachers' test creation processes differ, with the majority of teachers adjusting already existing tests rather than creating brand new ones; the test creation process takes on average, between 30 and 60 minutes. For all of the teachers, the aim behind administering the test is to verify the ability and knowledge of the students, with the secondary goal being the need to obtain grades and practice for the students. Tools such as scholarly literature on field testing and AI are used only by a minority of teachers. The grading scales of the tests vastly differ between schools and teachers, with the differences between them widening the worse the grade gets.

The theoretical part covered all the essential information regarding the qualities of the test, the creation process, and the design choices for the grammar and vocabulary tasks. The practical part was intertwined with these topics through the construction of the questions and their consequent analysis. Thus, it is possible to state that the aim of the thesis was successfully achieved.

Resumé

Tato práce se zabývá problematikou vytváření testů se zaměřením na gramatiku a slovní zásobu. Skládá se z úvodu, teoretické části, praktické části a závěru. Teoretická část slouží jako základ pro praktickou část, vysvětluje důležitost testování gramatiky a slovní zásoby, zabývá se principy, které by měl test splňovat, dále se zabývá typy testů, procesem vytváření testu, typy cvičení, bodování a konečného známkování. Praktická část pomocí poznatků z teoretické části analyzuje postupy vytváření testů učiteli středních škol, které byly získány za pomoci použití online dotazníku. Součástí praktické části je taktéž analýza testů, která zkoumá, zda informace získané z dotazníku skutečně korespondují s tím, jak učitelé testy vytváří, avšak tato analýza se pouze soustředí na typy cvičení.

První kapitola teoretické části se zaměřuje na základní otázky spojené s testováním slovní zásoby a gramatiky. Nejprve se zabývá testem, který je definován jako nástroj sloužící k ověření znalostí a schopností testované osoby v dané oblasti, s cílem o dané osobě udělat rozhodnutí korespondující s typem testu. Dále je důležité zmínit, že testování gramatiky i slovní zásoby má svůj specifický přístup, v rámci gramatiky si musí učitel stanovit, zda chce testovat vědomé či nevědomé vědomosti, zatímco problematika testování slovní zásoby se zabývá otázkou spojenou s porozuměním slova; tyto faktory ovlivňují obsah a strukturu testu z pohledu použitých cvičení.

Druhá kapitola zkoumá principy, které by test měl splňovat, aby mohl být považován za užitečný; mezi tyto principy patří: reliabilita, validita, autenticita a praktičnost. Učitel by měl rozumět těmto principům, aby mohl vytvářet co nejkvalitnější testy a tím pádem získat co nejpřesnější informace o znalostech a schopnostech svých žáků. Reliabilita se dá definovat jako konzistence v testovacích výsledcích, která může být ovlivňována nesrovnalostmi ze strany žáka, učitele nebo testu; z důvodu těchto nesrovnalostí se žák může oddálit od svého skutečného výsledku, který by získal nebýt těchto nesrovnalostí. Validita zkoumá, do jaké míry test měří oblast jazyka, kterou měl učitel v plánu ověřovat. V rámci validity je důležité zmínit, že argument pro to, zda je test validní je složen z několika důkazů validit. Posledním kladeným principem na test je v bakalářské práci autentičnost; v rámci autentičnosti se učitel snaží o to, aby jazyk použit ve cvičeních v testu byl co nejbliž použit v přirozených situacích.

Třetí kapitola rozebírá hlediska, která jsou potřebná zvážit při plánování a následné konstrukci testu. Zaprvé jsou představeny typy testů, se kterými se učitel přichází do styku, hlavně tedy achievement a progress; achievement testy jsou použity na konci lekce, nebo

učebního období, jedná se o testy, které jsou sumativní, tedy jejich cílem je zjistit, zda žák ovládá látku probranou za dané období. Zatímco progress testy jsou formativní, jejich cílem je zjistit slabá místa žáků, aby se na něj v budoucnu učitel mohl zaměřit ve výuce. Typem testu se tedy rozumí rozhodnutí, které učitel dělá o svých žácích na základě jejich použití, což znamená, že v určitých případech může být použit jako achievement i progress test zároveň.

Dalším bodem, který popsán v této části jsou specifikace testu, které fungují jako plán použití při jeho následovné konstrukci. Obsah specifikací testu se liší na základě, zda jsou použity v rámci školní třídy nebo pro standardizované testování, avšak zatímco pro standardizované testování je jejich použití nutnost, v kontextu školní třídy je jejich použití jen doporučením. Pro potřeby učitele tyto specifikace obsahují obsah testu z pohledu testované látky, časový limit testu, a cvičení, které použije. Testovaná látka tedy v kontextu gramatiky a slovní zásoba znamená všechny relevantní gramatické struktury a slovní zásobu. Z pohledu cvičení se učitel musí rozhodnout jaký vybere typ cvičení z pohledu typu odpovědi, může se jednat o typy cvičení u nichž žák vybírá ze předem stanovených odpovědí nebo typy cvičení založené na tvorbě, která u gramatiky může být od jednoho slova až po kratší větu. Oba typy těchto cvičení mají své výhody a nevýhody, které je důležité brát v potaz při jejich tvorbě. Cvičení zaměřených na výběr z mnoha možností jsou složitá na tvorbu z důvodu výběru adekvátních nesprávných odpovědí, ale výhoda tohoto typu cvičení spočívá v rychlé opravě, zatímco typ cvičení, kde žák vytváří odpověď mají opačný problém, tedy nejsou tak těžké na konstrukci, ale jejich proces opravy je náročnější na učitele. Ve chvíli, co učitel vybere typ a obsah cvičení, který chce použít, tak už jen stačí vytvořit klíč správných odpovědí a bodové ohodnocení testu.

Problematika bodování a hodnocení je posledním bodem teoretické části, ale není rozvedena do hloubky jako ostatní části, protože není integrálním bodem procesu tvorby testu, tak jako předem zmíněné specifikace. Hlavní myšlenky této části spočívají v tom, že body za chyby ve cvičeních by měly být strženy jen pokud je chyba v testovaném jevu, dále je důležité zmínit, že konverze bodů na známku je ovlivněna hlavně institucí a kulturou dané země.

Praktická část začíná s definováním cíle této bakalářské práce, tedy zjistit, jak učitelé na středních školách vytvářejí své testy se zaměřením na gramatiku a slovní zásobu. Tímto bylo dosaženo použití online dotazníku a následné analýzy vybraných testů od učitelů. Dále jsou představeny otázky této části, které zahrnují způsob vytvoření testů, cíl, s jakým učitelé vytváří své testy, čas, který potřebují na tvorbu testu, použití externích pomůcek pro výrobu

testu, preference typů cvičení, které používají na testování gramatiky a slovní zásoby, způsob, kterým konvertují body na známku a problémy, kterým čelí při tvorbě testů.

Odpověď na otázku teoretické části se dá odpovědět pomocí následujících zjištění. Hlavním rozdílem při tvorbě testů je, že většina učitelů místo vytváření zbrusu nových testů použijí předem vytvořené testy, které jen upraví dle svých potřeb. Proces tvorby testů učitelům zabere mezi 30 až 60 minuty. Hlavním cílem všech učitelů, kteří odpověděli na dotazník je ověřit znalosti a schopnosti svých žáků, zatímco sekundárním cílem je získání známek pro klasifikaci. Externí pomůcky jako odborná literatura na testování nebo AI je použito menšinou učitelů. Bodové hranice pro různé známky se výrazně liší pro různé školy a učitelé, čím horší je známka tím větší jsou rozdíly bodových hranic.

Teoretická část pokryla všechny důležité otázky vzhledem na principy aplikované na test, proces vytváření testu a preference v použití různých typů cvičení testující gramatiku a slovní zásobu. Teoretická část propojila všechny zmíněná témata pomocí otázek zaměřujících se na způsob, kterým učitelé vytvářejí testy a jejich následná analýza.

Bibliography

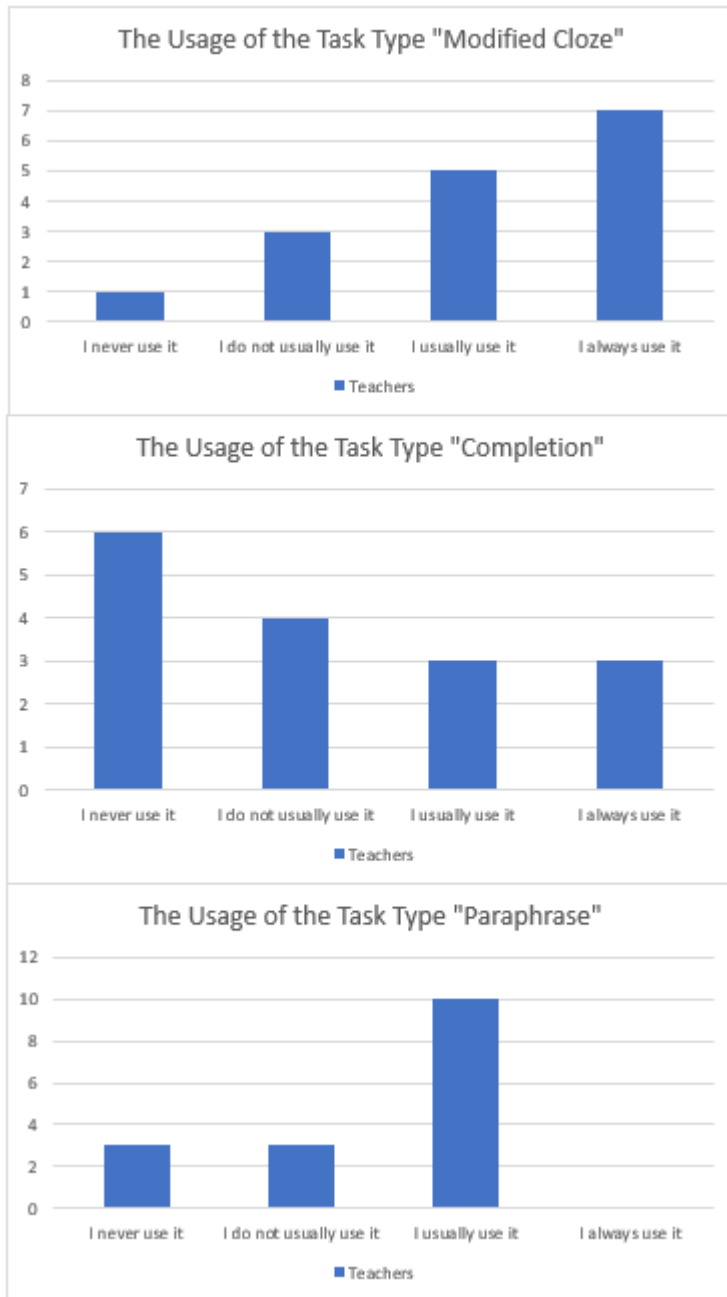
- Alderson, J. Charles, Caroline Clapham, and Dianne Wall. 1995. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Bachman, Lyle F., and Adrian S. Palmer. 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Brown, Henry Douglas. 2004. *Language Assessment: principles and classroom practices*. San Francisco: Longman.
- Cambridge University Press. n. d. "Test." Cambridge Dictionary. Accessed May 2, 2024. <https://dictionary.cambridge.org/dictionary/learner-english/test>.
- Cambridge University Press. n. d. "Verify." Cambridge Dictionary. Accessed May 2, 2024. <https://dictionary.cambridge.org/dictionary/learner-english/verify>.
- Cambridge University Press. n. d. "Vocabulary." Cambridge Dictionary. Accessed January 5, 2024. <https://dictionary.cambridge.org/dictionary/learner-english/vocabulary>.
- Carr, Nathan T. 2018. *Designing and Analyzing Language Tests*. Oxford: Oxford University Press.
- Davies, Alan. 2007. *An Introduction to Applied Linguistics*. 2nd ed. Edinburgh: Edinburgh University Press.
- Douglas, Dan. 2014. *Understanding Language Testing*. New York: Routledge.
- Ellis, Rod, Shawn Loewen, Catherine Elder, Rosemary Erlam, Jenefer Philp, and Hayo Reinders. 2009. *Implicit and Explicit Knowledge in Second Language Learning, and Testing and Teaching*. Bristol: Multilingual Matters.
- Ellis, Rod. 2014. "Grammar Teaching for Language Learning." *Babylonia* 2 (14): 10-15.
- Fulcher, Glenn, and Fred Davidson. 2007. *Language Testing and Assessment*. Milton Park: Routledge.
- Harris, Michael, and Paul McCann. 1994. *Assessment*. Oxford: McMillan.
- Hughes, Arthur. 1992. *Testing for Language Teachers*. 4th ed. Cambridge: Cambridge University Press.

- Kubiszyn, Tom, and Gary Borich. 2003. *Educational Testing and Measurement*. 7th ed. New Jersey: John Wiley & Sons Publishing.
- Lívía, Nagyné Foki. 2006. "From Theoretical to Pedagogical Grammar: Reinterpreting the Role of Grammar in English Language Teaching." PhD diss., University of Pannonia.
- Ma, Anne. 2008. *A Practical Guide to a Task-Based Curriculum*. Hong Kong: City University of Hong Kong Press.
- Paribakht, T. Sima, and Marjorie Bingham Wesche. 1993. "Reading Comprehension and Second Language Development in a Comprehension-Based ESL Program." *TESL Canada Journal* 11 (1): 9-29.
- Pearson, P. David, Elfrieda H. Hiebert and Michael L. Kamil. 2007. "Vocabulary assessment: What we know and what we need to learn." *Reading Research Quarterly* 42 (2): 282-296.
- Purpura, James. 2004. *Assessing Grammar*. Cambridge: Cambridge University Press.
- Read, John. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Stahl, Katherine A. Dougherty, and Marco A. Bravo. 2010. "Contemporary Classroom Vocabulary Assessment for Content Areas." *The Reading Teacher* 63 (7): 566-578.
- Wang, Yunai. 2008. "Authenticity in Language Testing." *Hawaii Pacific University TESOL Working Paper Series* 6 (2): 31-36.

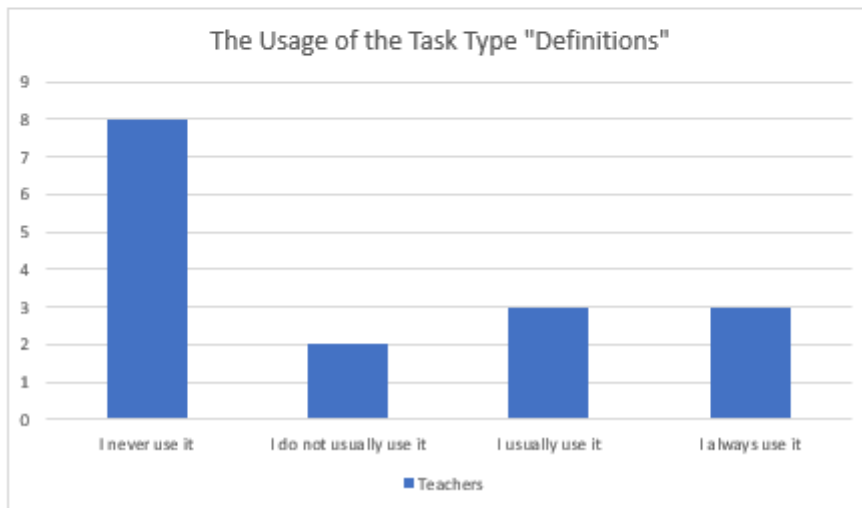
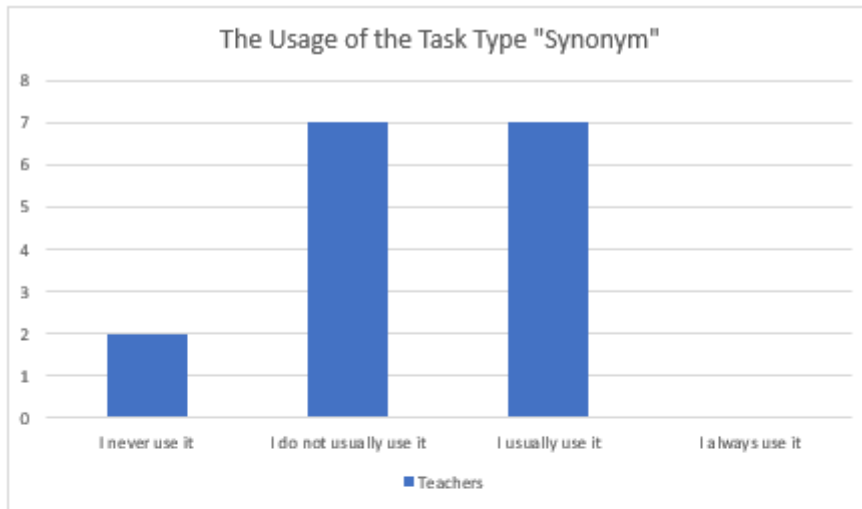
Appendix

Appendix A <i>The usage of various grammar tasks</i>	57
Appendix B <i>The usage of various vocabulary tasks</i>	58
Appendix C <i>The usage of various vocabulary tasks 2</i>	59
Appendix D <i>Test 1</i>	60
Appendix E <i>Test 2</i>	61
Appendix F <i>Test 3</i>	62
Appendix G <i>Test 4</i>	63
Appendix H <i>Test 5</i>	64
Appendix I <i>Test 6</i>	65
Appendix J <i>Test 7</i>	66
Appendix K <i>Test 8</i>	67
Appendix L <i>Test 9</i>	68
Appendix M <i>Test 10</i>	69
Appendix N <i>Test 11</i>	70
Appendix O <i>Test 12</i>	71
Appendix P <i>Test 13</i>	72
Appendix Q <i>Test 14</i>	73
Appendix R <i>Test 15</i>	74
Appendix S <i>Test 16</i>	75
Appendix T <i>Test 17</i>	76
Appendix U <i>Questionnaire – respondent’s side</i>	77

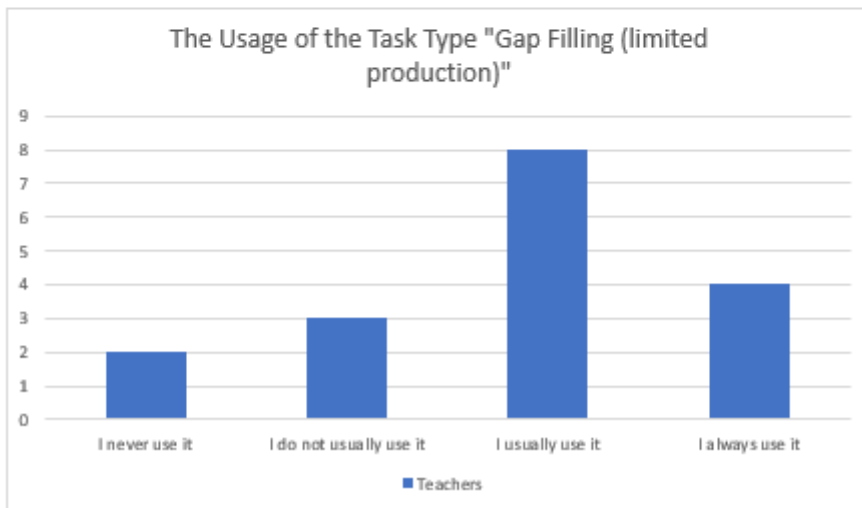
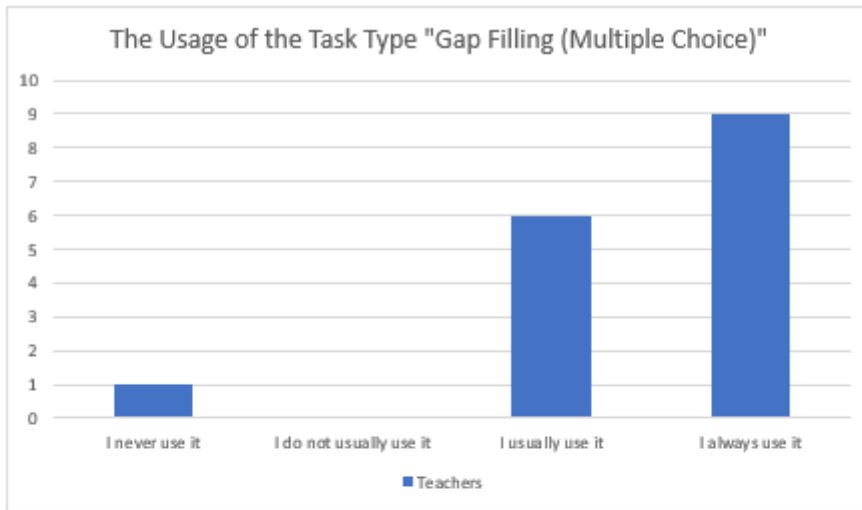
Appendix A *The usage of various grammar tasks*



Appendix B *The usage of various vocabulary tasks*



Appendix C *The usage of various vocabulary tasks 2*



Appendix D Test 1

Test B: Unit 4 Name [REDACTED] Class AL3 30 points MAX

1. Complete the sentences with the correct missing words. (5 points)
 - a) When Mr Watson retired, the school voted for a new head teacher to replace him. 1
 - b) She is very sensitive to other people's feelings.
 - c) Before you apply for a job, you should get some practice of answering questions.
 - d) A person who can speak two languages equally well is bilingual.
 - e) On the school website you can find the topics for the presentations. You can choose any of them.

2. Complete the sentences with the appropriate prepositions in the correct place (5 points)
 - a) Joe is preparing a maths exam now. 2
 - b) I looked for the house for my phone up around.
 - c) I wanted to learn French, but finally I ended studying Japanese. have up
 - d) I used to play a band called Crazy Mushrooms. in
 - e) It's based a true story. on 2

3. Complete the sentences with the correct form of the verb in brackets (past simple/continuous, past perfect simple/continuous or used to) (10 points)
 - a) I realized I had forgotten (forget) to turn off the heating at work when I got home.
 - b) There was used to be (be) a good pub in this house. But there's a bank there now.
 - c) Yesterday at 6 PM I was buying (buy) a new pair of shoes.
 - d) Jerry used to take (take) pictures in the old-fashioned way. But eventually, he's learned to use the digital instead.
 - e) I discovered there was nobody at the gym when I got there. The waterpipes in the showers had frozen (freeze) up overnight.
 - f) While Mark was going (go) home, he saw a terrible accident.
 - g) Did you keep (you, keep) a pet when you were little?
 - h) Between the years 2015 and 2016 I was attending (attend) an English course in Brisbane.
 - i) My mum found out that someone had stolen (steal) her purse. She was shocked.
 - j) Granny had had (have) a problem with her TV before we arrived.

4. Circle the correct verb in each sentence (10 points)
 - a) What books did you use to read / were you reading when you were little? 9
 - b) I had helped / was helping Granny all day on Saturday.
 - c) When Jane did / had done the shopping, she picked her children up from school.
 - d) Jim crashed into a car because he wasn't realising / hadn't realised the light was red.
 - e) Yesterday, in the first lesson, the teacher used to test / was testing our spelling.
 - f) When Jenny was at university, she met / used to meet her future husband.
 - g) Who was living / had lived in the flat before you bought it?
 - h) Jim had never tasted / has never tasted paella before he went to live in Spain.
 - i) Mum didn't use to travel / wasn't travelling as much as she does now.
 - j) You looked so deep in concentration. What were you thinking / had you thought about?

Appendix E Test 2

Test B: Unit 8 Name XXXXXXXXXX

Class 104 30 points MAX

1. Complete the sentences with the correct missing words. (5 points) 5

- When you travel out of Schengen you must have a passport.
- We've invited 60 guests to our wedding.
- When you travel to the UK you can use either a ferry or a train under the Channel.
- My parents got home four hours later because their plane had been delayed.
- How many visitors come to the castle a year?

2. Complete the sentences with the appropriate prepositions in the correct place. (5 points)

- We checked at 10 a.m. and then caught the train home. in 3
- My parents arrived at the airport just in time.
- Do you think Tom Cruise has ever been on a cruise? on
- I usually go jogging on the river bank. on along
- Fred dislikes travelling in coach. He doesn't have enough space for his long legs there. in BY

3. Complete the sentences with the correct form of the verb in brackets. (5 points) 4

- I wouldn't allow that behaviour if I was (be) his father. were
- What will you tell him if he chooses (choose) the most expensive suit of all?
- If I had listened (listen) more carefully at school, I wouldn't have had to study at home.
- If you heat ice it melts (melt).
- What would you do if you found (find) a wallet in the street?

4. Complete the second sentence in each pair so it has the same meaning. (5 points) 3

- I didn't order it online. I didn't get it in time.
If I had ordered it online, I would have got it in time. 4
- You didn't tell me. I didn't buy any milk.
If you had told me, I would have bought milk.
- You forgot to water the flowers. They died.
If you hadn't forgotten to water the flowers, they wouldn't have died.
- It started raining. We had to stop playing.
If it hadn't started raining, we wouldn't have stopped playing.
- They trained hard all year. They won the cup final.
If they hadn't trained hard all year, they wouldn't have won the cup final.

5. Circle the correct word in each sentence (10 points) 10

- These T-shirts are fine. Neither / Both of them look good.
- There's a train to Tábor every / each hour.
- The boys have prepared a quiz both / each.
- Forget about it. We have another / other problems.
- Neither / either team is good enough to win the competition. They're both terrible.
- Can I have another / other cup of tea, please?
- Where's another / the other sock? I only have one!
- I spoke to Tom and Ben. The other / others have not arrived.
- Which do you prefer, blue or red? I like both / each.
- I have not invited either / both of them.

Appendix F Test 3

Test A

File 7A 30 points MAX

Name

Class: DL2

1. Fill in the gaps with a word to complete the sentences (10 points)

- a) Boarding school is a type of school where students study, sleep and spend their free time. Boarding
- b) Watch your manners! If you misbehave in class like that again you might get expelled. And that will be the point of no return. 8
- c) A school subject where you learn about rivers, mountains and important cities is called gography.
- d) The learners at the secondary school are called students.
- e) The entrance exam for university was quite difficult. I hope I will pass it.
- f) A school term usually lasts twelve weeks. term
- g) We studied the French Revolution in our history class. That was breathtaking.
- h) You have to finish the primary school before applying for the secondary level.
- i) Phil feels anxious as he is afraid of failing the exam again. That would be a disaster. 10

2. Fill in the gaps. First conditional and future clauses. (10 points)

- a) After I graduate I will go travelling around the world. (graduate, go)
- b) If we don't book the tickets now, I am afraid there will be any more left later on. (not book, not be)
- c) I think he will be really surprised when he hears about it. (be, hear)
- d) Gregory will continue saving until he has enough money to buy a new car. (continue, have)
- e) If I am late for work, my boss will be angry with me. (be, be)

3. Complete the sentences with a conjunction (5 points)

- a) They will not invite us until they finish the reconstruction. 4
- b) Watch the road ahead carefully before you start overtaking.
- c) Shall I take an extra sandwich with me if we finish late? IN CASE
- d) I can't go unless you pay for my ticket. I am broke.
- e) I will look for a job in September after I come back from holiday.

4. Finish the sentences (5 points)

- a) You can travel anywhere you like as long as you can take me with you
- b) If you press the red button, you will win \$1000
- c) If I can get the tickets for the concert, you'll take you with me
- d) We will miss the train unless you speed up!
- e) The children won't go to sleep until you read them Bedtime story 3,5

Appendix G Test 4

FSA - Sports

Complete the test following the instructions given.

ODJHJ

Prepara te sporturi (Cămin și țigăni și din vîrta

1. meară /mei/ [jocet boad: 1]

Zădăre năi adăonec:

2. țigăni /jocet boad: 1]

Zădăre năi adăonec:

3. meară /jocet boad: 1]

Zădăre năi adăonec:

4. țigăni /jocet boad: 1]

Zădăre năi adăonec:

5. meară vîrta /jocet boad: 1]

Zădăre năi adăonec:

6. țigăni /jocet boad: 1]

Zădăre năi adăonec:

7. meară /jocet boad: 1]

Zădăre năi adăonec:

8. țigăni /jocet boad: 1]

Zădăre năi adăonec:

9. meară /jocet boad: 1]

Zădăre năi adăonec:

10. țigăni /jocet boad: 1]

Zădăre năi adăonec:

11. ampara /jocet boad: 1]

Zădăre năi adăonec:

12. țigăni /jocet boad: 1]

Zădăre năi adăonec:

13. meară /jocet boad: 1]

Zădăre năi adăonec:

14. țigăni /jocet boad: 1]

Zădăre năi adăonec:

15. meară /jocet boad: 1]

Zădăre năi adăonec:

16. țigăni /jocet boad: 1]

Zădăre năi adăonec:

17. meară vîrta /jocet boad: 1]

Zădăre năi adăonec:

18. țigăni /jocet boad: 1]

Zădăre năi adăonec:

19. meară /jocet boad: 1]

Zădăre năi adăonec:

20. țigăni /jocet boad: 1]

Zădăre năi adăonec:

Multiple choice

Choose the correct option.

21. ampara _____ [jocet boad: 1] court track path

Zădăre năi adăonec:

22. meară _____ [jocet boad: 1] court course court

Zădăre năi adăonec:

23. țigăni _____ [jocet boad: 1] court course path

Zădăre năi adăonec:

24. meară _____ [jocet boad: 1] court course court

Zădăre năi adăonec:

25. țigăni _____ [jocet boad: 1] path arena arena

Zădăre năi adăonec:

Irregular verbs

Write the past and past participle of the irregular verbs.

26. ampara

26. țigăni /jocet boad: 1]

Zădăre năi adăonec:

27. meară /jocet boad: 1]

Zădăre năi adăonec:

28. țigăni /jocet boad: 1]

Zădăre năi adăonec:

29. meară /jocet boad: 1]

Zădăre năi adăonec:

30. țigăni /jocet boad: 1]

Zădăre năi adăonec:

Translation - sentences

Translate the sentences from Czech into English.

31. Vyrval ji měkký děr! [jocet boad: 2]

Zădăre năi adăonec:

32. Jindy je tu! [jocet boad: 2]

Zădăre năi adăonec:

33. Jak se jmenuje tenhle sport? [jocet boad: 2]

Zădăre năi adăonec:

34. Když jí naposledy viděl? [jocet boad: 2]

Zădăre năi adăonec:

35. Jak často chodí káždě? [jocet boad: 2]

Zădăre năi adăonec:

Appendix I Test 6

OEF - U10 - Sport

Complete the sentences with the correct form.

Vocabulary - Translation

Translate the words from German into English and vice versa.

1. jährling hat die gleiche Bedeutung wie ...

Zielstreife nicht übernommen

2. Kern ...

Zielstreife nicht übernommen

3. vertikal ...

Zielstreife nicht übernommen

4. gefällig ...

Zielstreife nicht übernommen

5. jährling ...

Zielstreife nicht übernommen

6. vier ...

Zielstreife nicht übernommen

7. computer ...

Zielstreife nicht übernommen

8. Spieler ...

Zielstreife nicht übernommen

9. auf ...

Zielstreife nicht übernommen

10. hat ...

Zielstreife nicht übernommen

Vocabulary - Gaps

Complete the sentences with the correct word trying with the new given. Mark the whole word. Use the new words.

11. The footballer was sent off for arguing with the referee.

Zielstreife nicht übernommen

12. Usain Bolt is one of the most famous sprinters in the world.

Zielstreife nicht übernommen

13. For an important game you need to be in good form.

Zielstreife nicht übernommen

14. Peter Jackson won the gold medal at the Winter Olympics in 2018.

Zielstreife nicht übernommen

15. Realising that we were finishing the effort, she won the game.

Zielstreife nicht übernommen

16. Peter Jackson won the gold medal in Vancouver twice, in 2011 and 2014.

Zielstreife nicht übernommen

17. After the game all my friends were gone.

Zielstreife nicht übernommen

18. The winner received the cup and some money.

Zielstreife nicht übernommen

19. Great! We won in the end! Our team scored a goal in the last minute.

Zielstreife nicht übernommen

20. You have to take part in the test day if you want to be a player and be part of a team.

Zielstreife nicht übernommen

Multiple choice

Choose the correct word to complete the sentences.

21. I have _____ the number of times I've seen records from 15 to 20.

increased
 decreased
 stayed
 changed

22. That's great! These exercises will _____ your fitness over six months.

lower
 raise
 stay
 show

23. The goalkeeper tried to _____ the goal, but he had no chance.

save
 shoot
 miss
 show

24. The football player _____ the ball straight into the net.

scored
 caught
 missed
 shot

25. Sports programmes on TV are watched by millions of _____ people.

performers
 viewers
 fans

Verbs

Complete the sentences with the correct verb.

26. In basketball, you need to _____ the ball through the hoop.

Zielstreife nicht übernommen

27. In football, you need to _____ the ball into the goal.

Zielstreife nicht übernommen

28. In tennis, you have to _____ the ball over the net.

Zielstreife nicht übernommen

29. In the 800 metres race, you have to _____ round the track three times.

Zielstreife nicht übernommen

30. He wants to be stronger, that's why he _____ weight.

Zielstreife nicht übernommen

Translation - Sentences

Translate the sentences from German into English.

31. Das Gewicht meines Aufwärtlers ist groß.

Zielstreife nicht übernommen

32. Zwei Spieler haben sich gegenseitig einen Elfmeter verschossen.

Zielstreife nicht übernommen

33. Der Fußballer schoss ein Tor.

Zielstreife nicht übernommen

34. Der Fußballer fuhr auf dem Rasen.

Zielstreife nicht übernommen

35. Ich bin ein Fußballfan.

Zielstreife nicht übernommen

Appendix J Test 7

English file pre-intermediate 8

name: _____

1 Make first conditional sentences.

- 1 If Mark / ring / me / I / tell you _____
- 2 Yasuko / be / upset / if she / not pass / the exam _____
- 3 I / not be able / to go to work / if it / snow _____
- 4 Maggie / arrive / on time / if she / take / a taxi _____
- 5 They / not go / to the beach / if it / rain _____
- 6 If we / not take / a map / we / get / lost _____
- 7 If Pablo / miss / the meeting / his boss / be / angry _____

7

2 Circle the correct word.

- 1 Is this book **your** / **yours** or your brother's?
- 2 This isn't our camera; it's **their** / **theirs**.
- 3 Have you seen **her** / **hers** new computer?
- 4 Dom and Kate have got a dog. **Its** / **Their** name is Scooby.
- 5 That pen is **my** / **mine**! Give it to me!

5

3 Make statements and questions with *should* and the words in brackets.

- 1 _____ so fast. The roads are very icy. (you / not drive)
- 2 If he wants to learn Spanish, _____ in Spain. (he / study)
- 3 This computer is old. _____ a new one? (we / buy)
- 4 _____ to bed late. You've got an exam in the morning. (you / not go)
- 5 _____ nicer to each other. (people / be)
- 6 What _____ about my noisy neighbours? (I / do)

6

4 Circle the correct word.

- 1 Is Fiona going to **wear** / **carry** her new dress for the party?
- 2 I never **borrow** / **lend** money from my friends.
- 3 Nicole is a famous singer. She **wins** / **earns** a lot of money.
- 4 We **knew** / **met** Juan's wife for the first time yesterday.
- 5 I hate **waiting** / **hoping** for people when they're late.

5

5 Complete the sentences with the correct phrase. There are three extra phrases.

get a ticket get better get divorced get emails
get home get on well with get nervous get up
get presents get fit get married

- 1 I don't have any problems with my neighbours. I _____ them.
- 2 If they leave work now, they'll _____ by five o'clock.
- 3 Do I have to _____ before I get on the bus?
- 4 Mary and Joe have decided to _____. The wedding is in June.
- 5 If we _____ late, we'll miss the train.
- 6 I'm not very well at the moment, but I hope I'll _____ before my holiday.
- 7 People often _____ before interviews.
- 8 He's doing more exercise because he wants to _____.

8

Appendix K Test 8

GRAMMAR

1 Choose the correct form of the verb.

Example: I used to love / was loving meeting up with my cousins on holidays.

- 1 Where did you grow up / did you use to grow up?
- 2 We prepared / were preparing lunch when the phone started ringing.
- 3 Sandra used to live / lived in Paris for three years.
- 4 A Did your children use to like / Were your children liking secondary school?
B My daughter was fine, but my son used to hate / hated going at first.
- 5 How old were you when you were starting / started primary school?
- 6 I wasn't watching / didn't use to watch so much TV a few years ago.
- 7 A What were you doing / did you do this time last year?
B I was travelling / travelled round the US with friends from university.
- 8 They didn't have / weren't having so much money when we first met them.
- 9 I was meeting / met a lot of interesting people when I was a student.
- 10 What job were you wanting / did you want to have when you were younger?

	12
--	----

2 Complete the sentences with the correct preposition.

Example: My father is looking forward to his retirement.

- 1 They flew _____ the Alps on the way to Italy.
- 2 What are you waiting _____? Let's go now.
- 3 I opened my front door and my cat ran _____ me.
- 4 Their country is particularly famous _____ its food.
- 5 Do you know what this photo reminds me _____?
- 6 It was late when we finally arrived _____ San Francisco.
- 7 You'll find the plates _____ the second shelf over there.
- 8 I don't know who's responsible _____ updating the website.

	8
--	---

Grammar total	20
---------------	----

VOCABULARY

3 Match the sentences with the correct word / phrase in the box. There is an extra word / phrase you do not need.

toddler child mid-twenties late twenties pre-teens
early thirties baby pensioner early twenties
late thirties teenager mid-fifties

Example: My youngest daughter started primary school last week. child

- 1 My brother's 21st birthday party was fantastic.
- 2 Maria learned to drive when she was 17.
- 3 Our 90-year-old neighbour has an extraordinary memory.
- 4 I had my first child when I was 25 years old.
- 5 Sam is 14 months old and is starting to walk.
- 6 I went to work abroad when I was 32.
- 7 Clara's 12 and always wants to stay up late.
- 8 Our son refused to sleep until he was ten months old!
- 9 She became self-employed when she was 38.
- 10 My brother is 29, but he looks much younger.

	10
--	----

4 Complete the sentences. Write one word in each gap. They are all related to describing or taking photos.

Example: The old-style cameras with film were better than digital cameras.

- 1 There are some cypress trees in the b _____.
- 2 If you want to take a photo and the light's poor, use a l _____.
- 3 Some people are standing on t _____ of a tall building.
- 4 A picture can be bl _____ if you move when you take it.
- 5 In the b _____ right-hand corner you can see a ruin.
- 6 A photo can be o _____ o _____ f _____ if you use the wrong setting.
- 7 In the f _____ there's a terrace with a table and chairs.
- 8 If you want to take a photo of a person, use the p _____ setting.
- 9 There's a group of children standing i _____ fr _____ of the temple.
- 10 You should z _____ in when you are far from something and want a close-up.

	10
--	----

Vocabulary total	20
------------------	----

Appendix L *Test 9*

Put in the infinitive or the -ing form of the verb in brackets.

1. I was pleased _____(see) you at the party.
2. The nuclear station is not likely _____(blow up) although it may shut down.
3. It's possible _____(work) all day without being disturbed.
4. You are free _____(leave) any time you want to.
5. You should be able to practise _____(sing) every day.
6. Scientists are hoping _____(discover) a new chemical element.
7. I don't mind people _____(ask) me questions.
8. He works too slowly _____(be) any use to me.
9. He can 't stand _____(wait) for other people.
10. I guess there is no point _____(sit) around here any longer.
11. The lemonade was cool enough for us _____(drink).
12. Please stop _____(interrupt) me when I am trying to explain something.
13. I've forgotten _____(buy) flowers for my girlfriend's birthday.
14. I simply can't afford _____(waste) time trying to explain this to you.
15. The thief admitted _____(enter) the house but says he didn't take anything.
16. It is difficult _____(find) a place to park in this town.
17. Peter prefers _____(drive) rather than travelling by train.
18. I've given up _____(try) to communicate with John.
19. If you put off _____(explain) it to her, then it will only become worse.
20. Many dream of _____(win) the lottery so that they never have to work again.

Complete the second sentence so that it means the same as the first.

- 1 He doesn't help me with my maths homework any more.
He's given up _____ with my maths homework.
- 2 My wife thinks I should take more time off work.
My wife wants me _____ more time off work.
- 3 Their boss said they could go home early.
Their boss let _____ home early.
- 4 She said she would call Max tomorrow.
She agreed _____ Max tomorrow.
- 5 Tim said he would take Helen to the station.
Tim offered _____ Helen to the station.

Appendix M Test 10

GRAMMAR

Mark A, B, or C to complete the sentences.

- Ayesha likes wearing _____.
A colourful clothes B colourful clothes C clothes colourful
- This book is _____ interesting than the one I read before.
A not B less C least
- Her name's Miranda and _____ is Sergio.
A he's B him C his
- I got my mother some flowers. I got _____.
A them for she B they for she C them for her
- Lend him your pen. Lend _____.
A it to him B he it C it to he
- What are _____ children's names?
A theirs B their C they
- The beach wasn't _____ in the photo, so I was disappointed.
A beautiful than B as beautiful as C as beautiful than
- Zara is much _____ than her sister.
A friendlier B more friendly C friendly
- My husband's parents usually visit _____ once a month.
A us B we C our
- That was the _____ meal I've ever eaten.
A baddest B worse C worst
- My idea is good, but yours is much _____.
A better B best C good
- I'm afraid we don't sell black trousers, only _____.
A greys B grey one C grey ones
- My boyfriend sent me an email yesterday. He _____.
A sent to me it B sent it to me C it me sent
- This tablet's _____ than the one you showed me.
A less expensive B least expensive C as expensive
- My new laptop is _____ of all the ones I've ever had.
A much faster B the faster C the fastest
- I need to call Erika. What's _____ phone number?
A her B hers C she's
- Felipe and I often see _____ at the tennis club.
A us B each other C to we
- It is available online in a _____ range of colours than in the shops.
A much wider B more wide C much wide
- Last winter was _____ since records began.
A the colder B most cold C the coldest
- Work is _____ busier now than it was last month.
A many B very C much

VOCABULARY

a Mark the correct adjective form, A, B, or C.

- rely
A relyible B reliable C relyful
- glamour
A glamorous B glamourous C glamourish
- create
A creative B createless C createful
- envy
A envious B envyish C envyful
- sense
A sensible B sensative C sensible
- passion
A passionate B passionful C passionish
- response
A responseful B responsible C responsible
- addict
A addictless B addictable C addictive

b Mark the correct word A, B, or C.

- a woman's surname before she gets married
A pseudonym B nickname C maiden name
- waste materials such as glass and paper which can be used again
A recyclable B organic C complex
- a range of products made by a company under a particular name
A brand B product C project
- describes someone who has original and interesting ideas
A considerate B thoughtful C creative
- describes something that many people like
A unusual B popular C common
- Your parents gave you the name of a family member or famous person
A named after B named about C named from

Appendix N *Test 11*

Grammar

1 Translate sentences

- 1 K narozeninám jsem dostal počítačovou hru.
- 2 Táta včera nešel do parku.
- 3 Všechny peníze jsme utratili v obchodech.
- 4 Koupila si moc hezké šaty.
- 5 Včera večer jsme nejedli pizzu.

Mark: ___ / 5

2 Write questions using the words below.

- 6 where / you / go / yesterday afternoon
- 7 you and Jess / watch / anything on TV last night
- 8 what / you / buy / at the shops last weekend
- 9 what time / Matt / get up on Saturday
- 10 who / be / your favourite teacher last year

Mark: ___ / 5

Vocabulary

3 Complete the sentences with the adjectives below. There are two extra adjectives.

ashamed bored disappointed excited frightened relieved upset

- 11 You were very _____ when your best friend moved to America. You cried all day.
- 12 Megan was _____ of the big, angry dog.
- 13 My brother was _____ when the doctor said he was OK.
- 14 The old man talked about the same thing for over an hour – we were very _____.
- 15 Thanks for inviting me to your party next week – I'm really _____ about it!

Mark: ___ / 5

4 Choose the correct words to complete the sentences.

- 1 I'm a little bit / extremely hungry, so can I have something small to eat, like an apple?
- 2 This is extremely / rather hot, but you can hold it in your hand.
- 3 It's very / a bit cold outside, so wear a big coat when you go out.
- 4 I'm extremely / a bit tired, but I don't need to go to bed.
- 5 Gary is very / a little bit cross with me – he isn't speaking to me at all.

Mark: ___ / 5

Total: ___ / 20

Appendix O Test 12

Unit 2 Test A

Grammar

1 Complete the sentences. Use the past simple or past continuous form of the verbs in brackets.

- When it _____ (stop) raining, we _____ (go out) and _____ (put) the kayak in the water.
- While I _____ (wait) at the bus stop, I _____ (see) an accident.
- Suddenly, the man _____ (take) the purse from the woman's handbag and _____ (run off).
- At 5 p.m. yesterday, I _____ (read) a book about Captain Cook and my sister _____ (watch) a film about the explorer Robert Falcon Scott.
- I _____ (relax) in the sun when a cloud _____ (appear) in the sky.
- When the sailor _____ (open) the box, he _____ (find) an old map inside.
- While everyone _____ (argue), I _____ (realise) it was a good time to go.
- As I _____ (leave) the party, I finally _____ (remember) the man's name.
- The man _____ (wear) a raincoat and _____ (carry) a big suitcase.
- The girl _____ (run) through the forest when she _____ (slip) over.

Mark: ___ / 10

2 Read the text and find eight mistakes (an unnecessary word, a word missing or an incorrect word). Circle the mistakes and write the corrections. There are no mistakes in two of the lines.

Ann was walking along the river and was talking on her phone. 1 _____
 when suddenly she noticing there was a man in the river. 2 _____
 He was swimming to the side, but he was having problems. 3 _____
 Quickly, Ann did jumped into the water and swam up to the man. 4 _____
 But while she reached him, his head went under the water. 5 _____
 Ann swam under the water, but she can't find the man. 6 _____
 A few moments later, saw the man at the side of the river. 7 _____
 He was with some friends and they all laughing at her. 8 _____
 At that moment, she realised that it was a silly joke. But 9 _____
 shortly afterwards, two police officers arrive to talk to the men. 10 _____

Mark: ___ / 10

VOCABULARY

3 Complete the sentences with the correct words. The first letter of each word has been given.

- It was a bright day in July, but inside the Son Doong C_ _ _ it was very dark and cold.
- It's a s_ _ _ climb to the top of Ben Nevis in the Scottish Highlands, but the view is worth it.
- Angel Falls, on the Guaja River in Venezuela, is the highest w_ _ _ _ _ in the world.
- During the day, the temperature in the Sahara d_ _ _ _ can reach 50°C, but at night it sometimes falls to below 0°C.
- The beautiful White C_ _ _ _ s of Dover are very high and it's dangerous to walk too close to the edge.
- The edges of Lake Como in Italy are s_ _ _ _ _ but it's quite deep in the middle.
- Over 1,500 people lost their lives in the early morning of 15 April 1912 when the Titanic sank in the L_ _ waters of the North Atlantic Ocean.
- You can collect seashells from the s_ _ _ _ in the Maldives, but you can't take coral pieces.

Mark: ___ / 8

4 Complete the sentences with the words below. There are two extra words.

boots life jacket jet-skiing kayaking paddles poles rucksack safety harness

- If you go hiking in the countryside, take a _____ so that you can carry your things in it.
- Don't worry about falling off the climbing wall – you'll be wearing a _____
- I don't really like _____ because it's dangerous and too noisy!
- When you go for a long walk, the most important thing is to wear comfortable _____
- If you go out to sea on a boat, it's a good idea to wear a _____
- When I'm walking on rocky ground, my _____ stop me from falling over.

Mark: ___ / 6

5 Replace the simple adjectives in brackets with extreme adjectives. The first letter of each extreme adjective has been given.

- That helmet doesn't fit me – it's t_ _ _ _ _ (very small).
- Doing a bungee jump was t_ _ _ _ _ (very frightening).
- Breaking my leg while crossing that river was a_ _ _ _ _ (very bad).
- After swimming across the lake, I was e_ _ _ _ _ (very tired).
- The hill we had to climb looked h_ _ _ _ _ (very big).
- The countryside we were walking through was s_ _ _ _ _ (very beautiful).

Mark: ___ / 6

Appendix P Test 13

Unit 3 Test A

Grammar

1 Circle the correct words to complete the sentences.

- There are only a few / a little actors in that new play.
- It doesn't matter how much / how many action scenes there are in a film if there isn't a good storyline.
- I think some / any soundtracks are more interesting than the film itself.
- Have you got much / many money with you?
- There is much / a lot of work to do if you want to finish today.
- There weren't some / any people left in the building.
- There are not much / not many sitcoms that I like.
- Would you like some / a little new video games for your birthday?

Mark: ___ / 8

2 Read the advice for an actor on a film set. Rewrite the advice using one of the verbs in brackets.

- It's very important to bring some food and water with you.
I (must / have to) _____.
- You are not allowed to use your mobile phone on the set.
I (mustn't / don't have to) _____.
- Don't look at the camera when you're acting.
I (mustn't / don't have to) _____.
- It's important to learn the script.
I (must / mustn't) _____.
- If you are sixteen or under, it's compulsory to bring a parent with you.
I (must / have to) _____.
- It isn't necessary to wear special clothes.
I (needn't / mustn't) _____.
- It isn't necessary to stay on the film set after 8 p.m.
I (mustn't / don't have to) _____.

Mark: ___ / 7

3 Complete the advert with the words below.

don't have to have to must (x2) mustn't

Right! I¹ _____ eat something. Boil the water, open the packet, read the instructions ... Oh no, I² _____ prepare a lot of vegetables!
Does this sound like you in the kitchen? If it does, you need **READYLUNCH!**
Open the packet.
Don't worry – you³ _____ prepare any vegetables.
Just add hot water, sit down and enjoy!
But remember, you⁴ _____ tell your friends about this or they'll all want one!
Go on. You really⁵ _____ try it!

Mark: ___ / 5

Vocabulary

4 Read the definitions of films and TV programmes and write the correct words.

- I watch this every day to find out what's happening in the world. _____
- I don't really like this type of film because the actors are always singing. _____
- I like this type of programme, especially if I know the answers to the questions. _____
- In this type of film, a man and a woman meet, and after some problems they fall in love at the end. _____
- This is a film about events from the past; the actors always wear beautiful clothes. _____
- In this programme, real people perform a special skill and the audience chooses the winner. _____

Mark: ___ / 6

5 Complete the sentences with the correct words. The first and last letter of each word has been given.

- I think thrillers are g_____g – you never know what's going to happen next!
- Science fiction films usually have a lot of s_____e_____s to make you believe you're in a different world.
- It doesn't matter what kind of film it is – if the acting is c_____g, I usually love it.
- Film producers say that you can't make a good film from a bad s_____t.
- Many war films contain a lot of fighting and are too v_____t for younger children to watch.
- I really like watching horror films, especially if the c_____s are very scary.

Mark: ___ / 6

6 Complete the sentences with the words below. There is one extra word

concern decisions express get grades have provide research take part in

- The fact that children watch too much television causes _____ for many parents.
- Some people like to _____ online games so that they can play with their friends.
- Using a mobile phone can _____ positive and negative effects on people's lives.
- Some _____ are difficult to make.
- Scientists are doing _____ into how important it is for students to do homework.
- Students who work hard usually get good _____.
- My dad always encourages me to _____ my opinions about what is happening in the world.
- Eating fruit and vegetables can _____ a lot of health benefits.

Mark: ___ / 8

Appendix Q Test 14

Unit 7 Test A

Grammar

1 Match the sentence halves. Then complete the second conditional sentences using the correct form of the verbs in brackets.

- 1 If I _____ (live) in Australia.
 - 2 I _____ (not feel) lonely.
 - 3 I _____ (save) as much money as possible.
 - 4 If I _____ (not go) to school.
 - 5 I _____ (call) her.
 - 6 If I _____ (eat) more fruit and vegetables.
 - 7 If you _____ (know) the answer.
 - 8 If I _____ (can) play a musical instrument.
 - 9 I _____ (see) lions and elephants.
 - 10 If I _____ (be) on a reality TV show.
- A I _____ (join) a band.
 B I _____ (not learn) anything.
 C I _____ (become) famous.
 D If I _____ (have) more friends.
 E I _____ (not tell) you!
 F If I _____ (travel) to Kenya.
 G If I _____ (find) a good job.
 H I _____ (be) healthier.
 I I _____ (go) surfing every day.
 J If I _____ (know) her phone number.

Mark: ___ / 10

2 Rewrite the two sentences as one sentence using the word in brackets. Change one of the verbs into the past perfect and make any other necessary changes.

- 1 Tina left her credit card at home. She wanted to buy a new dress in the sale. (but)

- 2 Samantha learned to ski. Then she went on the skiing holiday. (after)

- 3 Andrea's computer broke. Then she stopped shopping online. (after)

- 4 Gemma saved her pocket money for three months. She could buy a skateboard. (because)

- 5 Sheila wrapped all the presents. Then the guests arrived. (after)

- 6 Howard bought a lot of things in the shop. He got a free gift. (because)

- 7 Helen and Mark got married. Then they went on a honeymoon to Mauritius. (after)

- 8 Anna didn't spend much money before Christmas. She bought some new shoes in January. (so)

- 9 Ben tried to repair his own bike. Then he borrowed his friend's bike. (after)

- 10 Gavin went to university. He found a well-paid job. (because)

Mark: ___ / 10

Vocabulary

3 Read the definitions of shops and services and write the correct words. The first letter of each word has been given.

- 1 This shop only sells fresh flowers. f _____
- 2 This is a shop where you can buy fresh fruit and vegetables. g _____
- 3 This shop sells clothes and other things to raise money for a good cause. c _____ s _____
- 4 If you want to buy or sell a house, you go to this place. e _____ a _____
- 5 If you don't have a washing machine, you can go to this place to wash your clothes. l _____
- 6 This is a great place to have a hot drink and meet your friends for a chat. c _____ s _____
- 7 This shop only sells beauty products. c _____ s _____
- 8 In this shop, you can buy special or unusual foods from other countries. d _____
- 9 If you want to decorate your home, you need to buy things from this shop. D _____ s _____
- 10 People go to this place if they want to invest or borrow money. b _____

Mark: ___ / 10

4 Write the numbers and currencies in words.

- 1 451# _____
- 2 £5,487 _____
- 3 \$3,500,000 _____
- 4 €522,364 _____
- 5 ¥32,975 _____

Mark: ___ / 5

5 Read the dialogue and circle the correct words.

- Steve Have you heard that there's a ¹sale / ²coupon on at WF Arnold's? They're moving to a bigger building so there are loads of ³bargains / ⁴price tags.
- Ben Yeah, but I've got enough stuff to wear and I can't really afford to buy anything at the moment.
- Steve I can lend you some money if you like. Yesterday I took back a coat that I'd bought and got a ⁵receipt / ⁶refund, so I've got plenty of cash now.
- Ben Don't worry. I'd prefer to ⁷owe / ⁸save up and spend my own money. I don't really like ⁹borrowing / ¹⁰lending from friends.
- Steve OK, no problem.

Mark: ___ / 5

Test B

File 6A 30 points MAX

Name:



class: DL2

1. Fill in the gaps with a word to complete the sentences (6 points)

- a) The movie was dubbed directed by Christopher Nolan...
- b) ...and it stars Cillian Murphy in the role of the scientist. 5
- c) The story is based on a book by Ernest Hemingway.
- d) The cast of the film includes actors and actresses from all around the world.
- e) The plot twist at end of the movie was quite surprising. I didn't expect that.
- f) Did you know the soundtrack of the movie had been recorded by the Prague philharmonic orchestra? 9

2. Turn active voice into passive voice (10 points)

- a) Frank gave his girlfriend a golden ring. A golden ring was given to his girlfriend (by Frank)
- b) The Simpsons are reconstructing their bathroom. Bathroom is being reconstructed (by the Simpsons)
- c) They have repaired your car. Car has been repaired by them.
- d) The company delivered the consignment with a delay. The consignment was delivered with delay by company
- e) Jim and Lucy showed us the downtown yesterday. The downtown was shown by Jim and Lucy to us yesterday 9

3. Fill in the verb in passive voice (10 points)

- a) All the flights were cancelled (cancel) due to bad weather conditions.
- b) The thief hasn't been caught (not catch) yet. He must be hiding somewhere.
- c) His latest novel has been translated (translate) into 16 languages so far.
- d) When I entered the area, another passenger in front of me was being scanned (scan) for metal objects.
- e) I hope the flight won't be overbooked (not overbook). That would be a trouble for all of us.
- f) The footage was recorded (recorded) in New York during an annual flower festival.
- g) The mystery of his disappearance has never been explained (never explain).
- h) The bus was being crowded (crowd) this morning. That was a pain.
- i) Our passports were checked (check) and we could proceed to the gate.
- j) I am afraid the road ahead is being reconstructed (reconstruct). Let's find another way.

4. Translate (4 points)

- a) Váš program se právě instaluje. 4

Your program is being installed

- b) Nesmí se to prodávat dětem.

It can't be sold to children. mustn't

Test B

File 3A

Name: [REDACTED]

Class: D1D

1. Fill in the gaps with a word (10 points)

- a) Slow down. There are s PEED c AMERAS all around the city. You might get a fine.
- b) A v AN is a medium-sized road vehicle, used especially for carrying goods.
- c) It is very difficult to travel around the city during r USH h OUR.
- d) We had to wait at the t RAFFIC LIGHTS as the crossroads seemed very busy.
- e) It is recommended to cross the road at the z EBRA c ROSSING to avoid any accident.
- f) There is a t AXI r AMK just outside the airport. You might get a cab there.
- g) The s PEED IMIT on motorways in the Czech Republic is set for 130 km/h.
- h) I guess we should s LOW DOWN early to avoid the traffic.
- i) You can't drive here. It is a p EDESTRIAN z ONE. Traffic is not allowed here.
- j) We are r UNNING o UT of fuel. We should find a petrol station soon.

2. Fill in the gaps with adjectives and adverbs (10 points)

- a) What is the most expensive (expensive) thing you have ever bought? 9
- b) The farthest (far) I have ever flown is to Delhi.
- c) I think Bob is much happier (happy) in his new job.
- d) This is the most luxurious (luxurious) hotel we have ever been to.
- e) Claire sang the most beautifully (beautiful) of all the choir.
- f) The sequel was even more boring (boring) than the first part.
- g) Listening is the hardest (hard) part of the exam.
- h) Granny is the most active (active) person in the village. You wouldn't believe that.
- i) What's the most common (common) word in English?
- j) I think skiing is easier (easy) than horse-riding.

3. Translate (10 points)

- a) Měla stejný účes jako její kamarádka. She had the same haircut as her friend. 7
- b) Autobus není stejně pohodlný jako vlak. The bus isn't as comfortable as a train.
- c) Mohl bys řídit pomaleji? Should you drive slowly? Could ... more s.
- d) Jsem vyšší než můj bratr. I am taller than my brother.
- e) To je ta nejromantičtější věc, kterou mi kdy řekla. This is the most romantic thing which she sold me. 10/5

Appendix U *Questionnaire – respondent's side*

<https://forms.gle/pqFpX8hgJVs4dK2N8>