Univerzita Karlova

Filozofická fakulta

Ústav anglického jazyka a didaktiky

Diplomová práce

Ivana Dodecká

Grading in English Lessons at Secondary Schools

Klasifikace ve výuce angličtiny na středních školách

Praha 2021   PhDr. Tomáš Gráf, Ph.D.

Prohlášení

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze dne 20.7. 2021   Ivana Dodecká

Abstrakt

Cieľom práce bolo zistiť, akým spôsobom učitelia angličtiny na stredných školách klasifikujú prácu svojich žiakov a určiť, do akej miery učitelia postupujú systematicky a aké sú v týchto systémoch slabé miesta. Na zber dát bol použitý dotazník rozoslaný na rôzne typy stredných škôl (gymnáziá, SOŠ, SOU). U respondentov sa zisťoval predovšetkým spôsob hodnotenia, tj. napr. aké kritéria pri hodnotení používajú, čo hodnotia, akými princípmi sa riadia, či majú na školách smernice známkovania a do akej miery sú záväzné, ako boli v tomto ohľade inštruovaní pri štúdiu na vysokej škole a pod.

Výsledky výskumu poukázali na nesystémovosť klasifikácie na hodinách angličtiny, čo môže mať za následok nevyváženosť v hodnotení žiakov naprieč systémom stredných škôl. Táto práca sa preto v závere venuje prípadnej náprave a navrhuje možné zlepšenia.

Klíčová slova: známkování, hodnocení, klasifikace, dovednosti anglického jazyka, střední školy, učitelé angličtiny

Abstract

The aim of this thesis is to find out how English teachers at secondary schools grade students' performance, to what extent teachers grade systematically and what the possible weaknesses in these systems are. A questionnaire sent to various types of secondary schools ('gymnázium', SOŠ, SOU) was used for data collection. Respondents were evaluated based on their grading approach, with questions covering, among others, the following areas: use of grading criteria and type of activities graded, grading principles followed, availability of grading guidelines within schools and the extent to which these were binding, teacher instruction on grading at university, and so on.

The results of the research pointed to a non-systematic grading approach in English lessons which may result in inconsistency in student performance assessment across the secondary school system. In its conclusion, the thesis identified critical areas of the non-systematic grading approach and sought possible improvements.

Keywords: grading, assessment, quantitative assessment, English languages skills, secondary schools, English teachers

Contents

List of Abbreviations

CERMAT	Centrum pro zjišťování výsledků vzdělávání (Center for Education Results)
EFL	English as a Foreign Language
ESL	English as a Second Language
MŠMT ČR	Ministerstvo školství, mládeže a tělovýchovy České republiky (Ministry of Education, Youth and Sport of the Czech Republic)
SAT	Scholastic Aptitude Test
SOŠ	Střední odborná škola
SOU	Střední odborné učiliště
TOEFL	Test of English as a Foreign Language

List of Figures

List of Tables

List of Charts

1 Introduction

Our society seems to rely on judgement fairly often. We assess and are assessed when applying for a job, when we apply for a mortgage, when we vote. With each and every decision we assess and judge our choices, we determine the scale for the best. Perhaps alongside occasional bias, we come across minor but also high-stakes tests that prove and refute we might be good enough. From an early age in the role of students, we tend to face such challenges as assessment appears to be a standard feature in our education, too.

Yet, there might be a certain degree of ambiguity when it comes to the perception of assessment in the classroom. While some believe assessing students could have a motivational character (Košťálová et al., 2008: 45), others claim assessment can foster anxiety and self-doubt in students (Brown, 2003: 1). There are voices that feel that to assess and be assessed is a skill that 'one can but doesn't have to manage well' (Slavík, 1999: 22). Be it as it may, classroom assessment seems like part and parcel of the learning experience and experts seem to agree that both society and teachers are unable to forgo it (Petty, 2009: 479).

I suppose such claims are what shape my own motivation for this thesis. As an English teacher myself, should I ever become involved in secondary education, I would no doubt come in contact with the grading system and assessment along with its issues. Yet, what might be didactically compelling is that, by all accounts, no prior research has been done into the area of grading English from the point of view of Czech teachers (see *2.10.8 Grading of the English Language in Czech Secondary Schools*). There appears to be a gap in the field which would examine such grading and its systematicity. What this creates then is a need to explore this area which has not been covered until now and ascertain whether there is a systematic approach in grading at Czech schools. This thesis will attempt to do just that.

In the following thesis we will therefore look closely at different aspects of assessment and what they mean for the English language classroom. We will especially consider the grading part of assessment and examine to what extent there are shared tendencies and possible bias in the grading system at Czech secondary schools. By evaluating similarities and differences in the teachers' approach, we will try to find out whether there might be a particular way for us, teachers, to 'manage well', to assess fairly and effectively.

2 Classroom Assessment

2.1 Defining Classroom Assessment

Based on Cheng and Fox, assessment refers to 'all those activities undertaken by teachers, and by their students in assessing themselves' (Black and William, 1998: 2 [in Cheng and Fox, 2017]). Studying assessment in English classes specifically, Troudi et al. (2009: 546) define it as a 'process of collecting information about a student to aid in decision making about the progress and language development of the student.' Measuring the breadth and depth of learning, assessment becomes then an umbrella term. It can range from daily classroom assessment practices to large-scale testing designed externally and administered to all students on one level (Petty, 2009: 479; Cheng and Fox, 2017: 3-4). It is thus multi-dimensional and complex, yet a good teacher never ceases to assess his or her students, no matter how incidental or intended the assessment would be (Cheng and Fox, 2017: 1-2; Brown, 2003: 4).

As a skill, assessment in the classroom can be intellectually very demanding and seems to involve distinguishing important and unimportant phenomena in the world around us. Within the important phenomena, further differentiation between 'the better and the worse' can occur (Slavík, 1999: 15). It can be 'an expression of our positive or negative attitude towards different aspects of student's work and achievement that can take on different forms' (Skalková [in Ježová, 2009: 11]). These forms can be as diverse as teacher's head nodding, stern looks, verbal comments, or grading (Ibid.) Such patterns can become reliable because they happen repeatedly, directly affecting student's learning (Pasch, 1998: 104). Assessment can be determined then simply as 'a systematic process that leads to evaluating a student's or a group of students' qualities and performance' (Kalhous et al., 2009: 404). This systematicity of assessment should reflect a teacher's own methodical behaviour when preparing, organizing, facilitating students' assessment (Pasch, 1998: 104). There is perhaps a certain regularity to these actions and there is often the support of norms, too. Thus, standardization and regularity together can help systematize assessment (Kolář and Šikulová, 2009: 18).

Yet, it appears the process relies on individuality made up of each and every teacher. Slavík supports this view and alleges that assessing, overall, is personal and therefore could often be prejudiced (Slavík [in Hozmanová, 2015: 5]). Kolář and Šikulová concur. They, too, believe

that assessment is highly subjective and add that such bias is ever-present (Kolář and Šikulová, 2005: 12-3). People in general tend to assess almost constantly, as Ježová puts it 'sometimes too often' without even realizing their own preconceptions (Ježová, 2009: 11). Still, Cheng and Fox argue that day-to-day assessment of students' progress 'is unquestionably one of the teacher's most important, complex and demanding tasks'. Teachers are seen as an essential part of the assessment process in the classroom, or as Cheng and Fox put it, they are the 'principal agents of assessment'. It is then their job to make sure the standards and objectivity of the assessing process are met because a good quality assessment practice can be conducive to a student's learning (Cheng and Fox, 2017: 1).

2.2 Objectivity in Classroom Assessment

In her thesis, Šťastná seems convinced that classroom assessment, while 'not easy', tends to be always subjective. She argues that a teacher's personality consistently affects the process of assessment (Šťastná, 2014: 44). Slavík would go even further and suggest that intertwined social relations have a profound influence on assessment. The author implies that a teacher's short- or long-term approval or dislike either towards the students or even their parents would impact teacher assessment of the student's performance (Ibid.; Slavík, 1999: 66-7). Among factors that could influence the teacher's decision, Kalhous et al. then, rather shockingly, mention 'the student's appearance, his behaviour, performance in other subjects, student's personality but even the parents' social status, nepotism etc.' (Kalhous et al., 2009: 213-4). Ježová lends weight to this. She suggests that not only could classroom assessment help develop social relations between teacher and students, but the quality of such relationships can directly influence how the student would be assessed. She urges teachers not to succumb to personal bias and to keep in mind that actually, assessment could potentially improve the mutual relationship with the assessed student (Ježová, 2009: 17). On the other hand, while accepting there is a certain amount of partiality found in classroom assessment, J. Slavík firmly believes it would be unfair to blame teachers for it. He considers any prejudice coming from teachers subconscious and most certainly unintentional. Slavík sees teacher bias as non-deliberate and in most cases teachers themselves completely unaware of it (Slavík, 1999: 67).

At this point one might wonder: taking into account the apparently constant level of subjectivity, can assessment ever be without prejudice? Kunnan in his *Evaluating Language*

Assessments demonstrates that it is absence of bias that is so critical to the assessment process (Kunnan, 2017: 193).

Even though they have the right to do so, Slavík notices that students hardly ever make their attitude known about their assessment results, whether the results appear fair or unfair. Interestingly, when teachers determine the final grade based on term grade average, students often equate this with fair assessment (Slavík; Kolář and Vališová [both in Šťastná, 2014: 44-5]).

Yet still the question remains, can assessment ever be 'fair'? And what is it exactly that 'fairness' might involve? If relying on McNamara and Ryan for a definition (2011: 161-178), then the concept of fairness can be found to closely relate to the technical quality of assessment, such as a test. In its 'zeal for fairness' the traditional approach is to aim for a standardized approach to assessment (Karami and Mok, 2013: 102). Interestingly, Troudi et al. in their study (2009: 546), contrary to the previous researchers, came to the conclusion that teachers do not actually play such a major role in assessment at all. The authors determined this was due to the top-down managerial approach to education but also and especially, due to the great validity of the assessment process.

2.3 Validity of Assessment

Validity and fairness appear 'tightly interwoven', indeed (Karami and Mok, 2013: 101). G. Camilli states clearly: (test) validation is an essential element in the analysis of fairness (Camilli, 2013: 104-120). A valid test tests what it is designed or intended for (Ur, 2004: 44), it simply tests what it is supposed to (Harmer, 2004: 322) and it yields accurate, useful and meaningful results (Brown, 2003: 22).

> The validity goal in classroom assessment is to meaningfully and accurately interpret assessment information (e.g., grades, scores, teachers' oral and written comments, student observations of errors, and students' recognitions about their learning – those remarkable eureka moments when students become aware of and/or acknowledge their learning)
>
> (Cheng and Fox, 2017: 65)

Pasch (1998: 109) stresses that teachers should never test their students on something that has not been previously discussed, negotiated or promised in class. For example, if we want to

test students' writing skills, the essay question should not require specialist knowledge, unless we know all students share this knowledge. If the essay, however, asked for such knowledge, the test would be regarded as invalid (Harmer, 2004: 322). Questions should also not be difficult for the students to understand or be culturally biased or validity will be compromised (Petty, 2004: 471).

Harmer finally brings up the concept of so-called 'face value'. He asserts that any form of assessment should not only attempt to be valid but also should try to look valid. Face validity is, Brown acknowledges, only in the eye of the beholder, the test-taker. This type of validity only considers the extent to which students themselves view assessment as 'fair, relevant, and useful for improving learning' (Brown, 2003: 26). Harmer further explains that, for instance, a test consisting of only three multiple choices would 'not convince students of its face validity however reliable […] teachers thought it to be' (Harmer, 2004: 322).

2.4 Reliability of Assessment

Hand in hand with validity then, comes reliability of assessment. P. Ur determines that a reliable test is 'one that produces consistent results when administered on different occasions' (Ur, 2004: 44). If we, for example, had two groups of students that were demonstrably similar taking the same test, we should expect the same marking range (Harmer, 2004: 322). Reliability can be undermined by various factors. These can relate to students themselves (due to psychological or physical causes such as anxiety, illness or tiredness), to the tests (test is too long, time limit too short, instructions unclear) or to test administration (Brown, 2003: 21-22). Finally, the teacher alone can contribute to unreliability of assessment and this consideration will become our main focus in this thesis.

2.4.1 Reliability of Teacher Assessment

'You will never amount to very much.'

- comment made by a Munich schoolmaster to a ten-year-old pupil called Albert Einstein

(Petty, 2004: 472)

According to Petty, reliability means that 'different examiners award the same mark to the same script, and each year's paper [awards] the same grade to a student of a given standard' (Petty, 2004: 471). Brown talks about so-called inter-rater reliability when two or more

assessors mark the same test consistently. Brown suggests this could be down to a number of factors. He quotes 'lack of attention to scoring criteria, inexperience, inattention or even preconceived biases' among some of these factors. Intra-rater reliability then, according to Brown, is common among classroom teachers due to 'unclear scoring criteria, fatigue, bias toward particular 'good' and 'bad' students, or simple carelessness' (Brown, 2003: 21). The author himself admits to partiality especially when overloaded with marking students' tests. What he suggests as a solution is to look at around half the tests before going back to all the tests and only then scoring them and awarding grades. This might perhaps make assessment more even or balanced. Though very difficult to achieve, rater reliability can be increased by the use of analytic scoring, Brown concludes (Ibid.).

Despite it being a challenge, teachers should award the same mark if they assess an identical performance twice on two different days for assessment rendered reliable. Yet Petty claims that in the 1960s schools were regularly retesting students in the same subjects but under different examiners and were always getting 'totally different results' (Petty, 2004, 471). To expose how tricky establishing reliability in assessment can be, Petty also points out the results of a research study from the 1950s. Based on this study around 10% of students, who sat for the same exam over a period of several days, 'passed' on one occasion but 'failed' on the other and vice versa. Petty does conclude that test questions might showcase more reliability than for instance, essay questions. Nevertheless, he admits that especially in practice, perfect reliability can be near impossible to achieve. That is why he recommends 'carefully designed marking schemes' where assessment is based on objective criteria rather than relying on the teacher's general impressions (Ibid.). Since the role of teacher can possibly become a reliability variable, Harmer recommends avoiding situations where the result of assessment depends largely on the teacher. Enhancing reliability in practice is then of paramount importance for Harmer. He further suggests teachers make assessment instructions 'absolutely clear, restricting the scope for variety in the answers, and making sure that test conditions remain constant' (Harmer 2004: 322).

Although assessment can be 'inaccurate and unreliable' requiring a high standard of validity and reliability, it certainly appears assessment is fundamental in our educational system (Petty, 2009: 479). It seems to be a crucial part of both learning and teaching. So much so that we might question the different ways in which teachers tend to assess. We might also then wonder whether or not all teaching involves assessment. Is this process indeed constant or can there also be some interaction that is assessment-free? (Brown, 2003: 4)

2.5 Approaches to Assessment

At this point, it might be vital to stress that assessment of the English language can be challenging at times. Šťastná (2014: 32), for example, asserts that there are stark contrasts in performance of different students. This can be due to a variety of reasons such as the subjective perception of the difficulty of the language, lack of support system, lack of understanding of the topic discussed or even due to finding the lessons boring (Bilanová et al., 2010: 17-8). It is then this variation in performance and skills that reflects how varied the teacher's assessment of their students can be. Šťastná consequently highlights the importance having a universal system in place which could help teachers assess performance. She, however, admits establishing such a system can be quite demanding and riddled with flaws (Šťastná, 2014: 36). To showcase this difficulty, authors Bilanová et al. offer an example when assessing a written test: if we have 10 multiple choice questions, 10 gap fills and 10 sentences to translate, we could give 1 point for any correct answer in the first task, 2 points for any correct answer in the second exercise (since the student must come up with their own answer) and finally 3 points for each fully correctly translated sentence in the last task. At first sight this point system might seem fair, the authors, however, point out that 50% of the points test translation skills only. This appears to prevent more rounded performance assessment which might favour some and disadvantage other students (Bilanová et al., 2010: 41-42).

The authors then argue that even under redistributing the points equally (1 point for any correct answer in any of the three tasks), further ambivalence might still emerge. Questions arise on how to assess potential mistakes in the last task, the translation exercise. Should a spelling mistake be equated to a grammar mistake? Should some of these errors be penalized with possibly half a point or rather no point at all? This confusion emphasizes how impossible it would be to assess by giving 'a quick glance' or making 'subjective decisions on a whim' (Ibid.). What the authors highlight here is how significant it is for the assessor to know exactly what they are trying to assess, what form of assessment tests the performance best and what criteria there can be for such assessment. In the end, however challenging it might be, it is a system put firmly in place that appears essential in assessment process (Ibid.).

The fact that assessment should be systematic rather than impulsive could be further confirmed by an experiment that was led by Cyril Weir. In 1993 Weir conducted an

investigation into 'impressionistic' assessment. For the study, Weir asked his postgraduate students to mark eight exam scripts out of a total 20 points using their first impression, or so-called 'impressionistic' marking. The results seemed quite shocking. What Weir found was an excessive range of assessment for many of the marked papers. While one assessor marked a script with five points, another marked the same paper with a score of 20. There were also scripts whose scale of assessment varied from one to 15. Weir found this particularly alarming. He declared: 'the worst scripts … if they had been marked by certain markers, might have been given higher marks than the best scripts!' (Harmer, 2004: 328)

To help resolve subjectivity of the assessor, Harmer proposes five solutions. First, he suggests that scorers be trained in assessing performance. This training should be done by scorers analyzing different examples of students' performance at various levels and discussing together their marking: 'If scorers are allowed to watch and discuss videoed oral tests, they can be trained to rate the samples of spoken English accurately and consistently in terms of the pre-defined descriptions of performance' (Harmer, 2004: 329).

Second, Harmer believes that having more than one assessor can enhance reliability to a great degree. 'The more people who look at a script, the greater the chance that its true worth will be located somewhere between the various scores it is given' (Ibid.). At this point, Harmer also mentions the role of moderators that exist within certain public examination boards. Their role is to check sample scorings to make sure the assessment conforms with the general standards of the exam (Ibid.).

Next, employing pre-defined descriptions of performance or so-called global assessment scales is advised. This means teachers could rely in their assessment on direct guidelines explicitly describing what students should be capable of doing to obtain a certain mark (see *Figure 1*). Relying on global assessment scales as a teacher can, however, potentially result in a fair amount of inconsistencies. The problem is that the scale descriptions might not actually accurately match students' capabilities in certain cases. If, for instance, the student's performance in terms of his/her pronunciation was on the poor side yet he or she was correct grammatically, the global scale would not match the student's skills. In that case, the scale could not provide adequate assessment support for teachers. Another problem that might arise is that various teachers might actually disagree on how to interpret the scale descriptors. For that reason, global scale would not perhaps work well on its own and would have to be implemented alongside other Harmer's solutions (Harmer, 2004: 329-330).

Score	Description
0	The candidate is almost unintelligible, uses words wrongly, and shows no sign of any grammatical understanding.
I	The candidate is able to transmit only very basic ideas using individual words rather than phrases or fuller patterns of discourse. Speech is very hesitant and the pronunciation makes intelligibility difficult.
2	The candidate transmits basic ideas in a fairly stilted way. Pronunciation is sometimes problematic and there are examples of grammatical and lexical misuse and gaps which impede communication on occasions.
3	The candidate transmits ideas moderately clearly. Speech is somewhat hesitant and there are frequent lapses in grammar and vocabulary use. Nevertheless, the candidate makes him/herself understood.
4	The candidate speaks fairly fluently, showing an ability to communicate ideas with not too much trouble. There are some problems of grammatical accuracy and some words are inappropriately used.
5	The candidate speaks fluently with few obvious mistakes and a wide variety of lexis and expression. Pronunciation is almost always intelligible, and there is little difficulty in communicating ideas.

Figure 1. Global assessment scale for oral ability (Harmer, 2009: 329)

To support teachers in their impartial assessment, Harmer further advises analytic profiles. He believes that generic assessment does not yield objective results, while, on the other hand, assessing closely each separate criterion of a performance might bring more neutrality. According to Harmer, the more the performance is analysed in greater detail, the more objective the assessment is likely to be. For example, in oral assessment, different components could be judged. These can be fluency, pronunciation, use of vocabulary and grammar, intelligibility, repair skills or task completion and so on. After that, each individual aspect would then be assessed on an analytic scale. In *Figure 2* the aspect of fluency is described and scored on a scale Thus, Harmer suspects marking becomes more reliable if the marks are given for different elements of a performance. However, he also states that the best way to assert unbiased and reliable results might be a combination of analytic as well as global scoring. Unfortunately, this approach would likely make assessment 'extremely lengthy and cumbersome' for teachers. As a result, Harmer encourages assessors to try and find a balance between reliability and practicality (Harmer, 2009: 330-331).

Score	Description
0	The candidate cannot get words or phrases out at all.
1	The candidate speaks hesitatingly in short, interrupted bursts.
2	The candidate speaks slowly with frequent pauses.
3	The candidate speaks at a comfortable speed with quite a lot of pauses and hesitations.
4	The candidate speaks at a comfortable speed with only an occasional pause or upset.
5	The candidate speaks quickly with few hesitations.

Figure 2. Analytic scale for fluency in oral assessment (Harmer, 2009: 330)

Finally, in order to boost the reliability of assessment, the role of an examiner could be separated from the role of interlocutor during oral tests. This strategy may 'cause practical problems, but it will allow the scorer to observe and assess, free from the responsibility of keeping the interaction with the candidate or candidates going' (Harmer, 2009: 331).

After examining all five of Harmer's suggestions, it appears that to provide both valid and reliable assessment it might take a lot of effort, time and commitment on the teacher's part. We might want to ask then what exactly the role of a teacher in assessment is.

2.6 Teacher as Assessor

According to Harmer (2004: 59), to be an assessor means to give feedback, correction and grades to students. A teacher becomes somewhat of an expert guide 'who provides students with ongoing guided direction thereby enabling students to keep moving forward with their work' (McDowell and Harman, 2008: 2). There is, however, much variability and a lack of systematic principles and procedures in assessment (Davison and Leung, 2009: 394). Yet teachers should always make sure that students know how and on what they are going to be assessed. Harmer advises that teachers should let their students know what exact criteria they would use to carry out the assessment. This way students would have a clear idea how to prepare in advance and what to focus on in their performance (Harmer, 2004: 59-60).

Fairness can be considered a substantial part of assessment, as the last thing students would want is to be judged unfairly. Treating all students based on the same set of criteria then becomes essential (Ibid.). Only when we encourage assessment validity and reliability, can then student expect an appropriate, useful and dependable result.

It is also of utmost importance for teachers to be sensitive when giving assessment. If the students' performance is good, it needs to be acknowledged. If any criticism is given, it must be constructive (Ibid.). 'The moment one person is placed in the position of having the right to criticize the performance of another, the relationship becomes asymmetrical, dominance being attributed to the assessor. […] this contributes to [teacher's] role as authority' (Ur, 2004: 295). Such relationship could be intimidating to students at times, so teachers might need to keep in mind that their fairness and sensitive feedback could go a long way.

When anticipating and receiving assessment feedback and results from the teacher, students can, indeed, feel quite vulnerable. Therefore, assessors might want to treat their students gently and with understanding.

In 2008, McDowell and Harman conducted interviews with lecturers at a UK university. Not only did they examine in what ways lecturers assess, but also how they feel assessing students. During the interviews, the notion of the assessor as a 'guide' was reinforced. The teachers stressed the necessity of 'care' and 'assistance' to their students, as one of them put it:

> And I keep saying to [my students] that in some ways I feel a bit like as if there's water flowing down a river and often, they're hitting rocks and don't quite know how to get round them and with experience we can just help them move round a problem so they're free to move onto the next crit.
>
> (McDowell and Harman, 2008: 2)

If the assessment is not delivered with enough tact and sensitivity, students can be susceptible to losing interest and incentive. Under all circumstances, there ought to be compassion and support when communicating a bad grade to make it more acceptable (Harmer, 2004: 60). Otherwise we risk the student possibly becoming unhappy, hurt or even resentful and this might affect their motivation in the long run.

2.7 Functions of Assessment

2.7.1 Motivation

There was a famous research done by Gardner and Lambert in 1959 which showed that highly motivated students produced better results than students without any motivation at all.

For the first time, motivation had been acknowledged as one of the most important factors for students' production (Harmer, *How to Teach English*, 2004: 8).

Students' desire to learn can come from many places. It could be simply the love for the subject or just pure practicality, since English is perhaps becoming a dominant language of research or even entertainment. Interestingly, Harmer believes a fear of failure can be a big motivator for students, too. But so can be necessity. Obligation and meeting requirements can often be a catalyst for learning in the classroom and Harmer hopes the teacher would 'provoke interest and involvement in the subject even when students are not initially interested' in it at all (Ibid.):

> It is by [teachers'] choice of topic, activity and linguistic content that [teachers] may be able to turn a class around. It is by [teachers'] attitude to class participation, their conscientiousness, their humour and their seriousness that they may influence their students. It is by [teachers'] own behaviour and enthusiasm that they may inspire.

> (Ibid.)

Harmer concludes that, even though teachers can inspire their students, they are, after all, not directly responsible for student's inner motivation (Ibid.).

Interestingly, Cheng and Fox explain they themselves use assessment for more than just providing feedback, determining final grades or formally documenting students' learning. They, too, employ assessment in order to motivate their students in their learning and to 'make' them work harder (Cheng and Fox, 2017: 10).

There are some practical approaches for teachers that could help raise student motivation. Making assessment criteria clear and visible for students could be one such means. Students would know what to focus on and their time and effort would be productive. Guidance notes can also help refocus students in the direction that the teacher considers most beneficial while possibly minimising students' later misunderstandings or omissions. Petty claims that 'if students know what to aim for, they are more likely to be successful' (Petty, 2009: 299-304).

Al Rifai also agrees that teacher's enthusiasm for English can be infectious. Factors such as teacher's attitude and teaching behaviour in class can positively affect student motivation. This can set off a chain of powerful and far-reaching effects. Al Rifai in his study found out that a highly motivated student is likely to change their learning strategy of the English language: 'the better the performance of teachers, the more the students will benefit with

class course and the more they will be motivated to change and adopt learning strategies toward English language learning' (Al Rifai, 2010: 5216-5227).

If the student's attitude towards the subject improves, this then consequently increases student's intake from the course and teachers during lessons, Al Rifai claims. As a result, the student will be exposed more to the teacher's enthusiasm during class which would subsequently, once again, boost student's motivation (Ibid.).

Al Rifai's next piece of advice for teachers to strengthen student motivation in the classroom, would be to make English courses more interesting. Although the author does not specify what the word 'interesting' entails, he states that such lessons would make students learn more readily (Ibid.: 5227).

Finally, praise and grades can increase motivation, too. Al Rifai in his research found out that students with above average grades in English were motivated to study the subject significantly more than those students whose grades were 'up to average' (Ibid.: 5224).

2.7.2 Other Functions of Assessment

Besides the motivational function, Kolář and Šikulová further determine 5 other functions of assessment. There is an informative function which updates the student on his or her progress, how close they are to achieving the desired goal and where the student stands in comparison to other students in the classroom (Kolář and Šikulová, 2009: 48-9).

Teachers also use assessment to oversee and manage student learning. This so-called regulatory function would help the teacher analyse student's pace, how and what he or she studies as well as help redirect the student if necessary, scrutinize his/her weaknesses and advise on improvement (Ibid.: 50-52; Šťastná, 2014: 17).

Next, educative function impacts the psychological development of the student's personality. Positive assessment helps formulate a student's own identity and aspirations. It enhances his or her confidence, reaffirms self-belief and promotes diligence and endurance. On the other hand, negative assessment affects student's sense of self, plants doubt in their mind and undermines their self-esteem (Kolář and Šikulová, 2009: 52-53; Šťastná, 2014: 17).

Assessment can possibly even function as a prognostics tool. The teacher could use assessment to predict students' progress and consequently his/her future prospects. The

teacher could then possibly advise the student in terms of personal growth and professional development (Kolář and Šikulová, 2009: 53; Šťastná, 2014: 17).

Finally, assessing could have a differentiating role. Following assessment, students could be separated into groups based on their level, performance, learning style, talents, etc. This assessment function can, however, sometimes further lead to labelling students (Kolář and Šikulová, 2009: 54). Labelling in education is portrayed as a divisive element. On the one hand, it can contribute to systematic improvements in the educational system, on the other, it tags students and judges their behaviour and performance based on the labels attached (Samkange, 2015: 1419).

Labelling can be used positively, for example when identifying gifted students or students with special needs. It presents itself as a useful way of classifying children through which the teacher can offer support to students according to their abilities and learning needs. Some argue that labelling has even led to 'the development of specialized teaching methods, assessment approaches, and behavioural intervention […] Labelling if correctly done, children are able to receive specialized help from teachers' (Samkange, 2015: 1422).

Unfortunately, there are also a number of limitations concerning labelling. Most significantly, it can lead to stigmatisation. The label could affect how the teacher interacts with the learner which can be detrimental to students' learning and development. This can further lead to peer rejection, ridicule, discrimination and even become a self-fulfilling prophecy (Ibid.).

2.8 Purposes of Assessment

Assessment can serve many different purposes, but Cheng and Fox direct our attention to three: instructional, student-centred and administrative. Instructional purpose, although often neglected in the literature, tends to be the most essential out of all of them. Assessment in general supports teachers in collecting information about students' learning, their understanding, and their skills via using assessment methods. Teachers take all this information into consideration when they plan, give, and adjust instructions. Through assessment teachers can check whether the students understood what is expected of them, what they should learn, while being aware of the teacher's support in their learning process (Cheng and Fox, 2017: 7-8).

Next, student-centred assessment delivers diagnostics about the student's performance, identifying his or her strengths and weaknesses. In this case, teachers use assessment to determine what learning opportunities would be best suited for the student's particular needs in order to support his or her learning. Such diagnostic assessment can be, for example, a placement test (Cheng and Fox, 2017: 8).

Lastly, using a scale to summarize a student's achievement would fall under the administrative purpose of assessment. This usually occurs based on a grading system that is often either numerical (1, 2, 3, …) or ordinal (A, B, C…). Grades are 'public statements' about student achievement and the process of obtaining them is complex and high-stakes. It requires a teacher making judgements with often serious consequences for the student. Not only can a student's motivation, expectations and self-perception be affected, so can his/her parental and social relationships. Since students obtain grades at the end of the instruction, it is usually their final performance that carries the most weight. As teachers assess what the students have learnt summarizing their achievement through a grade, grading is often closely linked to summative assessment (Ibid.).

2.9 Forms of Assessment

2.9.1 Formative and Summative Assessment

In consideration of assessment form, function is an important distinction we can make in thinking about how a procedure is going to be used. Among the most typically identified are formative and summative assessment. As far as classroom assessment is concerned, formative assessment is the most relevant. In this form of assessment, evaluation focuses on students 'in the process of "forming" their competencies and skills' and aims to assist students in the development of this process (Brown, 2003: 6). Feedback on performance is deemed crucial in this: that appropriate feedback is given by teachers in such a way that it is internalized by students. How future learning is to develop or 'form' is the main focus of formative assessment. In practical terms, informal assessment ought to tend towards the 'formative', with the main outcome being continuing progress and development. Feedback, then, is provided with a view to improving the learner's ability as they move forward (Ibid.).

As Petty describes, for teachers the main use of assessment is 'the ongoing or formative' aspect (2009: 479), utilized during the course of learning as a means of establishing 'whether,

and to what extent, learning has been successful'. Formative assessment is also useful, then, in identifying problem areas in learning which then informs decisions in the process. Petty finally stresses that the 'radically different ways' in which formative and summative assessment are conducted are due to their differing aims (Ibid.).

In summative assessment, the measurement and 'summarizing' of students' learning at the end of courses, lessons or particular sections of course books is the main focus (Brown, 2003: 6). In this sense, looking retrospectively at what has been learned and how far objectives have been achieved 'does not necessarily point the way to future progress'. We should, then, consider end of school year exams or 'proficiency' type exams as summative assessment (Ibid.).

In discussing the purpose of assessment, Petty (2009: 479) describes its multi-functionality: in grading the attainment of learners, in course placement as well as providing a long-term goal for learners. The sort of assessment applicable here is summative or final (a 'summing up of learners' achievement').

Brown asks that we consider the problem of viewing virtually all types of testing 'summative', whether exams, review tests or quizzes. Perhaps there is a tendency to think of exams and testing in a particular way, that the learning has taken place with the fixed purpose of being assessed rather than having any relevance to future development beyond the exam paper or test. The challenge for teachers, according to Brown, is in changing this view among learners. Rather than seeing their assessment in a summative way, can we 'instil a more formative quality' so that it is seen differently. Testing, then, might be seen as part of a broader 'learning experience' than merely just another hoop to be jumped through on the road to leaving school and classroom behind for something more 'meaningful' (Brown, 2003: 6-7).

Teachers, then, along with their students are better equipped in making choices and decisions when it comes to further classroom instruction (Cheng & Fox, 2017: 4). Formative assessment can be seen as collaborative in process whereby both teachers and learners are engaged in understanding the quality of learning taking place: strengths, weaknesses, areas to improve, and so on. For teachers, this process enables planning decisions based on such knowledge that may enhance learning experiences for their students who receive this understanding and improved knowledge (Ibid.)

Formative assessment is carried out with the aim of improving learning, its feedback to learners informative in their ongoing learning process. For Petty, to be 'truly formative', information given as feedback should be 'used by the learner to improve' (Petty, 2009: 480). Teachers are part of a process in providing the formative assessment that then enables and involves learners in the same process as they are able to carry out their own peer assessment or self-assessment activity formatively. According to Petty, this feedback-as-you-learn process has been shown to have more effect when it comes to student achievement compared to other factors (Ibid.).

Asian teachers, for example, have been shown to be more effective when compared to teachers in the USA. With 'aptitude and ability' the main focus of Western educators, in Asia teachers emphasized 'effort and persistence' more (*Scientific America,* 1992 [in Ibid.]). A review of research on formative assessment was carried out at King's College London in which it was found that student achievement could be improved by up to two grades through the use of certain formative assessment strategies, particularly so with weaker learners (Black and William, 1998 [in Ibid.]).

2.9.1.1 Strategies of Formative Assessment

> […] the key thing for me when I [meet with students] in terms of my formative feedback is that I want them to come out of those meetings with a real direction in which they feel confident. So I don't want them to leave feeling unsure what to do. My aim in all those meetings is for them to leave thinking, 'Great. I've got something that I can get my teeth into and I know where I'm going with it.'
>
> (McDowell and Harman's interviews with teachers, 2008: 2)

When in 2008 McDowell and Harman's interviewed teachers on their attitude towards assessment, many viewed their role as an assessor as 'one who provides students with ongoing guided direction thereby enabling students to keep moving forward with their work' (Ibid.).

Petty (2009: 481-2) goes on to emphasize the importance of informative assessment as a key dimension of formative assessment. With this in mind, he points to the fact that students should have particular information in order to improve their learning as much as possible. This includes:

a) Having clear goals that inform learners what they need to do, whether understanding task requirements or criteria attached to specific pieces of work which outline what is a 'good' piece of work - learners should know what to aim for.

b) Knowing what has been done well in tasks and what makes it 'good' provides an 'informative view' for learners regarding what they have done well - the sense that a goal has been achieved.

c) Providing information about what to improve constructively, with the emphasis on the positive and on further development - what and how to improve, rather than simply what is 'wrong'; targets for the next piece of work can be set this way; grading and giving marks in this sense do not amount to the same thing (Ibid.).

While these pieces of information can be given by teachers themselves, they may also come from peer or self-assessment.

Petty does, indeed, stress the role of self-assessment in the process of improving learning: improvement in a developmental sense, done while learning with the aim of 'fixing' problems before they are possibly made worse. Rather than errors in work being simply pointed out, a process of putting right any problems then following this up to check for improvement ('find faults, fix and follow up') is a way of encouraging development and fostering responsibility in learners, a sense that they themselves will have some accountability for their learning and work (Ibid.).

What Petty also points to here is that education's 'culture and set of assumptions' tend not to encourage teachers to approach assessment in this 'positive' way. We find instead a tendency to identify errors or examples of 'bad' work without the learners themselves being involved in the 'fixing' of errors or any checking of fixes (Petty, 2009: 483). The lack of encouragement given to teachers in using work diagnostically so as to establish targets in learning is viewed as a major problem. Teachers may give formative grades or marks but constructive guidance on how to improve is frequently missing. Teachers may also tend towards marking students' work themselves instead of the students being asked to have some role in the marking of their own work or being involved in peer-marking with models or criteria to aid them (Ibid.). Sometimes, however, a teacher delegating assessment to his or her students might reap certain benefits.

Language learners may often feel they make very little or no progress at all during the learning process. They may feel insecure and even doubt the whole process of learning. Although such feelings among students can be common, they could be reduced if only learners were made aware of their own progress regularly and explicitly. This is where the teacher can step in, helping their students see their progress (through verbal feedback, grading and so on) and thus motivate them to continue advancing. Yet, what if, instead of constantly relying on their teacher for encouragement, each student was taught how to assess themselves? (Worgan, 2010: 27) This way, teachers would be providing a 'powerful tool for students' future development' (Harmer, 2004: 102-104).

Providing what could be seen as accurate assessment, teachers can often be considered ideal assessors. Yet, students themselves can be just as competent. They can be 'extremely effective at monitoring and judging their own language production' since they often are aware of the stronger and weaker points in their performance. Teachers should then focus on developing this awareness as it seems to enhance students' learning (Ibid.) and make it more efficient. As a result of improved learning, good-quality self-assessment can actually save time in the classroom (Kratochvílová, 2011: 81; Petty, 2009: 492).

This means, teachers might want to gradually train their students how to self-assess. Self-assessment must, however, not be a 'random activity', a superficial process that a teacher can simply dismiss. On the contrary. Since it helps make students' learning more effective, self-assessment should be 'completely planned and systematic' and a regular part of students' education. Suitable conditions, such as mutual rapport and trust as well as student's feeling of safety in the classroom, must be paramount. Moreover, teachers must make sure they equip students with tools with which to assess themselves. Teachers should set up clear criteria and indicators of successful performance that students can lean against when self-assessing. Slavík advises to limit such criteria to 10-15 at a secondary school level, with any higher number becoming impractical and possibly overwhelming or confusing for the learner (Slavík, 1999: 122). So that students are aware of and fully understand their goals, teachers ought to make them transparent, clear and accessible (Kratochvílová, 2011: 79-84).

In the classroom, self-assessment could be done in a number of ways. Teachers could informally ask their students at the end of an activity how well they thought they did and/or even grade themselves (Petty, 2009: 484-7; Harmer, 2004: 102-104). Yet Kolář and Šikulová

(2009: 151-2) argue that teachers will not develop their students' assessing skills by simply asking them to grade themselves after their performance. Rather, educators should encourage learners to ask questions starting with 'what' ('*What can I improve in?*') or 'why' ('*Why did I not succeed?*'). Such questions might be much harder for students to answer since they require critical ability and analytical thinking competency. After all, the authors stress, self-assessment is a skill in itself and teaching and developing it must be pre-planned and systematic (Ibid.; Chin, 2016: 14).

As a result, some formal approaches to self-assessment may be acknowledged as more methodical and organized. Among such, students could be given materials that guide them towards making their own judgement by ticking off accurate self-descriptions or commenting on successfully attained objectives. Targeted surveys, proformas or checklists could all be used to record how the student feels about their own progress. Students can also write their own assessment of their success and difficulties through a 'record of achievement' (Harmer, 2004: 102-104). Afterwards, student's evaluation can always be compared with the teacher's own assessment (Ibid; Petty, 2009: 484-7).

Contrasting students' self-assessment with the teacher's might carry a special importance particularly where grades are concerned. Based on Ur's research, students 'outright disagree' if a teacher lets them mark their own tests and makes them solely responsible for assigning their own grades. Students apparently consider assessment part of a teacher's job and find it 'irresponsible and unprofessional' if the teacher 'opts out' (Ur, 2004: 293). Yet Ur strongly argues it is highly desirable to have students participate in decision making where grades are concerned as it gives students ownership of the learning process (Ibid.: 291-3). Thus, students could ideally be involved in creating the very criteria and indicators (measures to what extent such criteria are fulfilled) that would later help them assess themselves (Košťálová et al., 2008: 106-7). Arguably, such collaborative decision-making cannot be abrupt and immediate but calls for careful planning and needs to be taught and trained gradually (Ur, 2004: 291-3).

Although self-assessment can appear as a part of summative assessment, it seems to bring many more benefits if used formatively (Kolář and Šikulová, 2009: 152-3). When Kratochvílová ultimately offers, among other things, tips on how to improve the quality of assessment (see *2.10.6 Assessment Standards*), she encourages educators to strengthen precisely formative assessment. She strongly advocates for teachers to include the student and consider him/her a 'partner' in this process (Kratochvílová, 2011: 87-88). After all,

assessing oneself is 'a key element in formative assessment' (Andrade and Valtcheva, 2009: 12-13).

2.9.1.3 Peer Assessment

Similarly to self-assessment, peer assessment can serve as a beneficial tool in providing formative feedback. Just like self-assessment, peer assessment can provide a number of educational benefits. It can be intellectually stimulating improving students' critical ability and promoting better understanding of the tested subject as students reflect critically upon learning of their classmates (Chin, 2016: 13-18). 'Making judgements about their own or a peer's work clarifies [students'] understanding of the subject matter' because they can see different ways of performing one task (Petty, 2009: 491-2). On top of that, peer assessment can add to the learner's motivation as it puts him or her:

> in charge of their own improvement, that is, [both peer and self-assessment] develop in students a sense of 'ownership', responsibility and accountability. They increase students' effort and persistence. They develop a self-critical and reflective habit of mind.

> (Ibid.)

Peer assessment can strengthen students' independence, autonomy and confidence in the learning process. Yet again, this process must be managed appropriately by explicitly embedding it in the lesson plan. It should be defined clearly at early stages of the learning process where 'roles and responsibilities [are] laid out for all – including [the teacher]' (Chin, 2016: 13-18). Students must again be supported if they are to provide feedback to their peers in a critical and constructive way. It is the responsibility of the teacher to then ensure the process works: students are matched appropriately, there is enthusiasm to cooperate, social interaction is supported and so on. One way to help with peer evaluation is to pre-define assessment criteria and 'to provide open ended questions as guidance for students to respond to' (Ibid.).

Literature reports a wide range of methods on how to implement peer assessment. It can take either a holistic or rather a more structured approach using well-defined weighting algorithms. One such way could be multiplying a group grade by a weighting factor. 'Another commonly used method is the distribution of marks: the tutor provides a set of marks for the group and the students divide the marks according to individual efforts and

contributions to the work' (Ibid.). Chin emphasizes that peer assessment 'should assess the process of peer collaboration and not simply the product' (Ibid.).

Interestingly, students report they prefer self and peer assessment comments to the teacher's own (Bedford and Legg [in Chin]; Chin, 2016: 13-18). Students find peer assessment not only enjoyable and beneficial but also fair when judged by their classmates (Chin, 2016: 13-18). Based on research, it indeed appears that well-designed peer assessment is a 'reliable and valid method of assessment (Topping; Falchikov and Goldfinch [both in Chin]). The question arises, however, whether self-assessment can be equally as valid as peer assessing. After all, there seems to be a student tendency to over or underestimate their own performance . According to several researchers (Li; Tu and Lu [both in Chin]), both types of assessment can pose a risk of 'skewing grades.' The authors all agree that peer assessment might be a better measure of performance. Some (Dunning et al. [in Chin]) go as far as suggesting that 'self-assessment produces a wider range of scores and should therefore be excluded from grading.' Yet other research papers point out that there is no real evidence students were prone to over or underestimate performance. Having examined a number of different studies closely, Chin finally concludes that 'as long as the assessment criteria are well designed, there tends to be a closer correspondence between student grading than between [teacher] grading' (Chin, 2016: 13-18).

2.9.2 Norm-referenced and Criterion-referenced Assessment

When approaching assessment, it is necessary to clarify some of the terminology that covers the various dichotomy involved in describing it. Brown (2003: 7), for example, discusses the importance of norm-referenced and criterion-referenced testing.

Norm-referenced assessment involves the comparing of test-takers with the best among a group rewarded ('the best', 'top of the class' and so on). Marks are compared to reflect how well takers perform compared to the 'norm' (average) of all takers in a test. Test-takers are thus positioned by rank order and 'along a mathematical continuum' and these test scores are then given back to the test-takers in numerical form (i.e. 30 / 50) along with a percentile rank, showing the taker's score in relation to others. In terms of effectiveness, examinations involving larger numbers of takers are most suited to norm-referencing and are commonly found in the standardized tests of EFL such as TOEFL or in SAT tests. These are administered broadly to a large number of takers with the emphasis on efficiency when it

comes to the dissemination of results for those taking the tests. Such examination papers are formatted to require predetermined responses, with a format designed to be processed quickly and, administratively, at low cost (Ibid.).

When it comes to grading, the percentage of test-takers receiving each grade remains the same no matter their marks. The advantage here, according to Petty (2009: 479-80), is that when the difficulty level of exams varies from one year to the next this will not affect grades. It may not be, however, completely fair in that students can perhaps do better in an exam one year than in another. In terms of objectives, norm-referenced assessment is seen to be more appropriate in situations when the focus is on development as opposed to mastery (Ibid.).

By contrast, in criterion-referenced tests the emphasis is on the feedback given to test-takers on 'a specific course or lesson objectives' (Ibid.). While grading still features, such testing is typically found in one-off classroom tests with content related to curriculum learning. In this sort of testing, teachers or other test administrators may need more time and effort in order to deliver feedback appropriately and ensure that it is useful to test-takers. The 'instructional value' is thus more pointedly drawn out in the process of testing. Test-takers' scores may not be the main focus here, so the kind of continuum distribution described in norm-referenced testing is less useful. What becomes more important in criterion-referenced testing is the assessment of objectives that are appropriate to learners (Ibid.).

In criterion-referenced assessment, the reliability of tests depends on well-defined criteria, as in a checklist, marking scheme or list of competences (Petty uses the example of driving tests, for instance). In the absence of such criteria, different test markers will likely work to their own standard or else apply varying standards according to different candidates (or even what day it is). As for the appropriacy of criterion-referencing, Petty specifies its usefulness where 'mastery' is an objective (Ibid.).

Petty also cites student-referenced assessment as another alternative, something also referred to as 'medal and mission' assessment, in which a 'learner's own standard is taken as the reference' (Ibid.). Students are told the criteria that have been met along with other attainable criteria that must be fulfilled within a shorter time period (Ibid.).

Thinking contextually about language assessment, it seems clear that for the classroom language teacher involved in classroom-based assessment (as opposed to those involved in

more standardized and large-scale assessment), criterion-referenced testing has more relevance than norm-referenced tests (Brown, 2003: 7).

2.9.3 Description of Rubrics for Analytic and Holistic Assessment

Rubrics assist in assessing learning with teachers finding them a useful guide in both teaching and learning. Performance levels and assessment criteria are used in the organisation of rubrics, which typically have four or five levels. Depending on the type of assignment and skill focus of the assessment, there will be variation in the criteria applied (*What are rubrics?*[1]).

When it comes to format, there are two main forms to consider. Analytic rubrics feature a breakdown of objectives into specified components and evaluate individual components independently. The same criteria will be emphasized across multiple grades. In holistic rubrics, students' work is assessed as a whole and often focuses on specific descriptions of student performance by using 'anchor points' that assign them a value, contributing to the whole. This type of rubric has fewer details involved in the analysis, making it more simple in terms of integration. Holistic rubrics would not normally include any detailed information on student performance in specified areas. In addition, another, the so-called weighted rubric, is a type of analytic rubric that looks at some areas in more detail, such as when teachers have stressed particular aspects of learning areas more heavily. A weighted rubric would therefore focus attention on specifics of an assignment (Ibid).

2.9.3.1 Rubric Scores and Percentage Grades

As an example, a rubric based on the US system, might include four levels and four criteria. A student's performance at the bottom level (Level 1) meets the basic expectations in a limited or unsatisfactory way. As such, this is considered the minimum passing grade (or 60% in the US). A performance at the top of the scale (Level 4 in this case) means the student has met or possibly exceeded what was expected. The highest possible grade is given in this case (100%) (Ibid.).

There is, of course, the issue of error occurring in the conversion of such marks into a percentage. The 'misuse' or 'misinterpretation' of rubric scores is noted, with concern raised

---

[1] Based on 'What are rubrics?' *Deer Park School District directive* in Washington, US. October 2016. Accessed 12 December 2020. https://www.dpsd.org/wp-content/uploads/2016/10/rubric-scoring.pdf

over significant numbers of teachers equating the bottom level with outright failure. The example of grading from a score calculation converted to percentage in *Appendix I* illustrates a typical way of misinterpreting student's performance in a rubric (Ibid.).

Comparing 'traditional' and 'new' ways of interpreting scores from the rubric highlights the need for care in factoring together the number of levels in the rubric and the number of assessment criteria along with a minimum passing grade. In designing a rubric the following steps are recommended:

1. Determine the number of levels and assessment criteria in your rubric (4x4, 3x3, 5x4)
2. Decide on a grading scale for each level (90-100% = Level 4)
3. Assign the minimum and maximum passing grade percentage (minimum 60%, maximum 100%)
4. Divide the middle ranges so all levels are equal (60%, 80%, 100%)
5. Determine the overall grade by adding the percentages and dividing them by the amount of scores (87+73+73+100) = 333, 333/4 = 83%
6. Your set grading scale will determine what level the student scored (Ibid.)

2.9.4 Qualitative and Quantitative Assessment

The last major form of assessment looks at assessment in terms of its quality (qualitative assessment) and quantity (quantitative assessment). The former can comprise of comments and reports, the latter of grades, points, symbols and percentages (Slavík, 1999: 182; Harmer, 2004: 100-2; Kolář and Šikulová, 2009: 83).

2.9.4.1 Qualitative Assessment: Comments and Reports

Commenting on students' performance can happen at any stage during the learning process. Teachers see it as an 'inseparable part' of everyday learning (Šťastná, 2014: 65-67). It can be expressed explicitly or implicitly in a spoken, written or even non-verbal manner (e.g. nodding). Reports, on the other hand, would then act as written end-of-term summative assessment (Harmer, 2004: 100-2).

When assessing qualitatively, it is essential for teachers to balance positive and negative feedback. Harmer insists that it is the teacher's duty to praise students for a job well-done as much as it is their responsibility to point out problems in students' performance. Yet, indiscriminate praise and over-complimenting students can be counterproductive. So, too,

could excessive criticism, punishment and blame. Assessment is perhaps better 'handled with subtlety' (Ibid.). Moreover, this measured approval and disapproval should be combined with a 'teacher's genuine interest in and attention to a student's work' (Ibid.). Secondary students, for example, feel a strong need to understand the reasons behind teacher's assessment. Only then can the assessment be effective, and students' motivation increases (Ibid.).

While giving assessment through comments seems like second nature to teachers and has a potential to be very encouraging for students, the summative reports[2] seem to tell a slightly different, more complicated tale.

In 1988-89, an experiment took place that closely examined teacher assessment via summative reports. That school year, Czech primary schools were to replace all grades in the first grade and use summative reports exclusively instead. The experiment, however, was a failure. Based on Kolář and Šikulová (2009: 81-2), there were several reasons why this happened. The removal of grading was a directive which covered all schools with the same starting date without taking into account individual needs. On top of that, there had been a lack of groundwork and no preparation phase for this incentive. Teachers were left unsupported as there had been no methodology, nor reasons given for this transition. Many teachers, as well as parents, were simply not ready to adjust to the sudden shift. It was then no surprise the experiment was discontinued after just one year.

In general, reports should give a clear signal to the student, and their parents, on how well he or she has done in class. They help determine the student's future prospects (Harmer, 2004: 102), help suggest strategies on how to improve his or her performance and certainly have a potential to promote an overall 'social climate of learning' (Slavík, 1999: 132). In 1997, Číhalová and Mayer conducted a study in which reports replaced grades in several classrooms of a Czech primary school. Their efforts allegedly reaped many benefits. The authors reported a marked improvement in teacher-student communication. When disagreeing with an authority figure of their teacher, students were able to discuss and argue their case without fear of repercussions. Even more strikingly, students described feelings of joy while learning for the first time. Additionally, students' ability to self-assess as well as their ability to cooperate with each other improved significantly. In comparison to the previous experiment, this study had been carefully planned over the course of two years.

---

[2] so-called 'slovní hodnocení' in Czech (or final summative reporting as 'finální slovní hodnocení' in Czech) (Kolář and Šikulová, 2009: 84; Harmer, 2004: 102)

During this time, the idea was introduced to parents so they had enough time to get used to the initiative. Students likewise were gradually prepared for the transition for a year by newly focusing on the development of self-assessment and in-class cooperation. Teachers also had plenty of support from the management having a clear outline on how to proceed when implementing the programme. In addition, they all believed strongly in the concept of school reports. Their devotion to and confidence in the process became instrumental to the success of the transition (Číhalová and Mayer, 1997: 48-53). While Číhalová and Mayer seemed to insist that reporting could be a 'hopeful alternative to grading' (Ibid.: 28), others have not been as convinced. Kolář and Šikulová argue that the perks of reporting are simply overstated in the 1997 study. 'To swap one form of assessment for another most certainly does not pave the road to good-quality school assessment as educational means' (Kolář and Šikulová, 2009: 81). The teacher should perhaps assess 'using different ways, be inventive and avoid routine' (Ibid.). Kolář and Šikulová further assert that teaching is a complex, intertwined system of which reports are only a part. They point out that the form of assessment cannot be changed without the whole concept of teaching being reconsidered (Ibid.).

Summative reports might, indeed, present certain challenges. Work overload and time constraints can pose an issue. It might simply be unfeasible for teachers to write reports especially in higher grades in a state school where an average teacher can be responsible for assessing over 200 students. Likewise, the lack of shared, universal and clear criteria for reports can create complications. When opponents of reporting voice their concerns about report transparency, they point out the consequent inability to compare various students, schools or school districts. This can present further difficulties in obtaining summative statistical surveys and make it harder for students, and their parents, to choose a school or transfer to a different one. Reports, therefore, should abandon flowery language and rather compare performance based on the same criteria, using the same language that is comparable and easily interpreted (Slavík, 1999: 132-3).

Keeping in mind the possible disadvantages of reporting, educators like Slavík or Kolář and Šikulová then recognise the report as only a part of the assessing system. They rather tend to believe that 'a combination of extensive, subjective, purely qualitative reports with more content-dense, standardized quantitative assessment could, in a practical sense, be the most appropriate way to the comprehensive and satisfyingly informative assessment of a student' (Slavík, 1999: 132-3).

In her survey, Šťastná questions teachers in real life and their overall feelings on reporting as a form of assessment. She finds out that the teachers themselves are not convinced, either, by the concept of reports overall. The teachers consider this way of assessing ambiguous because they believe it lacks criteria and standards. The teachers also mention how demanding time-wise writing reports can be, adding an extra administrative task to their already busy schedule. Moreover, in the teachers' opinion, reports could lead to fixed, 'formulaic phrases' (Šťastná, 2014: 65-67). Such established, rigid clichés might be then harder to interpret for students in contrast to grades (Ibid.; Solfronk, 1996: 30).

Ježová then in her research (2009) polls students at 2 different primary/middle schools to examine their attitude towards assessment. She discovers that the students, as well as their parents, reject summative verbal assessment in favour of grades. Besides, Ježová observes that a report does very little to motivate students to improve their study habits. If summative reports replaced grades entirely, they would fail to provide enough incentive for students to study more, especially when it comes to primary students. Even more notably, implementing reports and excluding grades with older (middle school) students would have a mildly demotivating effect (Ježová, 2009: 63).

Nor could teachers imagine reports replacing grades (Šťastná, 2014: 58-60). For example, one teacher from Šťastná's study discloses: 'I wouldn't change grading. A report is often partial and long and also, does not motivate to improve as much' (Ibid.). Another teacher adds that grading summarizes assessment in a simple yet clear and understandable way (Ibid.).

Lastly, even in the absence of grades, teachers still create and rely on a grade-like system. Kolář and Šikulová describe an experiment where teachers for six weeks could not use grades to assess their students. However, the teachers very quickly came up with an alternative quantitative system with identical effects as grades, e.g. point scope, symbol scope, pictures and so on (Kolář and Šikulová, 2009: 82-83). Grading appears to have a stable place in the education process. Grades seem to be simply desired and expected from all - teachers, parents and students alike.

2.9.4.2 Quantitative Assessment: Grading

Despite the seeming preference for grades as a form of assessment today, the general attitudes towards grading were perhaps more ambivalent not so long ago. By the mid-1980s,

the critics of quantitative assessment started to point out the inherent subjectivity of grading as well as its demotivating aspect for students. Ziegenspeck comments that the more the pedagogical and psychological aspects of grading are considered, the more 'controversial' grading appears (2002: 29-31). He argues that grading is only a 'compressed' and 'mathematically thrifty, very economic' form of assessment (Ibid.: 31). In Ziegenspeck's words, grading using numbers is only 'clear, exact and unambiguous mathematical symbol; as an assessment tool, grade becomes unclear, non-transparent and ambiguous' (Ibid.: 61).

There are different functions that grading provides. Ziegenspeck speaks of three. First, there is a directory and informative function. This is followed by a pedagogical function. Thirdly, there is a selective, order-creating and authenticating function.

The directory function helps students navigate through the learning process and provides information for their parents about the process. It sees grades as a means of checking off completed performances as well as the student's behaviour. Besides fulfilling an informative function for parents, it also serves as a source of information for teachers and other official authorities. Ziegenspeck wonders however whether grades are actually the most feasible form of assessment. He claims that the grading number scale on its own might not be fully meaningful outside of the student-teacher relationship. The author argues that without a 'value to compare the grade to', grading can be quite reductive, especially to third parties. He casts doubt on whether a mere number scale can satisfyingly reveal enough information to fulfil its primary function. It is only the 'extreme values' (highest and lowest grade on the scale) that could be unambiguous. Grades overall speak nothing about the extent to which the learning goals were reached. Grades, Ziegenspeck claims, are nothing but 'mere symbols!' (Ibid.: 32-33)

Nevertheless, grades play an important role in student's life but their relevance changes as a student develops. For example, students between the ages 9-13 consider grades they receive objective, fair. They predominantly do not question a teacher's decision and thus any negative assessment is seen as their own failure. Secondary school students, however, reject the concept of grading more readily. These students, especially during puberty, seem to lose faith in teachers' ability to assess them objectively. Such rejection of authority could be occurring due to the teenagers' own self-development and changes. The students begin to question their own subjective sense of self, establishing their own self-identity and esteem. They might start to self-reflect and self-critique more frequently, too. Their sense of being

different yet unique can often lead to feeling misunderstood. (Ibid.: 34-35) This presumed subjectivity then brings perhaps a very distinct, a 'highly ambivalent' attitude towards grades (Ibid.: 34). Ziegenspeck speaks of a vast variety of emotions, both positive and negative, that grades can evoke in (not exclusively) pubescent students. There is 'fear and upset, absence of joy and anxiety, antipathy, lethargy and depression', on the one hand, with a sense of relief and a feeling of success on the other (Ibid.: 34). Yet, students in secondary schools, more than in any other age group, perceive grades as the teacher's 'manipulative tool of power' (Ibid.: 35). In the eyes of students, such teachers could very easily use grades exclusively as either a disciplinary or motivating instrument, while withholding help and dismissing any incentive for discussion and objectivization (Ibid.: 35).

Besides students, quantitative assessment carries a special meaning for parents, too. Grades have a potential to create a link between the family and the school. Affecting students emotionally, grades are likely to arouse passions and provoke sensitivities in the parents just as easily. Such feelings are, once again, largely subjective. Similarly to students, grades can invoke b7oth joy and fear in their parents who may possibly equate their offspring's performance to that of their own. For some, Ziegenspeck argues, grades uphold family honour (Ibid.: 36). To escape any subjectivity, Ziegenspeck urges teachers to include a commentary to add context to any grade. Without it, a grade could lose its actual value, mainly because parents cannot see where the particular grade stands in the bigger scheme of the whole class, and how it plays out in contrast to other students. In the Czech system a grade, for example grade '2', could cover a relatively broad spectrum of abilities and individualities among and within students. A '2' can describe students with varying and contrasting qualities: a talented but rather lazy student, an average yet diligent thinker or a dependent routine-loving student (Ibid.: 37). A grade covers a vast range of performances and therefore, ideally, may require a more specific differentiation coming from the teacher.

At this point, it might be compelling to consider the role of a teacher and what the process of grading actually means to him or her. First and foremost, grading can provide an opportunity for the teacher to assess themselves, to self-reflect and to examine their own performance with a slightly critical eye. Grading, according to Ziegenspeck, is the result of the teacher's just as much as the student's performance. It is the result of the teacher's chosen methodology, their pedagogic work and even their behaviour in the classroom. Yet, grading can remain a problematic area in their didactic work and a teacher should fully acknowledge the controversy and subjectivity: 'A well-informed teacher knows, that despite its apparent

exactness, a grade expressed via a number gives the wrong impression of reliability' (Ibid.: 38). This means that grade's actual value as well as its directory and informative function might be regarded by some as rather relative. Ziegenspeck admits that assessing using grades can be, indeed, quite controversial. Consequently, teachers might face a lot of pressure from not only students or parents but also from their own colleagues, school management and society in general while striving for accuracy and objectivity (Ibid.: 37-8).

Having discussed the first function of grading, it is time to quickly consider the remaining two. The pedagogical function means grades have the potential to 'reprimand, encourage and inspire' (Ibid.: 38). While a 'good' grade can have a very positive, motivating effect, a 'bad' grade can have the opposite effect. It easily can become a demotivating factor in the learning process, especially with students that receive it frequently, Ziegenspeck claims. 'Bad' grades can take away the joy from learning, create doubt and lower self-esteem. They supposedly have a power to perpetuate stereotypes about students if received repeatedly (Ibid.). Yet in contrast, Saňáková (2018: 77) proved in her thesis that even a 'bad' grade can have a motivating effect, actually more so than a 'good' one.

By and large, the pedagogical function understands grades as teachers' power tool to de/motivate, to edify with either an encouraging or punitive effect. 'Good' grades can become the absolute primary focus of the learning process, rather than the learning itself. Sycophants and teacher's pets are given way and an environment of superficial competitiveness among students can form. Such an atmosphere could embolden cheating, lying, finger pointing and talking behind one's back, Ziegenspeck warns. He calls grades the 'scourge of the school' and concludes: one-off praise or rebuke affects the student's performance positively in the long haul. Repeated criticism, on the other hand, lowers the student's long-term performance while repeated praise does not lead to an increase in performance in the long run. '[…] The effect of a primitive instrument can itself only be primitive' (Ziegenspeck, 2002: 41). Grades, Ziegenspeck argues, can only possibly affect students' mood and self-esteem (Ibid.).

Yet, what might prove more relevant to the student's development and more critical in his/her learning process, is that grades actually show students what they have mastered and what they need to improve in. 'But of course, our insufficiently divided grading system is unable to do that,' Ziegenspeck states, derisively (Ibid.). He rejects pedagogical function of grades in

the end because it points to the subjectivity of the teacher's assessment instead of arguing for an objective summary of the students' achievement (Ibid.: 39-42).

Finally the third function of grades, that of selective, order-creating and authenticating, can be briefly discussed at this stage. Ziegenspeck alleges that a final summative grade from a course no longer serves as a mechanism for school to assess a student. A final grade is on the contrary primarily a bureaucratic procedure for future selection on the basis of a student's past achievement. A final grade can positively or negatively impact the student's chances to succeed professionally and improve their social and material status. Final grades largely appear to be a quick and practical way to aid in the selection process and they often rely on grades received in the course of the learning process. This last third function argues, grades in general tend to have a significant impact on the students' future. Still, some argue that grades are just 'primitive tools' of assessment and the grading scale system is simply a mere guessing scale based on approximation (Ibid.: 44-5, 47). That is why experts agree it is of the utmost importance that the system of assessment be reliable and objective (Ibid.: 44-45; 85).

2.10 Grading

2.10.1 Problems with Grading: The Convention of the Grading scale

If grades are indeed just a mere 'primitive' instrument of assessment, what are some of the fundamental problems that the simplistic grading system can bring? Three problematic issues of grading could be identified: failure recognizing the complexities in the assessment process; grading being more an estimate rather than a precise value judgement and finally, its alleged subjectivity (Ziegenspeck, 2002). The system of grading thus remains a moot point and many scrutinize it with a rather critical eye.

Some experts, for example, comment on the continuity of a student's performance. To Ziegenspeck, performance increases and decreases on a continuum, rather than it would gradate in predetermined levels or degrees. Having studied teachers' tendencies regarding grading, Ziegenspeck notices an interesting phenomenon in the classroom. While comparing the frequency of different grades, he spots that the extreme endpoints in the grading system[3] are rarely applied in the classroom. He explains this by teachers' fairly careful and cautious attitude when grading. The author further polemizes that the lowest grade ('nedostatečně' or

---

[3] 'výborně' (or '1') and 'nedostatečně' (or '5') in the Czech system of grading

'5' in the Czech education system) is used so seldomly, that 'as a grading level loses its meaning altogether' (Ibid.: 49).

As F. Holzinger claims, extreme values on a grading scale are usually defined and recognized with more certainty. 'Extreme' students' performances, meaning their best and worst, tend to be therefore the most distinctive and unambiguous and teachers are likely to define them more easily. According to R. Falk, this supposed transparency, clarity and explicitness can consequently lead to a higher degree of objectivity in their grading. (Holzinger, Falk [in Ibid.: 65]). However, with more moderate performances and more 'moderate' grades (meaning '2', '3' and '4' in the Czech education system), the level of objectivity could become more precarious. The authors believe, the boundaries between each of these grades are slightly less defined and more blurred, which decreases objectivity in grading (Ibid.). After all, using numbers for grading is 'clear, exact and unambiguous only as a mathematical symbol, as a performance assessment a grade becomes vague, blurred and debatable' (Flitner [in Ibid.: 61]).

Besides its abstract meaning and a small informative value, author M. Simoneit also depicted the multitude of different meanings of the grade '3' in the Czech grading system. The grade can describe a student's real performance over the latest time period, present state of knowledge or aptitude for the subject, in this case History. Depending on a teacher, the grade can assess a number of different aspects. It can, for instance, assess students' chronological knowledge or thinking, their understanding of cultural development or even their interest in politics, historical figures and so on (Simoneit [in Ibid.: 60]). Overall, the numeral grading system appears to be just a 'terrible simplification' of students' performance (Wagenschein [in Ibid.: 52]).

On top of that, the less differentiated the grading scale is, the more unfair and subjective it appears to be. What seems to be the problem is that the grading scale used in schools is not an interval scale. If it were, the distance between each grade would be precisely the same. That would guarantee that the same performance would always lead to the same grade assuring absolute, scientific-like objectivity. However, the grading scale used in classrooms seems to be based 'only on bigger-smaller' principle (Ibid.: 51). Grades simply do not appear to be very accurate measuring values but rather simple guesses (Ibid.; Starch and Elliott, 1912: 442). That is why, Ziegenspeck argues, there is no reason to determine the final grade by counting an arithmetic average from individual course grades. Counting an average should be

recognized only as an 'emergency' solution due to the lack of more suitable mechanisms. After all, grades ought not to be observed in complete isolation. They should also recognize the student's overall performance in the course, consider the student's overall knowledge and acknowledge their abilities and effort (Ziegenspeck, 2002: 37; 49-52).

2.10.2 Problems with Grading: Institutional and Structural Issues

Indeed, there are a lot of various factors that play a role when awarding a grade. Perhaps surprisingly, the subject in which the student receives their grade, can affect the grade. A number of studies (Wolf, 1962; Sander, 1963; Weiss, 1964; Weiss, 1965; Hopp and Lienert, 1965 [all of them in Ibid.: 62-5]) have shown that teachers themselves consider certain subjects more 'prestigious' and more 'fundamental' than others (Wolf [in Ibid.]). As a result, the grading of these more 'essential' subjects is on average stricter than the grading of 'less important' subjects. More specifically, languages and mathematics allegedly belong to the former group while physical education or art classes would fall under the latter. Having pored over the dates from a German all-girls' secondary school, Sander comes up with concrete evidence to prove the hypothesis that 'non/essential' subjects have a propensity to alter grades. Sanders concludes that the average grade for foreign languages, including English, is between 3.20 - 3.49[4]. The 'less significant' subjects seemed to be assessed much more moderately at 2.16 - 2.49 (Sander [in Ibid.]).

In addition, Hopp and Lienert in their research also contrast average grades for different subjects at another German secondary school. What they discover is, again, that mathematics and languages, such as English, belong to 'main' subjects whose grading is on the stricter side. Students in these subjects tend to receive 'more bad than good grades' than from any other subject they take at school (Hopp and Lienert [in Ibid.]). One explanation for the stricter grading in the 'main' subjects could possibly be a higher requirement for precision in students' performances in these subjects. There might be a bigger need for accuracy and specific answers and possibly a certain preference for written testing, Ziegenspeck and Weiss polemize. The more written tasks there are to be graded, the stricter the grades tend to be. After all, mistakes in written tasks might be easier to spot and mark which supports preciseness, exactness and objectivity. This kind of precision can be harder to achieve when assessing oral performances (Ibid.: 65-6).

---

[4] based on a 6-point grading scale in the German school system where *1* means *excellent* while *6 insufficient*

What is more, grades seem to reflect the overall dynamic of a particular class, a specific Year as well as a type of school the student attends. First, the performance of an individual can be based on the performance of his or her class. This was statistically proven by Ingenkamp who compared parallel classes in the same school. Ingenkamp found out that the same teacher assesses the (objectively) same performance in parallel classes differently (Ingenkamp [in Ibid.: 67]). A similar American study done by Lietzmann at the beginning of the 20th century came up with comparable results.

The research confirmed that the same teacher awards different grades when he or she assesses the same performance at different times (Lietzmann [in Ibid.: 60]). Lietzmann's study even reports on how the same student receives different grades in different schools. Likewise, Hopp and Lienert's detailed analysis of grading at different 'Gymnasium'[5] levels observed the same student being graded more sternly in subjects such as English once he/she transitioned from the 'lower' to the 'upper Gymnasium' (Hopp and Lienert [in Ibid.: 68]). This finding is further confirmed by Weiss's research. Here, the author follows closely grading at Austrian secondary schools ('Mittelschulen'[6]) and notices how it becomes stricter and stricter with each passing Year. He thus observes a various level of strictness at different kinds of schools and even notes harsher grading systems in cities than anywhere else. After observing such disparity, Weiss concludes: 'Grades from different types of schools […] are not directly comparable' (Weiss [in Ibid.: 68-9]). Indeed, other numerous studies also show a significant variation and uneven tendencies among secondary schools when awarding grades. A close examination of their grading system reveals that grades are greatly affected by the type of a secondary school the student attends. For example, grades of students newly arriving from their primary school get generally worse at 'Gymnasiums'. On the other hand, grades of first-year students appear to improve or at least stay the same when attending other types of secondary schools (Ibid.).

2.10.3 Problems with Grading: Gender Stereotype

Finally, and perhaps somewhat surprisingly, the grade can also be shaped by student gender. In 1965, Hopp and Lienert decided to follow closely average grades received at secondary

---

[5] arguably the most academically demanding kind of German secondary school preparing students for a university education (possibly comparable to an English grammar school or a Czech 'gymnázium'). Hopp and Lienert's study focused on the so-called 'lower' and 'upper Gymnasiums' ('nižší a vyšší stupeň gymnázia')
[6] similar to a German 'Gymnasium'.

schools based on the student sex. For their research, Hopp and Lienert considered only 'Gymnasiums', especially those that specialized in modern languages. What the authors of the study eventually found out was that girls outperformed boys in the classroom significantly. With the final grade on average consistently higher or at least comparable to boys, female students showed a remarkably higher academic achievement of key subjects, mainly in the study of the 1st and 2nd foreign language (Hopp and Lienert; Knoche [in Ibid.: 70-72]).

While exploring the different propensity of various grading systems, experts have tried to also come up with a reason for the gender imbalance presented. It seems there is a strong consensus this problem cannot be explained by intellect or lack thereof. Educators seem to be in unison that the gender division related to grading may stem from a mixture of different reasons. Some blame teachers' favouritism towards girls to explain this uneven strictness in the grading system (Ziegenspeck, 2002: 72-3). Others (Rank [in Ibid.]) highlight the girls' alleged ability to adjust to new situations quickly and adapt to the school environment perhaps more easily than boys. This natural propensity of girls ostensibly gives them the advantage in situations where they are being assessed. Girls' seeming compliance and willingness to please others could be another potential source of the gender disparity in grading (Weiss [in Ziegenspeck, 2002: 72-3]). While boys are maybe perceived as somewhat rebellious and lacking in discipline, girls, in contrast, might benefit from a more approving behaviour. Arguably, female students appear to exhibit 'greater diligence, stronger academic motivation, longer endurance and higher reliability' (Ziegenspeck, 2002: 73). On top of that, as peculiar as it might seem, research indicates that clear, neat and legible handwriting can have quite a profound and positive effect on a grade. By extension, clean, good-looking, orderly and well-behaved students are more likely to be regarded as intelligent than students who disturb and misbehave in the classroom (Aebli [in Ibid.: 72-3]).

Finally, some assert (Meili; Knoche [both in Ziegenspeck, 2002: 72-3]). that girls possess a certain innate predisposition to develop verbal skills which could consequently help them reach language acquisition faster. This hypothesis could then readily support the results of W. Knoche's study. Through his extensive research and comprehensive analysis of 14,000 students from 50 'Gymnasiums', Knoche was able to determine that female students in secondary schools receive better grades especially in academic studies of foreign modern languages (Knoche [in Ziegenspeck, 2002: 72-3]).

2.10.4 Problems with Grading: The Subjective Element

When discussing a range of issues affecting a grade, surely, the role of teacher must not be left out of this debate. The teacher's capacity and competency to award grades brings about a sense of responsibility and dependability that ought not to be dismissed or taken lightly. To a certain extent, a grade may be a potential reflection of the assessor's judgement. But arguably, one's perception can be easily tainted, whether knowingly or unknowingly and this could then seriously jeopardize the objectivity of grading (Ibid.: 85).

As already witnessed, a grade can change for the same performance when the type of school, kind of subject or even gender category changes. It is then feasible to assume that a grade changes from teacher to teacher just as easily. A teacher in this scenario appears as just another variable. Research ultimately corroborates this theory. Starch and Elliott conducted an experiment in which they illustrated sharp fluctuations in grading among 152 teachers of the English language at 200 secondary schools. The teachers' objective was to mark two examination-answer papers from 0-100 points written by two first-year students. The results of the study revealed enormous and dramatic differences in the standards of the marking. While the teachers themselves believed their assessment could possibly differ by 10 points from one another at maximum, the researchers found out the actual range of marks was as large as 40 points. Starch and Elliott talk then about 'the tremendously wide range of variation' and how 'the promotion or retardation of a pupil depends to a considerable extent upon the subjective estimate of his teacher' (Starch and Elliot, 1912: 454). Having calculated the probable error, the authors determined that any one teacher's mark was 4 or 5 points off from the true[7] mark. Such a large probable error allegedly shows 'the absurdity of marking' (Ibid.: 456). The authors also noticed that small secondary schools marked somewhat more leniently than larger schools. Yet the range of variation stayed the same (Ibid.: 457). The authors finally asserted that 'even the standard of a given teacher is more or less variable and indefinite' (Ibid.: 454).

All considering, there certainly must be reasons behind such wide margins and apparent lack of reliability. It might be safe to say there are plenty of subconscious driving forces, personal impulses perhaps, when teachers grade students. The following will attempt to summarize

---

[7] under 'true mark' the authors understand the average mark given by a large number of teachers (Ibid.: 456)

teachers' six main tendencies that might cause bias, either positive or negative, knowing or unknowing.

2.10.4.1 'Logical Mistake'

Ziegenspeck describes this as connecting two unrelated traits based on faulty logic. In everyday language, it compares to the saying: *Show me a liar, and I will show you a thief*. In a school setting, a student excelling in one subject is expected to do so in another, or a student that aces a written test is supposed to pass an oral exam with flying colours, too. There is a transference of expectations without any logical basis (Ziegenspeck, 2002: 86-7).

2.10.4.2 The 'Halo-effect'

This is a transference based on the teacher's impression. If a teacher finds a student likeable in some way, the teacher will be more prone to give them a better grade than to an intelligent but a misbehaving student (Ibid.: 87). In a 2007 study of 13 high school English teachers, nearly all of them admitted to this bias. In case a student was 'on the line' between passing and failing, what decided the final result was the student's character. If a borderline student was judged to be of good character, the teacher would actually add a few points to their performance and pass him/her. A borderline student judged to be of bad character would still fail (Zoeckler, 2007: 96-7). Likewise, students' good/poor reputation and strikingly even a student's legible/illegible writing can affect their grade this way (Ziegenspeck, 2002: 87).

2.10.4.3 'Tendency to Persevere'

This bias means that the teacher's first assessment of the student does not change (and if so then only very slightly) with future assessments (Ibid.).

2.10.4.4 'Order and Contrast Effect'

There is a belief that when grading written essays of average quality one after another, the teacher can become convinced that all the essays cannot be of the same quality. He or she then tends to award a better grade to one of the essays (Ibid.: 88).

2.10.4.5 'Assessing Tendencies'

In this case, a grade can be influenced by the teacher's personal or professional life, e.g. their own past experiences from school. Teachers tend to adopt a habit of leaning towards always grading either strictly, leniently or moderately. A grade here can serve a motivational function. Teachers usually award a better grade to encourage learning but sometimes a worse

grade is awarded (based on a faulty assumption) in the hope it might motivate the student (Ibid.: 88-9).

2.10.4.6 'Mistakes Caused by Recognizing Repercussions'

There can be some serious negative consequences for students if they are awarded a bad grade (e.g. repeating a Year). Recognizing how the serious repercussions of a bad grade can influence the student's future, teachers sometimes avoid awarding such grades (Ibid.: 89-90). Moreover, some teachers believe there is an 'acceptable' number of failures in a class. The student's previous grades thus can subsequently affect the teacher's objectivity in their later grading (Zoeckler, 2007: 96).

2.10.5 Avoiding Subjective Elements Regarding Grading

To prevent making the six mistakes that underline teacher's own subjectivity, teachers are encouraged to acknowledge their own bias and always be willing to review any grade given. This, however, might require a fair share of self-criticism, P.H. Ludwig notes (Ludwig [in Ziegenspeck, 2002: 90]). Further, there perhaps needs to be a clear mental separation between the teacher's objective observation and their assessment, otherwise grading becomes just a 'loose interpretation' (Ibid.). Grading thus ought to take place only 'at the end of the systematic observation of the student's performance' when all the partial aspects of the performance have been considered (Ibid.). W. Sacher warns sternly that grading based purely on the overall impression can be 'dangerous' (Sacher [in Ibid.]) and should be abstained from. It might be also greatly desirable that teachers ignore student's personal details that are not directly related to the teacher's subject or classroom and which could potentially sway their grading. This may include e.g. any confidential information about the student or their performance in other classes. In addition, it has been proposed that correcting essays could be done anonymously without the teacher knowing the student's name to improve impartial assessment (Ibid.).

Finally, when in doubt, teachers should never hesitate to discuss a grade with their colleagues or ask them for any help and advice in this matter. Ziegenspeck ultimately insists that a systematic pedagogical diagnostics training on the possible reasons behind mistakes in grading is paramount. Such training would raise teachers' awareness of the issues, increase the likelihood of preventing them and therefore lead to a more objective outcome (Ibid.).

2.10.6 Assessment Standards

It is perhaps open for discussion how teachers could potentially improve the quality of their assessment. A number of authors attempt to answer it and give tips to educators. Kolář and Šikulová, for example, advise teachers to assess not only the result of the student's performance, but also the process of the student's learning. While devoid of interpersonal feelings, the teacher's language should be deliberate when assessing. The teacher's expression should be pragmatic, clear and focused as well as understandable to the learner. Descriptive style could be adopted when assessing. Teachers should depict what they can perceive with their senses (*I can see that all your letters in this word are correct*). They could focus on the student's progress when compared to the last performance (*You now spelt all the letters correctly*). Finally, teachers might want to describe the student's emotions (*You're probably happy, that you completed all the tasks correctly*) and give students space to correct themselves (*Look at the end of your sentence. Is everything alright there?*) (Kolář and Šikulová, 2009: 145-6).

According to Slavík, four criteria ought to be fulfilled if any effort to assess students well is to be successful. First, assessment must be targeted, in other words, valid. Students must be assessed based on previously discussed goals, requirements, and criteria. Second, assessment should be systematic, something Slavík understands as diverse as well as comprehensive. Here, he recommends different types of assessment (formative/summative, norm-referenced/criterion-referenced, analytic/holistic) be employed.

Teachers should also avoid focusing on assessing only one student's competence repeatedly (Slavík, 1999, 87-90). Third, assessment should be effective and purposeful. The amount of assessment should be enough to give teachers substantial information about students' learning while at the same time, it should not overwhelm neither the student, nor the teacher. Finally, fourth, assessment needs to be informative with predictive validity. This means, it is to report on students' progress in a comprehensible way for both students and their parents while helping predict students' future results (Ibid.).

Additionally, Košťálová et al. remark that assessment should never lead to a student feeling anxious or helpless or make him/her feel like a victim. They stress that assessment must take into account students' individual strengths and that continuous, formative peer and self-assessment can help pinpoint these aspects. Furthermore, when assessing the student summatively, teachers should bear in mind not only his/her test or exam results but also the

student's overall learning, too. Here self-assessment proformas, surveys and checklists, even talking to the student's parents and student's other teachers at school, can guide the teacher to assess the student well. What cannot be omitted with any assessment, Košťálová et al. insist, is the frame of reference. The authors argue, there must be a set of criteria, i.e. rubrics, that the teacher's assessment is compared against. Containing only well-defined, specific and easily checked off criteria, the rubric aids in assessing students' work independently from each other. They are to be set and known to students well before any assessment takes place (see also *2.9.3.1 Rubric scores and percentage grades*) (Košťálová et al., 2008: 15-23).

Kratochvílová believes that by formulating assessment criteria and familiarizing students with them beforehand, teachers can substantially enhance the quality of assessment. She suggests there are a number of strategies that can improve assessment quality. Among them is to encourage formative assessment and make the student a partner in the assessment process. It is important students are able to assess their own progress and results of their work. In agreement with Slavík, Kratochvílová also believes that utilizing different types of assessment might improve its quality. She also cautions against an abundance of assessment as quantity does not necessarily yield quality. To achieve effectivity in assessment, students must know what they are to learn and why they are to learn it. There ought to be a purpose to their learning process which should also be individualized. Students additionally need to know what to do to study. They must be fully aware and understand the assessment criteria completely. Thus, understandably, teachers should then only assess what has been taught in lessons (Kratochvílová, 2011: 87-88).

2.10.7 Differences Across the World: Grading Practices

In 2007, Cheng and Wang looked at classroom assessment by interviewing 74 ESL teachers from three different countries (China, Hong Kong, Canada). They found a considerable variation in their grading practices. They found significant differences in how the teachers prepared and chose their grading criteria, how much they involved students in their assessment or in how they reported the final course grades. For example, most teachers in Canada and China prepared their own grading criteria in contrast to less than half of Hong Kong teachers. Most Canadian teachers used analytic scoring, Chinese educators generally preferred holistic scoring and Hong Kong teachers were choosing rubrics the most. Surprisingly, only Chinese students seemed to be able to launch an appeal to their teacher concerning their grade. Finally, while Canadian teachers could not agree on a passing 'line'

for the final grade, Chinese teachers overwhelmingly agreed the passing threshold was 60 out of 100 (Cheng and Wang, 2007: 85-107).

Naturally, the question that one can ask next is what do grading practices look like in the Czech Republic? The National Institute of Technical and Vocational Education (*Národní ústav odborného vzdělávání* or *NUOV*) run by the Ministry of Education, Youth and Sports sets the standards for schools on how to grade student performance and what to focus on when grading.

Based on *NUOV* guidelines grade '1' ('výborně') in the subject of the English language, for example, is defined among others in the following terms:
- The student masters the required knowledge, data, facts and definitions comprehensively, accurately and fully while understanding the interrelationships
- The student thinks correctly following logic, they clearly exhibit autonomy and creativity. Their oral and written performance is correct, accurate and apt
- The student's results are of quality, only with minor inadequacies[8]

It may be argued, however, that some of the wording is slightly on the generic side since expressions like 'quality', 'inadequacies' or 'apt' could be considered rather ambiguous, vague and open to interpretation. Besides, Čapek, a renowned educator, wonders:

> Dear colleagues, teachers, do you really award grade '1' only to the students described this way? Do you know anyone like that? If teachers themselves achieved such characteristics, how better would it be in our schools at once. But not even teachers are like this, these characteristics are a nonsensical hodgepodge of an ideal state […]. Not even the most diligent girls, our favourite […] straight A students often do not have the skill of applying themselves or the skill of creativity. I still, however, award a '1' to them and other regular students. I find the prescribed grading divisions restrictive. They do not even capture suitably the specifics of my subject.
>
> (Čapek, 2010: 73)

Not only can the generic language of standard guidelines lead to more subjectivity, different Czech secondary schools establish different internal norms. While one differentiates each

---

[8] based on 'Stupně hodnocení prospěchu a chování v případě použití klasifikace a jejich charakteristiku, včetně předem stanovených kritérií.' Accessed 6 March 2021. http://www.nuov.cz/ae/stupne-hodnoceni-prospechu-a-chovani-v-pripade-pouziti?lred=1

grade according to its weight (from 1-6)[9], another uses percentages (from 0-200%)[10]. Tests, homework and participation are all graded based on different criteria at different schools[11]. Finally, some school management provide detailed instructions on how to grade for instance presentations/projects while teachers at other schools lack such support altogether[12].

In the end, all these biases and irregularities in grading seem to occur across different types of schools, subjects, teachers and even countries. It might be safe to say that the variation and subjectivity in grading is almost universal and expected. Grading itself stays non-systematic, inconsistent and fragmented. This may be exacerbated, in some cases, by an absence of sufficient support or clearer guidance on how to grade more uniformly. The aim of the following chapters is therefore to examine the lack of systematic approach in grading and its extent since it can leave a serious dent in desired objectivity and fairness in the classroom.

2.10.8 Grading of the English Language in Czech Secondary Schools

It is grading that is front and centre of this thesis. Four variables of grading will be considered:

- grading of a specific subject (the English language)
- grading that happens exclusively on a secondary level
- grading within the Czech Republic only
- grading from teachers' perspective

This focus raises the question as to whether research exists which has looked at the way grades are awarded under these conditions?

To the best of our knowledge, there appears to be only a fairly limited number of dissertations and scientific papers devoted to systematic grading using such variables. As far as can be determined, there seem to be just a few instances of university theses and academic papers of shorter character that try to map out grading this way. These research papers, however, ultimately set up different targets. They concentrate their efforts on:

- primary instead of secondary schools (Weisfeitová, 2007; Růžičková, 2008; Číško, 2009; Ježová, 2009; Musilová, 2012; Laufková and Novotná, 2014; Šťastná, 2014; Švarcová, 2017)

---

[9] Obchodní akademie, Střední odborná škola knihovnická a Vyšší odborná škola, Brno, Kotlářská 263
[10] Střední průmyslová škola stavební, České Budějovice. Anglický jazyk: Klasifikace předmětu.
[11] Ibid.; Střední škola oděvní a služeb, Vizovice
[12] Gymnázium Jiřího Ortena, Kutná Hora. Anglický jazyk: Klasifikace předmětu.

- other forms of assessment, e.g. peer and self-assessment, (Kovářová, 2007; Čuřínová, 2007; Hušková, 2010) or reports and their contrast with grading (Zedníková, 2006; Prudká, 2008; Martiník, 2010; Saňáková, 2018; Skutil, 2020; Šandová, 2020, Bláhová, 2021)
- various aspects of grading, such as motivation (Fričová, 2015; Větříšek, 2020)
- grading in subjects other than English (Lonská, 2011; Synková, 2013; Šenková, 2016; Švarcová, 2017; Nálezková, 2019) or without a specific subject in mind (Kluska, 2010)
- students' attitudes rather than those of teachers' (Langmannová, 2013; Bartuška, 2017; Zelníček, 2017; Horváth, 2020)
- a qualitative study of grading, detailing specific school policies and individual approaches (Horálková, 2013; Bureš, 2019; Král, 2020; Kučerová, 2020;) rather than carrying out a quantitative survey, i.e. any systematic, larger-scale analysis of grading in Czech secondary schools.

A quick look at the most notable literature does not seem to yield substantial results either. Ziegenspeck (1999: 26-31) only maps out a general history of grading in the Czech Republic while Kratochvílová (2011: 89-104) zeroes in on forms of assessment with the emphasis on self-assessment in Czech primary schools. Číhalová and Mayer (1997: 12-15) appear to be the only authors who conduct a survey examining the impact of grades on students' motivation at a Czech primary school.

Ultimately, literature does not seem to consider grading in the subject of English at Czech secondary schools from the viewpoint of teachers. It appears that any valuable insight into these particular areas of grading has been missing in both academic and didactic work. It might be then safe to say that such lack of coverage, alongside personal interest, could fuel a need for the following research.

The primary aim of this thesis is therefore to investigate the extent of a systematic approach to grading in the Czech Republic at a secondary school level. The focus will be on institutional and teachers' subjective biases that can cause differences in approaches when it comes to grading. In order to determine the level of variation in the system of grading at secondary schools, previous studies on grading irregularities will be considered. More

specifically, the research in this thesis will be based on Ziegenspeck's notes in *Chapters 2.10.1, 2.10.2, 2.10.3* and *2.10.4*. The grading scale, institutional issues, gender stereotypes and subjective element as potential problems that may lead to a non-systematic grading approach are all discussed in these chapters. Ziegenspeck's notes in *2.10.5* on how to avoid the problem of subjectivity in grading will also be taken into consideration. Also, some of the teacher approaches to assessment that Bilanová et al. introduce in *Chapter 2.5* will be examined. Further, *2.10.6*, where Šikulová, Slavík, Košťálová et al. and Kratochvílová comment on grading standards, will serve as a guide for further investigation to help explore the level of variation in the system of grading.

The research will thus try to find out to what extent these issues and biases are present in Czech secondary schools as well as examine how much the grading standards are being followed. This can help establish whether grading is mostly systematic on an individual (teacher) and/or institutional (school, *MŠMT*) level within different variables (teacher's length of experience, type of school and size of school). Directly reflecting on the aforementioned relevant chapters, this study is divided into five areas of research:

1. What are some of the problematic areas within the convention of the grading scale?

Do all grades on the scale carry the same weight or are they weighted or marked as better/worse (e.g. 1+, 3-, …)? Can some grades be 'erased', forgotten and replaced?

2. What are some differences and issues in grading on an institutional level?

What activities are being graded, and how, at different institutions? Who determines grading criteria? Are there clear guidelines on what and how to grade?

3. To what extent does the 'subjective element' play a role in grading?

To what level are teachers aware of their own bias? Do teachers award a better grade to motivate students? Are they influenced by the overall impression, grade repercussions or even by student gender when grading? Do teachers grade parallel classes in the same way?

4. Do teachers and institutions follow expert defined grading standards?

Do teachers announce grading criteria before assessing? Are, for example, essays corrected anonymously? Is peer and self-assessment included in grading and finally, are teachers open to a discussion before awarding a grade?

5. Are teachers supported in grading?

Are teachers given clear instructions on how to grade – for instance, in the form of a workshop? Are there explicit guidelines for teachers that determine grading criteria or grade weight?

These five areas of research into systematicity of grading are by no means an exhaustive coverage of the problem. However, the apparent gap in prior research on this topic, albeit a limitation in itself (see *5 Conclusion*), can perhaps give us an opportunity and relative freedom to decide what to primarily focus on. Any additional research questions will be tackled as they occur.

3 Methodology

For the purposes of this thesis 279 teachers were surveyed about their grading practices and approaches at different types of Czech secondary schools. Out of the 279 teachers, 154 were 'gymnázium' teachers, 110 taught at 'Střední odborná škola' (SOŠ) and 15 teachers worked at 'Střední odborné učiliště' (SOU). The survey was weighted in favour of 'gymnáziums' due to an initial shortage of responses from this type of a secondary school. The limited number of SOU responses reflected the relative low number of this type of school in the targeted location.

Ultimately, 75 teachers responded to our questionnaire, giving a return rate of 27%. Of these responses, 38 came from 'gymnázium' teachers (50.7%), 35 from SOŠ teachers (46.7%) and 2 teachers' replies were sent from SOU schools (2.7%). The teachers' age and gender were disregarded. Only their nationality (Czech) and teaching experience were noted. Under experience, teachers were divided into 3 categories: *not experienced* (teaching less than 2 years), *moderately experienced* (less than 10 years) and *experienced* (more than 10 years), with most of them falling under the last category. 66.7% of surveyed teachers were experienced, followed by moderately and not experienced teachers, with 22.7% and 10.7% respectively.

All teachers were contacted anonymously.
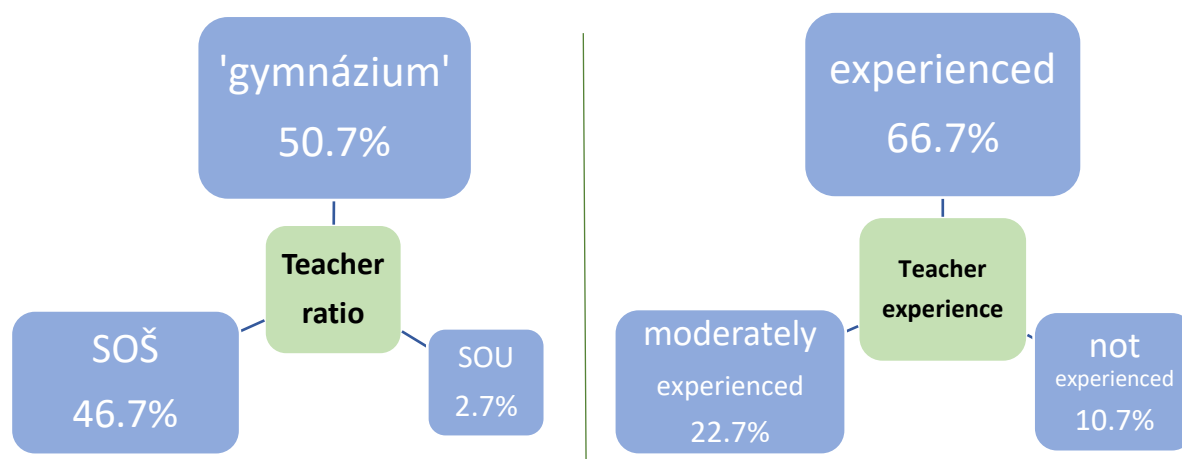

Chart 1: Surveyed teacher ratio and teacher experience

Altogether 59 schools were contacted encompassing the three main types of secondary schools in the Czech Republic. This included 30 'gymnáziums', 24 SOŠ schools and five

SOU-s. School selection was made randomly based on the secondary school lists at two websites: *www.seznamskol.eu/* and *www.stredniskoly.cz/skola/*.

Prague (97.3%) and its vicinity (2.7%) of approximately 30 km radius became the primary location for our research. This anticipated the inclusion of smaller schools (up to 500 students) with 44 responses (58.7%) but also bigger schools (between 500-1000 students) with 31 responses (41.3%).
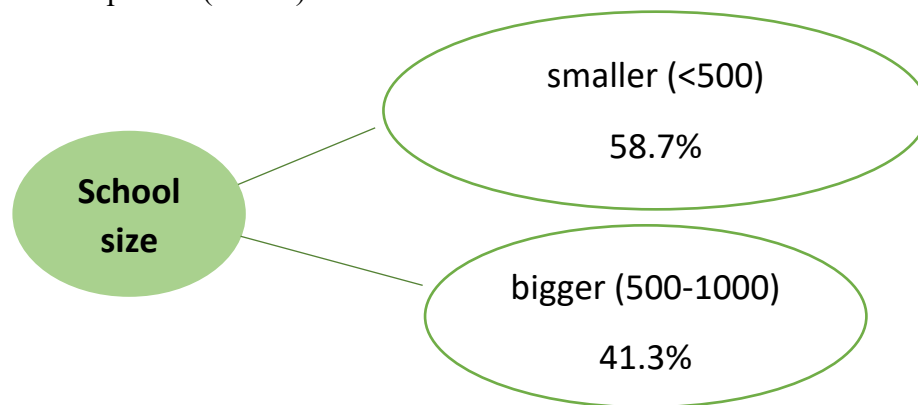
Chart 2: School size distribution

In order to examine a possible system to grading, a quantitative, rather than qualitative, study was chosen to better suit our purposes since it can generate a larger sample size and diversity among participants and thus produce good generalizability. This way a more comprehensive and systematic picture of the issue could be established. With that in mind, a Google Form-questionnaire (online survey) was chosen to help pursue this exact objective.

Our questionnaire format leaned heavily on Oppenheim's notions and suggestions (2001: 5-68, 100-149). We utilized a descriptive design of a survey containing mostly closed factual questions but with a possibility of a commentary. This cross-sectional study was conducted with a highly diversified sample of participants in terms of independent (experimental) variables (teacher's years of experience, school type and size). We strictly controlled the variables of teacher nationality (Czech), school location (Prague + 30 km) and school level (secondary).

38 questions were included in the survey (see *Appendix V: The Questionnaire*). Four demographic questions (years of experience, school type, school size, school location) were followed by a further 34 questions regarding teachers' grading approaches. A few of these questions served to introduce a new research area (e.g. 'Do you grade essays?'), several served as a series of internal checks to ascertain the reliability of the responses (e.g. 'Which

of the following factor(s) can influence the essay grade the most?') and a number of the questions were primarily used for the teachers' subsequent comments (e.g. What do you grade in presentations?)

The content of the questions reflected the five areas of research already mentioned in *2.10.8* This means the questions were based on the theoretical issues raised in chapters *2.10.1 - 2.10.6.* Focusing on oral and written skills to check for teachers' preferred forms of assessment (*4.4 Research Question no. 4: Question no. 2*) was purely arbitrary.

Grading guidelines in a number of Czech secondary schools had also an influence on the process of question formation (see *2.10.7*). Moreover, a consultation with a retired language teacher helped modify the survey questions appropriately.

The answer selection was then based on a series of semi-structured exploratory interviews with the same teacher. Assessment requirements for Maturita State Exams in the English language also served as guidance for some of the answer choices.

After its initial design, the questionnaire was piloted using a smaller number of volunteers (two teachers). This helped pinpoint any potential future problems caused by unclarity or ambiguity and adjust the questionnaire accordingly. While ordering the questions, we attempted to funnel them from the most generic down to the most specific as well as alternate longer and shorter / open and closed questions to keep the participants focused and involved (Oppenheim, 2001: 101). Afterwards we evaluated this email, self-administered questionnaire through descriptive statistics and charts/tables. All questionnaires returned to us were included in the study.

Nevertheless, not all answers to every question were included in our results. In *4.1 Research Question no. 1: Question no. 1*, the responses that offered a range for the lower (or upper) limit of a grade instead of an exact figure were dismissed to keep statistical accuracy. Any vague or contradictory responses that lacked clarity were dismissed, too. Also, any further inconsistency in teachers' responses resulted in rounding the percental sums to account for any error (e.g. *4.2 Research Question no. 2: Question no. 4* – the number of teachers who set their own grading criteria by themselves at least some of the time). Finally, the results could be subject to a possible margin of error due to the human factor.

4 Survey Results

In this chapter, we will look at the teachers' responses to the 34 questions used in our questionnaire. The questions will be grouped according to the five areas of research even though some questions might fall under more than one category, a factor that will be addressed in the discussion.

4.1 Research Question no. 1: Convention of the Grading Scale

There are five questions that deal with the controversies concerning the grading scale.

***Question no. 1*:**

- **What is the lower limit in percentages for grades 'výborně', 'chvalitebně', 'dobře' and 'dostatečně'? (e.g. 'výborně': 100% - 90%)**
- **What is the upper limit in percentages for the grade 'nedostatečně'? (e.g. 50% - 0%)**

Numerous studies show a significant variation in grading based on the type of secondary school (Weiss [in Ziegenspeck, 2002: 68-9]). Our open-ended question tried to find out from 'gymnázium' and SOŠ teachers where their limits for each grade lie on a percentage scale. The results were surprising in their great diversity, not only between the two types of schools but also within the same type of school. 67 teachers answered the questions, some responses being necessarily omitted for various reasons (lack of clarity, no numerical answer, a range rather than an exact lower limit).

In 'gymnáziums' the lower limit for the grade 'výborně' in percentages ranged anywhere from 91% to as low as 80%.

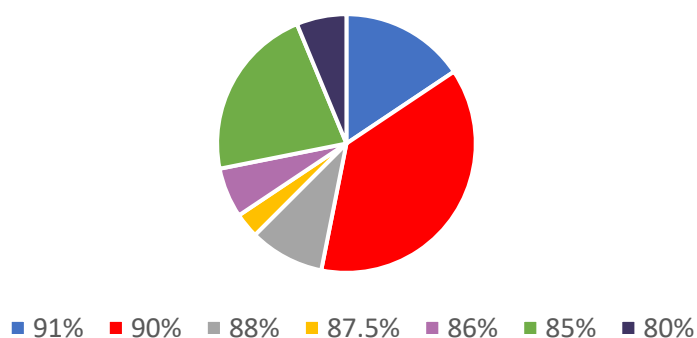Lower Limits for 'Výborně' - 'Gymnáziums'

■ 91% ■ 90% ■ 88% ■ 87.5% ■ 86% ■ 85% ■ 80%

Chart 3: Lower limit for grade 'výborně' in percentages at 'gymnáziums'

As seen in *Chart 3*, the most common lower limit for the grade 'výborně' was 90% with 12 responses. The lowest limit of 80% was represented by two teachers in two different 'gymnáziums'. Interestingly, both of these teachers were highly experienced (10+ years) and taught at large secondary schools (500-1000 students). One of the teachers expressed their unease and discomfort with grading in terms of percentages. The teacher stressed that various other factors, such as length and difficulty of the graded activity, played a significant role in awarding a grade to a student and these were harder to mathematically assess. Even more interestingly, this particular teacher revealed that they were the ones who set up the grading criteria exclusively, including the percentage scale for grading.

Quite noteworthy was also the fact that there were two teachers in our survey who differentiated between 'nižší gymnázium' and 'vyšší gymnázium' in their grading percentage scale, being stricter with the latter (85% vs 90% respectively). These findings only confirmed Hopp and Lienert's earlier results in the same area of research between different 'Gymnasium' levels ([in Ziegenspeck, 2002: 68]).

But perhaps most compelling was the response from a teacher who, according to his own words, was awarding grades based on a student 'meeting the deadline and task criteria' only. The quality of work was then assessed separately, formatively and interactively with students, oftentimes using peer and self-assessment. This particular teacher awarded the grade 'výborně' for purely submitting work on time or if the student apologizes in advance. The grade 'chvalitebně' was given for 'work submitted later than 24 hours [from the deadline]', grade 'dobře' then for 'work submitted later than 48 hours' and 'dostatečně' was given when the work was handed in later than 72 hours from the original time frame. Finally, 'nedostatecně' was awarded if no work or apology was produced.

Similarly, the situation was extremely diverse at SOŠ schools. When compared to 'gymnáziums', SOŠ schools' limit for the grade varied even more dramatically. The lower percentage for grade 'výborně' fluctuated wildly between 90% and 81%. A SOŠ teacher surveyed pointed out that 'grading criteria in Years 1 and 2 are milder while they become stricter in Years 3 and 4.' This reaffirms Weiss' research at Austrian secondary schools that observes how grading becomes stricter and stricter with each passing Year (Weiss [in Ziegenspeck, 2002: 68-9]). Even other SOŠ teachers concurred that grading criteria can slightly change in 'few cases' and that there were 'big differences' between classes in terms of percentage scales.

Still, other subtle idiosyncrasies between the two types of secondary schools could be detected. While for most 'gymnázium' teachers, 90% was the most common lower limit for the grade 'výborně', most SOŠ teachers, on the other hand, considered that limit to be 88%. If we were to suggest, however, that grading requirements at SOŠ were more benevolent than in 'gymnáziums', we might simply not be correct. As a matter of fact, the absolute lowest percentage for 'výborně' at SOŠ-s (81%) was still higher than the lowest limit for the same grade at 'gymnáziums' (80%). Not to mention, that the average lower limit for both school types was the same at 88%, thus establishing relatively corresponding standards in both.

Lower Limits for 'Výborně' - SOŠ

Legend: 90% | 89% | 88% | 87-87.5% | 86% | 85% | 83% | 81%
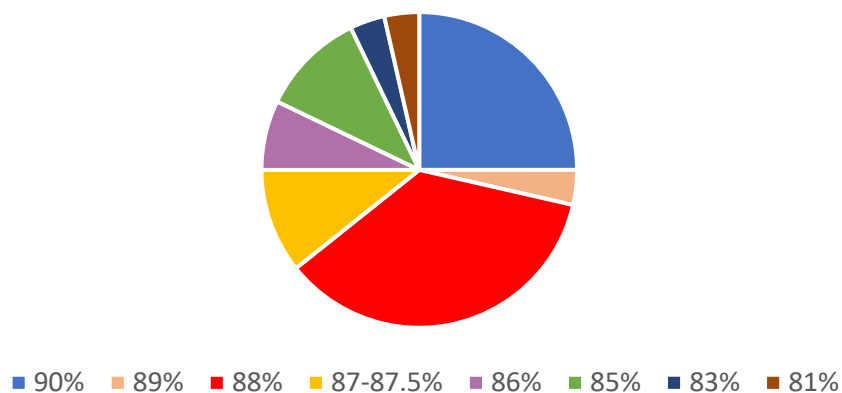
Chart 4: Lower limit for grade 'výborně' in percentages at SOŠ

What is most striking is how much flexibility some teachers have at their schools when determining the percentage for each grade. Many surveyed teachers had a clear vision and provided exact numbers to this question in a straightforward manner. They would often quote following their school management or CERMAT official recommendations, even though some teachers might disagree with such guidelines ('I personally consider this percentage range [from CERMAT] too benevolent'). On the other hand, there were a few teachers who had relative flexibility when it came to percental grading ranges. Some would disclose they had an option to adjust the scale (usually more generously for lower Years) or come up with their own. Such teachers would argue the question was 'difficult to answer', that they 'do not like percentages' and the grading scale would 'depend on the type/difficulty/length of the task'. Strikingly, this inconsistency in ground rules appeared to take place across both 'gymnáziums' and SOŠ-s, smaller and larger schools and among experienced or less experienced teachers.

Remarkably, we happened to obtain two responses from the same SOŠ where the percentage range for some of the grades differed substantially. Although the grade 'výborně' matched in terms of the lower percental limit in both cases (85%), it was quite striking that the ranges for all the remaining grades deviated from each other significantly. While the lower limit for 'chvalitebně' with Teacher 1 was 70%, it was only 60% with Teacher 2. The difference of 10% remained consistent between grades until 'nedostatečně'. To award the grade, Teacher 1 appeared stricter and expected their students to achieve at least 39%. Students in Teacher 2's classes, however, had to succeed only at 24%. The question remains why such a major variation is happening under the same school's roof. Did the two teachers know about this paradox? Also, did the school management know and if so, were such discrepancies allowed or even actively encouraged? The SOŠ itself was located outside of Prague and was of a smaller size of just under 500 students. This could suggest (although not necessarily) a more individual approach to students' needs and explain the diversity. Looking at the teachers' profile, we noticed several similarities. First, both teachers were relatively inexperienced having taught for less than 2 years. Both of them also relied on an 'overall impression' when they graded, e.g. presentations. Both teachers also set the same criteria for the parallel class they teach which further magnified the impact of the teacher's individual idiosyncrasies in grading. But, perhaps most interestingly, we read that both teachers set their own grading criteria, with one also relying on the advice of more experienced colleagues. It appeared then that the management did not seem to, for better or worse, interfere with students' grading at all and hence, this anomaly of subjectivity within one secondary school.
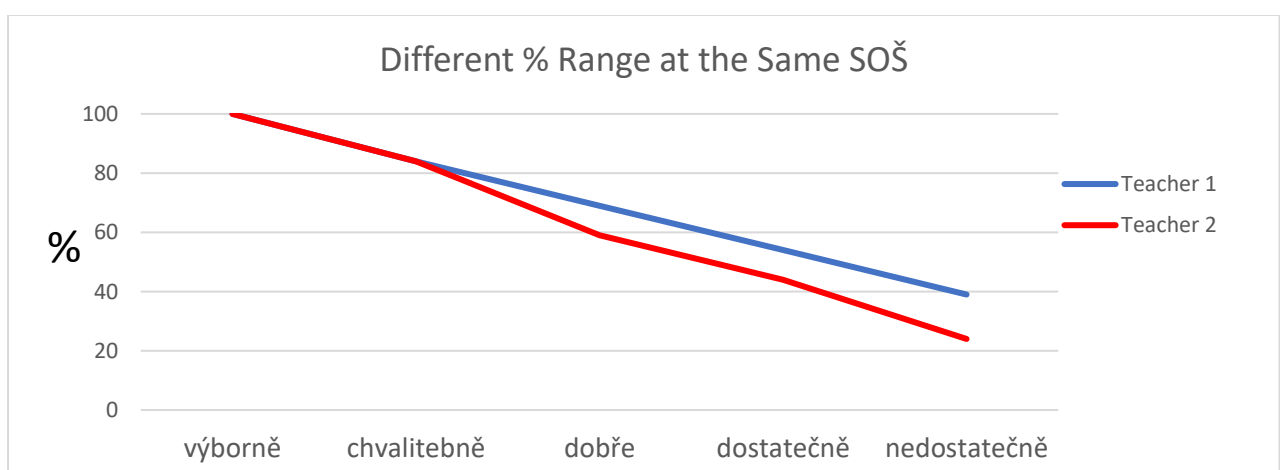

Chart 5: Different grade range in percentages at the same SOŠ

This brings us to consider the percentages for the lower end of 'chvalitebně', 'dobře' and 'dostatečně'. For both 'gymnáziums' and SOŠ the lower limits for grades were highly

diverse. It seemed each school had their own idea what the correct percental range for each grade should be and this created a lot of variety across both types of schools. Let us consider 'gymnáziums' first. Here, the lower limit for the grade 'chvalitebně' could range anywhere from 81% to as low as 65%. Ultimately however, the most common lower limit for the grade was 75% (nine responses) with the overall average for the lower limit also 75%.[13]

In comparison to 'gymnáziums', the lower limit range for a SOŠ grade 'chvalitebně' dipped even lower, from 80% down to a mere 60%. Somewhat surprisingly, the most common lower limit for this grade at SOŠ matched the 75% at 'gymnáziums' (six responses). However, the average lower limit was slightly lower than at 'gymnáziums', at 73%. This last result certainly fed into the stereotype that SOŠ school requirements are less demanding and grading might therefore be more benevolent. Even though sample sizes can always be larger, our 31 SOŠ responses should not be discarded as purely circumstantial. This sample number could arguably be considered representative enough to provide a relatively statistically significant and tangible conclusion confirming the common belief.

A very similar situation occurred with grades 'dobře' and 'dostatečně.' They also backed up the popular theory that at 'gymnáziums' grading is slightly stricter. At the same time, once again, the lowest limits for both grades in both types of school varied enormously. For example, the lowest limit for 'dobře' at 'gymnázium' extended from 71% to 45% almost mirroring SOŠ (70% - 45%). All considering, this could create ample opportunities for students of varying degrees of acumen and knowledge to obtain essentially the same grade. Both 'gymnáziums' and SOŠ shared the same most common limit of 60% but yet again, overall the average was higher in 'gymnáziums' (62%) than in SOŠ (59%).

'Dostatečně' followed the same trend established with the previous grades. Even here, we saw a lot of variation but SOŠ percental ranges were generally lower. For this grade to be awarded, the lowest limits oscillated between 61% - 30% ('gymnáziums') or 55% - 25% (SOŠ). The most common lower limit for this grade was 60% ('gymnáziums') or 50% (SOŠ). Yet, the overall average fell in both cases to a mere 49% ('gymnáziums') and 44% (SOŠ).

Once again, we saw several schools modifying the grade percentage limit making it stricter as the school Years progressed. An experienced SOŠ teacher reported: 'The passing line

---

[13] A 'gymnázium' teacher's response that indicated lower limits for grades 'výborně', 'chvalitebně', 'dobře', 'dostatečně' and 'nedostatečně' as 50%, 35%, 10%, 5% and 0% respectively was omitted from our analysis due to the extreme values that could skew the overall average distribution. These grade lower limits might imply a some form of non-mainstream approach to grading at that particular school.

between '4' and '5' is different in different Years (50-55-60% [getting stricter in the upper Years]).'

This takes us to the grade 'nedostatečně' and the upper limit at which it was still awarded. There was a wide scope of approximately 30% between the individual upper limits in different schools, stretching from 29% - 60% at 'gymnáziums' and 24% - 54% at SOŠ schools. This was a relatively broad scope which replicated the one in the previous grade, 'dostatečně'. The five most common upper limits for the 'nedostatečně' at both types of school can be seen in the chart below. While most 'gymnáziums' believed the grade 'nedostatečně' should be still awarded at 60% and lower, the passing line at SOŠ schools was lower, with an equal number of schools quoting 49% and even 43% (see *Chart 6*).
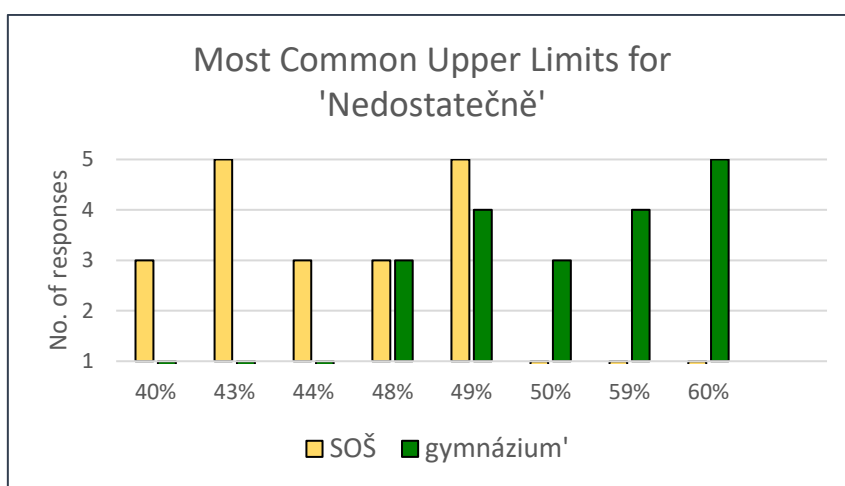
Chart 6: Five most common upper limits for grade 'nedostatečně' in percentages at 'gymnáziums' and SOŠ

Again we saw a stricter requirement for task passes at 'gymnáziums' on average by 9% with 54% and 45% at 'gymnáziums' and SOŠ respectively.

Overall, it seemed that the system of lower and upper limits set for grades by the secondary schools was fairly uncoordinated with a high level of individual approach. Each school, seemingly arbitrarily, presented their own criteria either collectively or, to a degree, relied on the teacher to decide for themselves.

Still, there was a consistency that revealed itself as the two school types were being compared. 'Gymnáziums'' standards appeared consistently stricter, with an increasing tendency with each grade. While with 'výborně', the average lower limit was the same for both school types indicating some correspondence in grading standards, the difference

between the two school types gradually increased with all the other grades, up to 9% for 'nedostatečně'.

GRADE	'GYMNÁZIUM' average lower limit	SOŠ average lower limit	DIFFERENCE in average lower limits
'výborně'	88%	88%	**0%**
'chvalitebně'	75%	73%	**2%**
'dobře'	62%	59%	**3%**
'dostatečně'	49%	44%	**5%**
'nedostatečně'	54%	45%	**9%**[14]

Table 1: Difference in lower limit for all grades between 'gymnáziums' and SOŠ schools

The results in *Table 1* then ultimately corroborate the stereotype of more demanding grading norms at 'gymnáziums' in comparisons to SOŠ schools.

Also consistent, though, was the increasing rate of the lower limit range in both school types. While the lower limit range for 'výborně' was approximately 10%, it was roughly 20% for the grade 'chvalitebně', around 25% for 'dobře' and circa 30% range for both 'dostatečně' and (upper limit range for) 'nedostatečně'. This could indicate that the lower the grade, the more inconsistency and disparity between schools there is since the range of all possible percental limits is spread wider among schools.

Yet, *in Chapter 2.10.1* Holzinger had claimed that an extreme value, such as with 'nedostatečně', was likely to be distinguished from others with more certainty. Teachers would supposedly have no problem in defining it. On the contrary, they would allegedly do so even more easily. This would subsequently create an objective and systematic approach to grading (Holzinger [in Ziegenspeck, 2002: 65]). However, this theory does not ring true when we look at our results. We witnessed that the boundaries not only between 'moderate' grades ('chvalitebně', 'dobře', 'dostatečně'), but also the extreme 'nedostatečně' were immensely blurred. Determining these boundaries with certainty proved to be particularly difficult which sadly, casts some doubt on any systematicity within grading.

---

[14] For the grade 'nedostatečně', only the five most common lower limits for each type of school were considered

From our research we could thus confirm grades can be affected by the type of secondary school but could the same be said about school size? Previous studies noticed that small secondary schools were prone to award more lenient grades than larger schools (Starch and Elliot, 1912: 457). We investigated this hypothesis in the context of the surveyed Czech secondary schools and, with some surprise, were able to confirm it. Indeed, the grading in smaller schools was softer with their percentage limits for all grades being lower than in bigger schools. Not only that, we also further inspected smaller and bigger 'gymnáziums'/ SOŠ. What we found was quite striking – smaller 'gymnáziums'/ SOŠ schools were on average more lenient in their grading. Bigger 'gymnáziums'/ SOŠ, on the other hand, were stricter, with their grade percentage limits being higher for all grades (see *Appendix II*). It seems that a student for the same performance would get better grades, should he/she choose a smaller secondary school. Surely, more research is needed to try to clarify why this phenomenon occurs.

Question no. 2:

- **Do you use plus/minus grading? (e.g. 2+, 3-, …)**

As a mathematical element, using numbers for grading should convey a clear, unambiguous and exact message. Yet, when utilizing numbers as a performance assessment, grades become vague and debatable (Flitner [in Ziegenspeck, 2002: 61]). Since the grading scale is not an interval scale with the distance between each grade precisely the same, no scientific-like objectivity could be achieved. Differentiating the grading scale, however, could balance the unfairness and subjectivity (Ibid.; Starch and Elliott, 1912: 442).

Symbols, such as 'plus' (+) or 'minus' (-) written as a suffix after a grade oftentimes discern a 'higher' or 'lower' form of performance within each grade. In our research, we were interested whether the culture of plus/minus grading remains the same across different classrooms. Are there teachers who award grades that are slightly 'higher' or 'lower' than their corresponding 'neutral' numerical value? How do such grades then translate in the final grade? How does *1-* compare to *2+* in different classrooms and are both signs even equally used? Could using symbols get us into even murkier waters if used inconsistently across secondary schools? Surely after examining the inconsistency in percental grade limits in *Question 1*, any unpredictable or slightly erratic trend in grading could further magnify the already shaky value of a grade. Since all 75 teachers answered our question, we hoped to get

a good idea of how popular plus/minus grades are among teachers and how systemically regulated these symbols are.

Looking at the response data, there seemed to be an overwhelming need from teachers to accompany grades with plus/minus signs. 76% of 'gymnázium' and 81% of SOŠ teachers used plus/minus grades at least sometimes (29 and 30 responses respectively). Perhaps more interesting for us was to find out what kind of teacher would prefer plus/minus grading and so we reviewed the figures based on the teacher's teaching experience. What we found out was major differences in the two types of schools when experience is recognized.

Taking 'gymnáziums' first, the less experienced the teacher was, the more likely they were to use the plus/minus grading system. While 'gymnázium' teachers with less than two years under their belt fully promoted the use of +/- grading, the most experienced teachers at 'gymnáziums' were, however, most reluctant to use it, with every 3rd rejecting the practice. In SOŠ schools, the situation was reversed. It was the teachers with 10+ years of experience who were implementing such grading the most. Almost 90% of them approved of a plus/minus grading system. Yet, almost every other moderately experienced teacher (less than 10 years of experience) was not ready to extend the scope of a grade through +/- signs.
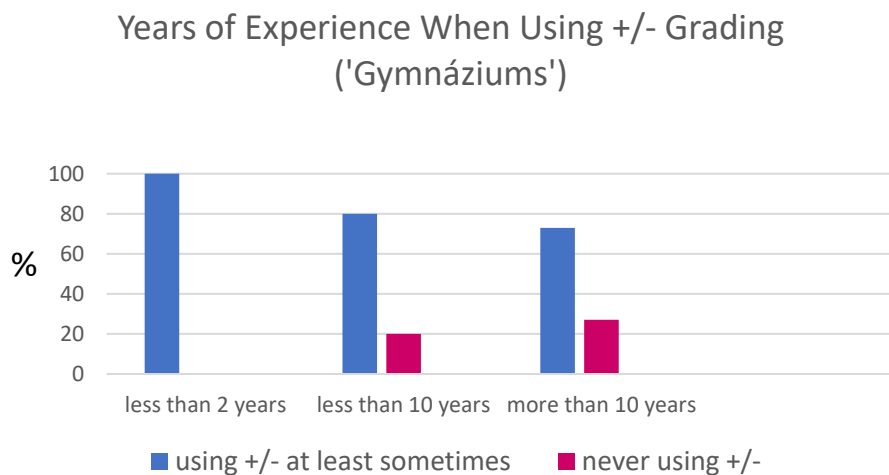

Chart 7: Years of experience when using +/- grading at 'gymnáziums'
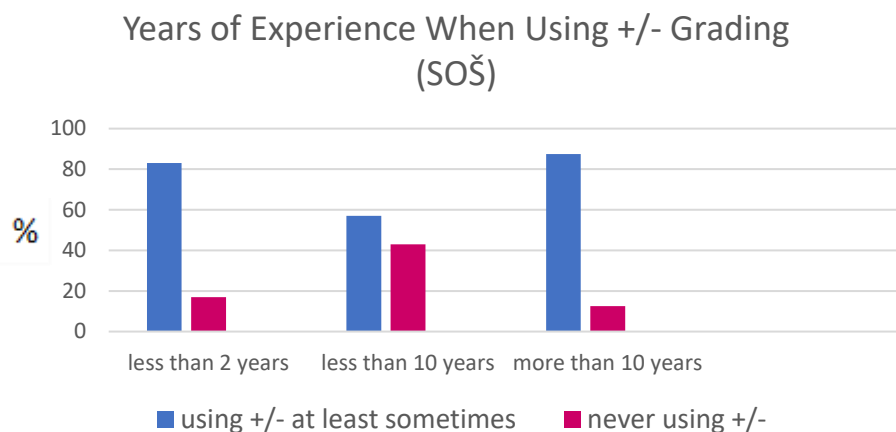
Chart 8: Years of experience when using +/- grading at SOŠ schools

In order to understand why teachers do or do not take advantage of this grading scheme, we scrutinized their written commentaries to learn more. Most of them quoted percentages as a precise enough tool to capture the slightest nuances in student performance thus eliminating the need for further differentiation through symbols. In some classrooms, firm rules about how to transform a percentage scale directly onto a +/- grade had been established. In one school, 90% - 94% would always directly translate to grade *1-*. In another, 85% - 90% would become *1-*. Finally, there was a secondary school in our survey which saw round percental values, such as 90%, 80%, 70% and 60% as directly corresponding to grades *1-*, *2-*, *3-* and *4-*. This partial overlap means that grade *1-* would spell three different concepts to three different sets of students. Yet, all three schools were 'gymnáziums' and with each clearly having their own grasp on the +/- grading system, bearing no resemblance and even clashing with one another.

Such precise and detailed standards were, however, not observed in other schools. As a matter of fact, there were some teachers who were set on avoiding plus/minus grading completely. Rather, these teachers strongly preferred written commentaries or giving extra tasks to students in case the final grade was hard to determine.

Interestingly, those teachers using +/- grading frequently seemed to favour the '-' sign. This conviction appeared to be exacerbated by teaching predominantly online at the time or by the technical inability of the online grading platform 'Bakaláři' to recognize a '+' sign. Some teachers were treating +/- as a personal assist and, yet, would never enter grades embellished this way into official school documents, platforms or electronic systems. There were some educators that even felt guilty about using plus/minus grading: 'I use 'minuses' when I grade.

But I think I should not be doing so – a five-grade scale should be sufficient for grading.' It was quite the opposite for other teachers who fully embraced plus/minus grading simply because, in their mind, 'a five-grade scale does not sufficiently capture subtle differences in performance.'

What we found out is that the phenomenon of using symbols to expand the value of grades has become extremely common which could indicate that teachers feel the need for such scale differentiation. Based on theoretical knowledge, any extra differentiation of the grading scale can only be a good thing, promoting transparency and objectivity in grading. Inconsistency in the usage of plus/minus grading across secondary schools might be the only issue.

Question no. 3:

- **Are students allowed retakes? If so, what happens to the original grade?**

It was the first part of the question that produced some of most uniform and consistent results in our whole survey. Out of 75, only three teachers would not grant a retake to a student. All three were experienced teachers (teaching for more than 10 years) and came from bigger schools (500-1000 students). Two of them taught at 'gymnáziums' and one at SOŠ. The rest of the surveyed teachers saw eye to eye and at least sometimes, allowed retakes in their classrooms.

What stirred up major controversy, however, was the second part of our question. There was quite a deep division among our surveyed teachers on what to do with the original grade after a retake. Some wanted to replace the original grade with the new one. Others would rather add the new grade next to the original one. However, more than 40% of teachers admitted that they, in fact, did not follow just one practice on how to proceed after a retake. In their classrooms, the new grade could either replace the original one or be added next to it. Such a mixture of different approaches might be confusing. Some students could feel disadvantaged, others might reap benefits. It most certainly can be perceived as conflicting. This ambivalence could also very easily have an impact on final grade results, especially if an arithmetic average from individual course grades is counted to determine the final grade. In that case, if a new grade is added to the original one, will both be counted towards the final grade? If, however, the older grade is replaced, will it be forgotten? What makes teachers decide one way or the other? Is it the grade itself, its value or weight, the type of activity or

teacher's benevolence? Does one teacher's behaviour follow a pattern or is it decided on a whim? Further research into this problem would have to take place to answer these questions.

New Grade after Retake

- it replaces original grade
- it is added to original grade
- sometimes replaces, sometimes is added to original grade
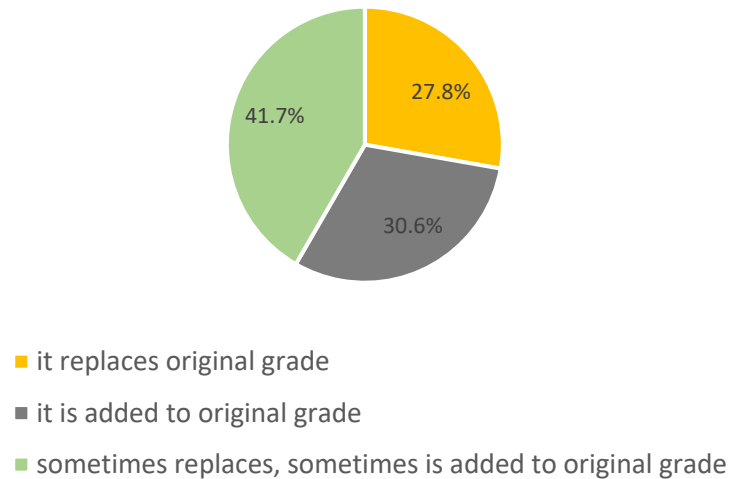
Chart 9: New grade after a retake

Finally, *Chart 9* clearly underlines how diverse their approach is when secondary schools deal with this issue. This indicates a complete lack of unified, coordinated strategy across secondary schools and, until tackled systematically, will only result in serious inconsistencies that have the potential to hurt students. There is some food for thought in what an experienced teacher from a smaller 'gymnázium' wrote in her questionnaire: 'I don't award a grade until the student becomes adamant that the performance was his/her maximum and that he/she cannot improve it further. I always leave the option open to redo their work (even several times).' How realistic this suggestion might be in practice, in terms of time, classroom management and work overload, is uncertain. But perhaps only grading the student's 'best possible version of himself/herself' might save teachers (and students) some grief around retake grades.

Question no. 4:
- **How many grades do you award a student per semester?**
- **Do you count an arithmetic average from individual course grades to determine the final grade?**

To discuss the counting of an arithmetic average when determining the final grade, we first should pay attention whether teachers actually have enough grades for the average to be relevant. After examining all 75 responses, we found that all but one teacher awards more than five grades per semester. Only a single experienced teacher working at a smaller Prague 'gymnázium' marked in the survey that he/she were awarding students with fewer than five grades in a single semester. Interestingly enough, he/she also did not count an average to conclude the final grade. As a result we could safely remove the participant from further analysis of this issue.

As for the other responses, the majority of teachers, that is 68%, gave out a relatively large number of grades (eight and more) while 31% kept the number of grades awarded anywhere between five to seven. Focusing on the first group, out of the teachers that award eight and more grades, most of them were experienced (almost 70%) and most (just over 60%) came from smaller schools with less than 500 students.

Since grades are notoriously inaccurate measuring values (Ibid.; Starch and Elliott, 1912: 442), some argue strongly against the counting of an arithmetic average to obtain a final grade. Ziegenspeck talked about this grade average counting only as an 'emergency' solution due to the lack of any other mechanisms (Ziegenspeck, 2002: 37; 49-52). However, after examining the questionnaire responses, we discovered only 18 examples (24%) when teachers/schools did not use the counting of a grade average. The majority relied on this mechanism at least some of the time. Moreover, Ziegenspeck kept reminding that each grade should encompass not only student knowledge but also their overall performance and even effort (Ibid.). A lot of teachers in our survey saw the counting of an arithmetic grade average as a stepping stone, a 'starting position' as one teacher put it, in determining the final grade. They disclosed that a student's effort, classroom participation or rigour could positively influence the arithmetic average. Even a student's personal problems or teacher's own impression of the student could have an impact on the grade average counted (see *Chart 10*). Two teachers kindly revealed to us some of the above-mentioned factors were behind their improving the grade from an average of 2,65 ('dobře') to '2' ('chvalitebně'). Other teachers mentioned having the freedom of +/- 15% to bend the grade 'if necessary'.

Factors Influencing Grade Average
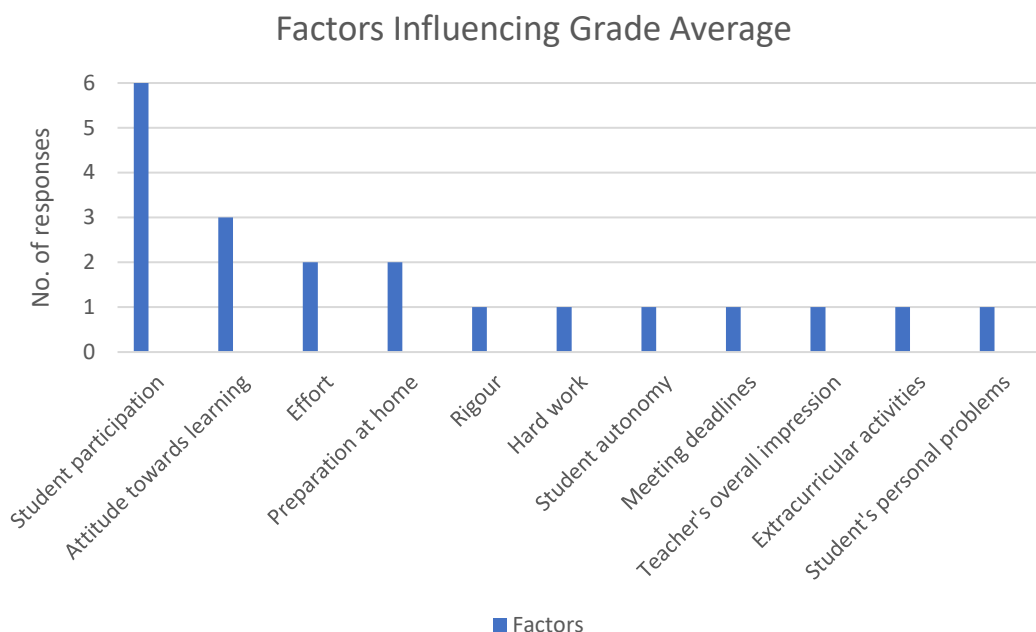
Chart 10: Factors influencing an arithmetic grade average (as disclosed by some teachers)

Only one teacher acknowledged that factors, such as the ones in *Chart 10*, could either improve or lower the student's final grade. Yet, most polled teachers who had the option to adjust the final grade from the counted average, admitted to regularly and consistently improving it.

Not everyone surveyed, however, had this much choice or such flexibility in the matter. A few teachers were commonly expected to count either the arithmetic average or a so-called weighted average ('vážený průměr') by their school management and rigidly stick to it. The electronic system 'Bakaláři' also required no deviation from the average counted. Either way, not having a last say in the student's final result might be frustrating to some teachers and may even go against what they believe in as educators. One of our surveyed teachers voiced their grievances about having to count the arithmetic average 'unfortunately, because we are told to do so by the school management, [but] personally I don't agree with it.' Yet until a better assessment mechanism to determine the student's final grade is designed, counting an arithmetic average from individual grades might be here to stay. Still, the fact that most teachers take into account other factors beside the numerical value of a grade average, indicates that they are aware of how grades can misrepresent students' overall performance. In following expert advice, they should be credited for their determination to make grades more objective and fair.

***Question no. 5*:**

- **Do all your grades carry the same weight?**

This question, which produced the most uniform response of all, was included to find out whether a grade conveyed the same message regardless of the activity from which it was obtained. Does 'výborně' from a dictation equates to 'výborně' from the final test? All but one teacher of 75 answered 'no', confirming our hypothesis that grades vary in their weight. What does it, however, say about the systematicity of grading? How is it affected? If 'výborně' in one context (activity/teacher/school) does not mean the same 'výborně' in another, how could a systematic approach in grading be established while using the same grading scale? The available data will allow us to try to paint a picture of how different teachers/institutions assign different weight to various activities in English language classrooms.

4.2 Research Question no. 2: What Are Some Differences and Issues in Grading on an Institutional Level?

Four questions were asked to cover this section dealing with problematic areas, controversies and insufficiencies coming from or associated with the institution of the school, namely school management and alternatively The Ministry of Education, Youth and Sports in Czech Republic (*MŠMT ČR*).

***Question no. 1*:**

- **Who determines the weight of a grade?**

We already determined that most teachers assign weight to grades. Now we needed to find out whether appointing weight to grades was allied to any formal and stringent ground rules. Or rather, was weighting grades an arbitrary exercise executed solely by the teacher? To what extent did teachers have a free hand in the decision making?
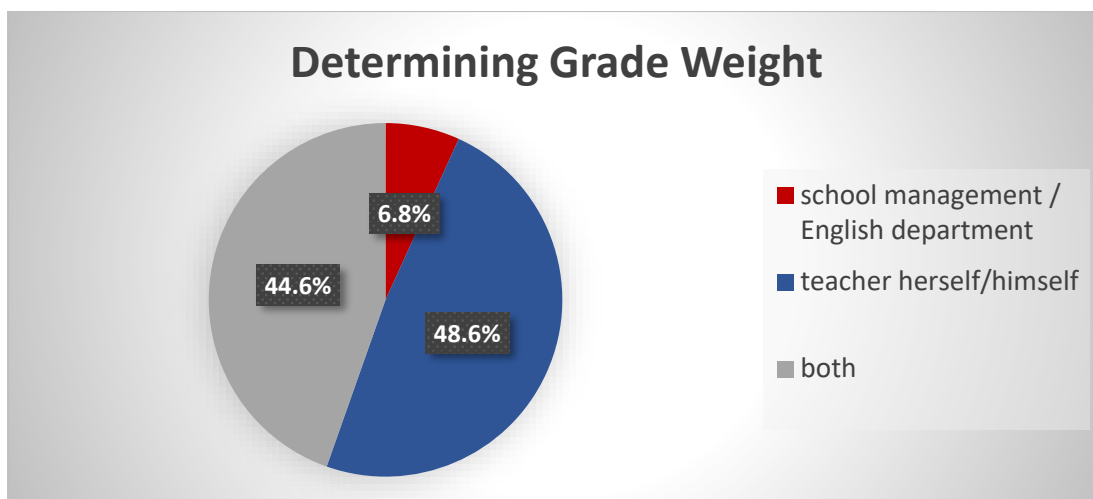
Chart 11: Determining grade weight

As seen in *Chart 11*, more than 93% of teachers from our poll assigned weight to grades themselves at least some of the time. The percentage of those teachers that could rely fully on school management did not even reach double numbers. A meager seven percent had the support of their management or their school English department to formulate official weighting rules.

Among teachers who created exclusively their own definitions for grade weight, some explained that although not mandatory, 'everyone is giving greater weight to summative tests and essays.' Another shared her own design for grade weight in which her every grade contains a weight coefficient from 0.1 to 1.0. Other teachers opened up about receiving at least partial support from their management: 'Head of the English department establishes that the weight for the active student participation in the classroom is between 1-3, I will determine myself what weight exactly it is going to be within the given range.' Finally, some English departments step in where 'relevant' work in Maturita year is concerned. Otherwise they let teachers themselves freely designate weight to each grade.

Inspecting the results, we are confronted with a pressing problem. If weighting grades is such a widespread practice and if each individual teacher is made largely responsible for its precise design and implementation, would it not bring about a massive amount of inconsistency in grading? There seems to be quite an extraordinary diversity in assigning weight to grades. How possibly could a systematic approach in grading be ensured then? We will soon examine how serious this issue can be and to what extent individual teachers, as well as schools, vary in their own interpretation of grade weight.

Question no. 2:

- **What do you grade?**

To find out more, we encouraged teachers to tell us what activities they were grading. We wanted to learn whether there was a consensus about what they assessed with a grade. We anticipated that final tests would surely be the most commonly graded activity. To our surprise, that was not at all the case. Up to 16 teachers (21%) did not grade final tests at all and we set out to investigate why this might be. An experienced teacher from a larger SOŠ offered an in-depth explanation:

> Around a year ago (before the lockdown) I've changed my grading system to grading students basically every lesson. […] Rather than just accuracy in their answers, I grade the students' overall communicative participation (willingness to react in a foreign language, express themselves in full sentences, use fixed expressions and jolt down new ones, focus on the instructions, use repetition, share ideas, think in the language and guess/come up with answers, not just mechanically fill in something [but] check spelling, use the language in practice in a productive as well as in a passive way.
>
> – response from an experienced SOŠ teacher

Out of all the possible answers for this question, this particular teacher marked down a single option – 'student participation (in the classroom)'[15]. There were two other teachers in our poll that did not grade any written tests (final, vocabulary/grammar, unit). All three, however, had one thing in common in that they all focused on grading student's work in the lesson – through student participation. The remaining teachers who excluded final tests, graded at least some form of written exams, such as shorter tests on vocabulary, grammar and/or unit tests. These became the most common activities graded.

---

[15] in Czech 'aktivita na hodině.' We specified this option in the instructions. Under 'student participation', we included not only 'student work in the lesson' but also 'student preparation for the lesson at home' (for example, homework)

Percentage of teachers grading particular activity
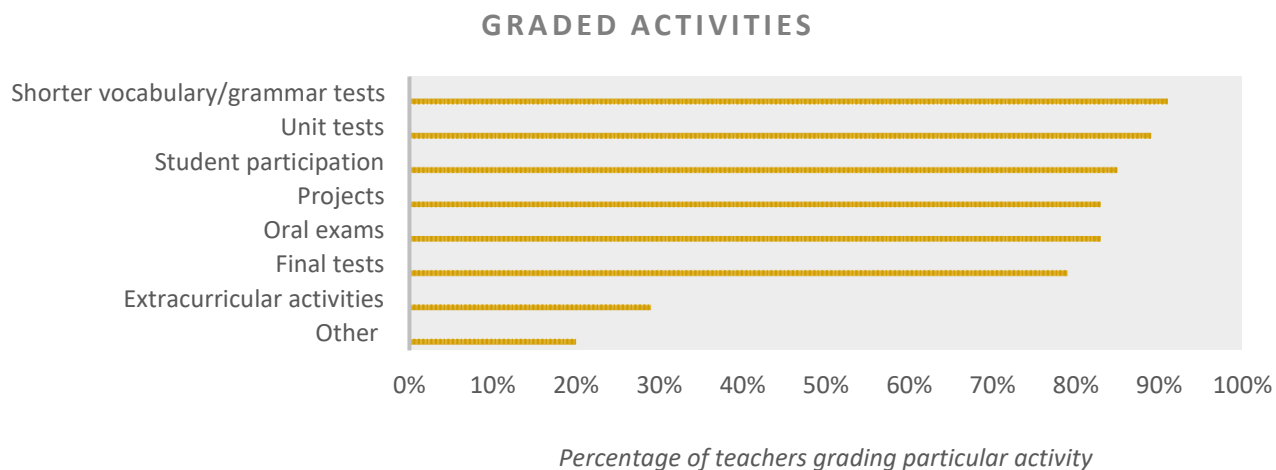
Chart 12: Most common graded activities in the surveyed secondary schools

Yet, perhaps somewhat surprising was the overall high number of participants who opted for 'student participation', with 85%, which was higher than grading of final tests, oral exams or projects. This result again illustrates that teachers do take notice of students' overall performance, that they see and recognize his/her effort just like literature recommends. It positively shows teachers recognize students' attitude and hard work beyond a pure numerical value reached in a test.

Question no. 3:

- **Grades for which activities carry the most amount of weight at the end of the academic year?**
- **Grades for which activities carry the least amount of weight at the end of the academic year?**

Due to the surprisingly high number of teachers controlling the grade weight fully (see *4.2 Research Question no. 2: Question no. 1*), this following point of discussion could certainly be incorporated within the next research area where teacher bias is discussed. We have, however, decided to include *Question no. 3* among the issues on an institutional level. After all, we would like to check for any inconsistencies in the perception of a grade awarded for different activities and to put the responsibility behind such a hypothetically haphazard situation solely on the teacher could be seen as secondary, if not somewhat unfair.

The investigation into the *Question no. 3* queries proved extremely productive and with a few surprises along the way. As expected, most teachers, 42 out of 74 (57%), put most weight on the final test grade (yellow ellipsis in *Chart 13*). What was less anticipated was the revelation

that the same number of teachers put equal emphasis on unit tests grades (green ellipsis). This meant the same amount of teachers weighted both grades equally. However, the two groups of teachers were by no means identical. They only partially overlapped. Just 4% of both groups weighted final and unit test grades equally. 30% of all surveyed teachers put most weight on final test grades while excluding grades from unit tests. On the other hand, 23% of all teachers stressed only unit tests grades rejecting grades from final tests.
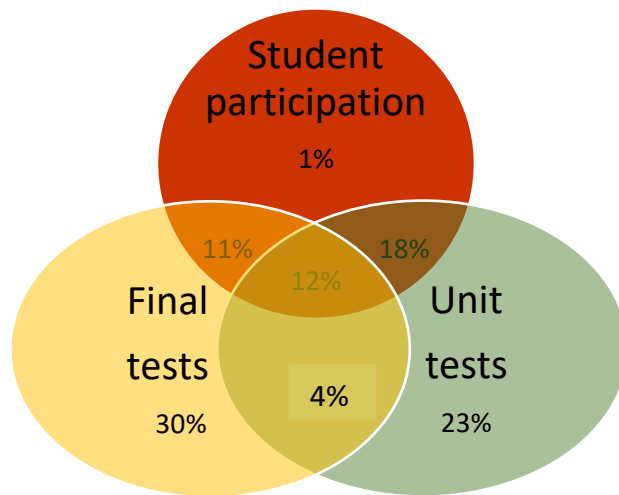
Chart 13: Teacher overlap of the three most weighted graded activities

Perhaps slightly unexpectedly, a classroom aspect that received a lot of attention in terms of weight was student participation. Quite astonishingly, the teachers assigned more importance to it than to shorter grammar/vocabulary tests, oral examination, presentations, essays or projects.

What transpired from our analysis was the sheer number of different combinations of activities the teachers weighted the most. There was little, if any, uniformity or trend. Almost every teacher appeared to focus the most on a different set of graded activities. One of them tried to explain: '[What I weight is] a combination of all [activities], depends on the magnitude of the given task. A big oral examination has a high value, a small one has a low value, [similarly] big projects carry a high value, [while] small projects carry a low value.' Another added: '[What I weight is the] combination of all the factors. But overall participation in lessons, projects and group work plays an important role.' A third commented further: 'Combination: classroom participation + completing homework or individual work + grades from oral exams/tests.' Some teachers, indeed, weighted graded activities individually and with each student separately. One teacher from a bigger 'gymnázium' with 10+ years of experience, according to his own words, weighted even student progress throughout the

semester before deciding on the final grade. This nicely complements Kolář and Šikulová's recommendations on improving assessment techniques. The authors advise teachers to make a note of the student's progress and compare it to the last performance rather than just focusing on the latest performance result in isolation (Kolář and Šikulová, 2009: 145-6).

In addition, some teachers noted they were putting most weight on activities that were not even listed in our options: tests focusing on listening/reading skills, seminar papers, mock Cambridge English exams. In the end, an SOŠ teacher with more than 10 years of experience summed up the situation well: '[Which graded activity is weighed the most] depends on the teachers.'

As for the graded activities that carry the least amount of weight, some consensus was actually reached. An overwhelming 60% of teachers in our poll agreed that grades awarded for shorter grammar/vocabulary tests carried the least amount of weight. Accounting for 30% of responses was then student participation, especially homework, because it was 'small', 'carries a supportive weight' or 'is something that students can copy from others'. Although not everyone felt the same. A 'gymnázium' teacher with moderate experience of teaching strongly voiced their discontent about having to follow the school guidelines which clearly stated that student participation was to be weighed the least. The teacher was adamant: 'But I don't agree with it.'

Finally, what was highly intriguing was the fact that the experienced teacher from a bigger 'gymnázium' who had been weighting individual student progress to add value, had also had a highly individual and unique approach to assigning the lowest weight. He/she had introduced dropping the lowest grade for each individual student, crossing it out completely from the gradebook.

Overall, it was striking how student participation found its place among the most weighted graded activities with some teachers and how at the same time it was the most unimportant activity for teachers. This general attitude happened with or without the teacher's consent. Ultimately however, the ample opportunity teachers had to choose their own weighting criteria when deciding on the final grade, generated many an individual and unique way to weight grades.

Question no. 4:
- **Who decides on grading criteria?**

- **Do you sometimes decide on grading criteria by yourself? If so, for which activities?**

Having already delved into such relevant areas as grade percental limits and activities graded, it is perhaps time to examine who is ultimately responsible for imposing rules behind grading. Who decides what is being graded and how? Who selects the grading scale? Who chooses what is going to be tested and how it is to be assessed? The following considers the absolute cornerstone of the grading process, namely the grading criteria.

Arguably, analysis in this area might give us a deeper insight into the inner workings of the grading system in Czech secondary schools. We were curious about how consistent and standardized grading could be by analyzing who is behind the creation of grading guidelines. We were expecting the school management or English department to assume control in this area which could reveal a systematic approach at least within the particular school. If, however, a teacher themselves was in charge of setting up the grading criteria, then we monitored for any student/colleague involvement or consultation with teaching materials. Experts like Ur (2004: 291-3) and Košťálová et al. (2008: 106-7) urged teachers to ideally involve students in the process of creating grading criteria. Students were supposed to be viewed as teacher's 'partners' and so we expected to see their presence in the teachers' responses. More experienced colleagues or even instructions in teaching materials might also help 'compare notes' so to speak and help offer perhaps a more detached approach to composing the grading criteria. If the responses, however, indicated no such interactions and the grading criteria would be rested solely on the teacher's shoulders, the risk of subjectivity and inconsistency would increase.

Despite prior expectations, the school management or English department exclusively imposed the grading standards in only roughly 7% of the cases (five out of 70). Most teachers, 24%, had some form of help from the management/English department but otherwise were fully responsible for delivering grading criteria on their own. Yet, in accordance with expert recommendations, all the teachers unanimously engaged their students in the process, too. The second largest group of teachers (21%) was fully in charge of creating their grading criteria without any help from the school officials. Still, this group of teachers also involved their students in the process.

What is ultimately quite striking is the large percentage (41%) of teachers who were given absolute free range at all times to develop their own grading criteria without the help of

management or students. The results could imply that schools just advocate for teacher autonomy but surely keeping checks and balances might become problematic here. After all, grading criteria in these schools seemed completely unregulated by the management/department and teachers were left not only unsupervised but also unsupported, not to mention the inevitable increase of variation within grading criteria.

With the results in, it became obvious very quickly that a majority of teachers (over 90%) set their own grading criteria by themselves at least some of the time. Nevertheless, while poring over the data, we noticed another detail which soon became a serious cause for concern. The variety of different activities for which teachers would set their own criteria was simply astounding. When questioning the systematicity in grading, we did not expect diversity in grading criteria of such magnitude. Not only would teachers, as expected, set their own criteria for student participation (83%), they would also do so for unit tests, grammar/vocabulary tests, oral exams, essays, presentations and projects, all with a strong 50%-and-above support. For example, 7 in 10 teachers (or 70%) would determine their own grading criteria for oral examination while 8 in 10 teachers (or 80%) would do so for presentations (see *Chart 14*).

Keeping this in mind, the chart below could perhaps be visually representative of the diverse and unexpected mixture of activities in apparent disarray.
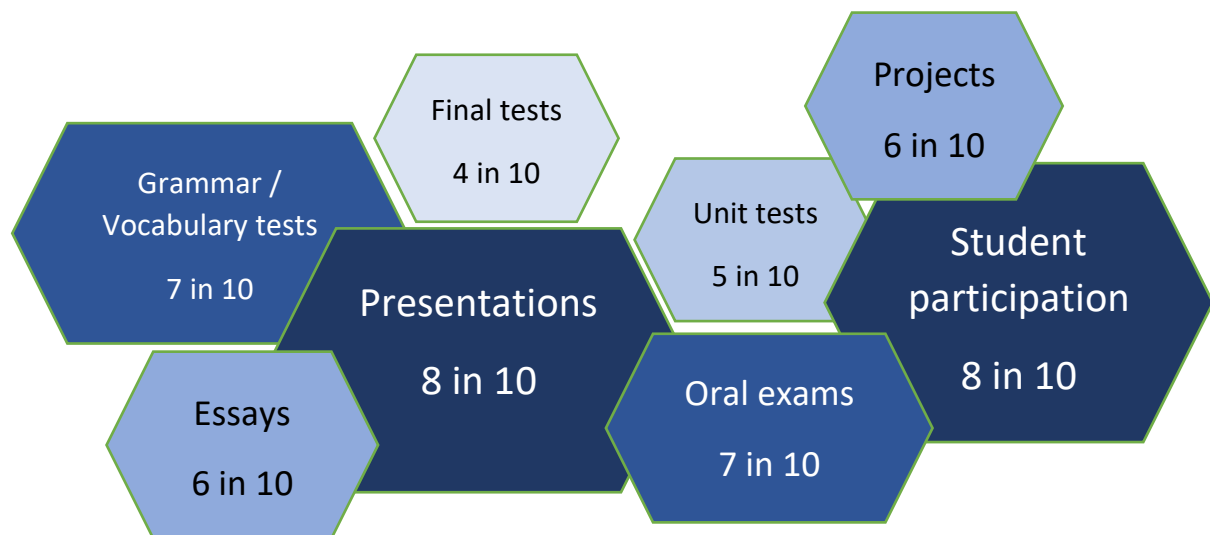

Chart 14: Number of teachers determining their own grading criteria themselves for specific activities at least some of the time

Perhaps the most surprising discovery was that around 4 in 10 teachers (40% or 27 out of 71) were creating their own grading criteria for the final test at least some of the time. For more

than half of these teachers, (56% or 15 teachers out of 27), the final test was always the most weighted activity in their classroom. From these 15 teachers, 60% (or 9 out of 15) were the ones who had decided to always put the most weight on the final test themselves. That means that (at least sometimes) the following scenario would happen: the nine teachers would themselves choose the criteria regarding the test and how it would be graded and then they would put the most weight on this particular grade to determine, with no help, supervision or outside consultation, the student's overall, end-of the-year grade.

Some teachers told us that they had freedom to design their own grading criteria except for the final tests in Maturita Year where the English department had the last say. Another commented that although their English department set up the same criteria for all the teachers, 'if we want to, we can adjust them.' Finally, an experienced teacher from a smaller 'gymnázium' shared this with us: 'The school management determines only that we must use grades and then issues recommendations. I myself determine grading criteria, which is great because then I can award students good grades.'

Teacher autonomy surely has its rightful place in the grading process. Entrusting a teacher with it can be essential in building their confidence and help them in their professional development. But perhaps even this autonomy needs to be clearly defined otherwise we can witness a disarray of different grading approaches that are not consistent for even one teacher, let alone a school. If we strive for systematic and objective grading processes across different platforms, then conditions should be created for this to happen. If we expect transparency and explicitness from teachers, then we surely should expect the same coming from the grading guidelines. Demanding teachers' work be more systematic and impartial could certainly be a fair point. But in that case, we need to make sure a teacher has access to much clearer directives than they seem to have at the moment. This might call for more assistance and coordination on an institutional level so that teacher autonomy could be methodically guided and efficiently supported.

4.3 Research Question no. 3: To What Extent Does the 'Subjective Element' Play a Role in Grading?

Having just identified some discrepancies in the grading process on an institutional level, the next section will tackle some of the issues of subjectivity stemming from perceived teacher bias.

Question no. 1:

- **Do you use the same criteria in parallel classes?**

Although Ziegenspeck still places this question under the previous research area of institutional prejudice, this thesis will deal with parallel grading within the following chapter on teacher bias. The reason behind the decision is quite simple. By proving earlier that a majority of teachers designed most of the grading criteria at least some of the time on their own, we were interested in how that would apply to parallel grading criteria. Already, Ingenkamp found out that the same teacher assessed one (objectively the same) performance in parallel classes differently (Ingenkamp [in Ziegenspeck, 2002: 67]). From a study done by Lietzmann, we also learnt, that a teacher awarded different grades if they graded the same performance but at different times (Lietzmann [in Ziegenspeck, 2002: 60]). How would this account for the teachers in our survey? To what degree might they actually be aware of any bias on their part?

Seldom did we see such a definite and favourable response as in the case of this question. More than 85% of teachers were in agreement their grading in parallel classes matched. Additionally, another 12% concurred because their standards correlated at least some of the time. This would mean virtually everyone from our 75 surveyed teachers, apart from 2, firmly believed they awarded corresponding grades to parallel classes due to using the same criteria.

Yet, if the observational studies by Ingenkamp and Lietzmann point to definite teacher bias in parallel grading, how do we explain our results? Is it possible teachers are simply not aware of their own partiality in some instances? If so, can we see other example of the discrepancy between perceived and imagined teacher objectivity to support this claim?

Question no. 2:

- **Do you award a better grade in order to motivate the student to keep applying themselves or to motivate them to try harder?**
- **Do you award a worse grade in order to motivate the student to try harder?**

In order to find out the level of the teachers' awareness of their own potential bias, we felt the need to ask them directly whether there might be some inner motives driving them to adjust a grade ever so slightly. Since grades have the ability to motivate, we exploited Ziegenspeck's

insinuation that some teachers might award better (or worse) grades to their students. Such a tendency would be carried out in the genuine hope of increasing a student's effort in the future. Although the questionnaire was fully anonymous, we did have some reservations about the level of honesty we were about to receive from our surveyed teachers. After all, our two questions alluded unabashedly to teacher conscious bias.

This, however, was never the case. Rarely were the teachers so remarkably candid in their answers than here. Almost 90% of the polled teachers replied with a straightforward 'yes / yes, sometimes' to the first query. Most teachers simply viewed awarding better grades to be a positive motivational tool. A 'gymnázium' teacher with a wealth of experience (10+ years) confessed to us: 'I sometimes adjust […] a grade (always to a better one)'. Another explained that it was the student individual background and their personal growth that can prompt a teacher to choose a better grade:

> I sometimes have students of different levels in my classroom and then I also grade their development and progress. What can then happen is that students receive the same grade for different performances. I comment on everything to my class ('A' made this or that mistake, he/she will get '2', although 'B' made more mistakes, he/she came with a much lower level from primary school and has improved a lot, so he/she deserves '2' as well. This way of grading can't be used all the time, but it works as motivation and other students perceive these differences and I don't get negative feedback on it.
>
> - an experienced teacher from a smaller 'gymnázium'

The high number of conclusive, affirmative responses to our first question might heavily imply that teachers did not see anything controversial about adjusting grades for the better. There could be a number of reasons for such a mindset: an internal conviction that this practice can be justified, that rewarding student effort and progress is fair, that the practice ultimately works by having a positive impact on students; hence the level of teachers' openness about their grading tendency. Be this as it may, the final numbers distinctly showed how common and widespread the practice was across different types of school.

On the other hand, tampering with grades in the opposite direction brought on much more uncertainty from teachers. Still, almost every third teacher (32%) admitted to modifying the grade for the worse at least some of the time if they believed this would motivate students to try harder next time. Unfortunately, whether a worse grade could have a positive influence on

a student at all, remains highly debatable. While a 'bad' grade has been proven to have a highly motivating effect (Saňáková, 2018: 77), awarding a 'worse-than-what-it-should-be' grade in order to motivate students remains quite polarizing. Experts were, however, able to explain carefully how this particular assessing tendency is based on a false premise. They proved that the assumption a worse grade could ever encourage students to perform better is nothing but faulty (Ziegenspeck, 2002: 88-9). Interestingly, even though literature concludes that awarding a worse grade happens only in 'exceptional cases', our research, with the frequency of three in ten teachers, reveals this practice is far from uncommon (Ibid., 89). When wishing to discuss teacher bias, this assessing tendency might be therefore a matter of concern.
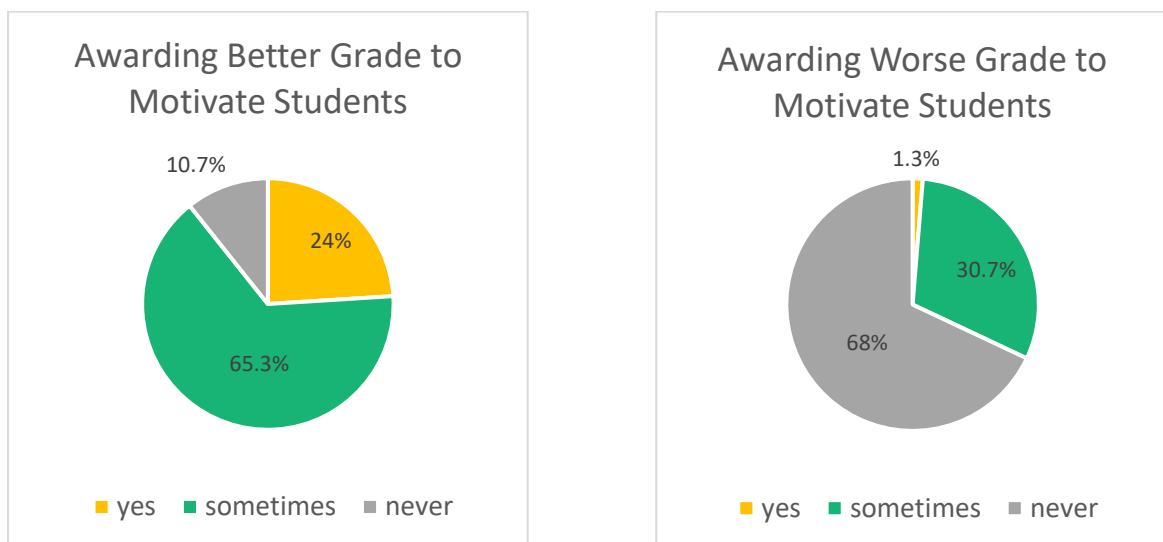
Chart 15: Percentage of teachers awarding a better grade (on the left) and a worse grade (on the right) to motivate students

Question no. 3:

- **Which of the following factor(s) can influence a grade?**

This question referenced extensively *Chapter 2.10.3 Problems with Grading: Gender Stereotype* and *Chapter 2.10.4 Problems with Grading: The Subjective Element*. Its primary objective was to establish the degree of systematic teacher bias based on Ziegenspeck's scheme.

The teachers' responses were checked against 12 factors such as the student's previous grade, his/her behaviour in the classroom, a teacher's relationship with the student and so on (see *Chart 16*). These factors were carefully chosen to cover the seven different types of bias they

could display: 'gender stereotype', 'logical mistake', 'tendency to persevere', 'mistakes caused by recognizing repercussions', 'order and contrast effect', 'halo-effect' and finally, 'assessing tendencies' (for 'assessing tendencies' see also *4.3 Research Question no. 3: Question no. 2*). These biases considered all the systematic subjective issues related to grading and the role of the teacher as proposed by Ziegenspeck (2002: 70-74, 85-90).

First, let us consider student gender as a potential source of teacher bias. Looking at the data in *Chart 16*, it appears that the student's gender did not play any role whatsoever in grading with the teachers we surveyed. Yet, studies in literature show different results. According to research done by Ziegenspeck, Knoche and others (Ziegenspeck, 2002: 70-3), female students in secondary schools received better grades especially in the study of modern foreign languages. Ziegenspeck saw this as an example of teacher favouritism towards girls. That is why we set out to investigate whether there might be any truth to his claims. However, our surveyed teachers did not tend to make any distinction between awarding grades to boys or girls. Of course, this bias could be subconscious, but the fact that none of the 75 teachers polled marked this answer or commented on this issue, made Ziegenspeck's theory extremely unlikely. Therefore we are inclined to believe there are other factors that would explain girls receiving better grades then just their gender alone.

Chart 16: Factors influencing grades

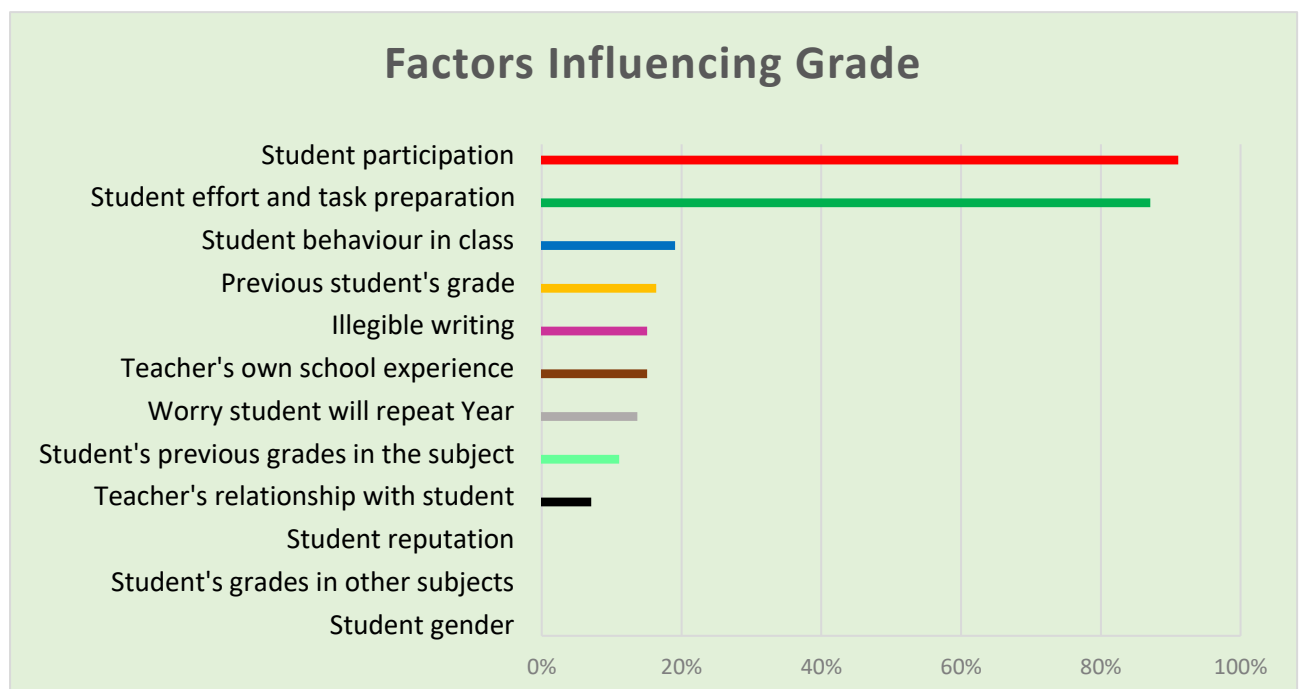Neither a student's grades in other subjects ('logical mistake') nor a student's reputation ('halo-effect') seemed to play any role in teachers' grading. Other factors such as the teacher's relationship with the student ('halo-effect') or the student's previous grades in the subject ('tendency to persevere') had also very little, if any, influence on the grade. However, 10 teachers (13.3%) admitted being swayed if their grade could cause the student to repeat the Year ('mistakes caused by recognizing repercussions'). What came as an even bigger surprise was that some teacher grading was influenced by student writing legibility ('halo-effect'), although previous studies had already suspected this (Aebli [in Ziegenspeck, 2002: 72-3; Ziegenspeck, 2002: 87). An experienced teacher from a bigger 'gymnázium' summarized the reasons for this decision neatly: 'What cannot be read, cannot bring home the points ('Co nelze přečíst, nemůže přinést body')'.

With a testimony from almost 15% of teachers, we were also able to confirm the bias of 'assessing tendencies' in which the teacher's personal life, for example their own past school experience, could impact their grading decisions. Yet, even in this case, the number of teachers implicated was relatively small (11 out of 75 participants).

Next, the bias of the so-called 'order and contrast effect' was reviewed. It presupposes that the grade(s) of the previous student(s) could prompt the teacher to award the next student a dissimilar grade even if the performances objectively deserved the same mark. 16% of the teachers contested this bias. One of the most compelling answers came from a teacher (10+ experience, bigger 'gymnázium') who admitted somewhat reluctantly and with a fair amount of guilt that he/she might be displaying this bias 'subconsciously […] even though it should not happen'. This teacher presented a refreshingly honest and self-critical view of their own grading practice and, as P.H. Ludwig put it, to acknowledge one's own bias was the first step to improving assessment (Ludwig [in Ziegenspeck, 2002: 90]). Such insight, shown by this teacher, might therefore prove invaluable in the long run.

Finally, we checked the grading process for the 'halo-effect', examining the degree to which student behaviour in the classroom could affect teacher grading. Although one might argue this factor goes hand in hand with student reputation, teachers seemed to be influenced by the former much more than the latter. Both factors, however, describe the 'halo-effect', a bias based on the teacher's impression and subjective opinion. If a teacher finds a student likeable, they will be more prone to give them a better grade than to an intelligent but misbehaving student (Ziegenspeck, 2002: 87). In the survey, 14 different teachers (18.7%) let student

behaviour in the classroom impact their decision making when grading. This reveals how sensitively teachers can perceive the persona of the student (their behaviour, participation) during their own lessons, but might be less inclined to judge the same student for the same factors outside of their classroom. This observed teacher reaction in fact fully adheres to experts' guidelines calling to ignore any student's personal details that are not directly related to the teacher's subject or classroom (Sacher [in Ziegenspeck, 2002: 90]).

This slowly brings us to the top two most grade impacting factors as quoted by our surveyed teachers: first, student participation and second, student effort and preparation for the task.

Already Kolář and Šikulová (2009: 145-6) as well as Košťálová et al. (2008: 15-23) heavily underlined how important it is to assess the overall process of the student's learning, not just the individual performance result. Grades, after all, should not be observed without context.

On the contrary. It is the student overall performance in the course, his/her effort and work put into the learning process that must be awarded (Ziegenspeck, 2002: 37; 49-52). This recommended strategy, however, was employed by nearly every teacher we questioned. 'I grade improvement and effort,' one teacher told us. Teachers across the secondary school system appeared to listen to experts and apply their advice. One stated: 'I'm trying for maximum objectivity,' and another added: 'I've ticked off those boxes [in the questionnaire] that, in my opinion, can be taken into account when grading while still remain objective and just.' Finally, an SOŠ teacher confided in us: 'After so many years [of teaching], I pay attention to and am careful about [any bias].' With such level of awareness, self-criticism and honesty about their own potential subjectivity, teachers have so far seemed to have followed the expert advice on avoiding bias almost to the letter.

Question no. 4:

- **Which mistake in essays are students penalized more for – a grammar or spelling mistake?**

This question was inspired by Bilanová et al. (2010: 41-42) and their suspicion that there might be some ambivalence about teacher assessment of grammar and spelling mistakes in writing. The authors assumed there might be some teacher inconsistencies in penalizing students for the two types of mistakes. Having set teacher bias as one of our main research questions, we decided to try to shed some light on it. We were especially motivated since any

meaningful research into this area in the context of Czech secondary schools seemed to be missing.

The question in itself was, however, problematic, even contentious. Teachers themselves told us straightforwardly how their grading criteria changed based on what their primary objective was: 'It depends on the instructions', 'it depends on the purpose of the essay', 'all hinges on what the text focus is and what [skills] it is testing' were just some of their answers.

Other respondents pointed out that mistake penalization was contingent on a Year and student/class language level: 'There are different requirements for a "lower" and "upper" "gymnázium" ('nižší a na vyšší gymnázium'). With beginners, I will check for the correct usage of the past tense, with more advanced students I will then scrutinize absolutely everything.' Another teacher confirmed that as soon as they were preparing their students for a formal written essay, 'my grading is stricter [than] in lower Years […]'. Many teachers then stressed the importance of the context and 'severity of the error in the context.' As long as the mistake did not decrease understanding of the text, it tended to be tolerated: 'The mistake that hinders/makes it more difficult to understand the text is [penalized more]' or 'It depends on how much the mistake affects the meaning.'

In spite of a high level of relativity, but in the light of no available prior data, we decided to proceed with our analysis. The following conclusions are, by all means, only preliminary and more research or a repeated, more detailed, poll on this topic is needed. The results might give us a general idea of teachers' preferences with regard to grading different mistakes and answer Bilanová et al.'s question at least to some extent.

60% of our surveyed teachers expressed collectively that they marked students down more for a grammar mistake. One teacher was certain: 'If spelling mistakes don't block communication and aren't a result of extreme carelessness, they can be tolerated more than e.g. creating wild phrases in a Czenglish style and insisting on them.'

For 14% of the teachers, the question was too relative and depended on too many factors. One such factor could be the type of essay: 'If we are talking about formal writing, spelling mistakes caused by poor spell checking are more serious than in other kinds of essays.' This teacher in particular (moderate experience, bigger 'gymnázium') was among the other 14% (10 out of 70) that avoided answering the question directly and rather marked the option 'Other'.
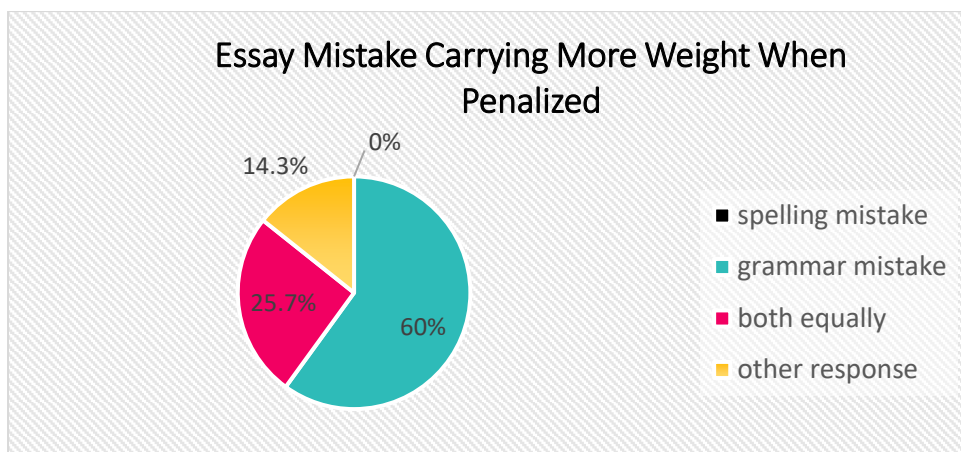
Chart 17: Essay mistake carrying more weight when penalized

Very striking was the fact that not a single teacher believed a spelling mistake should be penalized more than a grammatical one. An experienced SOŠ teacher knew exactly why they were more lenient with student spelling errors: 'Maybe it's because I myself also have problems with spelling.'

Finally, 26% of teachers were of the opinion both mistakes carried the same weight and so should be marked down equally: 'All mistakes have the same influence on the grade, whether we're talking about a grammatical mistake or poor cohesion.' There were also others for whom the origin of a mistake did not matter. They penalized errors according to whether they were 'big [or] small, not depending on whether they are grammatical, linking or spelling.'

Ultimately, the results and subsequent commentaries proved to be more heterogenous in nature than first anticipated. A highly experienced teacher from a bigger SOŠ offered her own solution to this somewhat divisive issue: 'Grammar is strongly linked to vocabulary and phraseology in English. That's why it's not appropriate to separate them even during the grading process and [rather] grade comprehensively, e.g. under the blanket term *Use of English*.'

4.4 Research Question no. 4: Do Teachers and Institutions Follow Expert Defined Grading Standards?

As precarious as grading most certainly can be, experts have identified some general standards that can improve the overall quality of grading. We will reveal which standards and to what extent they are being followed in the surveyed secondary schools. After all, grading

bias could lead to favouring some and disadvantaging other students so ensuring that the experts' recommendations are in place could lower the risk of it happening.

Question no. 1:

- **Do you grade essays anonymously, without knowing the student's name beforehand?**

Among the principles that can improve grading standards is a clear mental separation between the teacher's objective observation and their grading. This could be encouraged in a number of ways, for instance grading essays anonymously without the teacher knowing the student's name beforehand (Sacher [in Ziegenspeck, 2002: 90]). Unfortunately, this practice did not seem to be commonly used in the schools we surveyed. Up to 60% of the teachers never graded essays anonymously.
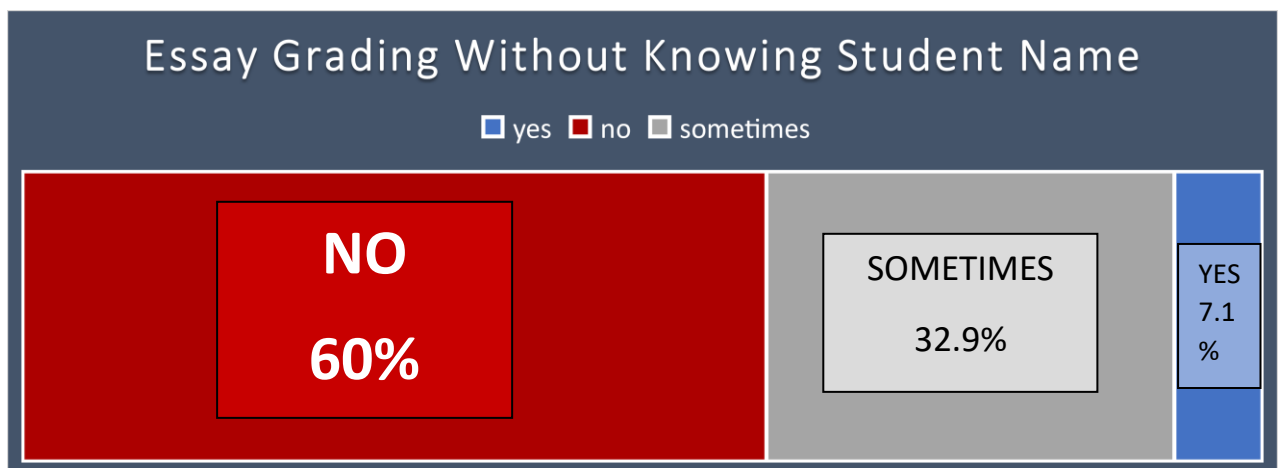
Chart 18: Essay grading without knowing a student name

However, the teachers had several underlying and sound reasons for their choice: 'I've been teaching my students for several years and I recognize their writing,' they revealed. 'I recognize [my students'] writing, I've known it since Year 1, so [grading anonymously] cannot be done,' they said. Another agreed that grading anonymously just simply was 'not possible'. Even a teacher doing her best to apply anonymous essay grading in practice, was struggling for the same reason: 'I try to do it but I remember their writing very well.' The perks of online teaching, however, made identifying a student's writing perhaps easier for some: 'I recognize (my students) based on their writing but now online, I usually [grade anonymously] – not intentionally but I mostly just don't look at the name, it's not important.'

Then there was a certain number of the respondents (7.1%) who could simply not afford grading essays anonymously. These teachers needed to know the student's name because they had several students with special needs in the classroom who were supported by different individualized grading. This naturally led to the need to know whose essay the teachers were grading.

However, the teachers seemed sufficiently self-aware and also mindful of the possible bias:

> I see it in myself that the author's name plays a role. Mostly, where straight '1' students are concerned, so then I have a tendency to give a better grade. Also, when I know that it is a nice, diligent student… Human factor, so it is better not to know the names. But with time, you recognize [students] based on their writing… So I have to be careful about it.

> - an experienced teacher from a smaller SOŠ

Question no. 2:

- **How do you grade essays?**
- **How do you grade presentations?**

We have already learnt that one of the most important strategies that could help create standards in grading is to employ various forms of assessment. Experts, such as Slavík and Kratochvílová, urged teachers to administer different forms of grading: norm-referenced/criterion-referenced and analytic/holistic (Slavík, 1999, 87-90; Kratochvílová, 2011: 87-88). More still, Košťálová et al. invited teachers to always use a frame of reference when grading (i.e. set of well-defined and specific criteria such as rubrics) rather than use their own overall impression (Košťálová et al., 2008: 15-23). Grading based on the overall impression could be 'dangerous' and ought to be completely abstained from (Sacher [in Ziegenspeck, 2002: 90]).

Since different forms of assessment could support grading standards and thus reinforce objectivity and warrant a more systematic approach, we needed to check how diverse the teachers' grading methods were. To achieve this, we chose to arbitrarily check their grading approaches in two learning skills, written and spoken, represented randomly by essays and presentations. It was essential to look for two aspects in the responses: first, whether teachers used more than one form of assessment; and second, whether they ever graded based on their 'overall impression'.

To begin with, we focused on essays. This activity was graded by a majority of the teachers (more than 93%) so we were able to collect fairly extensive data that could give us reliable results. The data analysis showed that most of respondents, 55% (37 out of 67) were choosing just a single form of grading. From this group, the majority, almost 50% (18 out of 37), used exclusively just an analytic rubric to grade. The explanation for such lack of variety in forms of assessment and a clear preference for a specific rubric could be found directly in the teachers' responses. In their questionnaire comments, the teachers talked about how they reflected and adhered to the CERMAT essay grading instructions because they felt the need to prep students for the essay part ('písemná práce') of Maturita State Exams. Some did this with compliance ('the CERMAT rubric is sufficient'), others with more cynicism: '[I grade essays] just like in Maturita State Exams, although I'd been doing it even before this famous period in the history of our education.' Many trained their students to get used to the Maturita grading criteria 'from Year 1' and continued to 'teach the students [these criteria] throughout their studies.'

The second biggest (27%; 10 out of 37) group, out of the teachers who turned to only one form of essay assessment, adopted (less detailed) holistic rubrics, 'although not always. Some tasks cannot be graded this way.' Due to Maturita exam requirements, the teachers' preference for rubrics might have been understandable. Still, there was occasional skepticism: 'However, CERMAT doesn't take into account the overall impression, and so in preparation for graduation I use, and explain to students, this form of assessment.'

There was, indeed, a faction of teachers who not only used a single form of assessment at all times, but also completely trusted their own perception while grading essays (most came from smaller SOŠ schools and had more than 10+ years of experience). The sheer number of teachers who relied on their 'overall impression' when grading was quite unexpected – 12% (8 out of 67) of all the respondents marked this option. That would denote that approximately one teacher in a group of ten uses purely their impression to grade essays at any given time.

This was even more striking when compared to the number of teachers who, on the other hand, followed experts' advice. Only approximately two teachers out of ten (22% or 15 out of 67) used two or more different forms of grading and never depended on their 'overall impression' to guide them. These were mostly highly experienced (10+ years) teachers coming primarily from either smaller SOŠ schools or bigger 'gymnáziums'.
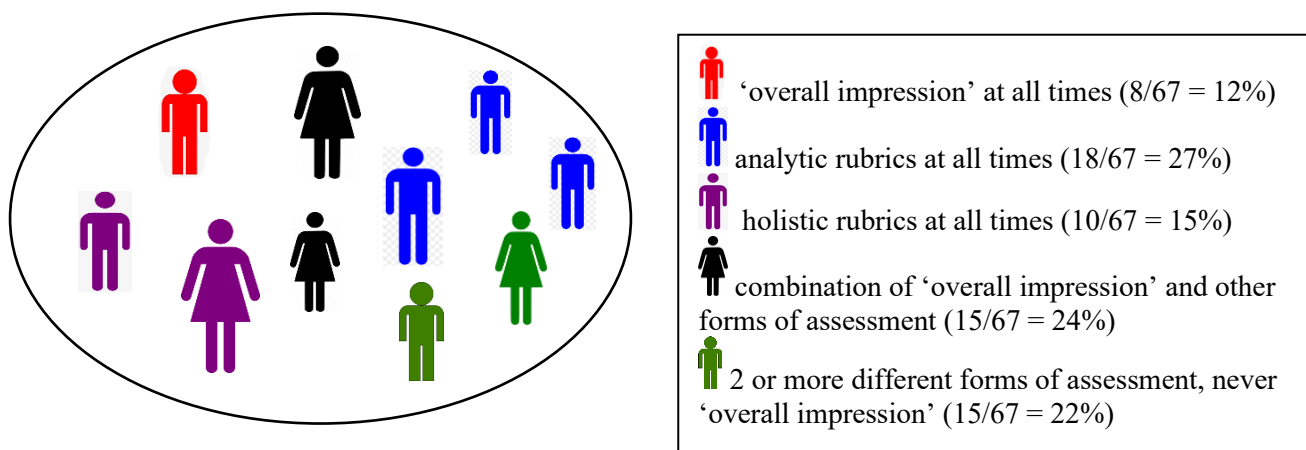
Chart 19: Forms of assessment when grading essays in a group of 10 teachers[16]

By and large, from *Chart 19* above, we can read that roughly one in three (red and black figures) practiced essay grading using their own individual impression at least some of the time.

Let us now consider how the results for written essays compare to the assessment of oral skills. We chose presentations as the leading representative of a graded spoken activity because we hoped a substantial number of teachers was assessing this activity in practice. Subsequently we would be able to obtain a large amount of responses and produce solid, tangible results.

Indeed, almost 90% of the teachers graded presentations. Again, we were looking for variety in the forms of assessment and total, or at least notable, absence of 'overall impression'-based grading. The results, however, mirrored the ones collected in relation to essays and were able to confirm the tendency to use a single form of assessment as well as the inclusion of teacher 'overall impression' in grading.

As 55% of the teachers adopted just one form of assessment when grading essays, similarly, 53% (34 out of 64) of teachers used only a single form of grading for presentations. Out of this group of teachers who used just one assessment method, most, 32%, leaned equally towards either analytic or holistic rubrics (11 out of 34 teachers).

---

[16] One response, from a teacher who used a 'criterion-referenced' form of assessment exclusively, could not be accounted for in *Chart 19* due to lower statistical significance

When taking all 64 responses into consideration, we saw norm-referenced grading, just like with essays, used slightly more often than criterion-referenced assessment (25% vs 22%). Both types of presentation assessment were predominantly used together with other forms of grading. Used on their own as a single grading method, their relevance was, overall, statistically negligible (1 and 2 teachers out of 64 for norm- and criterion-referenced assessment respectively).

Further, while 12% out of all the surveyed teachers depended only on their own impression assessing essays, it was 14% in the case of presentations. These teachers, just like with essay assessment, came mainly from SOŠ schools but tended to be moderately experienced.

At this point however, we came across the only key difference between essay and presentation assessment forms used. The number of teachers who used 'overall impression' exclusively or in alliance with other forms of assessment, was significantly higher with presentations (41%). That meant that while 3 out of 10 teachers would use their 'overall impression' to grade essays at least some of the time, it was 4 teachers out of 10 when grading presentations (note the red and black figures).
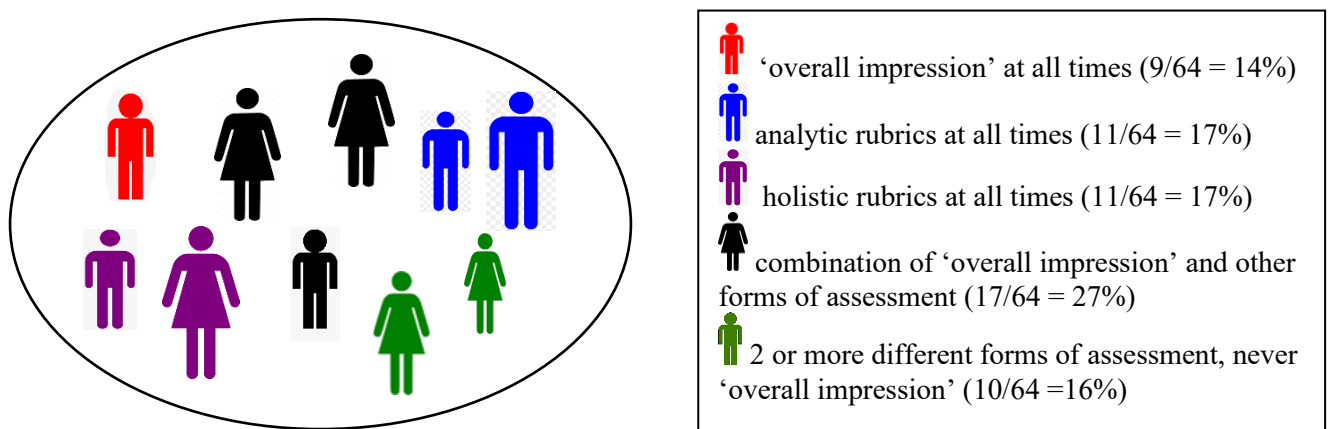
Chart 20: Forms of assessment when grading presentations in a group of 10 teachers[17]

In many cases, teachers' presentation grades recognized a combination of factors besides the content and lexical/grammatical competence. Among these were: preparation at home, non-verbal communication, clarity of meaning, handouts for other students, visual aids, time distribution or other students' feedback. A great number of teachers reckoned that the most

---

[17] We omitted responses that marked exclusively as a form of assessment the option 'other'

important aspect of the presentation grade was its the motivational value: 'I try to give only nice grades,' disclosed an experienced SOŠ teacher.

Question no. 3:

- **Do you let students know the grading criteria ahead of testing them?**

'In my opinion it is necessary for students to be familiarized with the grading criteria,' insisted an experienced SOŠ teacher in our survey.

As resolute as this teacher and many others were, the answer to *Question no. 3* was not fully guaranteed. From 'necessary' to 'usually' to 'not always' to a clear 'no', teachers' responses varied sharply. Yet, literature is unanimous in its endorsement of familiarizing students with grading criteria before testing them. Students must be aware and understand perfectly the criteria based on which they would be graded. Learners must simply know what to do to study. That is why teachers are advised to make their grading criteria as transparent and accessible as possible (Kratochvílová, 2011: 79-84; 87-88).

And yet, as imperative as this practice might seem, not every teacher would always familiarize their students with the grading criteria. 31% of those surveyed would only do so sometimes. There was also a single case where a teacher unequivocally rejected the practice. The teacher had also answered that the work of creating all the grading criteria would fall on their shoulders alone and that they themselves could determine the different grade weight. It appeared that this experienced teacher was given full autonomy by their school (a smaller SOŠ) when it came to grading criteria. Still, he/she was more the exception.

Most schools (68%), in fact, seemed to instill quite rigorous ground rules: 'The teachers at our school have a duty to familiarize students with grading criteria at the beginning of each school year' (response from a bigger 'gymnázium'). Teachers would frequently show their grading criteria 'in the first class of a new school year [and] even throughout a school year, especially every time [students] ask' (response from a smaller SOŠ).

Once they shared the criteria (often based on CERMAT guidelines) with their students, the teachers would follow two distinct paths. The majority would adhere to the criteria completely. Others, however, would have the option to alter fine details. An experienced teacher from a smaller SOŠ told us: '[I grade according to CERMAT] - but I don't always grade everything [although] students know it beforehand.'

Sometimes, the sense of fluctuation was more palpable: 'I usually follow the rule that anything under 44% is a "5",' revealed another (moderately) experienced SOŠ teacher.

Finally, some inconsistency hinted at possible lack of support from the management at this 'gymnázium': 'Not always [do I let students know the grading criteria ahead], because this is my first year teaching at this secondary school and I'm just learning how to grade, so it's difficult for me to determine criteria beforehand.'

We also noticed something peculiar. Those teachers who were never determining grading criteria by themselves (4/75 = 5%), would always familiarize students with them and would always use the same criteria for parallel classes, too. This means these teachers complied with expert advice on all fronts. That shows that some schools do have uncompromising standards, a system in place that is 'by the book' in relation to grading criteria. But considering the percentage of these schools, they are truly few and far between.

Question no. 4:

- **Do you let students grade themselves?**
- **Do you let students grade each other?**

Throughout the thesis, self- and peer assessment has been promoted as a fundamental and indispensable part of the grading process (Košťálová et al., 2008: 15-23; Ur, 2004: 291-3). How does this claim stand up in the surveyed Czech secondary schools?
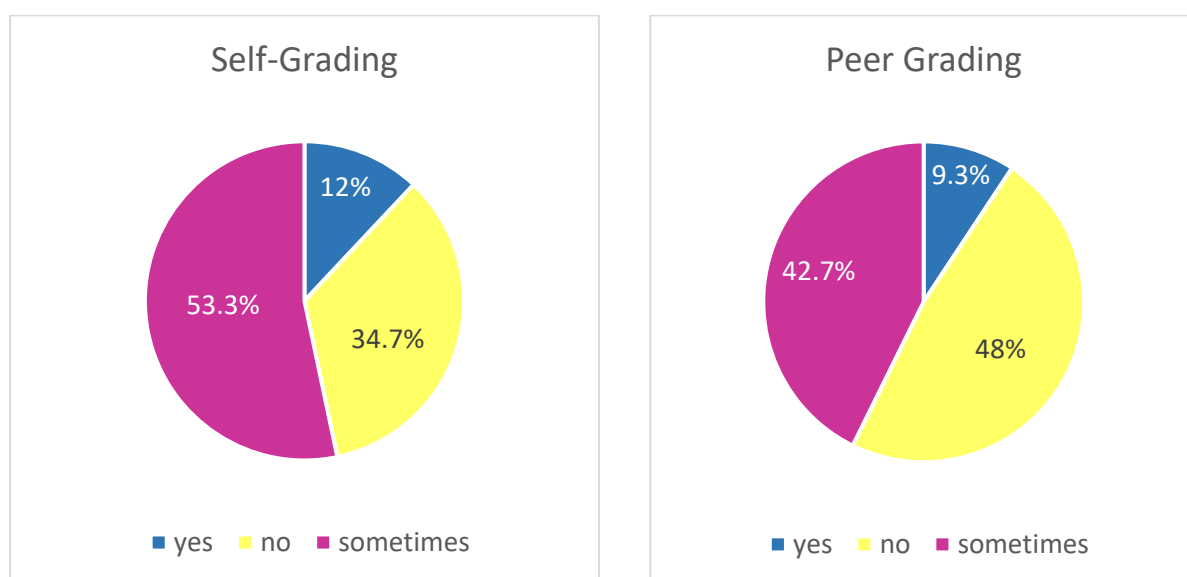
Chart 21: Self- and peer grading

From the chart above, it is obvious both self- and peer grading was a crucial part of classroom assessment. More teachers allowed students to grade themselves than they allowed other students to grade each other. Below, a 'gymnázium' teacher describes an elaborate approach to integrating student self-grading in essay writing:

> Students will be given a layout and a writing guide with grading criteria [(e.g. topic, length)], tips and a 'no-no list' of recurring errors. They write the essay, I will upload my commentary and based on it, they will grade each criterion themselves with a separate grade. [Based on their grades] I will then give them one summative grade.

> - moderately experienced teacher from bigger 'gymnázium'

All in all, teachers tended to prefer self-grading (65.3%) to peer grading (52%) at least some of the time. Although peer grading did not seem as popular, several teachers reported that they graded presentations either via peer grading alone or in a combination of teacher and peer grading. An experienced teacher (10+ years) from a smaller 'gymnázium' also revealed their strategy for grading essays using peer grading:

> Students often grade one another's criteria fulfillment. They write mutual comments and can edit or rework their own work even after reading it. This system has worked very well for me. On the one hand, they learn better, [on the other,] they are satisfied because they have better grades.

Question no. 5:

- **Do you discuss or consult the grade before awarding it to the student?**
- **If so, whom do you discuss or consult the grade with?**

Some teachers already described how they graded a student's presentation by enlisting the help of other students. Most teachers, indeed, discussed the grade before they awarded it – 80% at least some of the time. Experts would commend such behaviour because it exhibits willingness to review a grade and perhaps, to some extent, acknowledge own bias. As this requires a certain amount of self-reflection and could be considered a difficult task, that teachers employ this approach is creditable (Ludwig [in Ziegenspeck, 2002: 90]).

Having evaluated the results, we came to the conclusion that the teachers who participated in our survey mainly debated grades with either the student or a colleague or both of them.[18] With a slim majority, the student being graded was consulted overall more often than the teachers' colleagues (70% vs 65%). Again, giving a grade a second look before awarding it is something that literature actively encourages and, clearly, teachers seemed perfectly consistent with the official recommendations (Ziegenspeck, 2002: 90).

Discussing Grade Before Awarding

| 65% | 70% | |
| with colleagues | with the student | 26.7% with whole class |
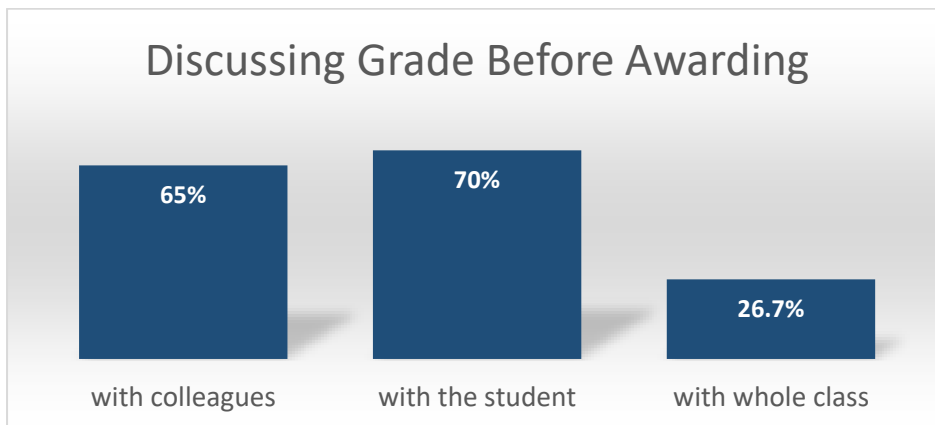
Chart 22: Discussing a grade before awarding

4.5 Research Question no. 5: Are teachers supported in grading?

As already observed, achieving a more rounded performance assessment can be a challenge at times. A systematic teacher approach to grading might require systematic institutional support that would equip educators with clear directives on specific grading standards in a range of graded activities.

***Question no. 1*:**

- **Has a workshop/seminar/class on the topic of grading ever been provided for you by your employer/university?**

An overwhelming 'no' answer to this question had been quietly fancied before the poll. We had felt and speculated that most teachers had had no previous opportunity to revise their (future) grading methods based on some formally organized, methodical discussion. However, this was not exactly the case. While around 17% of the teachers could not

---

[18] The respondents were able to mark more than one option.

remember any workshop/seminar on grading, only 44% answered our question with a resounding 'no'.

Most of the remaining participants (≈ 39%) attended a CERMAT training session on how to grade the written and oral part of Maturita State Exams. Maturita is, surely, a unique type of exam and so it is perhaps disputable how easily the CERMAT seminar would navigate teachers in assessing other graded activities (especially projects, student participation or extracurricular activities).

On the one hand, we already saw many teachers following/having to follow CERMAT instructions while grading common student essays, so such a seminar could have helped them set up the grading criteria. On the other hand, we did not see the surveyed teachers grade presentations based on CERMAT guidelines at all, even though an oral exam is also a part of the state exams.

What is more, at least two teachers (one was personally providing CERMAT training), were quite vocal about their participating in the yearly CERMAT seminar. The two teachers were also actively using CERMAT guidelines when grading classroom essays and presentations. Nevertheless, both teachers still included their own 'overall impression' when grading the two activities in direct contradiction to the explicit CERMAT protocol.

This would indicate that the support the *CERMAT* seminar gives to teachers might be limited to the grading for Maturita only and that using CERMAT as formal guidance for overall English classroom grading may be insufficient. If there are seminars for grading a specific Maturita exam supported by the *MŠMT*, why is there not a seminar covering grading in general?

Not all our positive responses specified the seminar attended. We wonder how many of these answers referred to the restrictive CERMAT grading seminar. Although a few teachers had participated in other workshops (all focusing on formative assessment), further research is required to provide evidence that a systematic pedagogical training on grading is offered to teachers. After all, such formal systematic training could curb mistakes in grading, increase their prevention and lead to a more objective outcome (Ibid.). A demand for such seminars is already there. As one teacher pondered: '[No, I've not attended any seminars on grading], but I would love to.'

5 Conclusion

'Grading is demanding.'

- an experienced SOŠ teacher

The aim of this thesis was to find out whether there is a systematic approach to the grading of English language skills at Czech secondary schools. The thesis looked for the presence or possible absence of conformity within the grading process. In case of absence, the paper examined the extent of disparity and inconsistency. Weak spots in the grading process were identified and the effects of such drawbacks explained.

The research, however, was subject to several limitations that should be borne in mind. First, the lack of prior research made it impossible for us to compare our results to any reference. Next, restricting the research location to Prague and its close vicinity became necessary due to the scope of our thesis.

Despite our curiosity, due to the lack of responses from SOU teachers, we could not statistically account for them. Their responses were, however, indispensable anecdotally. We also did not differentiate between different types of SOŠ schools. When preparing the questionnaire, neither regular/alternative nor private/state schools were distinguished which meant there was no consideration given to students who followed individual study plans.

Finally, even though our questionnaire was distributed at the time of distance learning, the survey questions did not discriminate between teacher approach while teaching online during the COVID pandemic and the 'standard' teacher approach in the classroom. If the grading style had shifted somewhat during online teaching, the teachers tended to comment on the changes themselves. Still, this could be an area of further research.

Our result analysis confirmed some and refuted other grading-related biases suggested in literature (see *Appendix III*). Based on our data, we were able to conclude the teachers had a natural tendency to follow expert advice (e.g. differentiating the grading scale or including self-/peer grading). In cases where the teachers' answers departed from the recommended guidelines, we saw a strong link between such bias and the teachers' underlying, honest motives. For instance, the action of awarding a better grade was taken due to its motivational value and the conspicuous absence of grading essays anonymously was needed to apply individual grading criteria. This could imply that some instances of bias stemmed from the

teacher's belief there were valid reasons for such actions and that these actions were done with the best intentions.

Further inconsistencies affecting grading standards seemed out of teachers' control since being firmly established by educational institutions. These would include, for example, the disparate spectrum of percental grade limits that varied considerably across different school types/sizes.

Finally, some problematic areas could not rely on any formal expert advice that would guide teachers or institutions. Still, they visibly created disparity in grading. Although the literature mentions the controversy of counting an arithmetic average, it does not offer any solutions and thus the teachers relied heavily on the counting as our survey showed. In addition, no expert confronts the dilemma of what happens to the original grade after a retake grade is given. Grade weight and its potential for a versatile interpretation goes also unaddressed in the literature although it certainly deserves a mention. The lack of clear guidelines (either expert or institutional) appeared to lead to a greater level of teacher autonomy which only further caused a very mixed collection of results (e.g. large number of individual teachers being solely responsible for determining grading criteria).

On the basis of the aforementioned problematic areas, we have attempted to devise a 15-point list of recommendations to formally tackle systematic inconsistency in grading with the emphasis on possible prevention. The first 10 suggestions correlate with expert advice, the last five reflect on the conclusions we came to after analyzing teachers' responses:

The recommendations are as follows:

1. different aspects of a performance to be graded (using rubrics), never 'overall impression'

2. self- and peer assessment made a regular part of grading

3. grades regularly consulted after a student/class discussion

4. in cases of uncertainty, grade discussed with colleagues

5. anonymous grading of written work whenever possible

6. mental separation of student performance and their persona (their behaviour in and outside of classroom / any antipathy / student's previous grades in the subject / writing legibility / previous student's grade)

7. use of different forms of assessment, both for an activity and overall (analytic/holistic rubrics; norm-referenced/criterion-referenced)

8. students informed ahead of testing re. grading criteria and included in the process of creating grading criteria if possible

9. same grading criteria in use for parallel classes (if possible)

10. grading to account for student effort, preparation for the task and participation in the classroom within a particular grade and separately

11. differentiation in the grading scale with both symbols +/-

12. clear institutional (ministerial) recommendations for schools/teachers covering a range of different graded activities, especially in terms of:

- lower/upper percental grade limits for different types of schools
- grade weight for different activities
- counting of the arithmetic average
- grading criteria for different activities

(are grades weighted? is the arithmetic average counted? who determines weight/criteria? what are most/least weighted activities? what do the criteria for different activities entail?)

13. unifying of institutional (school) guidelines for teachers in relation to:
- retake grades
- grammar/spelling mistake penalization

14. teacher autonomy clearly defined (including in any of the above areas of 12., 13.)

15. regular support and training provided for teachers on these guidelines (in 12., 13., 14.)

Finally, a couple of other close observations could be made. First, it became apparent from their responses and commentaries that the teachers greatly valued student effort and participation in the classroom. To such an extent in fact that in-class activity (working hard in the lesson, doing their homework, preparing for the task) has been shown to have a significant influence on the grade besides his/her actual tested performance. Students would surely benefit from increased focus on in-class work/activity.

Second, we noticed that while some teachers were resolute and unswerving in their responses, others were much more hesitant. While the teachers from the first group found answers to some of our questions unequivocal and definite, other teachers' replies to the same questions were flexible, variable-dependent and generally broader. This revealed a fragmented and diverse spectrum of grading at secondary level. Our data showed this fragmentation to be substantial, in some cases driven by different school policies, other times driven by teacher autonomy.

We wonder whether in some cases the teacher's individual approach might be a default for the lack of systematic approach on the institutional level.

Similarly, we wonder whether in some cases individual school policies might come from the absence of unifying guidelines on the ministerial level.

This might then create a need for individual teachers/schools to create their own systems to follow which, as we discovered, overlap only to some degree.

Furthermore, many of the teachers confessed their discomfort about grading. They recognized its relativity, finding it 'dependent on circumstances' but also 'counterproductive', often preferring written reports to grading. An experienced teacher from a smaller 'gymnázium' summed it up: 'I'd rather not grade at all, but I have to.'

Through this study, what we can confirm is that teachers put a lot of effort into being systematic and impartial in their grading. Due to reasons beyond their control, grading meanwhile, is not.

6 References

Books

Čapek, Robert. 2014. *Odměny a tresty ve školní praxi*. České Budějovice: Grada.

Číhalová, Eva, and Ivo Mayer. 1997. *Klasifikace a slovní hodnocení*. Praha: Agentura STROM.

Harmer, Jeremy. 2004. *How to Teach English*. Malaysia: Longman.

Harmer, Jeremy 2004. *The Practice of English Language Teaching*. Malaysia: Longman.

Kalhous, Zdeněk, et al. 2009. *Školní didaktika*. Praha: Portál.

Kolář, Zdeněk, and Renata Šikulová. 2009. *Hodnocení žáků*. Havlíčkův Brod: Grada.

Košťálová, Hana, et al. 2008. *Školní hodnocení žáků a studentů*. Praha: Portál.

Kratochvílová, Jana. 2011. *Systém hodnocení a sebehodnocení žáků.* Brno: Sabing.

Slavík, Jan. 1999. *Hodnocení v současné škole*. Praha: Portál.

Solfronk, Jan. 1996. *Zkoušení a hodnocení žáků*. Praha: Pedagogická fakulta UK.

Pasch, Marvin, et al. 1998. *Od vzdělávacího programmeu k vyučovací hodině*. Translated by Milan Koldinský. Praha: Portál.

Ur, Penny. 2004. *A Course in Language Teaching*. Cambridge: University Press.

Ziegenspeck, Jörg. W. 2002. *K problému známkování ve škole*. Translated by Gabriela Štrynclová. Pardubice: EFFE.

Journal Articles

Al Rifai, Nada. 25 January 2010. "Attitude, motivation, and difficulties involved in learning the English language and factors that affect motivation in learning it." *Procedia Social and Behavioral Sciences* 2: 5216–5227. www.sciencedirect.com.

Andrade, Heidi, and Anna Valtcheva. 2009. "Promoting Learning and Achievement Through Self-Assessment." *Theory Into Practice* 48, no.1: 12-19. DOI: 10.1080/00405840802577544.

Brožová, Stanislava, et al. 2007. "Angličtina v (odborném) vzdělávání." Zpravodaj – odborné vzdělávání v zahraničí, příloha V. http://www.nuv.cz/uploads/Periodika/ZPRAVODAJ/2007/Zp0711pVa.pdf.

Camilli, Gregory. April 2013. "Ongoing issues in test fairness." *Educational Research and Evaluation* 19, no. 2: 104-120. DOI: 10.1080/13803611.2013.767602.

Cheng, Liying, and Xiaoying Wang. June 2007. "Grading, Feedback, and Reporting in ESL/EFL Classrooms." *Language Assessment Quarterly* 4, no. 6: 85-107. DOI: 10.1080/15434300701348409.

Chin, Paul. 2016. "Peer assessment." *New Directions in the Teaching of Physical Sciences* 3: 13-18. https://www.researchgate.net/publication/328340643_Peer_assessment.

Davison, Chris, and Constant Leung. September 2009. "Current Issues in English Language Teacher-Based Assessment." *TESOL Quarterly* 43, no. 3: 393-415. https://www.jstor.org/.

Karami, Hossein, and Magdalena Mo Ching Mok. March 2013. "The quest for fairness in language testing." *Educational Research and Evaluation: An International Journal on Theory and Practice* 19, no. 2-3: 158-169. http://dx.doi.org/10.1080/13803611.2013.76761.

Laufková, Veronika, and Katerin Novotna. June 2014. "Školní hodnocení z pohledu žáků." *Orbis Scholae* 8, no. 1: 111−127. DOI: 10.14712/23363177.2015.8.

McDowell, Liz, and Kerry Harman. 9-11 December 2008. "'The designer', 'the teacher' and 'the assessor': changing academic identities." Paper presented at the SRHE Annual Conference: *Valuing Higher Education*, Liverpool, UK. http://nrl.northumbria.ac.uk/3316/.

McNamara, Tim, and Kerry Ryan. 25 May 2011. "Fairness Versus Justice in Language Testing: The Place of English Literacy in the Australian Citizenship Test." *Language Assessment Quarterly* 8, no. 2: 161-178. DOI: 10.1080/15434303.2011.565438.

Samkange, Wellington. 30 September 2015. "The Role of Labelling in Education: A Focus on Exceptional Learners." *Global Journal of Advanced Research* 2, no. 9: 1419-1424. www.gjar.org.

Schinske, Jeffrey, and Kimberly Tanner. Summer 2014. "Teaching More by Grading Less (or Differently)." *CBE Life Sciences Education* 13, no. 2: 159–166. https://www.lifescied.org/doi/full/10.1187/cbe.cbe-14-03-0054.

Starch, Daniel, and Edward C. Elliott. September 1912. "Reliability of the Grading of High-School Work in English." *School Review* 20, no. 7: 442-457. https://doi.org/10.1086/435971.

Terfa, Amenu Miressa. July 2017. "An Assessment of the Practice of Vocabulary Teaching Strategies in EFL Classes: Kellem Secondary School Grade 9 and 10 English Teachers in Focus." *International Journal of Scientific and Research Publications* 7, no. 7: 581-9. https://www.researchgate.net/publication/335189809_An_Assessment_of_the_Practice_of_V ocabulary_Teaching_Strategies_in_EFL_Classes_Kellem_Secondary_School_Grade_9_and_ 10_English_Teachers_in_Focus.

Troudi, Salah, Christine Coombe and Mashael Al-Hamly. September 2009. "EFL Teachers' Views of English Language Assessment in Higher Education in the United Arab Emirates and Kuwait." *TESOL Quarterly* 43, no. 3: 546-555. https://www.jstor.org/stable/27785037

Zoeckler, G. Laurence. Spring 2007. "Moral Aspects of Grading: A Study of High School English Teachers' Perceptions." *American Secondary Education* 35, no. 2: 83-102. http://www.mrshutt.org/uploads/4/6/3/8/46385679/moral_aspects_of_grading__zoeckler_.pdf

Worgan, Michelle. 2010. "Promoting learner autonomy." *Modern English Teacher* 19, no. 3: 25-28.

E-books

Berry, Rita. 2008. *Assessment for Learning: Basic Concepts of Assessment*. Hong Kong: Hong Kong University Press. https://www.jstor.org/stable/j.ctt1xcs68.6.

Bilanová, Markéta, Eva Lorencovičová and Jan Netolička. 2010. *Metodika výuky anglického jazyka na 2. stupni základních škol a středních školách z pohledu pedagogické praxe – náměty pro začínajícího učitele*. Studijní materiál pro distanční kurz: Přípravný seminář k profesní praxi. Ostrava: Ostravská univerzita. https://projekty.osu.cz/synergie/dok/opory/bilanova-metodika-vyuky-anglickeho-jazyka-na-2-stupni-zs-a-ss.pdf

Brown, H. Douglas. 2003. *Language Assessment: Principles and Classroom Practices*. California: Longman University Press. https://www.academia.edu/26575645/H_Douglas_Brown_Language_Assessment_Principles _and_Classroom_Practice.

Čapek, Robert. 2010. *Třídní klima a školní klima*. Praha: Grada. http://www.zstgm-steti.cz/uploads/tridni_klima_a_skolni_klima.pdf.

Cheng, Liying, and Janna Fox. 2017. *Assessment in the Language Classroom*. London: Palgrave Macmillan. https://www.researchgate.net.

Dostálová, Iva, and Stephen Douglas. 2016. *ANGLIČTINA - Příprava na státní maturitu*. Praha: Grada. https://obalky.kosmas.cz/ArticleFiles/219519/auto_preview.pdf/FILE/Anglictina-priprava-na-statni-maturitu_Ukazka.pdf.

Kunnan, Antony. 2017. *Evaluating Language Assessments*. New York: Routledge. https://www.researchgate.net/publication/320537835_Evaluating_Language_Assessments

Maturitní zkouška z anglického jazyka – písemná práce. 2013. Praha: CERMAT. https://www.gymnaziumhello.cz/Files/SDzWMqM5I_3_o=/MZ%20z%20AJ%20-%20PP.pdf.

Oppenheim, N. Alexander. 2001. *Questionnaire Design, Interviewing and Attitude Measurement*. London and New York: Continuum.

Petty, Geoffrey. 2004. *Teaching Today: A practical Guide*. 3rd edition. Cheltenham: Nelson Thornes. https://books.google.cz/books?id=QfSPDVcauFUC&pg=PA471&lpg=PA471&dq=geoffrey+petty+reliability+of+assessment&source=bl&ots=BrHHcTyT5d&sig=ACfU3U26x-41mucpNlunOb8vx-hwo0aFTg&hl=en&sa=X&ved=2ahUKEwiRgOOMopPqAhXOPsAKHffaBJIQ6AEwAXoECAoQAQ#v=onepage&q&f=false.

Petty, Geoffrey. 2009. *Teaching Today: A Practical Guide*. 4th edition. Cheltenham: Nelson Thornes. https://www.academia.edu/22851576/G_Petty_Teaching_Today_A_Practical_Guide_4th_Ed.

Tierney, D. Robin. April 2016. "Fairness in educational assessment." In M. A. Peters (Ed.), *Encyclopedia of Educational Philosophy and Theory*. Singapore: Springer Science+Business Media. DOI 10.1007/978-981-287-532-7_400-1. https://www.researchgate.net/publication/304999512.

Theses

Bartuška, Václav. 2017. "Hodnocení ve vyučování zpohledu žáků středních škol." BA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky a psychologie.

Bláhová, Kristýna. 2021. "Postoj pedagogů k hodnocení žáků na ZŠ." MA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky.

Bureš, Pavel. 2019. "Hodnocení žáků na středním odborném učilišti." MA thesis, Masarykova univerzita, Pedagogická fakulta, Katedra speciální a inkluzivní pedagogiky.

Číško, Petr. 2009. "Hodnocení v současné škole se zaměřením na druhý stupeň základní školy." MA thesis, Univerzita Karlova v Praze, Husitská teologická fakulta.

Čuřínová, Zuzana. 2007. "Formy hodnocení žáků na základních školách v České republice /s poukázáním na problematiku hodnocení žáků s SPU/." MA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky a psychologie.

Fričová, Martina. 2015. "Hodnocení žáků na 1. stupni základní školy." MA thesis, Univerzita Karlova v Praze, Pedagogická fakulta, Katedra primární pedagogiky.

Horálková, Lenka. 2013. "Hodnocení žáka učitelem při odborném výcviku vzdělávacího oboru kuchař – číšník." BA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky a psychologie.

Horváth, Petra. 2020. "Postoje žáků 1. ročníků středních škol k hodnocení." BA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky a psychologie.

Hozmanová, Veronika. 2015. "Hodnocení ve výtvarné výchově." MA thesis, Západočeská univerzita v Plzni, Fakulta pedagogická, Katedra výtvarné kultury.

Hušková, Petra. 2010. "Hodnocení žáků ve vyučování." BA thesis, Technická univerzita v Liberci, Fakulta přírodovědně-humanitní a pedagogická, Katedra pedagogiky a psychologie.

Ježová, Marcela. 2009. "Metody a způsoby hodnocení žáků základní školy." MA thesis, Masarykova univerzita, Pedagogická fakulta, Katedra pedagogiky.

Kluska, Petr. 2010. "Představy žáků ZŠ a student SŠ a VŠ o správném a nesprávném zkoušení a hodnocení (známkování)." MA thesis, Masarykova univerzita, Filozofická fakulta, Psychologický ústav.

Kovářová, Radka. 2007. "Hodnocení ve výuce anglického jazyka na 1. stupni základní školy pomocí Evropského jazykového portfolia." MA thesis, Technická univerzita v Liberci, Fakulta pedagogická, Katedra primárního vzdělávání.

Kučerová, Karolína. 2020. "Hodnoticí úsudek učitele a jeho vnímání žákem na gymnáziu." BA thesis, Masarykova univerzita, Filozofická fakulta, Ústav pedagogických věd.

Langmannová, Pavlína. 2013. "Hodnocení ve výuce." BA thesis, Masarykova univerzita, Pedagogická fakulta, Katedra Fyziky, chemie a odborného vzdělávání.

Lonská, Julie. 2011. "Hodnocení a klasifikace ve školní TV." MA thesis, Univerzita Karlova v Praze, Fakulta tělesné výchovy a sportu.

Martiník, Karel. 2018. "Školní hodnocení jako forma podpory učení." BA thesis, Univerzita Karlova v Praze, Pedagogická fakulta, Katedra andragogiky a managementu vzdělávání.

Musilová, Miroslava. 2012. "Hodnocení žáků II. stupně základní školy a jeho vliv na motivaci." MA thesis, Univerzita Karlova v Praze, Filozofická fakulta, Katedra pedagogiky.

Nálezková, Kateřina. 2019. "Hodnocení žáků v tělesné výchově pohledem učitelů základních a středních škol." MA thesis, Masarykova univerzita, Fakulta sportovních studií, Katedra pedagogiky sportu.

Prudká, Alena. 2008. "Problematika hodnocení na 1. stupni ZŠ." MA thesis, Univerzita Karlova v Praze, Pedagogická fakulta, Katedra primární pedagogiky.

Růžičková, Jana. 2008. "Hodnocení na 1. stupni ZŠ." MA thesis, Univerzita Karlova v Praze, Pedagogická fakulta, Katedra primární pedagogiky.

Saňáková, Bronislava. 2018. "Vliv slovního hodnocení a známkování v hodinách AJ na motivaci žáků druhého stupně základní školy." MA thesis, Masarykova univerzita, Pedagogická fakulta, Institut výzkumu školního vzdělávání.

Skutil, Martin. 2020. "Školní hodnocení z pohledu studentů učitelství 1.stupně ZŠ." MA thesis, Univerzita Karlova, Filozofická fakulta, Katedra pedagogiky.

Synková, Karolína. 2013. "Hodnocení a klasifikace žáků zmatematiky na střední škole." BA thesis, Masarykova Univerzita, Pedagogická fakulta, Katedra matematiky.

Šandová, Lucie. 2020. "Hodnocení žáků ve čtení sporozuměním na 1. stupni základních škol." MA thesis, Univerzita Karlova, Pedagogická fakulta, Katedra primární pedagogiky.

Šenková, Radka. 2016. "Hodnocení vpředmětu tělesná výchova na střední škole." MA thesis, Univerzita Hradec Králové, Přírodovědecká fakulta, Katedra tělesné výchovy a sportu, Pedagogická fakulta.

Šťastná, Lada. 2014. "Hodnocení žáka učitelem v předmětech český a anglický jazyk." MA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky a psychologie.

Švarcová, Kateřina. 2017. "Relativita známkování na 1.st. ZŠ." MA thesis, Univerzita Karlova, Pedagogická fakulta, Katedra primární pedagogiky.

Větříšek, Jakub. 2020. "Souvislosti mezi školní výkonovou motivací, self-efficacy, sebekontrolou a školním prospěchem u středoškolských studentů." BA thesis, Masarykova univerzita, Filozofická fakulta, Psychologický ústav.

Weisfeitová, Jana. 2007. "Hodnocení a klasifikace na 1. stupni ZŠ." MA thesis, Jihočeská univerzita v Českých Budějovicích, Pedagogická fakulta, Katedra pedagogiky a psychologie.

Zedníková, Martina. 2006. "Vnímání známkování a slovního hodnocení učiteli, žáky a rodiči." MA thesis, Univerzita Karlova, Pedagogická fakulta, Katedra primární pedagogiky.

Zelníček, Luděk. 2017. "Hodnocení žáků v praktickém vyučování vybrané střední školy." BA thesis, Mendelova univerzita v Brně, Institut celoživotního vzdělávání, Oddělení sociálních věd.

Website content

"Anglický jazyk: Klasifikace předmětu." Střední průmyslová škola stavební, České Budějovice. Accessed 21 March, 2021. https://stavarna.com/skola/predmety/anglicky-jazyk-anj-19.html.

"Angličtina v (odborném) vzdělávání." Accessed 21 March, 2021. http://www.nuv.cz/uploads/Periodika/ZPRAVODAJ/2007/Zp0711pVa.pdf.

"Hodnocení a klasifikace v předmětu anglický jazyk na GJO." Gymnázium Jiřího Ortena, Kutná Hora. Accessed 21 March, 2021. https://www.gymkh.cz/dokumenty/SVP/D%C5%A1_ANJ.pdf.

"Hodnocení ve školní praxi." Accessed 21 March, 2021. https://www.pf.ujep.cz/obecna-didaktika/pdf/Hodnoceni.pdf.

"Hodnocení žáků ve školním roce 2020/2021." Střední škola oděvní a služeb, Vizovice Accessed 21 March, 2021. https://www.souviz.cz/cardfiles/card-6220/card-7913/files/kriteria-hodnoceni-zaku-2020-21.pdf_130e30ce6ed23a38dbcbb030f4a867c11595753357.pdf

Král, Petr. "Hodnocení žáků na střední škole." Seminární práce – Pedagogika II. 2020. Accessed 8 May, 2021. https://kdf.mff.cuni.cz/vyuka/pedagogika/seminarky2020/Kral.pdf

"Kritéria hodnocení." Ministerstvo školství, mládeţe a tělovýchovy. 11 October 2013 Accessed 20 March, 2021. https://www.msmt.cz/file/21151.

"Kritéria hodnocení písemné práce z Anglického jazyka profilové části maturitní zkoušky." SOŠ Karlovy Vary. 16 October, 2020. Accessed 20 March, 2021. https://soskv.cz/uploads/dokumenty/maturita/Krit%C3%A9ria%20hodnocen%C3%AD%20p%C3%ADsemn%C3%A9%20pr%C3%A1ce%20z%20AJ.pdf.

"Kritéria hodnocení písemné práce z cizího jazyka." Akademie řemesel Praha, Střední škola technická. Accessed 21 March, 2021. https://www.zelenypruh.cz/download.php?fid=3249.

"Kritéria hodnocení prezentace." Základní škola Plešivec. Accessed 16 March, 2021. https://www.zsplesivec.cz/data/files/770/1479kriteria-hodnoceni-prezentace---inf.pdf.

"Kritéria hodnocení žáků, obor vzdělání 63-41-M/02 OBCHODNÍ AKADEMIE, 1.–4. ročník vrámci průběžné klasifikace školní rok 2018/2019." První cizí jazyk (anglický jazyk). Obchodní akademie, Střední odborná škola knihovnická a Vyšší odborná škola, Brno, Kotlářská 263. Accessed 21 March, 2021. https://www.oabrno.cz/data/2018_19/OA.pdf.

"Kritéria hodnocení žáků, obor vzdělání 72-41-M/01 INFORMAČNÍ SLUŽBY, 1. –4. ročník vrámci průběžné klasifikace 2019/2020." Anglický jazyk (první cizí jazyk). Obchodní akademie, Střední odborná škola knihovnická a Vyšší odborná škola, Brno, Kotlářská 263.

Accessed 21 March, 2021. https://www.oabrno.cz/data/2019-2020/dokumenty/Informan_sluby_kritra_hodnocen_2019-2020.pdf.

Needell, Claire. 27 June, 2014. "With Writing, the Problem Is in the Grading." Accessed 10 May, 2021. https://www.nytimes.com/roomfordebate/2014/03/10/can-writing-on-a-college-entrance-exam-be-properly-assessed/with-writing-the-problem-is-in-the-grading.

"Normativní a kriteriální hodnocení." Accessed 13 March, 2021. https://www.studentske.cz/2008/03/1-normativn-kriteriln-hodnocen.html.

"Norm-referenced, criterion-referenced and ipsative assessment." Robert Gordon University Aberdeen. Course: Postgraduate Certificate - Tertiary Level Teaching. Educational & Staff Development Section. Accessed 25 March, 2021. http://www2.rgu.ac.uk/celt/pgcerttlt/assessing/assess5.htm.

Ostertágová, Alexandra and Michal Rehúš. January, 2021. "Správa z reprezentatívneho prieskumu o priebehu a dopade dištančnej výučby v školskom roku 2020/2021." Accessed 8 July, 2021. https://www.vzdelavacieanalyzy.sk/wp-content/uploads/2021/06/Prieskum-distancne-vzdelavanie_final.pdf. g

"Pravidla hodnocení v předmětech." Gymnázium Jiřího Ortena, Kutná Hora. Accessed 21 March, 2021. https://www.gymkh.cz/pravidla-hodnoceni-v-predmetech/.

"Pravidla hodnocení žáků." Střední pedagogická škola FUTURUM, Hodnotící řád. Accessed 10 March, 2021. https://spgsfuturum.cz/pro-rodice/.

"Přílohy ke sdělení kritéria hodnocení." Sdělení MŠMT – kritéria hodnocení zkoušek společné části maturitní zkoušky a výběrové zkoušky Matematika+ v roce 2020. Ministerstvo školství, mládeže a tělovýchovy. 2013 – 2021. Accessed 12 March, 2021. https://www.msmt.cz/file/52831/.

"Přílohy ke sdělení kritéria hodnocení písemné práce." Sdělení MŠMT – kritéria hodnocení zkoušek společné části maturitní zkoušky a výběrové zkoušky Matematika+ v roce 2020. Ministerstvo školství, mládeže a tělovýchovy. 2013 – 2021. Accessed 12 March, 2021. https://www.msmt.cz/file/52834/.

"Research 101: Generalizability." 5 November, 2015. Accessed 10 July, 2021. https://www.hydroassoc.org/research-101-generalizability/.

Shepard, A. Lorrie. February 2000. "The Role of Classroom Assessment in Teaching and Learning." CSE Technical Report 517. CRESST/University of Colorado at Boulder. Accessed 10 January, 2021. https://cresst.org/wp-content/uploads/TECH517.pdf.

"Střední školy." Accessed 10 April, 2021. www.seznamskol.eu/.

"Stupně hodnocení prospěchu a chování v případě použití klasifikace a jejich charakteristiku, včetně předem stanovených kritérií." Národní ústav odborného vzdělávání. 2012. Accessed 21 March, 2021. http://www.nuov.cz/ae/stupne-hodnoceni-prospechu-a-chovani-v-pripade-pouziti.

"Školy online." Accessed 10 April, 2021. www.stredniskoly.cz/skola/.

"Vysvětlivky k průběžné klasifikaci." Fakultní základní škola PedFUK. Accessed 22 March, 2021. http://www.fzsmeziskolami.cz/dokumenty2018/ezk.pdf.

"Vyučující v epidemii covid-19. Závěrečná zpráva z dotazníkového šetření na základních a středních školách." March, 2021. Accessed 10 May, 2021. https://www.jsns.cz/nove/projekty/pruzkumy-setreni/vyucujici_v_epidemii_covid_19_zz_jsns.pdf?fbclid=IwAR05bao5bcN-GMCbVDOnIyNZTMu_8hG3yf5COlgbKYJtqJdmXUsUIRjC8Q0.

"What are rubrics?" Deer Park School District directive in Washington, US. October 2016. Accessed 12 December, 2020. https://www.dpsd.org/wp-content/uploads/2016/10/rubric-scoring.pdf.

Interview

Naďa. January-February, 2021. Skype Interview. Prague, Czech Republic.

7 Résumé

Školská klasifikácia má nepochybne svoje úskalia. Jedným z nich je dozaista povaha systémovosti pri známkovaní. Tá sa stala hlavným terčom nášho výskumu. Pozornosť sme bližšie zaostrili na klasifikáciu anglického jazyka a opýtali sa, do akej miery sa systém známkovania zhoduje naprieč rozličnými strednými školami v Českej republike.

Po teoretickej stránke boli spracované základné teórie vzťahujúce sa k hodnoteniu anglického jazyka. Poukázali sme na to, aké náročné môže byť dosiahnutie objektivity v hodnotení (Harmer, 2004) a ako sebahodnotenie (*self-assessment*) a vzájomné hodnotenie študentov (*peer assessment*) môže dopomôcť k väčšej systémovosti a objektívnosti známkovania. Nesústredili sme sa ale iba na kontroverznosť hodnotenia, priblížili sme i jeho ciele a funkcie (Al Rifai, 2010). Následne sme sa venovali podrobnej analýze rôznych foriem hodnotenia, opierajúc sa o viaceré zdroje: Brown (2003), Petty (2009) a Cheng & Fox (2017). Rozobrali sme rozdiely, výhody a slabiny formatívneho a sumatívneho hodnotenia, definovali sme kriteriálne a normatívne hodnotenie žiakov i jeho využitie v praxi. Nevyhli sme sa ani bodovacím tabuľkám, jednoduchým i tým detailným a sústredili sa na ich silné stránky a ich zúročenie v triedach.

Nakoniec sme sa pristavili pri poslednej forme hodnotenia, ktoré sa delilo na kvalitatívne (slovné hodnotenie) a kvantitatívne (klasifikácia). Klady a zápory oboch typov hodnotenia sme vzájomne porovnali na základe niekoľkých štúdií (Šťastná, 2014; Ježová, 2009). Zároveň sme sa plne od tohto momentu sústredili na kvantitatívne hodnotenie.

Pre potreby našej práce sme chceli zistiť, aké problémové oblasti sa najčastejšie vyskytujú pri tomto type hodnotenia a ako sa im vyhnúť. Podrobne sme preto rozobrali štyri najproblematickejšie oblasti známkovania podľa odborníkov (Ziegenspeck, 2002; Zoeckler, 2007; Starch and Elliot, 1912), medzi ktoré patrili: malá diferenciácia hodnotiacej stupnice, nesystémovosť a nejednotnosť klasifikácie v rámci inštitúcií, diskriminácia na základe pohlavia a nakoniec subjektivita samotného učiteľa. Do úvahy sme následne vzali štandardy v známkovaní, doporučené expertmi v obore (Slavík, 1999; Kolář and Šikulová, 2009; Kratochvílová, 2011; Košťálová et al., 2008) a na ich základe postavili náš výskum. Ak sa odborníci zhodujú, že dodržaním stanovených noriem sa môže dosiahnuť istá jednotnosť a väčšia objektivita v známkovaní, a teda sa podporí celková systémovosť klasifikácie,

potrebovali sme zistiť, do akej miery sú tieto štandardy dodržané na českých stredných školách.

Pri pohľade na doterajší výskum ohľadom klasifikácie sme skonštatovali, že podobná štúdia, ktorá by mapovala systémovosť známkovania anglického jazyka na českých stredných školách chýba. Keďže táto výskumná oblasť sa zdala neprebádaná, stanovili sme si päť výskumných otázok, pomocou ktorých sme sledovali rozsah systematického prístupu v klasifikácii angličtiny na rôznych typoch českých stredných škôl (gymnáziá, SOŠ, SOU).

Výskumné otázky odzrkadľovali štandardy a zároveň sporné oblasti v známkovaní a boli nasledovné:

1. Aké problematické oblasti existujú v rámci hodnotiacej stupnice?
2. Aké existujú rozdiely a problémy v klasifikácii na úrovni inštitúcií?
3. Do akej miery zohráva pri klasifikácii úlohu subjektivita učiteľa?
4. Dodržiavajú učitelia a inštitúcie klasifikačné štandardy?
5. Majú učitelia pomoc a podporu pri klasifikácii?

Následne sme vytvorili cielené špecifické podotázky, ktorých úlohou bolo pomôcť nám zodpovedať jednotlivé otázky výskumné.

Na preskúmanie danej problematiky sme zvolili formu online dotazníka, ktorý sme rozoslali učiteľom stredných škôl v rámci Prahy a blízkeho okolia. Zahrnuté v ňom boli všetky vytvorené podotázky. Výsledky boli nakoniec spracované v empirickej časti, prevažne s použitím diagramov a tabuliek.

V rámci prvej výskumnej otázky sme pozorovali vysokú mieru nesystémovosti a nejednotnosti v klasifikácii angličtiny naprieč školami a to vzhľadom na jednak typ ale i veľkosť školy. Napríklad horná hranica pre udelenie známky „nedostatečně" bola v gymnáziách stanovená v rozmedzí 29% - 60%, čo vytvorilo obrovské spektrum roztrieštenosti v systéme známkovania na tomto type školy. Nesúlad sme takisto zaznamenali pri oprave známok. Väčšina učiteľov totiž pôvodnú známku niekedy nahradí novou, inokedy zasa novú známku k tej pôvodnej pridá.

Výsledky analýzy druhej výskumnej otázky len potvrdili fragmentáciu systému klasifikácie. Učitelia vykazovali vysokú mieru autonómie pri určovaní váhy pre jednotlivé známky.

K rovnakým záverom sme došli i pri ďalšom rozbore, kde sa zistilo, že vyše 90% učiteľov si aspoň niekedy stanovuje hodnotiace kritériá pre väčšinu žiackych aktivít samostatne.

Tretia výskumná oblasť sa zamerala na rolu učiteľa a jeho možnú subjektivitu pri klasifikácii anglického jazyka. Bolo zaujímavé sledovať, že až 32% učiteľov udeľuje žiakovi horšiu známku, než by si zaslúžil, za účelom motivovania k väčšej snahe do budúcnosti. Odborníci jednoznačne odrádzajú od tejto praxe, pretože v skutočnosti k väčšej motivácii nevedie. Tiež bolo zistené, že žiakova aktivita na hodine a jeho snaha a príprava na zadanú úlohu má veľký (prevažne) pozitívny vplyv na známku.

Pri štvrtej oblasti výskumu sme opäť zaznamenali nejednotnosť v systéme hodnotenia písomných prác a prezentácií z angličtiny. Napriek jasným odporúčaniam odborníkov sme boli svedkami relatívne častého spoliehania sa učiteľov na vlastný celkový dojem pri klasifikácii týchto dvoch aktivít. Taktiež navzdory odborným radám, každý tretí učiteľ zverejňoval kritériá známkovania pred skúšaním/testovaním iba z času na čas. Na druhej strane, učitelia jednotne využívali - v odbornej literatúre tak vysoko cenené - sebahodnotenie a vzájomné hodnotenie študentov.

Napriek tomu, že nesystémovosť v klasifikácii angličtiny na stredných školách sa potvrdila, analýza poslednej výskumnej oblasti nás utvrdila v tom, že učiteľom chýba dostatočná podpora a pomoc inštitúcií akokoľvek systém známkovania zjednotiť. Preto sme sa na záver práce pokúsili navrhnúť možné zlepšenia pri klasifikácii, ktoré vychádzajú jednak z odborných rád expertov a jednak z odpozorovaných výsledkov nášho výskumu, ktoré by možno mohli určitý systém do klasifikácie vniesť.

The Traditional Way

PRESENTATION

	LEVEL 4	LEVEL 3	LEVEL 2	LEVEL 1
	Exceptional	Effective	Acceptable	Developing
Knowledge / Understanding				
Demonstrates an understanding of the topic	thorough understanding	considerable understanding ✓	moderate understanding	emerging understanding
Inquiry / Thinking				
Develops and supports an original idea or opinion about the topic	thorough development and support	considerable development and support	moderate development and support ✓	emerging sense of development and support
Communication				
Addresses audience and speaks clearly with fluency, structure, and purpose	high degree of fluency, structure, and purpose	considerable fluency, structure, and purpose ✓	moderate fluency, structure, and purpose	emerging fluency and sense of structure and purpose
Application				
Exercises rhetorical skills such as emphasis, timing, pacing, reasoning, and questioning	high degree of skill	considerable skill	moderate skill	emerging skill ✓

Overall Grade: $\frac{9}{16}$ = 56%

The New Way

PRESENTATION

	LEVEL 4	LEVEL 3	LEVEL 2	LEVEL 1
	Exceptional	Effective	Acceptable	Developing
Knowledge / Understanding				
Demonstrates an understanding of the topic	thorough understanding	considerable understanding ✓	moderate understanding	emerging understanding
Inquiry / Thinking				
Develops and supports an original idea or opinion about the topic	thorough development and support	considerable development and support	moderate development and support ✓	emerging sense of development and support
Communication				
Addresses audience and speaks clearly with fluency, structure, and purpose	high degree of fluency, structure, and purpose	considerable fluency, structure, and purpose ✓	moderate fluency, structure, and purpose	emerging fluency and sense of structure and purpose
Application				
Exercises rhetorical skills such as emphasis, timing, pacing, reasoning, and questioning	high degree of skill	considerable skill	moderate skill	emerging skill ✓
	100%	87%	73%	60%

Overall Grade: 9 (77%)

source: 'What are rubrics?' *Deer Park School District directive* in Washington, US. October 2016. March 2020. Available on-line from https://www.dpsd.org/wp-content/uploads/2016/10/rubric-scoring.pdf

108

Appendix II: Grading Differences Between Different School Sizes/Types

1. Average Lower Limits for Smaller and Bigger Secondary Schools (in %)

GRADE	SMALLER schools	BIGGER schools	DIFFERENCE
'výborně'	87.6%	87.8%	0.2%
'chvalitebně'	73.1%	75.4%	2.3%
'dobře'	58%	62.5%	4.5%
'dostatečně'	42.9%	50.6%	7.7%
'nedostatečně' (percentage upper limit)	41.4%	49%	7.6%

2. Average Lower Limits for Smaller and Bigger 'Gymnáziums' (in %)

GRADE	SMALLER 'Gymnáziums'	BIGGER 'Gymnáziums'	DIFFERENCE
'výborně'	88.3%	87.6%	0.7%[19]
'chvalitebně'	74.3%	76%	1.7%
'dobře'	59.9%	63.6%	3.7%
'dostatečně'	46.1%	52.4%	6.3%
'nedostatečně' (percentage upper limit)	45.1%	50.4%	5.3%

3. Average Lower Limits for Smaller and Bigger SOŠ Schools (in %)

GRADE	SMALLER SOŠ	BIGGER SOŠ	DIFFERENCE
'výborně'	87.5%	88%	0.5%
'chvalitebně'	72.8%	74.3%	1.5%
'dobře'	57.4%	60.7%	3.3%
'dostatečně'	41.8%	47.6%	5.8%
'nedostatečně' (percentage upper limit)	41%	46.7%	5.7%

---

[19] 'Výborně' is the only grade where the percentage lower limit in smaller 'gymnáziums' is stricter than in bigger 'gymnáziums'. However, this slight irregularity can perhaps be dismissed as it does not seem to affect the general trend of stricter grading in bigger schools, be it 'gymnáziums' or SOŠ schools.

Appendix III: Biases Confirmed and Refuted

* Ts – teachers, Ss – students

RESEARCH QUESTION	CONFIRMED BIAS *(what needs improving)*	REFUTED BIAS *(what Ts do well)*
1. Grading scale bias	high variety in percental grade limits	differentiating grading scale using +/-
	inconsistent handling of original grade after retake	
	counting arithmetic average	
2. Institutional bias	Ts* determine grade weight alone	variety of graded activities
	inconsistency in most/least graded activity weight	
	Ts determine grading criteria alone	Ss* involved in determining grading criteria
3. Subjective bias		grading criteria same for parallel classes
	awarding better grade	no awarding worse grade
	'halo-effect' - tendency to persevere - mistakes caused by negative repercussions - assessing tendencies - order and contrast	- no logical mistake - no gender stereotype - Ts self-aware of own bias - Ts ignore Ss' personal details outside of classroom - grading overall learning process
	inconsistency in marking down mistakes	
4. Following standards	no anonymous essay grading	inclusion of - self-grading - peer grading
	- relying on overall impression when grading - always using only one form of assessment when grading an activity	discussing grade before awarding
	high number of Ts not letting Ss know grading criteria beforehand	
5. Systematic support	not enough systematic institutional support of teachers for range of different graded activities	

Appendix IV: Overview - Five Research Questions and Their Sub-Questions

Research Question	**Sub-questions**
1. What are some of the problematic areas within the convention of the grading scale?	*Question no. 1*: • What is the lower limit in percentages for grades 'výborně', 'chvalitebně', 'dobře' and 'dostatečně'? (e.g. 'výborně': 100% - 90%) • What is the upper limit in percentages for the grade 'nedostatečně'? (e.g. 50% - 0%)
	Question no. 2: • Do you use plus/minus grading? (e.g. 2+, 3-, …)
	Question no. 3: • Are students allowed retakes? If so, what happens to the original grade?
	Question no. 4: • How many grades do you award a student per semester? • Do you count an arithmetic average from individual course grades to determine the final grade?
	Question no. 5: • Do all your grades carry the same weight?
2. What are some differences and issues in grading on an institutional level?	*Question no. 1*: • Who determines the weight of a grade?
	Question no. 2: • What do you grade?
	Question no. 3: • Grades for which activities carry the most amount of weight at the end of the academic year? • Grades for which activities carry the least amount of weight at the end of the academic year?
	Question no. 4: • Who decides on grading criteria? • Do you sometimes decide on grading criteria by yourself? If so, for which activities?
3. To what extent does the 'subjective element' play a role in grading?	*Question no. 1*: • Do you use the same criteria in parallel classes?
	Question no. 2: • Do you award a better grade in order to motivate the student to keep applying themselves or to motivate them to try harder? • Do you award a worse grade in order to motivate the student to try harder?
	Question no. 3: • Which of the following factor(s) can influence a grade?
	Question no. 4: • Which mistake in essays are students penalized more for – a grammar or spelling mistake?
4. Do teachers and institutions follow expert	*Question no. 1*: • Do you grade essays anonymously, without knowing the student's name beforehand?

defined grading standards?	*Question no. 2*: • How do you grade essays? • How do you grade presentations?
	Question no. 3: • Do you let students know the grading criteria ahead of testing them?
	Question no. 4: • Do you let students grade themselves? • Do you let students grade each other?
	Question no. 5: • Do you discuss or consult the grade before awarding it to the student? • If so, whom do you discuss or consult the grade with?
5. Are teachers supported in grading?	*Question no. 1*: • Has a workshop/seminar/class on the topic of grading ever been provided for you by your employer/university?

Známkování anglického jazyka

Dotazník se týká známkování v hodinách anglického jazyka. Jeho vyplněním nám pomůžete zmapovat systém známkování v tomto předmětu na českých středních školách.

Označte, prosím, odpověď, která se nejvíc hodí pro Vaše hodiny angličtiny. Budete-li chtít, můžete vyplnit i komentář.

Za vyplnění dotazníku Vám předem děkujeme.

1. Na jaké škole momentálně učíte?

- a) gymnázium
- b) střední odborná škola
- c) střední odborné učiliště
- d) jiná

2. Jak dlouho učíte?

- a) méně než 2 roky
- b) méně než 10 let
- c) více než 10 let

3. Kolik žáků navštěvuje Vaši školu?

- a) méně než 500
- b) 500 až 1000
- c) více než 1000

4. Kde se nachází Vaše škola?

- a) v Praze
- b) jiné

5. Měli jste někdy možnost absolvovat školení / seminář / workshop v rámci zaměstnání nebo během studia na téma známkování?

- a) ano
- b) ne
- c) nepamatuji si

Možný komentář:

…………………………………………………………………………………………………...………
…………………………………………………………………………………………..………………
…………………………………………………………………………………………..

6. Co hodnotíte známkou? Označte, prosím, všechny možnosti, které se hodí:

- a) závěrečné testy
- b) testy za jednotlivé lekce

c) kratší testy zaměřené na slovní zásobu a gramatiku
d) ústní zkoušení
e) projekty
f) aktivitu v hodině (domácí příprava žáka, jeho práce v hodinách…)
g) mimoškolní činnosti (např. olympiády)
h) jiné

7. Mohou se žáci oznámkovat sami?

a) ano
b) ne
c) někdy

8. Podílejí se na známkování spolužáka ostatní žáci ve třídě?

a) ano
b) ne
c) někdy

9. Používáte při známkování znaménka „plus“, nebo „minus“? (např. 2+, 3- ….)

a) ano
b) ne
c) někdy

Možný komentář:

………………………………………………………………………………………….………
………………………………………………………………………………………….………
……………………………………………………………………………………….

10. Je výsledná známka na konci školního roku určena na základě výpočtu průměru ze všech známek?

a) ano
b) ne
c) někdy

Možný komentář:

………………………………………………………………………………………..………
………………………………………………………………………………………..…………
……………………………………………………………………………………..

11. Kolik známek obvykle dostane žák za pololetí? Prosím, doplňte:

a) méně než 5
b) 5 až 6
c) 7 a více

12. Doplňte, prosím, procentuální hodnocení ke korespondujícím známkám: (např. výborně: 100–88 %)

1. VÝBORNĚ:
2. CHVALITEBNĚ:
3. DOBŘE:
4. DOSTATEČNĚ:
5. NEDOSTATEČNĚ:

Možný komentář:

…………………………………………………………………………………………..………

…………………………………………………………………………………..………………

………………………………………………………………………………..

13. Udělujete lepší známku, abyste žáka motivovali k tomu, aby se i nadále snažil nebo své úsilí zvýšil?

   a) ano
   b) ne
   c) někdy

14. Udělujete žákovi horší známku, abyste jej motivovali k větší snaze?

   a) ano
   b) ne
   c) někdy

15. Známkujete písemné práce?

   a) ano
   b) ne

16. Co hodnotíte při písemné práci? (Můžete označit i více odpovědí.)

a) zpracování zadání a obsah
b) organizaci a kohezi textu
c) slovní zásobu a pravopis
d) mluvnické prostředky a srozumitelnost textu
e) celkový dojem
f) jiné

Možný komentář:

………………………………………………………………………………………………..………
………………………………………………………………………………………………..………………
………………………………………………………………………………………………..

17. Co má v písemné práci na známku největší vliv? (Můžete označit i více odpovědí.)

a) zpracování zadání a obsah
b) organizace a koheze textu
c) slovní zásoba a pravopis
d) mluvnické prostředky a srozumitelnost textu
e) celkový dojem
f) jiné

18. Která chyba má větší váhu při hodnocení písemné práce?

a) pravopisná chyba
b) gramatická chyba
c) obě mají stejnou váhu
d) jiné

Možný komentář:

………………………………………………………………………………………………..………
………………………………………………………………………………………………..………………
………………………………………………………………………………………………..

19. Opravujete písemné práce svých žáků anonymně, bez toho aniž byste znali jméno žáka?

a) ano
b) ne
c) někdy

Možný komentář:

………………………………………………………………………………………………..………
………………………………………………………………………………………………..………………
………………………………………………………………………………………………..

20. Jak známkujete písemné práce? (Můžete označit i více odpovědí.)

a) pomocí jednoduché bodovací tabulky (např. obsah = 5 bodů, styl = 4 body…)
b) pomocí detailní bodovací tabulky (kritéria dílčích částí jako obsah nebo styl jsou podrobně popsána)
c) pomocí kriteriálního hodnocení (splnil / nesplnil)
d) pomocí normativního hodnocení (výkon žáka je hodnocen ve vztahu k výkonům ostatních, kteří plní stejný úkol)
e) na základě vlastního celkového dojmu
f) jiné

Možný komentář:

……………………………………………………………………………………………..………
……………………………………………………………………………………………..………………
………………………………………………………………………………………..

21. Známkujete prezentace?

a) ano
b) ne

22. Co hodnotíte při prezentaci? (Můžete označit i více odpovědí.)

a) splnění zadání
b) obsah projevu
c) odbornou úroveň
d) strukturu
e) slovní zásobu
f) gramatiku a přesnost
g) plynulost projevu
h) výslovnost
i) neverbální projev a styl
j) vizuální pomůcky
k) distribuci času
l) celkový dojem
m) jiné

Možný komentář:

……………………………………………………………………………………………..………
……………………………………………………………………………………………..………………
………………………………………………………………………………………..

23. Co má při prezentaci na známku největší vliv? (Můžete označit i více odpovědí.)

a) splnění zadání
b) obsah projevu
c) odborná úroveň
d) struktura
e) slovní zásoba
f) gramatika a přesnost

g) plynulost projevu
h) výslovnost
i) neverbální projev a styl
j) vizuální pomůcky
k) distribuce času
l) celkový dojem
m) jiné

24. Jak známkujete prezentace? (Můžete označit i více odpovědí.)

a) pomocí jednoduché bodovací tabulky (např. obsah = 5 bodů, styl = 4 body…)
b) pomocí detailní bodovací tabulky (kritéria dílčích částí jako obsah nebo styl jsou podrobně popsána)
c) pomocí kriteriálního hodnocení (splnil / nesplnil)
d) pomocí normativního hodnocení (výkon žáka je hodnocen ve vztahu k výkonům ostatních, kteří plní stejný úkol)
e) na základě vlastního celkového dojmu
f) jiné

Možný komentář:

…………………………………………………………………………………………..………
…………………………………………………………………………………………..………………
………………………………………………………………………………………..

25. Zveřejňujete kritéria známkování před zkoušením / testováním?

a) ano
b) ne
c) někdy

Možný komentář:

…………………………………………………………………………………………..………
…………………………………………………………………………………………..………………
………………………………………………………………………………………..

26. Používáte stejná kritéria známkování v paralelních třídách?

a) ano
b) ne
c) někdy

27. Kdo stanovuje kritéria známkování? (Můžete označit i více odpovědí.)

a) vedení školy / předmětová komise
b) já sám/a, ale do procesu zapojuji i studenty
c) řídím se doporučeními učebnice / učebních materiálů
d) opírám se o rady zkušenějších kolegů
e) jiné

28. Stanovujete někdy kritéria známkování jenom sám/a?

a) ano
b) ne

Už jenom 10 otázek a je to!

29. Při kterých aktivitách určujete kritéria známkování sám/a?

a) závěrečné testy
b) testy za jednotlivé lekce
c) kratší testy zaměřené na slovní zásobu a gramatiku
d) ústní zkoušení
e) projekty
f) prezentace
g) diktáty
h) aktivita v hodině (domácí příprava žáka, jeho práce v hodinách…)
i) mimoškolní činnosti (např. olympiády)
j) jiné

Možný komentář:

………………………………………………………………………………………...………
………………………………………………………………………………………..……………
………………………………………………………………………………..

30. Mají všechny Vámi udělené známky stejnou váhu?

a) ano
b) ne
c) někdy

31. Kdo určuje váhu známky?

a) vedení školy / předmětová komise
b) já sám/a
c) kombinace obou
d) jiné

Možný komentář:

………………………………………………………………………………………...………
………………………………………………………………………………………..……………
………………………………………………………………………………..

32. Co má nejvyšší váhu pro udělení výsledné známky na konci roku? (Můžete označit i více odpovědí.)

a) závěrečný test
b) testy za jednotlivé lekce
c) kratší testy zaměřené na gramatiku a slovní zásobu
d) ústní zkoušení
e) prezentace

f) projekty
g) písemné práce
h) diktáty
i) aktivita žáka v hodině (domácí příprava, práce v hodinách…)
j) mimoškolní aktivita (např. olympiády)
k) jiné

Možný komentář:

………………………………………………………………………………………………..………

………………………………………………………………………………………………..………

…………………………………………………………………………………..

33. Co má nejnižší váhu pro udělení výsledné známky na konci roku? (Můžete označit i více odpovědí.)

a) závěrečný test
b) testy za jednotlivé lekce
c) kratší testy zaměřené na gramatiku a slovní zásobu
d) ústní zkoušení
e) prezentace
f) projekty
g) písemné práce
h) diktáty
i) aktivita žáka v hodině (domácí příprava, práce v hodinách…)
j) mimoškolní aktivita (např. olympiády)
k) jiné

Možný komentář:

………………………………………………………………………………………………..………

………………………………………………………………………………………………..………

…………………………………………………………………………………..

Už jenom pět otázek a je to! Děkujeme :-)

34. Diskutujete nebo se radíte o známce předtím, než ji udělíte žákovi?

a) ano
b) ne
c) někdy

35. S kým o známce diskutujete nebo se radíte před jejím udělením žákovi? (Můžete označit i více odpovědí.)

a) s Vašimi kolegy
b) s daným žákem
c) s celou třídou

36. Které z následujících možností mohou mít vliv na udělení známky? (Můžete označit i více odpovědí.)

    a)   chování žáka v hodině
    b)   aktivita žáka v hodině
    c)   Váš vztah k žákovi
    d)   snaha a příprava žáka na daný úkol
    e)   předchozí známky žáka ve Vašem předmětu
    f)   známky žáka v jiných předmětech
    g)   Vaše vlastní zkušenost ze školy
    h)   žákova reputace
    i)   známka předchozího zkoušeného žáka/žáků
    j)   nečitelnost písma
    k)   jestli je žák chlapec nebo dívka
    l)   obava, že žák bude opakovat ročník
    m) jiné

Možný komentář:

……………………………………………………………………………………………..………
……………………………………………………………………………………………..………………
…………………………………………………………………………………………..

37. Mají žáci možnost si známku opravit?

    a)   ano
    b)   ne
    c)   někdy

38. Pokud si žák známku opraví:

    a)   nová známka nahradí původní
    b)   nová známka se přidá k původní
    c)   někdy nahradí, někdy přidá

39. Chtěli byste nám ještě něco sdělit?

………………………………………………………………………………………….………
………………………………………………………………………………………….………………
…………………………………………………………………………………….....

Moc děkujeme za vyplnění dotazníku, hodně jste nám pomohli. Přejeme hezký den :-)