

**Univerzita Palackého v Olomouci**

**Pedagogická fakulta**

**Katedra anglického jazyka**

# **Comparison of formative and summative assessment in language testing**

Nela Neduchalová

**Anglický jazyk se zaměřením na vzdělávání (maior),**

**Speciální pedagogika pro 2. stupeň ZŠ a SŠ**

Bakalářská práce

Olomouc 2022

Vedoucí práce: dr. hab. Konrad Szcześniak

Prohlášení Prohlašuji, že jsem bakalářskou práci vypracoval samostatně s využitím uvedených pramenů a literatury a souhlasím s uveřejněním této práce.

V Olomouci dne

.....

vlastnoruční podpis

## **ABSTRAKT**

Cílem této práce je navrhnout efektivní formu hodnocení každé jazykové dovednosti na základě srovnání formativního a sumativního hodnocení. Cílem této práce je prozkoumat výhody a nevýhody těchto typů hodnocení a poskytnout přehled příkladů technik pro hodnocení. Tato práce se zaměří na testování a především hodnocení jazykových dovedností (mluvení, psaní, poslech, čtení). Tyto návrhy budou vycházet z literární rešerše a přehledových studií provedených v minulosti, týkajících se hodnocení ve vzdělávání. Uvedená doporučení mohou zjednodušit zavádění sumativního a formativního hodnocení do vzdělávacího procesu.

**KLÍČOVÁ SLOVA:** formativní hodnocení, sumativní hodnocení, hodnocení, testovací metody, známkování, motivace, vzdělávání, jazykové dovednosti, proces učení

## **ABSTRACT**

The aim of this thesis is to suggest an effective form of assessment for each language skill based on a comparison of formative and summative assessments. This thesis aims to investigate the advantages and disadvantages of these types of assessment and provide an overview of examples of assessment techniques. This thesis will concentrate on testing and more importantly assessing language skills (speaking, writing, listening, reading). These suggestions will be based on literary research and reviewing studies conducted in the past regarding assessment in education. Said recommendations may simplify implementing summative and formative assessments in the educational process.

**KEYWORDS:** formative assessment, summative assessment, evaluation, testing methods, grading, motivation, education, language skills, the learning process

### Acknowledgements

I would like to thank Dr hab. Konrad Szcześniak, the supervisor of my thesis, for his guidance and kind words of advice that enabled me to write this thesis.

# TABLE OF CONTENTS

<b>INTRODUCTION .....</b>	<b>6</b>
<b>1 BASIC PRINCIPLES OF LANGUAGE TESTING .....</b>	<b>7</b>
1.1 INTRODUCTION TO LANGUAGE EXAMINATIONS .....	7
1.2 CRITERIA OF EFFECTIVE EXAMINATION .....	8
1.2.1 Test validity.....	8
1.2.2 Test Reliability .....	9
1.2.3 Test usefulness – other test qualities .....	10
1.3 SPECIFIC METHODS OF EXAMINING INDIVIDUAL LANGUAGE SKILLS .....	11
1.3.1 Oral exams .....	12
1.3.2 Written examinations .....	13
1.4 ASSESSMENT OF LANGUAGE TESTS.....	14
1.4.1 Defining the terms.....	14
1.4.2 The scoring methods .....	17
<b>2 FORMATIVE AND SUMMATIVE ASSESSMENT IN LANGUAGE TESTING .....</b>	<b>20</b>
2.1 FORMATIVE ASSESSMENT .....	20
2.1.1 Introduction to Formative assessment.....	20
2.1.2 Strategies of formative assessment .....	22
2.2 SUMMATIVE ASSESSMENT .....	30
2.2.1 Introduction to summative assessment.....	30
2.2.2 The effects of summative assessment .....	31
<b>3 SUGGESTING AN APPROPRIATE FORM OF LANGUAGE ASSESSMENT.....</b>	<b>35</b>
3.1.1 Assessing speaking.....	35
3.1.2 Assessing writing .....	36
3.1.3 Assessing listening .....	38
3.1.4 Assessing reading.....	39
<b>CONCLUSION .....</b>	<b>41</b>
<b>BIBLOGRAPHY .....</b>	<b>42</b>
<b>LIST OF TABLES.....</b>	<b>45</b>

## **INTRODUCTION**

“Teachers will not take up attractive sounding ideas, albeit based on extensive research if these are presented as general principles, which leave entirely to them the task of translating them into everyday practice.” (Black and Wiliam, 1998, p. 15)

This quote brings us to the realization that formulating the ideas and strategies properly and therefore simplifying their usage is vital. Since the COVID-19 pandemic started, the ministry of education in the Czech Republic passed an ordinance. The ministry ordained ways of assessment to be implied in Czech schools that were used primarily only in schools for children with special needs, before. Without knowing what an assessment consists of, its strategies, benefits, and drawbacks, it can be difficult to implement it.

This thesis will concentrate on testing and more importantly assessing language skills (speaking, writing, listening, reading). The objectives of this thesis are to investigate the advantages and disadvantages of summative assessment and formative assessment and provide an overview of examples of assessment techniques. However, the main aim is to suggest an effective form of assessment for each language skill based on a comparison of these types of assessment

The first part of this thesis will provide an overview of basic principles of language testing along with some examples of testing methods, the usefulness of tests, and a brief introduction to assessment in education and its types.

The second chapter describes more thoroughly formative and summative assessments, some strategies as well as studies concentrating on the effects these assessments have on teachers and students.

The last chapter will contain suggestions and recommendations for testing and assessing language skills based on research on this topic provided by literary sources and empirical studies conducted by other authors in the past.

This thesis will provide insight into different forms of assessment. Hopefully, it may help future and current teachers and everyone who reads this thesis to familiarize themselves with this topic and lower the obstacles to implementing it in the educational process.

# 1 BASIC PRINCIPLES OF LANGUAGE TESTING

## 1.1 Introduction to language examinations

According to Bachman and Palmer (1996), there are a few crucial concepts that we have to concentrate on when developing a language test. This also applies to testing the English language in general. The first step to creating a complex and appropriate way of testing is to shed all expectations that there is such a thing as ‘the perfect test’ which would suit everyone. The level of our speaker or group, what is their future primary use of language, and the context of their learning should always be taken into account. The most used and the most popular type of testing English for non-native speakers is examining the four language skills – listening, speaking, reading, and writing. However, after putting in perspective all of the above-mentioned concepts, we see how the needs and expectations of the exam may differ from the original idea. Another important problem exam developers may face is the connection to the content of what our users have learned, it has to be put in the right context and it has to test their abilities in such a way that relates to what they have been taught. This can be achieved by creating a conceptual framework that makes the exam developer answer all the crucial questions regarding:

- Length of lectures or courses
- Discipline
- Student’s level in terms of topical content
- Student’s work with the input (the lecture)
- Type of questions (length, complexity, focus)
- Type of answers (multiple choice, constructing their own answers; written, spoken)

Furthermore, Bachman and Palmer (1996) claim that we have to take into consideration also the students themselves. They speak of humanizing the testing process which means we are supposed to try to find a way of including the students in the process, treat them as responsible individuals, and most importantly provide them with as much information about the testing procedure as possible. This and how we design the test will influence the anxiety level of our test takers. Our overall goal is to encourage and enable the students to execute their abilities at their highest level.

## 1.2 Criteria of effective examination

### 1.2.1 Test validity

Henning (1987) differentiates 5 types of validity in language tests, including written, or oral exams.

- a) **Content/Face Validity:** Indicates if the test measures, what it was designed to measure. Also, if the content is appropriate, comprehensive, and representative of this measure.
- b) **Response Validity:** Indicates if the score and manner in which exam takers responded, are consistent with the true ability, which the test aimed to measure.
- c) **Concurrent Validity:** Determines how accurately the score on said exam corresponds with performance on another exam. Provided this other exam has proved to measure successfully the same ability or skill we aim to measure.
- d) **Predictive Validity:** Measured usually by the academic or professional performance serving as criteria. It demonstrates how the scores of the exam correspond with future performance in the targeted aspect of behaviour.
- e) **Construct Validity:** Determines if the test really measures the hypothetical construct. For instance, a test measuring personal success should not have questions aiming only at financial independence, since success can have several different criteria.

He claims that tests can be criterion-referenced or content-valid. The criterion-referenced are used to measure the skills or abilities taught in a particular course or lesson. The content-valid tests aim to estimate the level of general expertise in language skills. The contents tend to be of medium difficulty for the target population (Henning, 1987, p.180). He states some of the threats to validity, for example, irrational constructs, inappropriately chosen criteria, poor selection of contents, or invalid application of tests.

Referring to validity Fulcher and Davidoso (2007) stated crucial aspects of validity. The structural aspect means that the exam's contents and way of evaluating it should correspond with the items we aim to test. The goal of the test is to provide information about the knowledge or skill in the desired area. Even though it is impossible to fit the whole



curriculum or the lesson's content. The score should possess the aspect of generalisability. Does the test predict abilities that extend beyond just the contexts displayed in the test? Finally, the aspect, which is an equivalent of the concurrent validity, already explained by Henning. This aspect provides the exam developer with confidence in the test.

### 1.2.2 Test Reliability

„Reliability is the consistency of test scores across facets of the test.“ That is what Fulcher and Davidoso (2007) claim. Bachman and Palmer (1996) define reliability as a 'consistency of measurement'. Therefore, when taken again in consistent conditions, if the test results are stable and consistent, the test is considered to have high reliability. Brown (1996) explains that reliability can be calculated as a reliability coefficient ( $r_{x,x'}$ ) and the range where it moves is + 1.0 (totally reliable) – 0 (no reliability). This also measures variance or measurement error, which also needs to be consistent. He provides the following example:

“For instance, if the scores on a test have a reliability coefficient of  $r_{x,x'} = .91$ , by moving the decimal two places to the right, the tester can say that the scores are 91% consistent, or reliable, with 9% measurement error (100% - 91% = 9), or random variance. If  $r_{x,x'} = .40$ , the variance on the test is only 40% systematic and 60% measurement error.“ (Brown, 1996, p.193)

Brown (1996) describes three often used strategies that can be used to determine test reliability. These include test-retest reliability, equivalent forms and internal consistency strategies.

- a) **Test-retest reliability:** This is an indicator of how the test results change in time. The test has to be given to the same group of students. The time between the two administrations should be sufficient, as to rule out the possibility that test takers remember it well enough to repeat their answers. However, the time period must not be too long, since the learners could acquire additional knowledge of the topic. This method rules out the factors and conditions that may have influenced the test takers and it evaluates the constants in the sample.
- b) **Equivalent-forms reliability:** This is measured by using two equivalents of the same test and administering it to the same group. Provided these two forms of the test are parallel and contain the same items and the number of these items is identical. However, giving the same test to students, who have not quite mastered the topic will

not give us much information. To prevent this, a third measure can be included to confirm that one of these forms is reliable.

- c) **Internal consistency reliability:** This indication is said to be less complex and demanding since it measures reliability by administering a test once to only one group. It determines how reliable a test is based on the correlation between answers to individual items. The most common methods are Cronbach alfa and the split-half method. Cronbach alfa calculates the correlation between every possible pair of items. The split-half method is essentially creating a set of measures, dividing it into two equal parts, and creating an equivalent of the “parallel forms”. Usually, one form is made of the odd-numbered items and the other of the even-numbered ones. Then the two sets of scores are calculated to see how they correlate. However, this reliability only applies to one half of the test (either half).

### 1.2.3 Test usefulness – other test qualities

Regarding reliability, Fulcher (2007) mentions the term “test authenticity”, which is how the tasks on the test correlate to the tasks in real life. He also describes another feature - test interactiveness that determines the degree to which the test taker engages their abilities when taking the test.

According to Bachman and Palmer (1996), there are also other test qualities. Practicality is one of them. If a test is practical it is economical and takes an appropriate amount of time and manpower or skills. Crucial is also impact. The quality of a test can be among other things determined by its impact on society, the educational system, and individuals – teachers and students. The effect is described as ‘washback’. The test takers are influenced by three aspects of taking a test: 1. The experience of preparing for the test and taking it 2. The feedback or a grade they receive 3. The decisions that will result based on their performance (for example an admissions test). This relates to fair test use – a term used to describe if the test was used according to what the takers have been told. For a test to be used fairly, the examiner should state what will be tested and how decisions will be made based on the score. Teachers can also be directly impacted by tests. Their method may change to a “teaching for the test”. To avoid this the contents of the instructional program should correspond closely to the test tasks and their characteristics.

Bachman and Palmer (2006) describe the term “test usefulness“, this attribute combines reliability, construct validity, authenticity, interactiveness, impact, and practicality. They

claim that rather than maximizing individual qualities, test developers should maximize the overall test usefulness. These individual qualities should be evaluated by their combined efficiency and not individually. They add that none of these qualities or the usefulness of the test can be prescribed generally, distinction is to be made between tests for different groups and specific situations.

Of course, authors such as Bachman and Palmer (1996), Fulcher (2007), Douglas Brown (2004), James Dean Brown (1996), and many more, writing about language testing and similar topics, deal with other factors influencing testing language. For example, designing the tests, scoring methods, measuring errors and variance using arithmetic operations. Although these factors influence mostly written tests, most of them are complex and their explanation would be comprehensive, they are nevertheless worth mentioning. Further, specific methods of testing the four language skills, beyond written tests, will be discussed.

### **1.3 Specific methods of examining individual language skills**

According to Fulcher and Davidoso (2007) language, testing is a complex process. Its influences and the impact it has spreads beyond the limits of the educational process. It can affect society in terms of equality and opportunity. The perception of exams is mostly based on personal experiences, some students may perceive testing as harmful to their self-esteem, and others may give tests credit for opening doors to lifelong opportunities, academic and professional.

There are different methods of examining language skills. These methods each have their process. Preparing students for specific tests can be challenging since every test requires particular preparation. Choosing the method of examining language skills is demanding. Usefulness, ethical approaches, suitability, and effects on students and teachers – all of these qualities should be considered before realizing an exam. The study was conducted by Maria Assunção Flores and her colleagues in 2014. They researched undergraduate students of two universities in Portugal. They gathered evidence suggesting that written tests are according to students the most anxiety-inducing. However, they also found out that these tests are the most commonly used assessment methods. They distinguished two types of methods: traditional and learner-based. They include written and oral examinations in the first type, and the second type includes portfolios and group projects. The respondents split into two groups – one tested more frequently by the learner-centered methods and the second one tested more frequently using the traditional methods. Their research indicates that students

tested more frequently using learner-centred methods, feel as though the assessment is more fair and effective. Feedback, as a reliable source of information, is perceived as significantly important by both groups of students. These effects and qualities differ with each method. This overview aims to briefly introduce some of the commonly used methods in language testing.

### **1.3.1 Oral exams**

Oral exams play a crucial role in testing English. This term refers to an exam of a student's speaking ability while also discriminating between students who acquired the skills or knowledge and those who did not. Luoma (2003) revisits Och's 1979 theory of planned and unplanned speaking. Planned speech is rehearsed and prepared, unplanned speech tests the ability to react and produce language in a conversation. When assessing speaking, Luoma (2003) makes an intriguing point. Examiners tend to excuse errors commonly made by all speakers when examining a native speaker. There are mistakes made typically by learners of the language. However, the prejudice makes it harder for examiners to differentiate between these mistakes and mistakes made typically by all speakers. Therefore, disadvantaging the learners in comparison to native speakers. Joughin (2010) describes the disadvantages of oral assessment. It can be time-consuming since the teacher is able to assess a limited number of speakers at a time. Oral examination can be anxiety-inducing because of the lack of anonymity or fear of being judged based on their gender, clothing, or race. Disadvantaged are also students with hearing or speech difficulties. On the other hand, pupils with reading and comprehension difficulties may prefer an oral exam to a written one. Speaking is not the only skill that can be assessed by oral exams. Oral exams seem to be reliable sources of information about learners' vocabulary, grammar, unplanned speech also listening skills. Here is an overview of a few chosen examples of how an oral exam may look.

**Role-play:** Budden (2004) describes role-play as a fun exercise adding variety and encouraging learners to produce language. This method can also be used to assess unplanned speech. The students are given a description of an imaginary situation (at the restaurant, going to a shop, ordering a pizza) and demonstrate their skills to communicate in said situation. This can test their level of preparation for these possible scenarios. Another option is assigning students roles of imaginary people (roommates fighting, a popstar accepting an award, a job interview) and they play the role, so they can explore and react to situations not common in their real lives.

**Interviews:** An interview can be used as a method of gathering information about the exam taker's ability to respond to questions posed by a teacher. This method is commonly used as a practice in the elementary level of English. It can be used to practice questions and positive and negative responses.

**Presentations:** Assigning students with presentations is a method of eliciting evidence about a student's mastery of planned speech. This approach to oral testing differs from the previously mentioned mostly in because it enables students to prepare. They can be given any amount of time the teacher decides to be sufficient. (Weir, 2005, p. 67)

Other methods include Debate, storytelling, speeches, discussions, description of pictures, interpreting an event or a written/recorded text, improvisation, reading aloud, and many more.

### 1.3.2 Written examinations

Writing is indubitably dominating traditional assessment. It includes essays, reviews, reports, and written examinations (Joughin, 2010, p.3). Wiegle (2002) believes that writing is becoming a crucial skill in peoples' lives. As technology advances, global communication in a second language becomes essential for education, business, and personal reasons. She further states that writing demands the use of cognitive activities, such as planning and information processing. The writer is expected to manage the audience, form, and appropriateness of the information about the topic. Writing can test vocabulary, spelling, grammar, structure and organization abilities, and students' ability to work with literary sources. Again, students with disabilities relating to reading and comprehension can be disadvantaged. Here are some examples of writing assignments that can be used to test language.

**Essays:** An essay is a piece of writing based on a question, statement, or an argument. Duigu (2002) names some of the abilities and skills that an essay requires. By assigning it, teachers can recognize to which extent did their students master said skills and abilities. These include systematic organization, understanding the topic, presenting an argument, using formal/appropriate English, and getting a grasp of grammar and vocabulary.

**Portfolios:** "... in terms of writing assessment a portfolio is a collection of written texts written for different purposes over a period of time" (Wiegle, 2002, p.198). Brown (2004) on the other hand, believes that portfolio can contain, besides written texts, tests, projects,

and notes also audio and/or video recordings. Wiegler (2002) claims that the goal of portfolio assessment is gathering evidence of the development of students writing skills which is the reason why it has to contain a collection of writing samples. The implementation of portfolios has a potential positive impact. Likely, discussions about writing that arise between students and teachers because of portfolios (mainly while giving or receiving feedback) can be beneficial. Even though, these discussions do not depend solely on portfolios and can be held during revision or different types of testing. She also points out the advantage of reduced anxiety of creating a portfolio compared to timed writing tests. Weir (2005) questions the validity of portfolios since the students may have done some of their writing at home. There is no way to determine the extent to which their parents could have helped them.

**Written tests:** Tests can be separated into two groups based on questions: open-ended and multiple-choice questions. Fulcher (2010) claims that multiple-choice tests are more easily scored because open-ended questions usually require a human judgment. Another difference between the two types is that multiple-choice questions usually offer the student four answers (in most cases, only one is correct), so answering becomes a process of elimination. When answering an open-ended question, the students have to construct the answer by themselves. Some authors argue that the latter is more similar to real-life situations regarding language usage. Tests can also be true or false, fill-in-the-blank, matching terms or pictures, error correction, and others. Answers can be based on knowledge of the topic, and listening or reading skills.

Other methods include projects, a summary of a text, literature review, research papers, writing a letter/email, writing a description, and many others.

## **1.4 Assessment of language tests**

### **1.4.1 Defining the terms**

The terms assessment, evaluation, and testing are often incorrectly understood, due to them being often used interchangeably. Even some experts do not agree on definitions of these terms. This chapter will provide examples of definitions and discuss their meaning with other definitions of these terms.

Brown (2004, p.3) defines the term ‘test’ as: “a method of measuring a person’s ability, knowledge, or performance in a given domain“. It also can be described as a tool or

instrument enabling teachers to measure students' abilities, knowledge, or performance. Crucial is also a measurement of a given domain. As discussed in previous chapters, the test is only a sample of performance and a proper test has to measure the overall language proficiency.

He claims that tests are a part of the assessment. Assessment is, according to Brown (2004), an ongoing process, sometimes even incidental. He claims that teachers assess performance. Information about students' performance can be gathered from tests but also from activities in class. Although a different term covers practice classroom activity - teaching. Teaching requires students to not be assessed and therefore enables them to practice and learn. This depends on a person's perspective. Brown (2004) claims that even a slight error correction during class conversation can be considered an assessment. Yet, he states that practice exercises and non-assessed situations in class are teaching. This raises the question, of how does a teacher explain and practice a skill or knowledge in a lesson without pointing out, describing, or correcting the errors. Moreover, he mentioned that the assessment does not have to be intentional, then how is it possible for a teacher to stop assessing?

Referring to assessment Moss (2003, p.16) claims that it is a non-stopping ongoing process. She describes that she seizes every opportunity for learning about her student's abilities and knowledge as well as about the efficiency and quality of her teaching, thus far. Moreover, she always responds to or anticipates her student's actions with her own actions. She admits that many of these exchanges are in fact, not conscious. Her definition implies that it is not possible to stop assessing when one decides to do so. This suggests that teaching and assessment happen simultaneously.

Evaluation and assessment in education are by some authors described as completely different terms and some authors do use them interchangeably. According to Burke (2010), the terms assessment and evaluation differ. Simply put, she states that assessment is the process of gathering evidence (using tests, presentations, etc.) and evaluation is the procedure of collecting data and judging it. She labels the two types of assessment (summative and formative) as based on purpose. This would be understandable if she did not then give examples of other types of assessment. For example, based on grading standards – norm-referenced assessment, and criterion-based assessment. As mentioned grading and judging the results of the assessment is by Burke (2004) considered to be an evaluation. Other authors consider these terms to be tests. Brown (1996) compares norm-

referenced tests (NRT) to criterion-referenced tests (CRT). Norm-referenced tests determine the level of student's knowledge in relation to the group. The result will be in percentile, for example, the 79<sup>th</sup> percentile means that the student performed better than 79 test-takers out of 100. This does not give us any information about the correct number of successfully fulfilled tasks or correctly answered questions. Criterion-referenced tests measure students' performance without relating it in any way to other students. It is compared to the criterion, which is usually expressed by a passing grade (for example 70% of questions answered correctly).

Brown (1996) and Burke (2004) view these terms differently and there are many authors with even more perspectives from which this web of terms can be looked at. Some claim that formative and summative are types of evaluation, others believe that any assessment can be used summatively or formatively. Discussing all of these approaches is unrealizable since the deciding point is a perspective, and who is to say which perspective is wrong and which is right? For the purpose of this thesis, the term assessment will be used as a process of collecting information about a student's knowledge, abilities, and skills as well as a way of judging the results. The reason is the way the assessment is usually defined and the logical conclusion arising from these definitions.

“Formative assessment is carried out during the learning process as an intervention that is designed to encourage further learning and change. It is frequently used in contrast to summative assessment. “(Fulcher, 2007, p.372).

“Summative assessment is conducted at the end of a programme of study to assess whether and how far individuals or groups have been successful “(Fulcher, 2007, p.372).

This means that there is a demand for the teacher to evaluate the results of assessment to be able to apply the assessment. In summative assessment, one is not able to determine the progress without judging the results and in formative assessment, it is not possible to alter (formulate) the future teaching without knowing if the change is necessary.

There is an extensive amount of different types of assessment. However, for the purpose of this thesis following overview will cover only the ones most commonly associated with formative and summative assessment and language testing. As disclosed, these can be in some literary sources described as types of evaluations.



*Informal assessment* can be given in form of comments, impromptu feedback, or by putting a sticker on an assignment or homework. This assessment include also advice on compensating strategies, taking notes, written comments on an essay, and so forth.

*Formal assessment*, on the other hand, uses methods of systematic observation. These can include tests, but also oral presentations, and portfolios. This assessment happens periodically (Brown, 2004, p.5-6).

These types of assessment can generally be used in the formative or summative way. This depends on these factors: at which point of the learning process is the assessment happening and how will be the results used. The topic of results and their interpretation can be controversial since there are many methods and each educator may have a different opinion on which is the most effective.

#### **1.4.2 The scoring methods**

Bachman and Palmer (1996) split the scoring methods of language exams into two groups. First are tests, where the result (usually a number) discloses how many tasks did the student complete successfully. With the other method, teachers rate student's language abilities presented by task responses, using scales. These scales start at zero evidence of knowledge and end at the mastery of the tested language ability. They claim that both of these approaches require two steps:

- 1) Criteria specification: what will be considered correct or what will determine the quality of a response
- 2) Deciding which procedures will enable the teacher to arrive at a score

Brown (2004) also adds that the examiner should decide on the weight of a grade. This decision may be based on the difficulty of the task or on what type of language skill the examiner aims to measure. He also provides arguments against grading by using letters (this also concerns the numerical grading in the Czech Republic). Some of these include that the number of skills and qualities students have to display to receive a grade is not representative by a letter. Since it does not disclose any information about strengths and weaknesses. He and other authors (for example Burke, 2010) encourage educators to consider alternatives to this type of grading. Wiliam (2011) writes that grades cannot be abandoned entirely since school systems (especially in America) rely on them. He also encourages smarter grading that should occur on fewer occasions. He suggests stopping grading children in elementary

schools entirely since they are still learning. And claims that learning stops with receiving a grade so teachers should be sparing with giving them. This opinion may be considered controversial, although more authors believe that the grading system should be changed in some way. That includes less frequent grading and alternative grading.

Brown (2004) offers an alternative way of grading in relation to both types of assessment:

- a teacher's marginal and/or end of exam/paper/project comments
- a teacher's conference with the student

If the examiner aims to use formative assessment he recommends these alternatives:

- a teacher's written reaction to a student's self-assessment of performance,
- a teacher's review of the test in the next class period,
- peer-assessment of performance,
- self-assessment of performance, and

If the teacher aims to assess summatively, he recommends these modifications:

- a teacher's summative written evaluative remarks on a journal, portfolio, or other tangible product
- a teacher's written reaction to a student's self-assessment of performance in a course
- a completed summative checklist of competencies, with comments
- narrative evaluations of general performance on key objectives

The main advantages of narrative evaluation are „individualization, evaluation of multiple objectives of a course, face validity, and washback potential” Brown (2004, p. 296). However, this method is time-consuming and students tend to not pay much attention to the written narrative since the summative assessment is usually still attached.

Brown (2004) also sees student-teacher conferences as effective. They may require more time, for that he recommends holding these conferences with individual students during a lesson when they complete assigned tasks.

Another alternative method includes using a checklist. These usually contain questions and the student or a teacher then checks boxes that match the student's performance (options range from excellence to unsatisfactory). Checklist evaluations and rubrics are not as time-

consuming. Brown (2004) recognizes the advantages, of using checklists in scoring. He claims they are practical and reliable and students are more likely to absorb the information from checked boxes of criteria than from a paragraph about their performance. However, the reduction of individualization may be a disadvantage. Burke (2010, p.96) claims that checklists can be easily transformed into rubrics by changing the questions about performance into short statements she argues that it requires skill and practice to create them, but when they are done properly they can be of tremendous help. She claims that both rubrics and checklists can be used as summative or formative assessments, it depends on whether the teacher bases a grade on them or on feedback.

In the following chapter, formative and summative assessments will be introduced and discussed more in depth. The discussion will revolve around published definitions and research findings.

## **2 FORMATIVE AND SUMMATIVE ASSESSMENT IN LANGUAGE TESTING**

### **2.1 Formative assessment**

#### **2.1.1 Introduction to Formative assessment**

The formative potential of assessment has been discussed for a long time before really rising to the surface as an approach similar to the one we know today. It is believed to particularly come to the fore in the 1980' in the United Kingdom. Authors Torrance and Pryor (1998) claim that this was first about secondary education. Assessment theorists have been struggling with regard to the structure and application of formative assessment since its results can be perceived as less transparent or informative than the results of summative assessment.

The Council of Chief State School Officers (CCSSO) is a crucial organization in American education, its officials who are voted or appointed by governors, oversee and direct education in each state of the U.S. This organization formed the FAST (Formative Assessment System for Teachers) State Collaborative on Assessment and Student Standards (SCASS) with the purpose of providing educators with tools that help them screen and monitor assessment progress. Another purpose of the FAST SCASS is to provide guidance and resources to state-level policymakers on formative assessment. Even though their primary focus is pupils from kindergarten to 5<sup>th</sup> grade, they have published a definition of formative assessment reaching a wider audience.

“Formative assessment is a planned, ongoing process used by all students and teachers during learning and teaching to elicit and use evidence of student learning to improve student understanding of intended disciplinary learning outcomes and support students to become self-directed learners” (CCSO, 2022, p.2).

Laufková (2017) defines formative assessment as a continuous process which helps teachers, students and parents. The information gained from the formative assessment may help decide if the currently used learning/teaching style is effective. She puts more emphasis on the ability of the formative assessment to appraise students' progress, and shortcomings and

guides them toward ways that could remove the unwanted results and help them understand the topic more easily.

FAST SCASS also state in their document: “Effective use of the formative assessment process requires students and teachers to integrate and embed the following practices in a collaborative and respectful classroom environment:

- Clarifying learning goals and success criteria within a broader progression of learning;
- Eliciting and analysing evidence of student thinking;
- Engaging in self-assessment and peer feedback;
- Providing actionable feedback; and using evidence and feedback to move learning forward by adjusting learning strategies, goals, or next instructional steps.”

(CCSO,2022, p. 2-3)

The Assessment Reform group of the United Kingdom argues similar points that are crucial for assessment to improve learning:

- The provision of effective feedback to students
- The active involvement of students in their own learning
- The adjustment of teaching to take into account the results of assessment
- The recognition of the profound influence assessment has on the motivation and self-esteem of students, both of which are crucial influences on learning
- The need for students to be able to assess themselves and understand how to improve

(Wiliam, 2011, p.39)

Essentially, both of these sets of requirements overlap in feedback, adjusting learning/teaching, and self-assessment. However, when implementing this type of assessment, a teacher should aim to execute the supplementary practices formed by a group from both countries. Firstly, the reason being that the active involvement of students is essential for their improvement. Secondly, the student motivation and their performance in tests and during lessons are directly influenced by their self-esteem, which is often overlooked in relation to assessment. Lastly, the clarity of learning goals can prove useful for both sides, the teacher and the learners. This suggests that clarity of success criteria if

discussed or provided to them before testing, allows students to practice and focus on particular problems. Regarding the teacher, with concrete criteria, they can might take the risk of inappropriate grading and avoid students' complaints about their grades. Above mentioned practices, when applied, should ideally reflect in strategies of formative assessment.

### **2.1.2 Strategies of formative assessment**

William (2011) recasts five key strategies, adapted from Leahy, Lion, Thomson, and William (2005), and offers research studies proving their importance and describes techniques for implementing them. To make the process of formulating strategies of formative assessment and therefore making it easier for teachers to apply in their classrooms. This chapter will provide their overview and summary of the basic ideas.

- A) Clarifying, Sharing, and Understanding Learning Intentions and Success Criteria
- B) Eliciting Evidence of Learners' Achievement
- C) Providing Feedback That Moves Learning Forward
- D) Activating Students as Instructional Resources for One Another
- E) Activating Students as Owners of Their Own Learning

#### **Clarifying, Sharing, and Understanding Learning Intentions and Success Criteria**

When it comes to sharing and understanding learning intentions, the goal is quite comprehensible. Intentions can be shared with students in a written form or only said at the beginning of a unit, understanding can be achieved through asking students questions or having them repeat the intentions. One technique that is particularly useful and has been lent support from several experimental studies with over 260 college students. This study showed that the success rate is higher when students create their own outline of what they have been learning and generate their own study questions and answer them in the test. The other students have been provided with study guides or left to be prepared on their own. The teacher then evaluates both, the questions and the answers. These students have proved to reach the learning goal and it is a tremendous help for the teacher to see what students think they have been learning (William, 2011, p.68).

Clarity of learning intentions shines a light on the language differences between teachers and students. Torrance and Pryor (1998) have found in their study of English schools that the language used when assessing pupils has a significant impact on children's performance in the present time but also in the future. The problem of language differences demonstrates itself not only in assessment, even so it begins when forming the goals. It is a challenge for the teacher to formulate an intention that would fulfil the curriculum and the students would understand it. In some cases, it is useful for the teacher to create a model of progression in their subject ahead of time and it can be a help when shaping for example a class discussion (Wiliam, 2011, p.66).

The correct formulation can be demanding and even though the School Education Programmes generally have anticipated outcomes in the school's curriculum, the demand appears with an attempt of adhering to them throughout the school year and in each class. However, in some cases, even the standard in the curriculum can be inappropriately constructed.

Table 1 Examples of confused and clarified learning intentions (Wiliam, 2011, p.61)

Context of learning	Confused learning objective	Suggestions for clarifying the objective
Assisted suicide	To be able to present an argument for or against assisted suicide	To be able to present an argument either for or against emotionally charged preposition
Changing a bicycle tire	To be able to write instructions on how to change a bicycle tire	To be able to write clear instructions

Above mentioned example demonstrates the importance of formulating intentions in such a way that emphasizes the difference between instructional activities and intended learning outcomes. Since the desired result is not for learners to be able to flawlessly execute one concrete task, yet to ensure that students will be able to put their newly gained skills into practice when faced with a similar task in the future of their learning process. On the other

hand, it might be considered easy to provide students with directions and show them directly what is the goal of your teaching process, however, this may result in unimaginative, unoriginal teaching. That is one of the challenges teachers have to overcome when designing a teaching plan since it has to be designed backwards (Wiliam, 2011, p.60-61).

### **Eliciting Evidence of Learners' Achievement**

This strategy of formative assessment is crucial since to improve the teaching strategy, it is necessary to discover the gaps in the students' knowledge. Wiliam (2011) claims that one way to approach this is by asking the right questions. He provides examples from his research and experience; all of these examples point toward the hidden information that questions can reveal if asked correctly. Concerning questions, he claims that they provide a window into students' thinking, however, they are not easy to generate, but they are crucially important if teachers are to improve the quality of students' learning (Wiliam, 2011, p.77). Questioning students effectively has its difficulties it is not just about framing the question properly authors Sadker and Zittleman (2008) believe that questioning is the foundation of effective teaching and they add:

“Although teachers rely most heavily on lower-order questions (e.g. their answers require memory and simple facts), higher-order questions are associated with critical thinking and should be an important part of classroom instruction. Effective teachers use intentional strategies, such as proper wait time, to allocate questions fairly among all students. When providing feedback, teachers typically use neutral acceptance, while praise, remediation, and criticism are more precise and helpful reactions” (Sadker, 2008, p.402).

They raise the point that there are many factors forming an effective question Wiliam (2011) besides discussing similar problems regarding time management and neutral language, describes the term “achievement gap” which is a phenomenon that seems to appear when teachers allow students to choose whether they want to participate or not. Since he claims that “high-engagement classroom environments appear to have a significant impact on student achievement” (Wiliam, 2011, p.81).

Littleton et.al (2005) in an article, written based on research, support this claim by finding that children who engaged in class and used language and talking as a tool for thinking outperformed children in similar schools without the ‘Thinking Together’ program. They scored higher in teacher-constructed exams and even standardized science tests. Wiliam (2011) also refers to the importance of mistakes and how the character of mistakes



and their frequency can also help teachers gather evidence about their students' level of understanding.

### **Providing Feedback That Moves Learning Forward**

Feedback is a crucial part of the learning process. It is a tool the teacher and the student both expect to receive since it can provide an idea about the current situation in the educational process. It is used in ways that can improve learning, however, to work with feedback and provide it correctly can be a challenge. Wiliam (2011) argues that feedback is highly effective with a study conducted in 1985 by Maria Elawar and Lyn Corno. The research took place in Venezuela and they focused specifically on written feedback in classes for 6<sup>th</sup> grade. The students were separated into three groups, one group received constructive feedback, one group was split into halves and one half received constructive feedback and the other only received scores without any further remarks and the third group received only scores. With this research, they provided evidence suggesting that constructive feedback worked as a helpful way to reduce the time needed to learn a given topic. Moreover, the results also showed that the student's attitude towards the subject improved, and the achievement gap between the sexes of the students, favouring males over the course of study, was reduced. Feedback also needs to be given in a particular way and under specific conditions to reach its full potential. These conditions were described by a group of Australian educators and researchers in 2019 who conducted research on this topic in 35 universities. The project had 4 stages including a large-scale survey, analysis of seven rich case studies through interviews, creation of a framework of conditions and principles of effective feedback, and the last stage involved workshops. The results indicated that successful practices of feedback are influenced by the design of the feedback itself, the number of participants, and the culture of the institution. All of these case studies have revealed that effective feedback has been influenced by a wide range of factors throughout history.

This has already been found in other studies focusing on the relationship between factors and the learning design of higher education. In addition, this particular project acknowledges that effective feedback is not influenced only by designs but also by learner and educator capacities and their dispositions, as well as by the competitiveness and multi-layered requirements of the university, faculty, and classroom environment but also by disciplinary contexts. These researchers identified 12 conditions that allow for effective feedback. However, this does not mean that all conditions must be met for the feedback to

be effective. The survey suggests that six or more of these conditions have proven to be significant factors in feedback success (Henderson et al., 2019).

Table 2 Conditions of feedback (Henderson et al., 2019, p.115)

Capacity for feedback	1. Learners and educators understand and value feedback
	2. Learners are active in the feedback process
	3. Educators seek and use evidence to plan and judge the effectiveness
	4. Learners and educators have access to appropriate space and technology
Designs for feedback	5. Information provided is usable and learners know how to use it
	6. It is tailored to meet the different needs of learners
	7. A variety of sources and modes are used as appropriate
	8. Learning outcomes of multiple tasks are aligned
Culture for feedback	9. It is a valued and visible enterprise to all levels
	10. There are processes in place to ensure consistency and quality
	11. Leaders and educators ensure continuity of vision and commitment
	12. Educators have the flexibility to deploy resources to the best effect

While being mindful of these conditions, one has to also bear in mind the protentional results. Wiliam (2011) points out possible results of feedback, he provides 4 types of response: change of behaviour, change of goal, abandonment of goal, and rejection of feedback. The change in behaviour is probably the one desired most by the feedback provider. This type of change can manifest itself in two ways if feedback indicates that the performance exceeds the goal, the recipient will make less effort, if feedback indicates that the goal has not been achieved, the recipient will make more effort. Changing a goal has two possible outcomes if the goal is achieved the recipient ideally changes their goal to one that is more demanding otherwise they will reduce their aspirations. Abandoning the goal can manifest either by a decision that the goal is too simple, if it has already been achieved it

might be considered too challenging and therefore unachievable. Usually, student decides that he or she is not smart enough or simply not good at a certain subject and they stop trying to improve. The last potential result is the rejection of feedback, which in case of both not reaching and reaching the goal will manifest itself as ignoring the feedback altogether. This is most common in the work environment. Wiliam (2011) also claims that the donor of the feedback has to ensure that the reaction will be cognitive rather than emotional, he adds that feedback has to be focused, relate to the study goals, known by the recipient, and most importantly it should increase the extent to which students take ownership of their learning. According to these authors and studies all these factors, conditions and outcomes have to be considered and the person giving feedback should pay close attention to them if they aim to give effective feedback.

### **Activating Students as Instructional Resources for One Another**

Wiliam (2011) in terms of embedding formative assessment emphasizes the usefulness of cooperative and collaborative learning. He provides quotes from interviews with fifth-grade students. The students admit that they prefer to seek assistance amongst other students since they have trouble understanding the teacher's language. Others admitted that they pretend to understand after a second attempt at explanation, this is allegedly due to the students being aware that the teacher is busy and not wanting to take up more time of the lesson. Others did the same for different reasons, for example, the fear of appearing unintelligent in front of the teacher or their classmates or simply wanting the lesson to be over as soon as possible. However, when working with peers, students often do ask for a change of pace or multiple repetitions until they understand properly. Peer tutoring may prove to be a more effective way of learning, this level of effectiveness would not be possible in the learning process between an adult and a student, due to the "change in power relationships". The efficiency of cooperative learning can be ensured by the presence of two requirements: group goals and individual accountability. Group goals force students to work 'as a group' rather than 'in a group', and individual accountability ensures that every student does their part.

He believes that peer assessment can be used as a tool for improvement rather than evaluation. Students are generally more straightforward with their peers than any teacher could probably ever dare to be. It is important to realize that the individual helping their classmate also receives benefits. Internalizing the learning intentions and success criteria in the context of someone else's work seems to be less anxiety-inducing.

Robert E. Slavin (1980), the author of the four main factors of collaborative learning: motivation, social cohesion, personalization, and cognitive elaboration, broadened the knowledge of the effects of collaborative learning with extensive research and several studies. He defines collaborative learning as a term referring to “classroom techniques in which students work on learning activities in small groups and receive rewards or recognition based on their group's performance” (Slavin, 1980, p.315). His findings imply that cooperative learning proves to be more effective in lower-level outcomes such as knowledge or application of principles. However, the techniques proved useful when containing the following:

- a) A structured, focused, schedule of instruction;
- b) Individual accountability for performance among team members;
- c) A well-defined group reward system, including rewards or recognition for successful groups.

His study implies, among other points that cooperative learning can positively influence students' self-esteem. Moreover, students who were taught using cooperative learning techniques generally reported greater liking for school than students in traditionally taught classes. Wiliam (2011) claims that activating students as learning resources for one another successfully leads to students taking ownership of their learning.

### **Activating Students as Owners of Their Own Learning**

Wiliam (2011) writes about the disconnection between general knowledge and the way learning is grasped in most classrooms. It stands to reason that only learners create learning, not teachers. Yet, in most schools, teachers, not students are held accountable for doing the learning. Accountability regimes that threaten the teaches with sanctions for unsuccessful learning, not the students, exist in many institutions. One of the ways to help students take ownership of their learning is self-assessment.

According to MacMillan (2008, p.40), in today's standard-focused education, student self-assessment is one of the few approaches that promise improvement of students' motivation and involvement in the learning process. If properly implemented in the teaching, self-evaluation can significantly support students' intrinsic motivation, increase their effort, support goal orientation, and more meaningful learning. It has a huge impact on student performance as it makes it easier for students to internalize success criteria and control their

own learning. He defines student self-assessment as “a process by which students 1) monitor and evaluate the quality of their thinking and behaviour when learning and 2) identify strategies that improve their understanding and skills” (MacMillan, 2008, p.40). Self-assessment is when students assess their work in order to improve their performance. Their goal is to identify differences between current and intended performance. Self-assessment can be facilitated for students, thanks to standards-based education, as it sets clear criteria and goals. In addition, the knowledge needed for self-assessment is provided by clear guidelines for measuring performance. Students also identify learning strategies and learning goals. It is essentially a cycle of three components: 1) students define their performance and learning strategies, 2) provide feedback, based on criteria they understand well, and 3) identify the next steps on the way to improve their performance.

Both William (2011) and MacMillan (2008) agree that one of the key components of self-assessment is metacognition a term referring to one’s knowledge and awareness of one’s cognitive process. Although the usefulness of this ‘self-awareness’ depends on another important factor which is motivation, which should be primarily intrinsic. William (2011) argues that motivation is not something students either have or do not. It is not something that a teacher can provide. From a psychological point of view, it can be considered a consequence of achievement.

Besides self-assessment, there is another technique that can result in students taking ownership of their own learning – self-regulated learning. Zimmerman (2002) argues that students who learn using self-regulation make more effort to succeed. This is due to the fact that they are aware of their strengths and weaknesses. They monitor their own behaviour and increase of efficiency. As they observe their own progress, they tend to be more self-satisfied and motivated. This superior motivation, together with adaptive teaching methods, may result in self-regulated students having a higher chance of success in the academic field. Furthermore, they generally perceive their futures with more optimism. He claims that self-regulation is important since acquiring lifelong learning skills is a major function of the educational process.

## **2.2 Summative assessment**

### **2.2.1 Introduction to summative assessment**

Some authors claim that summative assessment originates from the classical test theory, where the term true score embodies the summative value. This theory has roots in behaviourist learning theories from the beginning of the twentieth century.

The process of summarizing the student's understanding of the curriculum and its content with a grade may seem as an easy way to assess students. One of the ways of distinguishing what type of assessment is being used is realizing what is going to be achieved with it. Is the assessment helping with discovering what students had learned? Is the assessment taking place at the end of the unit or after the teaching process has ended? Is it a high-stakes exam?

By answering these questions one can discover what type of assessment is used since they provide us with necessary information about the summative assessment. The first question relates to eliciting evidence which was also mentioned concerning formative assessment, nevertheless, summative assessment uses different approaches, which will be described later in more detail.

The second question presents the fact about summative assessment that is taking place after the learning or teaching process has ended. This strategy can be critiqued because it seems that if the student gets a low score there is nothing that can be done, so this approach is unsuitable for day-to-day teaching and it focuses on the product rather than the process of learning. However, some types of tests, which are supposed to uncover the current level or depth of students' understanding, provide the evaluators with data. Furthermore, sometimes tests even help them compare students or make important decisions, which relates to the last question.

The third question points to the fact that a summative way of assessing is typical for the end of the year exams mandated high-stakes tests or tests that determine promotion/retention, help students gain scholarships, membership in societies, can win school awards, and provide funding. These tests can put pressure not only on students but also on the teachers that have a dual goal to fulfil, to help students achieve a thorough understanding of the standard and to help them score high (Burke, 2010, p.46). Summative assessment aims to measure or summarize, what a student has grasped, and typically occurs

at the end of a course or unit of instruction. A summation of what a student has learned implies looking back and taking stock of how well that student has accomplished objectives, but does not necessarily point the way to future progress” Brown (2004, p.6).

According to Garrison (2013), summative assessment is used to determine students’ knowledge at a particular point in time. She claims that many people associate summative assessment with standardized state-level tests, but this assessment is also important for district and classroom programs. It is used in classes to measure accountability which is a crucial part of the grading process. Although it is important to determine what students know at a given point in time, this type of assessment evaluates only specific aspects. Summative assessment is generally used after a finished segment of the learning process, such as the end of the school year or after finishing a unit in the textbook. It can provide the teacher with the idea of how the curriculum is being fulfilled thus far if students are placed in a suitable programme, and how effective these programmes might be. On the other hand, summative assessment lacks the means of helping teachers intervene effectively or adjust the instructions when they are relevant – during the learning process.

### **2.2.2 The effects of summative assessment**

Harlen, (2002), with her colleagues in the Assessment and Learning Research Synthesis Group conducted a systematic review of the impact of summative assessment and tests on students' motivation for learning. She reviewed 183 potentially relevant studies, eventually, 19 of these were identified as truly relevant to the research question. All of these studies concentrated on children aged 4 to 19. The main findings provided evidence that when preparing for a high-stakes test, teachers adopt a teaching style, disadvantaging pupils, who thrive in active and creative learning conditions. This style is called transmission teaching and it is an approach where the teacher puts themselves in a position of the source of knowledge and truth. Furthermore, the evidence suggests tests can possibly influence a teacher’s assessment in a way that the teacher becomes more focused on performance than the learning process. This can impact the way in which students perceive classroom assessment. They tend to misconstrue the teacher’s intentions and interpret the assessment as purely summative. Moreover, repetition of practice tests has a negative impact on lower-achieving students, because it lowers their self-image.

In relevance to age, the studies showed that students younger than 11 focus more on the learning process than its outcomes in comparison to older students. Younger students generally credit external factors and practice for success while older students believe that success is achieved due to effort and ability. Harlen (2002) assumed that older students would be a more comfortable taking tests, since they should be used to taking tests and are familiar with the process. Surprisingly, the opposite seems to be the truth. Older students experience more anxiety, cynicism, resentment, and mistrust of standardized tests. This results in these students guessing or randomizing the answers more often than younger students. The reason for this is supposedly minimalization of effort. In regards to the level of achievement, Harlen (2002) claims that for lower-achieving students, the negative impact of summative assessment doubles. Their low self-esteem further lowers when they are labeled as a failure. Furthermore, this does not impact only their self-image but also endangers their chances of future success. She adds that one way to avoid these serious negative consequences is a high level of support that comes with suggestions for improvements. This type of support can be based at home or school.

Students' motivation can be influenced by summative assessment directly or indirectly. Direct outcomes may include increased anxiety, how students perceive themselves, and how their self-esteem is affected by low scores. Teachers and the curriculum may be indirectly affected. This era of education concentrates on helping students find the most effective way of learning, the general goal should be to help them learn also in their future lives. The type of motivation that is essential for lifelong learning is called intrinsic. Intrinsic motivation differentiates from extrinsic motivation in the source of motivation. Intrinsic motivation comes from the students, they learn because it brings them satisfaction or they are interested in the topic, and this results in continued learning. This type of learner tends to overcome more obstacles on their way towards an achievement since they realize that it depends on their effort. Extrinsic motivation, on the other hand, initiates from the student's fear of failure. These students may also strive to receive rewards in the shape of grades or praise from the teachers or parents. The student's efforts are usually low and this type of motivation does not ensure the continuation of learning.

Mostly, the content of learning is not important to students, which is not productive. High-weight evidence shows that when evaluation and selectivity play a crucial role in the educational system, it makes the students motivated by grades and social status. It is crucial to try and avoid any negative impact on motivation for learning that can be avoided. It has



been alleged that present generations of students can be negatively influenced by negative outcomes of a rise in test scores, even though they are obviously in most cases unintended (Harlen, 2002, p.4-13). Harlen (2002) states that some studies suggest that intrinsic motivation can be reduced by using an assessment that does not provide students with control. Moreover, the learning seems to be superficial.

Even though feedback is a typical strategy for formative assessment, here it can be used to provide information about the finished assessment. This type of feedback has reportedly a strong influence on further learning. Teachers should be careful when giving this feedback, considering the negative ways that judgemental feedback can impact students' image of their capability and their chances to succeed. One way to prevent these consequences is by giving constructive feedback to learners and explanation of the purpose of the assessment. (Harlen, 2002, p.7, 48)

In regards to gender, according to Harlen's (2002) review, girls are prone to more test anxiety, as they blame themselves for a failure. Whereas boys tend to make more external attributions, which means blaming the class environment, teacher, or different circumstances for their errors. Obviously, the anxiety that both male and female students feel when taking tests may significantly increase if there is a pressure from parents, the community, or teachers. After tests and during the learning process, students regularly evaluate their performance. If the classroom assessment is predominantly formative, they are more likely to judge their learning, not just their performance.

What can be done to earn the benefits of summative assessment without the unwanted negative outcomes, influencing students' motivation? Harlen (2002) recommends based on her thorough research the following. There are many common practices that teachers should avoid. One of them is not letting the contents and form of the tests limit the content of teaching and the methods used in a classroom. Another counterproductive practice is the usage of drill tests or tests used for practicing, especially if the teacher suspects the knowledge of the learners is insufficient to enable them to succeed. She further states that the review revealed recommendations on what to do, and not just what to avoid which had proven to be beneficial in eliminating the undesired influences. For instance, goal-oriented learning has better results than performance-oriented learning. She advises teachers to share the goals with students and put emphasis on learner-centred approaches. Also, the demands of tests, teachers' expectations, and students' capabilities should be uniform. Tests should

take place when the teacher decides that students are ready to perform at a certain level. This precaution can minimize the risk of failure, and lowering their self-esteem. For tracking national standards, she recommends using a broad spectrum of tests. Summative assessment seems to be more useful if it is presented to students as a way of tracking the learning progress. Results could be reached also by creating a safe space in school to discuss assessments, these discussions should happen not just between students and teachers but also among teachers. This leads to another recommendation when discussing the assessment, students should become more involved in testing. Schools should also develop a self-assessment practice, where the teachers' assessment abilities are evaluated.

In practice, Harlen (2002) claims that research points to high-stakes tests (at the time) not being valid. The information they should provide about students' knowledge is biased since tests are narrowly focused. Many errors of measurement are supposedly often not revealed. Not to mention that students do not have to possess the needed skills or knowledge to pass the test if the teacher aimed the lessons toward taking/passing the test and not towards a real in-depth understanding of the topic. She also addresses the above-mentioned random answers and minimal effort typical for students convinced that they are going to fail no matter what and the test anxiety typical for girls and low-achieving students. Therefore, the results of these tests might exaggerate the achievement gap and provide teachers with unreliable data.

### **3 SUGGESTING AN APPROPRIATE FORM OF LANGUAGE ASSESSMENT**

In the previous chapters, the basics of testing language, methods of scoring as well as specific methods of testing language were introduced and discussed. Definitions of summative and formative assessment and studies conducted by researchers relating to these two types of assessment were also reviewed. An effective way of testing language skills will be suggested in this chapter. This suggestion will be based on the authors' claims and research findings. Student's language skills can be assessed in a number of ways. First, they have to be tested. Ideally using a specific way to gain comprehensive knowledge of the abilities, knowledge, and skill involved in producing language. Bear in mind that students will produce evidence and their knowledge can spread beyond that evidence, which is what is the aim of assessment essentially. Discover as much as possible of what students know and are able to do. Language skills intertwine. When producing or analysing language two or more at a time are generally used. When assessing language skills teachers should balance the formative and summative assessment in order to test in the most effective way. The following chapters will suggest testing and assessing approaches for each skill.

#### **3.1.1 Assessing speaking**

Speaking can be very stressful and anxiety inducing. Lowering this anxiety should be a priority when deciding how to test this language skill. As mentioned in the first chapter, there are two types of speeches – planned and unplanned. An effective way of examining a planned speech could be a structured interview. An interview as a basic concept was described, but interviews can also be structured, students knowing the questions they will be asked could be effective. The main reason is that some of the stress and anxiety might be reduced. If we aim to measure their public speaking skills this may not be the most effective way.

Another reason could be enabling the students to display the most advanced speech they can. Due to having time to plan their answers they may use more complex sentences. They could even feel more comfortable using newer vocabulary that is not really embedded in their lexicon yet. Therefore, students would probably not attempt to use the new word under pressure. They could mispronounce it or use it incorrectly. This could be prevented by

letting students plan their answers or speeches. What is a more effective way of learning a language than using it?

Nevertheless, some might argue that skill is not fully mastered unless used automatically and without effort. Therefore, students should be tested based on how effortlessly they use the skill and that will provide information about the extent to which they have mastered the said skill. This might be true for some people. However, some skills take a very long time to master. For example, many people have gotten a driver's license after taking their exam successfully. They have been tested and assessed and then received their license. This is not evidence that they have mastered driving a car. Nonetheless, it is safe to assume that more people have died due to insufficient mastery of driving than due to mispronouncing or misusing a word. And if people who have not mastered a potentially life-threatening skill can be assessed based on it, why should language be any different?

Regarding language, we should assess students based mostly on their progress and effort. Why not let them perform with the best material (ability, skill, knowledge) they have. However, the argument that they will not be able to prepare their speech in a real-life conversation might be a more convincing one. This decision is on the teacher who should gather evidence throughout a sufficient amount of time to be able to decide if their students are ready to be assessed. If they are ready they can be tested on their speaking skills also in a way where they need to think fast and respond. In the Czech Republic, some teachers tend to grade every little thing. Some may do it because they believe it increases motivation. However, the research suggests otherwise.

Perhaps there is no need for grades as we know them. Teachers assess their students all time even unintentionally. Maybe high-stakes assessment could be done once in a school year. That could be enough since the teacher would gather the evidence throughout the year and would give students feedback on what to work on more. Which is partially a formative assessment and at the end of the year the teacher would use the knowledge about the student's performance, progress, and effort to make a final summative assessment.

### **3.1.2 Assessing writing**

Writing, is an ability that can be seen as a privilege by some people still to this day. Writing is crucial for a modern-day person. Communication, relationships, finding a job and many more crucial things rely on one's ability to write. Writing is usually used in combination with reading, the two skills could not exist separately. That should be

considered when assessing writing. An effective way of testing writing abilities may be assigning text for students to write. It could be any type of text, for less advanced learners a shorter one (a paragraph, a short story, an article), and for advanced – a longer one (an essay, a literature review, detailed description). Most importantly, the teacher should decide what is to be assessed.

Writing is closely related to reading as mentioned. Moreover, the ability to write is composed of other individual skills. Sometimes the aim of the assessment is creativity, grammatical or syntactic correctness, reaching the assigned word count, meeting a deadline, staying on a topic or even the arrangement or appearance of the text. These criteria should be communicated to students. As to enable them to give their finest performance. They should then be assessed solely on these criteria. For instance, assigning a book review that should be submitted in two weeks. The students who are slow readers may struggle much more or even fail to meet the deadline. However, that does not provide the teacher with any evidence of their ability to write. For example, if the aim is to test the creativity in writing, then the time limit should be longer than when the testing ability to stick to a given topic. The previously mentioned assigned text also does not have to be graded. Teachers can assess it using peer assessment, and self-assessment in a formative way. That is to say, if a large portion of students has trouble staying on the topic, give examples of how to achieve that.

Portfolios concentrated on writing seem to be the most effective way to assess writing, all the texts and written assignments in one place. The students themselves can see their progress their strengths and weaknesses and how they developed in these areas. However, basing the final summative assessment purely on their effort may not seem fair. For instance, a student A starts their school year with writing abilities corresponding to a 1 on a grading scale and finishes their year with skills corresponding to 2. Student B starts with abilities graded as 4 and finishes with writing worth 2. Who did really learn and performed to deserve 2? This problem of fairness is difficult to solve. That is why the final grading should be accompanied by some sort of checklist or comments from the teacher.

Students can then be appreciated for their efforts or notified of their lack of efforts or shortcomings. That may motivate them or help them to focus on increasing their efforts or finding effective ways of decreasing the negative impact of their drawbacks.

### **3.1.3 Assessing listening**

Listening and comprehension are crucial parts of any conversation. Teachers can let students answer questions based on a recording or let them summarize what they heard. Again, the skills relate to one another. If a teacher decides to provide a written test with questions (open-ended or multiple choice) they are also assessing students' reading skills. If the teacher requires spoken answers or a summary they assess their speaking skills. Once again setting the criteria prior to the exam is crucial.

An effective way could be sufficient preparation for the listening exam. Explaining any confusion about the questions, reading them aloud to the students, and making sure that they understand what is required from them could be effective. The aim is to move learning forward. That can be done by informative non-judgemental feedback from the teacher, but also from the other students. Students may be put into groups and brainstorm effective ways of achieving the best possible results in listening exams. The teacher may facilitate their discussions, choose ways that may actually be effective (taking notes, concentrating on the questions etc.), write them on the board, and discuss them with the whole class. If the results of listening exams are still not satisfactory, the students can always ask the teacher to explain them again or practice them more.

This can help focus on the actual weaknesses that students have and not waste time explaining things they already know and are able to do. By using this approach teacher can make sure that not only, the students know the criteria of succeeding in an exam, but also how to fulfil them. Teachers should provide detailed instructions and rubrics (or a different set of criteria). If these criteria are discussed the students may assess themselves in the process of fulfilling the tasks. Assuming the students understand the criteria and know an effective way of achieving them, nothing prevents them from success. Obviously, there will be students who will never be able to give such a performance to receive 1. There is nothing wrong with that. The grade 3 is considered average and as commonly known, the average is where most of the people land.

Students in the class will most probably be split into high-achievers, lower-achievers, and "middle-achievers". The last group is the largest. The important thing is not making students feel as if there is something wrong with being average. Teachers should motivate students for future learning. As mentioned in previous chapters, grades influence motivation

and self-esteem rather strongly. By creating a non-judgmental, inspiring, and calm environment in a classroom the negative impacts may be reduced.

### **3.1.4 Assessing reading**

In this era of globalization, reading in English can be very useful; some learning materials or literary sources are not translated into Czech. There is usually a greater number of them in the English language. Therefore, reading in English can help students beyond the subject of the English language. Students can be assessed on many aspects of reading: reading aloud, comprehension, or reading speed. The issue is that every written test will in some way depend on reading abilities. Practicing it is therefore crucial.

Students with disabilities may struggle very much even if the test is aimed at their knowledge of the studied topic. An effective way of assessing the reading skill may be having students read a text (or a book in the case of advanced students), in a sufficient amount of time.

Then again, speaking or writing skills will be necessary for eliciting evidence of their reading skills. They could write a review or answer questions about the text. There may also be a criterion based on how the students interact with the text. If they can find answers in the text or utilize their short-term memory when recollecting what they have read. Assessment should be based on various aspects of reading.

Revisiting the theme of rubrics and pre-disclosed criteria. Teachers could assess individual skills that together create the ability to read. This may be done by assessing each skill individually or basing an assessment on the overall performance. An effective approach might be starting with assessing every criterion individually and then calculating an average. However, this approach should progress with time. When students are familiar with this assessing technique the final assessment could be based on the inappropriately executed criterion. This way may help students achieve success in other aspects of their education and their future life. It forces them to pay attention to every part that creates a whole.

In life, people tend to judge things based on every part of them. For instance, if a car was made, with beautiful paint and varnish, it had the fully functioning engine and rode smoothly, but the steering wheel was missing, the car would not be drivable. Every student, every person has their strengths and weaknesses. However, in education, every part of the

learning process is important. This evaluation of attention to crucial parts of fulfilling the task successfully may be effective for assessment.



## CONCLUSION

This thesis concentrated on comparing two types of assessment – formative and summative assessment. The aim was to suggest an effective way of testing and assessing language skills based on this comparison. The beginning of this thesis included studying and comprehension of information relating to the topic. Followed by processing this knowledge and organizing it into a logical order. This helped to analyse both of these assessments and investigate their benefits and their drawbacks.

Regarding formative assessment, the most noticeable disadvantage seems to be its lack of ability to inform educators about fulfilling standards. It could be difficult to compare schools or students solely on formative assessment. Additionally, the Czech educational system is based on grades, for example, high-school or college admissions. The advantage that appears to be worth mentioning is the potential of formulating the educational process. This formulation is based on the students themselves and their concrete needs, which could prove to be essential in their future learning.

The disadvantage of summative assessment may be some of the negative effects it has on students. The research suggested that summative assessment can have a negative influence on students' self-esteem and motivation. However, one of the most prevalent advantages of this assessment appears to be its practicality. Moreover, summative assessment is perceived as less time-consuming and allows teachers to assess a large number of students on one occasion.

The findings showed that an effective way of assessment is finding a balance between formative and summative assessment when assessing language. This thesis suggests that language skills are intertwined and all the suggested approaches for individual skills could be used for assessing any of them. Whether using summative assessment and implementing it formatively or assessing the results of students' learning formatively. Even though the learning process is finished and lowering the negative impacts mentioned previously.

The chapters provided an overview of basic principles of language testing, and descriptions of both types of assessment. In addition to research that has been done in the past relating to them and their effects on learners. The last part of this thesis offers examples and suggestions for effective methods of eliciting evidence of students' learning. In addition to the recommendation of approaches to assessing these methods.

## BIBLIOGRAPHY

- BACHMAN, Lyle F. a Adrian S. PALMER, 1996. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press. Oxford applied linguistics. ISBN 0194371484.
- BLACK, Paul a Dylan WILLIAM, 2010. Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan* [online]. Phi Delta Kappan International, 2010, **80**(2) [cit. 2022-04-14]. ISSN 0031-7217. Dostupné z: doi:10.1177/003172171009200119
- BROWN, Douglas H., 2004. *Language Assessment: Principles and Classroom Practices*. White Plains: Pearson Education. ISBN 0-13-098834-0.
- BROWN, James Dean, 1996. *Testing in Language Programs*. New Jersey: Prentice Hall. ISBN 0131241575.
- BUDDEN, Jo, 2004. Role-play. In: *Teaching English* [online]. British Council, 2004 [cit. 2022-04-13]. Dostupné z: <https://www.teachingenglish.org.uk/article/role-play>
- BURKE, Kay, 2010. *Balanced Assessment: From Formative to Summative*. Bloomington: Solution Tree Press. ISBN 9781934009529.
- CCSSO, 2022. Revising the Definition of Formative Assessment. In: *CCSSO: The Council of Chief State School Officers* [online]. Washington, DC: CCSSO, April 19 2022 [cit. 2022-04-19]. Dostupné z: <https://ccsso.org/sites/default/files/2018-06/Revising%20the%20Definition%20of%20Formative%20Assessment.pdf>
- DUIGU, Gabi, 2002. *Essay Writing for English Tests*. Cammeray: Academic English Press. ISBN 0-9578996-1.
- FLORES, Maria Assunção, Ana Margarida Veiga SIMÃO, Alexandra BARROS a Diana PEREIRA, 2014. Perceptions of effectiveness, fairness and feedback of assessment methods: a study in higher education. *Studies in Higher Education* [online]. Taylor & Francis, 8.4.2014, **40**(9) [cit. 2022-04-13]. ISSN 1470-174X. Dostupné z: doi:10.1080/03075079.2014.881348
- FULCHER, Glenn, 2010. *Practical language testing*. London: Hodder Education. ISBN 978-0-340-98448-2.

GARRISON, Catherine a Michael EHRINGHAUS, 2013. Formative and Summative Assessments in the Classroom. In: *AMLE* [online]. Columbus: AMLE [cit. 2022-04-08]. Dostupné z: <https://www.amle.org/formative-and-summative-assessments-in-the-classroom/>

HARLEN, Wynne, 2002. A Systematic Review of the Impact of Summative Assessment and Tests on Students' Motivation for Learning. In: *EPPI-Centre* [online]. London: EPPI-Centre, 30.5.2002 [cit. 2022-04-08]. Dostupné z: <http://eppi.ioe.ac.uk/cms/LinkClick.aspx?fileticket=Pbyl1CdsDJU%3d&tabid=108&portalid=0&mid=1003>

HENDERSON, Michael, Michael PHILLIPS, Tracii RYAN, David BOUD, Phillip DAWSON, Elizabeth MOLLOY a Paige MAHONEY, 2019. Conditions that enable effective feedback. *Higher Education Research & Development* [online]. Taylor & Francis, 9.10.2019, **38**(7) [cit. 2022-04-05]. ISSN 1469-8366. Dostupné z: doi:10.1080/07294360.2019.1657807

HENNING, Grant. *A Guide to Language Testing: Development, Evaluation, Research*. Michigan: Heinle & Heinle Publishers, 1987. ISBN 9780838426937.

JOUGHIN, Gordon, 2010. A Short Guide to Oral Assessment [online]. Leeds: Leeds Met Press [cit. 2022-04-13]. ISBN 978-1-907240-09-6. Dostupné z: <https://www.qub.ac.uk/directorates/AcademicStudentAffairs/CentreforEducationalDevelopment/FilestoreDONOTDELETE/Filetoupload,213702,en.pdf>

LITTLETON, Karen, Neil MERCER, Lyn DAWES, Rupert WEGERIF, Denise ROWE a Claire SAMS, 2005. Talking and thinking together at Key Stage 1. *Early Years: An International Journal of Research and Development* [online]. Taylor & Francis, **25**(2), 167-182 [cit. 2022-03-03]. ISSN 0957-5146. Dostupné z: doi:10.1080/09575140500128129

M. BAILEY, Kathleen et al., H. LONG, Michael a Catherine J. DOUGHTY, ed., 2009. *The Handbook of Language Teaching*. New Jersey: Blackwell Publishing. ISBN 9781405154895.

MCMILLAN, James H. a Jessica HEARN, 2008. Student Self-Assessment: The Key to Stronger Student Motivation and Higher Achievement. *Educational Horizons* [online]. Phi Delta Kappa International, 2008, **87**(1), 40-49 [cit. 2022-04-08]. ISSN 0013175X. Dostupné z: <https://files.eric.ed.gov/fulltext/EJ815370.pdf>

MOSS, Pamela A., 2003. Reconceptualizing Validity for Classroom Assessment. *Educational Measurement: Issues and Practice* [online]. Blackwell Publishing, 25 October 2005, **22**(4), 13-25 [cit. 2022-04-14]. ISSN 1745-3992. Dostupné z: doi:10.1111/j.1745-3992.2003.tb00140

SADKER, David M. a Karel ZITTLEMAN, 2008. *Teachers, Schools, and Society: A Brief Introduction to Education*. 2nd. New York: McGraw Hill Higher Education. ISBN 9780077287535.

SALVIN, Robert E., 1980. Cooperative Learning. *Review of Educational Research* [online]. SAGE Publications, **50**(2), 315-342 [cit. 2022-04-06]. ISSN 0034-6543. Dostupné z: doi:10.3102/00346543050002315

TORRANCE, Harry a John PRYOR, 1998. *Investigating formative assessment: teaching, learning and assessment in the classroom*. Philadelphia: Open University Press. ISBN 0335197345.

WEIGLE, Sara, ALDERSON, J. Charles a Lyle F. BACHMAN, ed., 2002. *Assessing Writing*. Cambridge: Cambridge University Press. ISBN 978-0-521-78446-7.

WILIAM, Dylan. *Embedded formative assessment*. Bloomington: Solution Tree Press, 2011. ISBN 978-1-934009-30-7.

ZIMMERMAN, Barry J., 2002. Becoming a Self-Regulated Learner: An Overview. *Theory Into Practice* [online]. Taylor & Francis, červen 2002, **41**(2), 64-70 [cit. 2022-04-08]. ISSN 1543-0421. Dostupné z: doi:10.1207/s15430421tip4102\_2

## **LIST OF TABLES**

Table 1 Examples of confused and clarified learning intentions (Wiliam, 2011, p.61) .....	23
Table 2 Conditions of feedback (Henderson et al., 2019, p.115) .....	26

## ANOTACE

Jméno a příjmení:	Nela Neduchalová
Katedra:	Ústav cizích jazyků
Vedoucí práce:	Dr hab. Konraz Szcześniak
Rok obhajoby:	2022

Název práce:	Porovnání formativního a sumativního hodnocení v testování jazyků
Název v angličtině:	Comparison of formative and summative assessment in language testing
Anotace práce:	Tato bakalářská práce se zaměřuje na formativní a sumativní hodnocení. Cílem této bakalářské práce je cíl prozkoumat výhody a nevýhody těchto typů hodnocení a poskytnout přehled příkladů technik hodnocení. První část je zaměřena na objasnění pojmů a definic. Druhá část popisuje efekty a strategie hodnocení. V poslední části jsou uvedeny návrhy efektivních způsobů hodnocení jednotlivých jazykových dovedností.
Klíčová slova	formativní hodnocení, sumativní hodnocení, hodnocení, testovací metody, známkování, motivace, vzdělávání, jazykové dovednosti, proces učení
Anotace v angličtině	This thesis focuses on formative and summative evaluation. The thesis explores the advantages and disadvantages of these types of evaluation. The thesis aims to investigate the advantages and disadvantages of these types of assessment and provide an overview of examples of assessment techniques. The first part focuses on clarifying terms and definitions. The second part describes the effects and strategies of evaluation. In the last part, suggestions are given for effective ways of evaluating individual language skills.

Klíčová slova v angličtině:	formative assessment, summative assessment, evaluation, testing methods, grading, motivation, education, language skills, the learning process
Rozsah práce:	47
Jazyk práce:	Angličtina