

DOCTORAL (PHD) DISSERTATION

Márta Barbarics

Alternative Assessment Methods in English as a
Foreign Language and in English Medium
Content Classes in Hungarian Public
Secondary Schools

2023

EÖTVÖS LORÁND UNIVERSITY
FACULTY OF EDUCATION AND PSYCHOLOGY

Márta Barbarics

**Alternative Assessment Methods in English as a
Foreign Language and in English Medium Content
Classes in Hungarian Public Secondary Schools**

DOI: 10.15476/ELTE.2023.290

Doctoral School of Education

Head of the Doctoral School: Prof. Dr. Anikó Zsolnai, DSc, habil.

PhD Programme in Language Pedagogy

Programme director: Prof. Dr. Krisztina Károly, DSc, habil.

Supervisor: Dr. Ildikó Lázár, PhD, habil.

Budapest, 2023

Abstract

The aim of the dissertation is to gain insights into the use of alternative assessment methods in English as a foreign language and in English medium content classes in public secondary education in Hungary. Twelve semi-structured interviews with teachers already using such methods map their views, motivations, purposes, and experiences of using alternative assessment in this context; in addition to a follow-up interview study presenting how they adapted their methods to remote delivery during the first wave of the Covid-19 pandemic. Based on an action research project, 44 students' perceptions of assessment in general and alternative assessment methods are presented both before and after the introduction of emergency remote delivery in English as a foreign language and in English medium content classes in public secondary education in Hungary. Results show that the participating teachers align their assessment methods with their educational goals and continuously improve their classroom practices, which makes it possible to face challenges such as adapting their methods to remote delivery. In line with the literature, participants also emphasize the need for a paradigm shift caused by the advancements of the 21st century and urge policy makers to reflect these changes in official policy documents. As there are central frameworks only for summative assessment and the one to five grading remained the only compulsory output requirement even during remote delivery, teachers would need autonomy and support in developing and implementing alternative assessment methods. Participating students' experiences confirm the disadvantages of traditional assessment in Hungary. Whereas the collected alternative assessment methods seem to have the potential to put students at the center and support their learning processes even in remote delivery.

Table of Contents

1. Introduction	1
2. Literature Review	4
2.1 A Review of Legislation and Official Policy Documents	4
2.1.1 International Policy Documents.....	5
2.1.2 Hungarian National Legislation.....	7
2.1.3 The Pedagogical Programme of the School of the Action Research Project.....	9
2.2 Categorizations and Definitions of Assessment Types.....	13
2.2.1 Categorization Based on the Purpose of Assessment	13
2.2.2 Other Categorizations of Assessment	15
2.2.3 Measurement, Test, Evaluation, and Feedback in Relation to Assessment.....	17
2.2.4 Alternative Assessment.....	20
2.3 Empirical Background	24
2.3.1 Assessment and Grades.....	25
2.3.2 Assessment that Supports Students' Development.....	27
2.3.3 Assessment in English as a Foreign Language and English Medium Content Classes	33
2.4 Further Factors Influencing Alternative Assessment.....	37
2.4.1 Nurturing Students' Intrinsic Motivation.....	37
2.4.2 Developing Complex Skills and Competences.....	39
2.4.3 Assessment Based on Gamification.....	41
2.5 Assessment in Emergency Remote Delivery	44
2.5.1 Terminological Issues	44
2.5.2 Emergency Remote Delivery in the World.....	47
2.5.3 Emergency Remote Delivery in Hungary	48
2.5.4 Recommendations of the Educational Authority Focusing on Assessment	50
3. Research Niche and Research Questions	62

4. Methods.....	65
4.1 Research Paradigm.....	65
4.2 Participants.....	67
4.2.1 The Researcher.....	68
4.2.2 Participants of the Interviews.....	69
4.2.3 Participants of the Action Research.....	71
4.3 Research Instruments.....	73
4.3.1 Interview Schedules.....	73
4.3.2 Instruments of the Action Research.....	75
4.4 Procedures.....	82
4.4.1 Student Questionnaires Between 2014 and 2017.....	82
4.4.2 Interview Study in 2018 and 2019.....	84
4.4.3 Action Research in the School Year of 2019 and 2020.....	85
4.4.4 Follow-up Interview Study in 2020.....	89
4.5 Methods of Data Analysis.....	89
4.6 Summary Table of Research Questions and Methods.....	98
4.7 Ethical and Legal Considerations.....	100
5. Results and Discussion.....	102
5.1 How Teachers' Pedagogical Views are Connected to Their Use of Alternative Assessment	103
5.1.1 Teachers' Competences for Democratic Culture.....	104
5.1.2 Teachers' Views on Conflicts and Sensitive Issues.....	118
5.1.3 Teachers' Views on Teaching and Learning.....	125
5.1.4 Teachers' Views Related to School Subjects, Materials, and Tasks.....	128
5.1.5 Teachers' Views on Assessment.....	130

5.2 Students' and Teachers' Experiences With Traditional Assessment.....	138
5.2.1 Teachers' Definitions of Assessment.....	139
5.2.2 Teachers' Views on Traditional Assessment.....	141
5.2.3 Students' Perceptions of Traditional Assessment.....	144
5.2.4 Issues With Traditional Assessment in the Emergency Remote Delivery	153
5.3 Teachers' Goals for Using Alternative Assessment Methods	157
5.3.1 Teachers' Assessment Related Goals	157
5.3.2 Teachers' Goals in Connection With Learning and School Subjects	160
5.3.3 Teachers' Goals Related to Students' Life Beyond School.....	163
5.3.4 Teachers' Personal Goals for Using Alternative Assessment.....	167
5.3.5 Teachers' Goals Related to Assessment During the Emergency Remote Delivery ..	168
5.4 Alternative Assessment Methods and Their Consequences	171
5.4.1 Solutions for Grading.....	172
5.4.2 Alternatives for Testing	182
5.4.3 Assessment Activities That Help Process the Material	191
5.4.4 Further Alternative Assessment Methods.....	198
5.4.5 Alternative Assessment Methods Used During the Emergency Remote Delivery....	214
5.5 Students' Perceptions of Alternative Assessment Methods.....	224
5.5.1 Reactions to Point Collecting.....	225
5.5.2 Reactions to Other Assessment Types.....	242
5.5.3 Reactions to Assessment During the Emergency Remote Delivery.....	259
6. Conclusion.....	268
6.1 The Findings Relating to the Research Questions	268
6.2 Pedagogical and Policy Implications and Suggestions.....	276
References.....	280

Appendix A: Hungarian-English glossary of related expressions as used in the dissertation.....	292
Appendix B: Links of websites and applications mentioned in the dissertation.....	294
Appendix C: Interview guides.....	297
Appendix D: Samples from the instruments of the action research.....	300
Appendix E : Student questionnaires.....	302
Appendix F: The consent form signed by the parents.....	320
Appendix G: Description of the point-collecting system for the 9th grade IT English group.....	322
Appendix H: Samples of alternative assessment methods mentioned in the teacher interviews.....	323
Appendix I: Samples of alternative assessment methods used during the action research.....	332
Appendix J: Data tables of students' questionnaire answers.....	344

1. Introduction

Assessment plays a major role in teaching and learning processes. According to the OECD (2013), promoting assessment that leads to the improvement of educational practices in addition to supporting student learning is essential in order to establish a high-performing education system; moreover, assessment frameworks should be the driving force of the reform agenda. The advancements of the 21st century necessitate a paradigm shift in education that has been accelerated by the Covid-19 pandemic resulting in emergency remote delivery worldwide as the Hungarian Educational Authority states in its Collection of digital pedagogical methodological recommendations (Farkas et al., 2021). The recommendations (Farkas et al., 2021) encourage teachers to replace some of their traditional assessment tools during remote delivery and consider ways of alternative assessment that can be utilized in face-to-face teaching and learning as well, while it is not the case the other way around: traditional assessment practices seem to be inadequate for remote delivery. As Báthory (1997) notes, assessment in education was not thoroughly researched in Hungary before the 1970s. According to Csapó (1992), it was due to political reasons; however, it is possible to explore what traditional assessment in Hungary connotes.

According to Hungarian governmental decrees and ministerial acts in force (Act CXC, 2011), teachers are obliged to evaluate students' knowledge, behavior, and diligence in the form of grades on a scale from one to five, as a consequence of which, assessment is mainly based on these numerical values. These grades have a substantial impact on students' lives as they define their enrolment possibilities in all levels of education (Act CCIV, 2011). Therefore, it is no surprise that stakeholders (schools, parents, and students) lay considerable emphasis on grades.

The one to five grading is deeply rooted in the Hungarian educational tradition. However, there were serious complaints about it as early as the beginning of the 20th century. There is a record from 1912 of dr. Gábor Kemény reading out his paper on "abolishing grades" at the annual meeting of the "central pedagogical department" in Budapest, Hungary, where teachers from various levels of education expressed their agreement with him (Kemény, 1912). For example, Samu Biró (as quoted by Kemény, 1912), an elementary school principal argues that if the use of grades makes no sense in secondary schools, it is even more harmful in primary schools. He says that students' knowledge is individual and cannot be fairly assessed with a number from one to five. He also supports his argument with rhetorical questions and real life examples, such as the fact that some students come to school tired after work while others come from wealthy families

supporting their education, so it is not their knowledge that is being graded. Middle school teacher Ágostné Martos (as quoted by Kemény, 1912) agrees that grades cannot reflect students' knowledge and they are used as tools of intimidation. She asked about 300 students what their feelings were in connection with their school report, and all of the students, even the exemplary ones with only perfect grades, were terrified and afraid waiting for their school reports containing their grades. The discussion ends with dr. Ödön Weszely, head of the department, requesting country-wide research and for their next meeting suggestions for alternative assessment that could replace grading (Kemény, 1912, p. 435). According to Csapó (1992), Gábor Kemény wrote the first comprehensive book on assessment in the Hungarian education system in 1934, and based on his empirical research, Kemény argued against the routine use of grades, outlining “the psychological principles of ‘proper evaluation,’ by which he meant a qualitative, verbal, motivating, and formative type of evaluation instead of the use of scales or numbers for grading” (Csapó, 1992, p. 6).

According to the governmental decrees and ministerial acts, school assessment should also serve the development of students, for example by “contributing to the harmonic mental, physical and intellectual development of children and young people through the conscious development of their skills, abilities, knowledge, proficiencies, emotional and volitional characteristics and cultural education corresponding to their age characteristics” (Act CXC, 2011, Section 1). However, as the above mentioned grading is the only compulsory output requirement and there are no guidelines on how other roles should be fulfilled, assessment serving the development of students depends entirely on teachers' individual assessment practices. Only few studies have been published in connection with assessment in Hungary focusing on ways that aim to fulfill other roles than grade giving (Hubai & Lázár, 2018), hence this research aims to gain insights into the use of alternative assessment methods in English as a foreign language (EFL) and in English medium content (EMC) classes in public secondary education in Hungary.

The research questions guiding the present dissertation aim to explore what teachers mean by assessment and alternative assessment in particular in EFL and in EMC classes in public secondary education in Hungary, what alternative assessment methods they claim to use, and what their views, motivations, purposes, and experiences of using alternative assessment are in this context. The exploratory nature of the research questions require qualitative data collection through purposive sampling of teachers who claim to use alternative assessment methods in EFL

and in EMC classes in public secondary education in Hungary. Twelve semi-structured interviews with innovative teachers already using alternative assessment methods took place to help answer these research questions. In addition to looking at teachers' perspectives, further research questions focus on students' experiences in connection with assessment in EFL and in EMC classes in public secondary education in Hungary, and what their perceptions of alternative assessment are in this context. In order to answer these research questions, based on the results of the interview study, an action research project was designed and carried out during the 2019/2020 school year. The action research project involved 44 students in three groups that I taught that year: a 9th grade beginner IT English group as an EFL class, an 11th grade mathematics in English group and an 11th grade elective advanced mathematics in English group as EMC classes.

Due to the first wave of the Covid-19 pandemic, emergency remote delivery was introduced during the action research project, which generated further research questions. As regards the students, I was interested in their experiences in connection with assessment during emergency remote delivery in public secondary education in Hungary and also their perceptions of the adaptation of alternative assessment methods to remote delivery in EFL and in EMC classes. In order to answer these questions, as part of the action research project data was collected through field notes, teaching journals, class observations, monthly student questionnaires, and focus group interviews with the participating students at the end of the action research project. I was also interested in how participants of the interview study could cope with the emergency remote delivery situation, for this reason a follow-up interview study took place with the twelve interviewees of the original interview study in order to explore how alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary can be adapted to remote delivery. The data was analyzed using the constant comparative method (Dörnyei, 2007).

As for the structure of the dissertation, the literature review begins with a review of educational acts and official policy documents concerning secondary school assessment from an international to a local level. The second section reviews the theoretical background including categorizations and definitions of assessment types and ending with the focus of the dissertation: alternative assessment. The literature review also presents empirical research in connection with alternative assessment and the context of the dissertation: assessment in EFL and EMC classes. The chapter ends by discussing further factors that can influence alternative assessment with a separate section dedicated to assessment in emergency remote delivery. Based on the structure of

the literature review, the research niche is presented, followed by the research questions. The methods chapter introduces the research paradigm, participants, instruments, procedures, data analysis, and ethical and legal considerations. Results and their discussion are presented according to the themes that emerged during the data analysis. The dissertation ends with the summary of the main findings by answering the research questions including limitations, further research possibilities, and pedagogical and policy implications and suggestions, followed by references, and the appendices which include a glossary, links, samples from all research instruments, samples from alternative assessment methods mentioned in the interviews and used during the action research project, and data tables of student questionnaire answers.

2. Literature Review

The literature review starts with a review of legislation and official policy documents from an international level to a school's pedagogical programme. As the context of the dissertation is Hungarian public secondary schools, rules regulating the assessment of students' achievement in this context are discussed in detail. They are followed by different categorizations and definitions of assessment types and terminology in relation to assessment; the chapter ends with the definition of alternative assessment that is the focus of the dissertation. The empirical background chapter presents research in connection with alternative assessment in general and in the context of English as a foreign language (EFL) and English medium content (EMC) classes. Some further factors that can influence the use of alternative assessment such as motivation, skills and competence development, and gamification are also discussed. The literature review ends with a last circumstance that impacted assessment: the Covid-19 pandemic resulting in emergency remote delivery of education worldwide.

2.1 A Review of Legislation and Official Policy Documents

Starting from an international level, first the OECD's (2013) Review on Evaluation and Assessment Frameworks for Improving School Outcomes is analyzed from the perspective of student assessment in public secondary education. Then national legislation regulating student assessment in Hungary is described. The last section of the review of legislation and official policy documents (2.1.3 The Pedagogical Programme of the School of the Action Research Project) presents an example of the local level of how individual schools specify the rules of student assessment in Hungary.

2.1.1 International Policy Documents

In 2009, the OECD Education Policy Committee launched a project in order to publish a Review on Evaluation and Assessment Frameworks for Improving School Outcomes analyzing evaluation and assessment practices internationally and providing recommendations for improvement (OECD, 2013). The resulting report (OECD, 2013) reviews different components of assessment and evaluation such as student assessment, teacher appraisal, school evaluation, and system evaluation on the primary and secondary levels of education. According to the context of the present dissertation, student assessment in secondary schools is reviewed on an international level as Hungary did not opt to take part in the OECD's country reviews. According to the report (OECD, 2013), although OECD countries take different approaches to evaluation and assessment, the following are clear policy priorities: taking a holistic approach to generate synergies between components, aligning evaluation and assessment with educational goals, focusing on the improvement of classroom practices and building on teacher professionalism, avoiding distortions such as teaching to the test, putting students at the center, building capacity at all levels of the education system, meeting local needs from regional to school levels, and designing successfully by building consensus among all stakeholders (pp. 14-15).

Looking at student assessment, several challenges emerged for which the review (OECD, 2013) recommends policy directions. The main challenges in evaluation and assessment in the domain of student assessment are the following: aligning and balancing factors such as educational standards and student assessment, summative and formative assessment, external assessments and teacher-based assessments in the assessment of learning, “developing fair assessments to all student groups, designing large-scale assessments that are instructionally useful, ensuring fairness in assessment and marking across schools, securing informative reporting of student assessment results” (OECD, 2013, Table 1, p. 20). Facing these challenges, the report recommends the following main policy directions for student assessment:

- ensure a good balance between formative and summative assessment,
- establish safeguards against an overreliance on standardised assessments,
- draw on a variety of assessment types to obtain a rounded picture of student learning,
- support effective formative assessment processes,
- ensure the consistency of assessment and marking across schools,
- ensure that student assessment is inclusive and responsive to different learner needs,

put the learner at the centre and build students' capacity to engage in their own assessment, maintain the centrality of teacher-based assessment and promote teacher professionalism, engage parents in education through adequate reporting and communication (OECD, 2013, Table 2, p. 21).

The report (OECD, 2013) reviews research showing the importance and impact of different types of assessment on student learning outcomes and arrives at the following conclusions. On the one hand, "carefully planned assessment interventions that are well aligned with learning goals and place students at the centre of the process have strong potential to raise achievement and reduce disparities"; however, "assessments that are not well designed and implemented may in fact contribute to alienating students (and teachers) from the education system and exacerbate inequity in education" (OECD, 2013, p. 144). Elaborating on different types of assessment, the report reviews research on the influence of formative and summative assessment (OECD, 2013, pp. 145-146) that will be detailed in chapter 2.3, Empirical Background of the dissertation. Issues in connection with fostering students' intrinsic and extrinsic motivation, developing competencies and 21st century skills are also reviewed in the report (OECD, 2013, pp. 146-148) and will be discussed in chapter 2.4, Further Factors Influencing Alternative Assessment. The impact, drivers and contextual developments of student assessment section concludes that "despite a high degree of interest in teaching and assessing new forms of learning, information collected in the OECD Review indicates that the use of innovative assessment approaches remains quite limited within the national assessment frameworks of OECD countries" (OECD, 2013, p. 149). Nevertheless, the following sections present some "promising approaches to achieving better alignment between competency-based curricula and assessment approaches, both large scale and classroom-based" (p. 150), none of which are connected to Hungary.

The section dealing with the governance of student assessment systems across OECD countries explores policy frameworks for both summative and formative assessment (OECD, 2013). In relation to Hungary, the report (OECD, 2013) finds the existence of central frameworks for internal summative assessment on all primary, lower secondary, and upper secondary levels of education (p. 154), the existence of standardized central examinations at the end of secondary education (p. 155), and the fact that there are no central frameworks for formative assessment on any of the primary, lower, or upper secondary levels of education (p. 156). In addition, "in

Hungary, elements of formative assessment such as verbal assessment and differentiated assessment methods are included in legal regulations and the national core curriculum” (OECD, 2013, p. 156). Hungary also appears in connection with the possibility of using central assessment for formative purposes on all educational levels. Although the report only states the existence of standardized central assessments with no stakes for students (OECD, 2013, p. 158), it might refer to the National Assessment of Basic Competencies (National ABC) at the end of grades 6, 8, and 10 and the international level assessments of PISA, TIMSS, and PIRLS (OECD, 2015). Due to the scope of the dissertation, these centralized assessments are not detailed further. The next section presents legislation of student assessment in Hungary.

2.1.2 Hungarian National Legislation

The official rules and regulations of student assessment in Hungary are described in the current Act on National Public Education (Act CXC, 2011). According to the chapter entitled “Fulfilment of Student Obligations”, teachers should “regularly evaluate the student’s performance and progress in form of grades throughout the teaching year and rate it in forms of marks at the end of the term and the teaching year” (Act CXC, 2011, Section 54, paragraph 1). As the wording of the act uses “grade” (in Hungarian “*érdemjegy*”) for assessment throughout the year, while “mark” (in Hungarian “*osztályzat*”) for assessment at the end of the term and the teaching year, it could refer to a formative-summative distinction; however, there are no other signs of such differences. In terms of meaning, grades and marks are used interchangeably, and in terms of official translations, English terminology is not consistent either (for example in Act CCIV, 2011, “marks” are used in English when “*érdemjegy*” features in Hungarian). For these reasons, in my translations, I will only use “grades” (for both “*(érdem)jegy*” and “*osztályzat*”). In the official translations of the acts, the terms “evaluation” and “assessment” are also used synonymously.

Act CXC, 2011, Section 54, paragraph 2, continues to detail the grading system by distinguishing three categories:

Grades and marks should be as follows:

a) evaluation and assessment of the student’s knowledge:

excellent (5), good (4), average (3), satisfactory (2), unsatisfactory (1),

b) evaluation and assessment of the student’s behaviour:

exemplary (5), good (4), variable (3), bad (2),

c) evaluation and assessment of the student's diligence:

exemplary (5), good (4), variable (3), negligent (2). (Act CXC, 2011, Section 54, paragraph 2)

Therefore, the grade category of knowledge is assessed on a scale from one to five, and the grade categories of behavior and diligence are assessed on a scale from two to five as there is no failing grade for behavior or diligence.

Admission to secondary and tertiary education is heavily dependent on grades. The Ministerial Decree 20/2012 (VIII. 31.) on the operation of public education institutions by the Ministry of Human Capacities states that secondary schools can decide whom to accept based on the given student's previous academic record, meaning the end of term grades, and the results of a centralized written exam on mathematics and Hungarian language. According to Act CCIV of 2011 on National Higher Education, admission to tertiary education, similarly to secondary education, depends on the earlier school performance of the applicants, meaning the end of term grades, and the results of the centralized secondary school leaving matura examination (Act CCIV, 2011, Article 40). Although in both cases "institutions may make admission subject to the fulfilment of reasonable and non-discriminatory requirements" (Act CCIV, 2011, Article 40), the decision is mainly based on students' grades and exam results. The OECD (2013) report of reviews of evaluation and assessment in education also notes that "Hungary, [along with many other European countries,] relies primarily on numerical marks for formal reporting" (p. 204). As a consequence of this, grades and exam preparation play an important role for all stakeholders (schools, parents, and students).

On the other hand, the act on education starts with the section entitled "Purpose and Principles of the Act", stating that schools should provide "comprehensive evaluation adjusted to the requirements and ensuring the development of children / students" (Act CXC, 2011, Section 1, paragraph 1), so according to the purpose and principles of the act, assessment should serve the development of students. Moreover, Section 64 deals with the promotion system of teachers that is based on eight competences of teachers specified in section 7 of Government Decree 326/2013 (VIII. 30.), where the third teacher competence is "providing learning support"; the fourth is "developing students' personality, [providing] individual treatment [...]"; the sixth is "ongoing evaluation and analysis of educational processes and the development of students' personality" are all in line with the purpose and principles of the educational act stating that teachers should develop

students' personalities with the help of evaluating and analyzing educational processes. Although section 62 (Act CXC, 2011) mentions that assessment should include explanation in the form of oral or written feedback, as Hubai and Lázár (2018) point out, without any guidelines or examples regarding the execution of it, teachers might not know how they should explain the grades they give to their students. Hubai and Lázár (2018) also draw attention to the following similar controversies in the National Core Curriculum (2012). In terms of assessment, it requires differentiation as a basic principle, taking into consideration the development of talented students, incorporating different types of assessment, while also applying assessment on unified grounds without any “direct guideline or practical advice [...] on what schools are supposed to do to succeed in terms of differentiation and uniformity at the same time” (Hubai & Lázár, 2018, p. 86). The National Core Curriculum (2012) also describes what the “educational program” of a school should contain, which Hubai and Lázár (2018) assume to refer to schools' pedagogical programmes, which in terms of assessment requires to reflect the “pedagogical concept, [...] the purposes, requirements, contents, and timeframes of the planned teaching and learning processes [...] that will help to control and assess students' knowledge, achievement and development” (Hubai & Lázár, 2018, p. 86). The following section reviews how assessment is present in a school's pedagogical programme.

2.1.3 The Pedagogical Programme of the School of the Action Research Project

It is each school's pedagogical programme that specifies the implementation of the rules and regulations detailed in the previous section. The Act on National Public Education (Act CXC, 2011, Section 54, paragraph 4) also allows for diversion in the following way:

the pedagogical programme of the school may prescribe a different marking system from those specified in paragraph (2) or use written assessments. Shall the school not apply the stipulations of paragraphs (1)–(2), but this form of assessment shall be needed due to a change of school or continuing studies, or on request of the parent or the student, mid-term and year-end assessments by marks shall be prepared. The rules of conversion from the marking system and evaluation applied by the school to marks and grades have to be specified in the local curriculum.

Looking at all the pedagogical programmes of the public secondary schools that the participants of the present research attended, there were no diversions from the one to five midterm and year-end grades required by the law. Only slight differences occur in the descriptions of how these

grades are acquired. Thus, this section reviews the assessment related parts of the pedagogical programme of the school of the action research. As the school is a vocational bilingual school, its pedagogical programme is called “vocational programme” (Gál-Berey, 2022); however, it includes the same sections as pedagogical programmes of general secondary schools in addition to sections specific to vocational and bilingual education in this case.

The introduction of the chapter on student assessment of the pedagogical programme (Gál-Berey, 2022) describes the content of the official rules and regulations quoted in section 2.1.2, Hungarian National Legislation. It reiterates that teachers should assess “the student’s performance and progress in form of grades throughout the teaching year and rate it in forms of marks at the end of the term and the teaching year” (Act CXC, 2011, Section 54, paragraph 1). It also includes how students and their parents should be notified of the grades and marks, in addition to the names of the grades in three categories of evaluating students’ knowledge (5: excellent, 4: good, 3: average, 2: satisfactory, 1: unsatisfactory), behavior (5: exemplary, 4: good, 3: variable, 2: bad), and diligence (5: exemplary, 4: good, 3: variable, 2: negligent). It is also stated that grades evaluating students’ knowledge are weighted and in order to determine the midterm and year-end marks weighted averages are calculated. It is teachers’ responsibility to provide at least three full grades (that weigh as one) each term.

Students’ rights in relation to their assessment are described in the next section of the pedagogical programme (Gál-Berey, 2022). Students have the right to possess the following information: “requirements imposed on them, indicators of performance levels, key points of the curriculum to be mastered”, “forms of assessment, methods that form the basis of evaluation (written, oral, practical task etc.), and their weight in the assessment”, and “retake possibilities, forms, and methods including its value and weight in the assessment” (Gál-Berey, 2022, p. 89). There are also restrictions on what can be graded. Students cannot receive grades for the following:

- lack of equipment required for a lesson;
- two grades cannot be given for the same test, except in exceptional cases that must be communicated to the student in advance (e.g. content and grammar for essays etc.);
- disciplinary purposes or behavioral or diligence failures;
- results of diagnostic assessments [...] (Gál-Berey, 2022, p. 90)

The following types of grades can be submitted in the national compulsory e-gradebook system with their weights following them in brackets (see Hungarian-English glossary in Appendix A):

written achievement test (1), written pop quiz (0.5), oral test (0.5), class work (0.25), homework (0.25), homework (0.5), home test (0.5), individual or group project work (1), exam (2), minimatura (2) (Gál-Berey, 2022, p. 90). There are additional grade types that are connected to vocational subjects, which are beyond the scope of the present research and for this reason will not be discussed here.

The pedagogical programme (Gál-Berey, 2022) also details rules that apply for the different types of grades. In connection with written achievement tests, rules and regulations are the following. It must take at least 45 minutes, which is the usual length of a lesson in Hungary, “or longer as specified in the subject’s local curriculum” (Gál-Berey, 2022, p. 91). In connection with the duration of achievement tests, students with special educational needs must be considered and provided with extra time if they possess an official decision requiring it. Some of the rules are highlighted in the pedagogical programme, such as “Teachers should inform students of the date of written achievement tests at least one week in advance. During one school day, maximum two written achievement tests can take place” and later “Written achievement tests must be corrected within two weeks and the results must be reported to students. After the deadline, the results can only be validated with the countersignature of the deputy director or director of the school” (Gál-Berey, 2022, p. 91). Further rules are that the boundaries for the grades given in percentages must be included in the local requirements of the subject. In addition, if a student is absent from a written achievement test, the student cannot get a failing grade; however, the teacher can require a retake test, and if the student does not use the opportunity, then “their work can be evaluated as unsatisfactory” (Gál-Berey, 2022, p. 91) which is the equivalent for grade one. The pedagogical programme also contains recommendations for the exact grade boundaries expressed in percentages to evaluate achievement tests (Gál-Berey, 2022, pp. 91-92).

The pedagogical programme defines written pop quizzes and oral tests as “the written [and oral] assessment of a smaller teaching unit, typically of the material of a few lessons prior to the test” (Gál-Berey, 2022, p. 92). They do not have to be announced in advance, and their number is not restricted on a school day either. The pedagogical programme also suggests taking into consideration the following criteria when assessing oral tests: “independent expression, expressiveness; realizing connections; processing the curriculum content; thinking skills, problem-solving skills, dynamism” (Gál-Berey, 2022, p. 92). In connection with the assessment of class work, it can be evaluated according to teachers’ own criteria. As an example, the pedagogical

programme lists “the performance of the entire lesson or even a single comment can be graded, but the grade for class work must always be justified” (Gál-Berey, 2022, p. 93). Different types of homework and project work are discussed as a common category. According to the pedagogical programme “homework is an important tool for continuous preparation, extracurricular practice, and independent immersion in the curriculum. Students’ weekly workload must be taken into account when determining and assessing it” (Gál-Berey, 2022, p. 93). Further recommendations about graded work that students should do at home are the following:

the smallest unit of homework is a task directly related to the material of the lesson, assigned from the textbook or other subject source;

students should have at least one working day to prepare written homework;

at least one week must be provided for the preparation of larger-scale homework or project work covering a unit. The completed project work must be assessed.

Students must complete assignments by the deadline set by teachers.

Teachers must inform students in advance if homework will be graded. In this case, an unsatisfactory grade can also be given for not completing the homework, or for forgetting it at home, or due to a missed deadline. (Gál-Berey, 2022, p. 93)

The last category that is connected to the focus of the dissertation is exams, which are worth a double grade. According to the pedagogical programme, the purpose of exams are to “confirm, improve, or make up for students’ grades in order to achieve a realistic midterm or year-end grade” (Gál-Berey, 2022, p. 94). The *minimatura* exam was named after the *matura* exam, which is the centralized nationwide secondary-school leaving examination. The pedagogical programme describes the purpose of the *minimatura* exam as “the time-proportional assessment of readiness for graduation, [...] written in 10th grade but can be deviated from in justified cases. The subject content and evaluation of the exam must be in accordance with the legislation in force for the *matura* exam” (Gál-Berey, 2022, p. 94).

The seven pages of detailed description of grading proves the prevalence of summative assessment. Written achievement tests have the most detailed description of rules and regulations, and they provide the unit weight for grades, in connection with which the requirement is to have three grades each term, showing the importance of achievement tests. There are no recommendations for formative assessment. Students’ self-assessment is mentioned once: “to determine and measure students’ knowledge, abilities, skills, and attitudes, in addition to methods

considered classic in pedagogical literature, we also use guided self-assessment tools, in which scientifically developed surveys and tests play an important role” (Gál-Berey, 2022, p. 88). Neither the methods considered classic nor the mentioned surveys or tests are elaborated on any further.

As it has been shown in the reviewed documents from international policy papers to local pedagogical programmes, legislation prescribes summative assessment. In Hungarian public secondary education, this entails the prevalence of grading on the one to five scale. As already mentioned in connection with Hungarian national legislation, terminology used in connection with assessment is not consistent, so it is necessary to review the categorizations and different definitions of assessment types. For this reason, the next chapter of the literature review synthesizes terminological issues arriving at definitions that will be used in the dissertation.

2.2 Categorizations and Definitions of Assessment Types

According to Sadler (1989), assessment “denotes any appraisal (or judgment, or evaluation) of a student’s work or performance” (p. 120). There are several categorizations of the types of assessment. Assessments can be categorized based on purpose, mode, intention, interpretation, and administration (Mihai, 2010). There are also expressions that are either used interchangeably with assessment or categorized as greater or smaller sets in relation to assessment: testing, measurement, evaluation, feedback, and alternative forms of assessment. The following sections review and clarify these categorizations and definitions.

2.2.1 Categorization Based on the Purpose of Assessment

One of the earliest categorizations is the formative and summative distinction first used by Scriven (1966). Although he argues for the primary importance of a “final evaluation of a project or a person” (Scriven, 1966, p. 5), which he defines as summative assessment, he also acknowledges that it is necessary to evaluate the process of development as well, which he calls formative assessment. Later, diagnostic assessment has been added to these two types (Golnhofer, 2003), meaning that this type aims to analyze a situation in order to gather detailed information before making a pedagogical decision (Vidákovich, 1990). The three types of assessment have different goals. Diagnostic assessment aims to diagnose a situation in order to make informed decisions about, for example, course content; formative assessment aims to keep track of students’ progress and identify ways of helping it along; and summative assessment aims to identify overall levels of achievement and measure students’ results against them (Rea-Dickins, 2000).

Bloom (1969), shortly after Scriven, also applies the formative-summative distinction to tests: “by formative evaluation we mean evaluation by brief tests used by teachers and students as aids in the learning process. [...] We see much more effective use of formative evaluation if it is separated from the grading process and used primarily as an aid to teaching” (p. 48). Thus, Bloom (1969) also separates two roles of assessment: “formative evaluation” (p. 48) to improve performance, aid learning and teaching, and summative assessment to give grades as a final evaluation of a product or project. Olechowski (2003) also emphasizes that grades represent summative assessment, and for this reason must be separated from the aims of formative assessment. Arguing for the use of alternative performance assessment, Olechowski (2003) highlights that assessment is inevitably a tool for managing development processes, so teachers must use it to support the development of students and humanize schools.

The different purposes of assessment are also emphasized in the categories of assessment *for* learning, assessment *as* learning, and assessment *of* learning (Earl, 2006). The goal of assessment *for* learning is to support students’ learning (Black et al., 2004; Leahy et al., 2005; Wiliam, 2011; Wiliam et al., 2004). Assessment *as* learning provides opportunities for students to monitor and critically reflect on their own learning processes, which develops their metacognitive skills (Earl, 2006). “Assessment *as* learning focuses on the explicit fostering of students’ capacity over time to be their own best assessors, but teachers need to start by presenting and modelling external, structured opportunities for students to assess themselves” (Earl, 2006, p. 42). Assessment *of* learning compares students’ proficiency to curriculum learning outcomes (Earl, 2006). Comparing the above mentioned two categorizations, formative assessment and assessment *for* learning share the same goals, while summative assessment and assessment *of* learning also have the same aims. The purpose of assessment *as* learning is to develop students’ self-assessment skills thereby increasing their autonomy and accountability to aid their overall learning process. In Scriven’s (1966) categorization, the “increase of realistic self-appraisal” (p. 47) appears among several educational objectives that can be assessed both formatively (reflecting on the process) or summatively (reflecting on the outcome).

In connection with the Hungarian context, Golnhofer (2003) highlights that diagnostic and formative assessment are rarely used, and the prevalence of grading results in the confusion of their different purposes: for example, this happens when students get grades for a diagnostic test at the beginning of a topic/unit and then that grade is counted into their average determining their

final grade at the end of the year (summative assessment). Vidákovich (1990) also raises questions in connection with using the same tests for both summative and diagnostic purposes. According to Golnhofer (2003), effective educational processes require the use of all three types of assessment (diagnostic, formative, and summative) in accordance with their goals. Thus, the right set of conditions must be created: professional tools (such as diagnostic tests, item banks for summative tests, formative methods, etc.) have to be available and it would be necessary for teacher training to help understand and appropriately use them according to their different purposes (Golnhofer, 2003).

Discussing assessment in Hungary, Báthory (1997) and Golnhofer (2003) both refer to Tyler's (1970) principles. According to Tyler (1970), when developing any type of curriculum or planning any instruction, first and foremost educational purposes must be determined. Then learning experiences that are likely to be useful in attaining these objectives must be selected. It should be followed by the organization of effective instruction and ended by the evaluation of the effectiveness of the learning experiences. So pedagogical assessment should reflect the relation between the set educational purposes and students' learning experiences accordingly. For these reasons, when designing assessment processes, it is crucial to be aware of the different purposes to be achieved (Báthory, 1997; Golnhofer, 2003).

2.2.2 Other Categorizations of Assessment

Categorizations of assessment can also be based on mode (oral or written), intention (formal or informal), administration (classroom-based or large-scale assessment), and interpretation (norm-referenced or criterion-referenced) (Mihai, 2010). Mode means whether students' oral or written products are assessed. "In terms of intention, an assessment can be a spontaneous comment (informal) or it can be systematic (formal)" (Mihai, 2010, p. 26). Administration refers to the scale of assessment whether it takes place in the classroom organized by the teacher or the school, or if it is "administered in the form of state or national examinations" (Mihai, 2010, p. 32). Categorization based on interpretation can be norm- or criterion-referenced, which is further explained below.

Norm-referenced tests mean that the test results of students are interpreted in relation to a norm: mean, median, standard deviation, and/or percentile rank; while criterion-referenced tests give feedback on the acquisition of a specific material, lesson, or course objectives (Brown & Abeywickrama, 2010). As Gipps (1994) highlights

norm-referencing grades an individual's performance in relation to that of his/her [*sic*] peers [...] Norm-referenced tests are designed to produce familiar proportions of high, medium and low scorers. Since students cannot control the performance of other students they cannot control their own grades; this is now widely considered to be an unfair approach for looking at pupils' educational performance. (p. 5)

This was the reason why Glaser (1963) created this distinction: "What I shall call criterion-referenced measures depend upon an absolute standard of quality, while what I term norm-referenced measures depend upon a relative standard [...]. Measures which assess student achievement in terms of a criterion standard thus provide information as to the degree of competence attained by a particular student which is independent of reference to the performance of others" (pp. 519-520). Gipps (1994) also raises the problem that in schools, the two types of assessment often get mixed. For example, there might be clear criteria describing what students should know; however, tests and grading are based on the level of the school (e.g., it is different in a vocational or a highly academic secondary school), and students' performances are compared to each other (and not to the criteria).

In the Hungarian context, Báthory (1997) dedicates a full chapter to such controversies of school assessment. Báthory's (1997) chapter is structured around two main contradictions: the subjective-objective spectrum of the assessor and the relative-objective spectrum of the situation. According to Báthory (1997), teachers tend to automatically gravitate grades to the class average resulting in the fact that in "better" classes, grading becomes stricter, and in "weaker" classes, grading becomes more lenient (p. 239). Research shows that the distribution of grades tends to be the same in different classes (about the same number of students get grades 1s, 2s... 5s) while their knowledge differs greatly (Báthory, 1997, p. 240), so assessment becomes norm-referenced and not criterion-referenced. It also follows that the best grade, a grade five, can have very different meanings in a prestigious secondary grammar school in the capital and in a disadvantaged vocational school in the countryside. Báthory (1997) infers that grading does not make assessment more objective; on the contrary, grading results in highly subjective, unfair, and unreliable assessment. Finally, he concludes that the greatest problem of assessment in Hungarian schools is the tendency for grades to dominate all discussions instead of students' learning or knowledge (Báthory, 1997, p. 241).

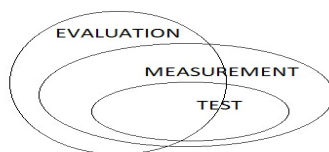
Gipps (1994) discusses the criticism of criterion-referenced assessment as well. “In order to meet the requirements for strict criterion-referenced assessment, criteria need to be specified in fine detail; however, this leads to over-specification and a focus on narrow, tightly defined objectives. [...] An example of an approach which aims to move away from norm-referenced assessment without going as far as full criterion-referenced assessment is standards-referenced assessment” (Gipps, 1994, p. 93), which term was introduced by Sadler (1987). According to Sadler (1987), criterion-referenced assessment is inappropriate for many subjects due to its tight specification of domains as in many cases the quality of a student’s work can be best assessed by direct qualitative human judgment. Sadler (1987) regards standards-referenced assessment feasible and credible through developed standards expressed in a combination of verbal descriptions (properties that characterize the designated level of quality) and exemplars (chosen to be typical of designated levels of performance). Sadler (1987) believes that this “combination of verbal description and concrete exemplar offers an efficient way of specifying standards, and if not kept centralized and secretive can help pupils to become self-monitoring learners” (Gipps, 1994, p. 94). Gipps (1994) introduces further categories of assessment types that others (such as Tsagari, 2004) label as alternative assessment, so they will be discussed in connection with alternative assessment.

2.2.3 Measurement, Test, Evaluation, and Feedback in Relation to Assessment

Before moving on to definitions of alternative assessment, some further expressions need clarification in relation to assessment. Bachman (1990) states that “the terms ‘measurement’, ‘test’, and ‘evaluation’ are often used synonymously” (p. 18); however, according to Bachman (1990), their definitions are the following: measurement is “the process of quantifying the characteristics of persons according to explicit procedures and rules” (p. 18); “test is a measurement instrument designed to elicit a specific sample of an individual’s behavior” (p. 20); and evaluation is “the systematic gathering of information for the purpose of making decisions” (p. 22). As Figure 1 shows: “not all measures are tests, not all tests are evaluative, and not all evaluation involves either measurement or tests” (p. 24); however, every test is a form of measurement.

Figure 1

The Relationship Among Evaluation, Measurement, and Tests (Based on Bachman, 1990, p. 23)



Bachman focuses on language testing and either does not use the term assessment (Bachman, 1990) or does not differentiate it from evaluation (Bachman, 1989). Cohen (2001) also focuses on language testing and does not differentiate “testing” from “assessment” using the two terms interchangeably in several publications (Cohen, 2022). Csapó (1992) does not differentiate among assessment, evaluation, and measurement, using them interchangeably and stating only that “tests and testing” can mean all kinds of systematic evaluation, but in his article he uses them in a narrower meaning as “formal measurement instruments to assess student academic achievement” (p. 5). Describing the Hungarian context, Báthory (1997) recommends using the English term “evaluation” for the Hungarian “értékelés” as it contains “value” (just as in Hungarian) implying that evaluation reflects what is considered valuable (p. 223). Báthory (1997) divides “pedagogical evaluation” into two parts: the collection and the interpretation of data. He defines “measurement” (in Hungarian “mérés”) as the process of data collection (p. 228) and details its possible methods and how validity and reliability can be assured.

Ur (1991) explicitly states using assessment and evaluation interchangeably (p. 244) and testing as a means for gathering information; “an activity whose main purpose is to convey how well the testee knows or can do something” (p. 33). According to Brown and Abeywickrama (2010), evaluation involves the interpretation of testing results to make decisions, for example, “if a student achieves a score of 75 percent (measurement) on a final classroom examination, he or she may be told that the score resulted in a failure (evaluation) to pass the course” (p. 5). Barret et al. (2014) define evaluation, assessment, and tests in the context of developing learners’ intercultural competence as follows. “Evaluation is the observation and measurement of the effectiveness of a lesson, course or programme of study” while “assessment is the measurement or systematic description of a learner’s degree of proficiency” (p. 34). According to Barret et al. (2014) both evaluation and assessment can happen through tests “used by institutions for both formative and summative evaluation purposes” moreover “there are many additional kinds of instrument to use in assessment – for example, portfolios and learner-diaries – and assessment can be carried out by teachers, or by learners themselves and their peers, through self-assessment or peer-assessment” (Barret et al., 2014, p. 34).

Baehr (2005) also makes a clear distinction between assessment and evaluation: “Assessment provides feedback on knowledge, skills, attitudes, and work products for the purpose of elevating future performances and learning outcomes. Evaluation determines the level of quality

of a performance or outcome and enables decision-making based on the level of quality demonstrated” (p. 441). So Baehr (2005) sees the difference in their purpose: that of assessment is “to improve the quality of future performances”, while that of evaluation is “to determine the quality of the present performance” (p. 442). As Baehr (2005) defines the formative-summative distinction differently, not focusing on the purpose but on the process (formative) versus the outcome (summative), she figures that both assessment and evaluation can be formative or summative whether the improvement (assessment) or judgment (evaluation) refers to a process (formative) or an outcome (summative). Other online sources, such as Surbhi (2017), use similar definitions to differentiate assessment from evaluation and agree that assessment focuses on the process in order to improve it, and evaluation focuses on a product making a judgment about its value, but they conclude that the purpose of assessment is always formative while evaluation is always summative.

Lastly, I have found that the definitions of feedback have a fundamental role in the different interpretations of assessment. In language teaching, Ur (1991) defines feedback as “information that is given to the learner about his or her performance of a learning task, usually with the objective of improving this performance” (p. 242). Ramaprasad (1986), in a more general context, defines feedback as the “information about the gap between the actual level and the reference level of a system parameter which is used to alter the gap in some way” (p. 4). Based on Ramaprasad’s definition, Sadler (1989) emphasizes that “information about the gap between actual and reference levels is considered as feedback only when it is used to alter the gap” (p. 121). He identifies three conditions that are all necessary to be fulfilled simultaneously: the learner has to “(a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action which leads to some closure of the gap” (p. 121). So in order to determine the effectiveness of assessment, improvement must follow it, which can be achieved through feedback. Sadler (1989) also emphasizes that if the information is simply recorded or it is too deeply coded, such as in the form of grades, it will not lead to appropriate action. The ultimate goal is to “facilitate the transition from feedback to self-monitoring” (Sadler, 1989, p. 122). Ur (1991) also expresses similar ideas, albeit with different terminology: “Feedback has two main distinguishable components: assessment and correction. In assessment, the learner is simply informed how well or badly he or she has performed. [...] In correction, some specific information is provided on aspects of the

learner's performance" (p. 242). However, Ur's (1991) definition and examples of correction still include only information on the performance and do not require further action from the students. Boud (2000) believes that the effectiveness of feedback lies in the actions for closing the gap, too. "This is one of the most often forgotten aspects of formative assessment. Unless students are able to use the feedback to produce improved work, through, for example, re-doing the same assignment, neither they nor those giving the feedback will know that it has been effective" (Boud, 2000, p. 158). In connection with the Hungarian context, Báthory (1997) also summarizes that pedagogical assessment has "three types of feedback cycles addressing three pedagogical domains - the target system, the teaching-learning process, and the results - which creates the basic structure of pedagogical assessment" (p. 228).

All in all, there are plenty of different variations of relations of measurement, test, evaluation, and feedback to assessment. I believe that it is necessary to choose an umbrella term that does not exclude any of the above mentioned functions. For this reason, I use "assessment" as the broadest category that includes all purposes (diagnostic, formative, summative, or others), modes, intentions, administrations, and interpretations. It follows that measurement, test, evaluation, and feedback will also be parts of assessment. As the context of the research conducted for the present dissertation is the Hungarian education system, I will translate "értékelés" as "assessment", "mérés" as "measurement", and I will use "evaluation" when "számonkérés" is used in Hungarian texts (Appendix A contains a glossary of translations I use in this dissertation).

2.2.4 Alternative Assessment

This section presents some definitions of alternative assessment in chronological order. Hamyan (1995) defines alternative assessment as "procedures and techniques which can be used within the context of instruction and can be easily incorporated into the daily activities of the school or classroom" (p. 213) and contrasts it to standardized testing, stating that alternative assessment does not provide a comparison of an individual to a larger group. She also describes different types of alternative assessment and discusses procedures for creating them. Kohonen (1997) emphasizes that assessment reflects what is valued in education, so using the term "authentic assessment", Kohonen's definition is the following: "forms of assessment that reflect student learning, achievement, motivation and attitudes on instructionally-relevant classroom activities [...] Its results can be used to improve instruction, based on the knowledge of learner progress" (p. 13). These definitions show that when the word "alternative" is used, it defines

assessment in relation to something else (e.g. standardized testing) and / or different characteristics of assessment are emphasized.

Looking at the history of alternative assessment, Alderson and Banerjee (2001) report that “the alternative assessment movement, if it may be termed such, probably began in writing assessment, where the limitations of a one-off impromptu single writing task are apparent” (p. 228). They contrast real life writing situations to the writing tasks of tests and exams concluding that assessment should reflect more, for example, a writing process that “involves much planning, editing, revising and redrafting, and usually involves the integration of input from a variety of (usually written) sources” (p. 228). They collect further definitions and examples of alternative assessment from all over the world and describe it as “procedures which are less formal than traditional testing, which are gathered over a period of time [...], which are usually formative rather than summative in function, are often low-stakes in terms of consequences, and are claimed to have beneficial washback effects” (p. 228). They argue that portfolio assessment, which is already used in subject areas such as art, could also be used in language teaching to overcome the drawbacks of traditional assessment. Synthesizing existing literature, they collect alternative assessment measures such as “checklists, portfolios, interviews, performance-tasks, [...] teacher-pupil conferences, learner diaries and journals, informal reading inventories, classroom reading aloud sessions, portfolios of books read, self-assessments of progress [...], process assessment, [...] continuous assessment, [...] peer evaluation using questionnaires” (p. 229). They also collect criticism of alternative assessment practices questioning their validity, reliability, and practicality. Critics suggest different wordings, such as “alternatives in assessment” (Brown & Hudson, 1998). Brown and Hudson (1998) also urge for taking into consideration “(a) the consequences of the washback effect of assessment procedures on language teaching and learning, (b) the significance of feedback based on the assessment results, and (c) the importance of using multiple sources of information in making decisions based on assessment information” (p. 653). Alderson and Banerjee (2001) conclude that research supports the effectiveness of alternative assessment procedures in informal, classroom-based situations; however, “standardisation of such procedures will be needed if they are to be used for high-stakes assessment, and that the financial and logistic viability of such procedures remains to be demonstrated” (p. 230).

Gipps and Stobart (2003) collect several definitions of alternative assessment, summarize the assumptions underlying them, describe the use of types of alternative assessment, and

“conclude by discussing the role of alternative assessment in systems in which accountability is one of the key functions of assessment, and propose ways forward” (p. 550). In their collection, alternative assessment offers “alternatives to standardized multiple-choice testing”, “Information Technology (IT)-based, ways of testing”, and “an alternative approach which makes assessment an integral part of the teaching and learning process” focusing “more directly on ‘performance’” (p. 549). They agree with the latter emphasizing that assessment should be an integral part of the teaching and learning process, and to achieve it, a paradigm shift is necessary, which moves assessment from “psychometric and measurement models to a model in which assessment is an integral part of the learning process [...] and so is not restricted to summative purposes” (p. 550). They detail “educational assessment” (defined by Gipps, 1994) in order to argue that “alternative assessment is not simply the use of alternative forms of assessment but is also an alternative use of assessment as part of the learning process”, and mention the following characteristics of “educational assessment”: it has an underlying constructivist view of learning, it uses tools and assistance in the assessment process (such as performance and portfolio assessments), and processes such as self-evaluation and feedback to enhance learning (Gipps & Stobart, 2003, p. 550).

According to Tsagari (2004), alternative assessment appeared as an answer to the following concerns about language testing. High-stakes, standardized tests affect the school curriculum by making teachers concentrate only on those subjects and skills that are included in the examinations, restricting teachers’ methods to employ exam preparation practices, and encouraging students “to adopt ‘surface’ approaches to learning as opposed to ‘deep’ approaches” (p. 3). On a psychological level, students take the role of “passive recipients of knowledge and their needs and intentions are generally ignored [which has] detrimental consequences on students’ intrinsic motivation, self-confidence, effort, interest and involvement [moreover, these processes] induce negative feelings [...] such as anxiety, boredom, worry and fear” (Tsagari, 2004, p. 4). Tsagari (2004) also quotes the above mentioned definition of alternative assessment by Alderson and Banerjee (2001), lists further terminology, and the following benefits of alternative assessment:

Researchers and practitioners in the field believe that alternative assessment can:

- a. Evaluate the process and product of learning as well as other important learning behaviours
- b. Evaluate and monitor instruction

- c. Produce meaningful results to a variety of stakeholders
- d. Relate to cognitive psychology and related fields
- e. Represent a collaborative approach to assessment
- f. Support students' psychologically
- g. Promote autonomous and self-directed learning
- h. Provide new roles for teachers. (Tsagari, 2004, pp. 7-10)

Mihai (2010) defines alternative assessment as “any type of assessment that requires students to perform, create, or do something by using tasks that correspond to meaningful instructional activities” (p. 125). According to Mihai (2010) there are three main types of alternative assessment: portfolio assessment, performance assessments, and curriculum-based measurement, and all three of them contain “checklists, rubrics, student self-assessment, peer assessment, interviews, and journals” (p. 126).

Kohn (2011) argues that even the presence of grades decreases students' interest in whatever they are learning, increases preferences for easier tasks, and negatively affects the quality of students' thinking, so he suggests abolishing grades and replacing them with “narrative assessments or student-teacher conferences - qualitative summaries of student progress offered in writing or as part of a conversation” (p. 30). Moreover, he also proposes inviting students and their parents to participate in constructing alternative forms of assessment.

As the examples above show, alternative assessment has multiple approaches and definitions emphasizing different segments of the learning and teaching process. The present research focuses on the Hungarian public secondary school context; thus, the definition of alternative assessment used in this study should be in relation to official rules and regulations of student assessment in Hungarian public secondary schools. Taking into consideration Sadler's (1989) definition of assessment in general: “any appraisal (or judgment, or evaluation) of a student's work or performance” (p. 120) and the Hungarian public educational context where teachers should evaluate students' knowledge, behavior, and diligence in the form of grades from one to five and also develop students with the help of assessment, my own definition of alternative assessment is as follows. The term “alternative assessment” is used in this study to refer to any appraisal, judgment, or evaluation of a student's work or performance that contains different elements in addition to or instead of grades with the purpose of supporting students' development.

The development of students is understood here in accordance with the expectations of the Hungarian educational act, which states that the

public education system [should] contribute to the harmonic mental, physical and intellectual development of children and young people through the conscious development of their skills, abilities, knowledge, proficiencies, emotional and volitional characteristics and cultural education corresponding to their age characteristics, thus educating people and responsible citizens who are virtuous and capable of independent life as well as achieving their objectives, while harmonising private interests with the interests of the public. (Act CXC, 2011, Section 1)

2.3 Empirical Background

The first component of the definition of alternative assessment used in the present dissertation is that the assessment of a student's work or performance should contain different elements in addition to or instead of grades. The second component is that it is carried out with the purpose of supporting students' development.

According to the first component, empirical research focusing on the following types of assessment are all relevant (for containing elements other than grades): formative assessment, assessment *for* learning, and within the "movement of alternative assessment" (Alderson & Banerjee, 2001, p. 228) authentic assessment, performance assessment, continuous assessment, ongoing assessment, informal assessment, descriptive assessment, direct assessment, dynamic assessment, instructional assessment, responsive evaluation, complementary assessment, portfolio assessment, situated or contextualised assessment, and assessment by exhibition as collected by Tsagari (2004). The list can also be continued with classroom assessment (Angelo & Cross, 1993), educational assessment (Gipps, 1994), process assessment (Gimenez, 1996), sustainable assessment (Boud, 2000), curriculum-based measurement (Mihai, 2010), rapid assessment (Yeh, 2010), and assessment based on gamification (Barbarics, 2015, 2016, 2017; Hubai & Lázár, 2018). Moreover, there is also extensive empirical research on the effectiveness of certain pedagogical methods used in order to carry out alternative assessment. Some examples listed by Tsagari (2004) are the following: conferences, debates, demonstrations, diaries or journals, dramatizations, exhibitions, games, observations, peer assessment, portfolios, projects, self-assessment, story retelling, and think-alouds. Instead of recording grades, Hamayan (1995) suggests recording alternative assessment information with the help of the following tools: anecdotal records,

checklists, learner profiles, progress cards, questionnaires, and rating scales (others, such as Mihai (2010), mention rubrics). As elaborating on all of them is beyond the scope of the present paper, only empirical research on the connections between assessment and grades will be discussed in further detail. Then empirical research focusing on the second component of the definition of what kinds of assessment support students' development is reviewed. The last section of this chapter looks at research concerning the context of the dissertation: assessment in English as a foreign language and English medium content classes.

2.3.1 Assessment and Grades

As seen in previous sections, many have discussed the disadvantages of grades and argued for a modified use of them, alternatives, or their complete abolition. This chapter looks at some empirical research exploring relations between the different aims of assessment and the application of grades.

Butler and Nisan (1986) compared the impacts of no feedback, task-related comments, and grades on student achievement in both problem-solving and “quantitative” tasks (e.g., those requiring quick, timed work to produce a large number of answers). 261 6th-grade students were randomly assigned to one of the three feedback conditions and were given the different tasks in three sessions. Then students' performance scores and the results of a questionnaire were analyzed. Butler and Nisan (1986) found that students receiving task-related comments (but not grades) on an initial assignment performed significantly better both on follow-up quantitative tasks and problem-solving tasks than students receiving grades or no feedback. Students receiving grades performed better on follow-up quantitative tasks than students receiving no feedback, but did not outperform those students on problem-solving assignments. In other words, giving grades does not appear to enhance students' future performance in problem solving. In addition to test results, questionnaire answers confirmed “the hypothesis that intrinsic motivation would be maintained after receipt of nonthreatening, task-related evaluation and undermined after repeated nonreceipt of feedback or receipt of controlling normative grades” (Butler & Nisan, 1986, p. 210).

Pulfrey et al. (2011) carried out three experiments to find relations between grades and students' goals. They differentiated between performance-avoidance goals focusing on avoiding incompetence (which are connected to extrinsic motivation) and performance-approach goals focusing on attaining competence (which are connected to intrinsic motivation). Participants were Swiss professional school students of EFL: 115 students ages 15-30 in the first experiment, 130

students ages 15-23 in the second experiment, and 97 students ages 16-24 in the third experiment. After the first experiment, they found that the “expectation of a grade for a task, compared with no grade, consistently induced greater adoption of performance-avoidance, but not performance-approach, goals” (Pulfrey et al., 2011, p. 638). The second and the third experiment added that this remained the case (grades inducing performance-avoidance goals) even if grading was accompanied by formative comments; furthermore, grades reduced autonomous motivation, while no grades decreased performance-avoidance goals in subsequent tasks. The connection between the two studies is that grades seem to reduce intrinsic motivation fueled by the need for autonomy (as described by self-determination theory of Deci and Ryan, 1985), and if a task is graded, the presence of formative comments does not seem enough to alleviate the negative effects of grading.

Schinske and Tanner (2014) carried out the meta-analysis of several further empirical research on different aspects of grading. They found that “high-achieving students on initial graded assignments appear somewhat sheltered from some of the negative impacts of grades, as they tend to maintain their interest in completing future assignments” (Schinske & Tanner, 2014, p. 161). Looking for positive impacts of grades, such as being incentives for students, studies found either very minimal or no evidence that grades motivated students to perform better. On the other hand, norm-referenced grading (producing normal distribution of grades in class) forces students to compete with each other and “can dissociate grades from any meaning in terms of content knowledge and learning” (Schinske & Tanner, 2014, p. 162). In connection with competition, they quote studies finding that cooperative learning environments make students “more interested in learning and find learning more worthwhile” (Schinske & Tanner, 2014, p. 162) than competitive learning environments. They conclude that while grading practices “have been driven by educational goals such as providing feedback to students, motivating students, comparing students, and measuring learning” (Schinske & Tanner, 2014, p. 163), empirical research suggests that these aims are often not being met using current grading practices. So they suggest some strategies for change: “balancing accuracy-based grading with effort-based grading, [...] providing opportunities for meaningful feedback through self and peer evaluation, [...] making the move away from curving [meaning norm-referenced grading...], and becoming skeptical about what grades mean” (Schinske & Tanner, 2014, pp. 163-164). Empirical research on what supports students’ development will be discussed in the next section.

Describing the Hungarian context, Golnhofer (2003) states that in performance-oriented education systems, it does not seem realistic to implement assessment without grading. Báthory (1997), after discussing the disadvantages of grading, lists some research projects focusing on assessment without grading in Hungary (pp. 239-242). In one study, ten secondary schools did not use any grades in any of the subjects for the four years of secondary school in the participating classes. Teachers developed detailed assessment criteria for their subjects and applied continuous textual feedback. Findings of the research prove that grading is not necessary for teaching and learning processes, and there is evidence that alternative assessment supports students' learning. However, as assessment without grading requires different attitudes and methods, it takes much more time and effort for teachers than their earlier traditional grading practices. After the end of the research projects as extra support was not provided anymore, assessment without grading remained just a pedagogical experiment. So Báthory (1997) concludes that working conditions of teachers (without extra support) do not allow quality assessment (p. 242).

Radnóti (2006) looking at teaching and assessment methods teachers use in Hungarian schools found that teachers aim to achieve several goals through assessment. Teachers find the following goals of assessment important: reinforcement, control, orientation, motivation, providing information on students' performance, feedback on how effectively goals have been reached, how to improve, and so on. However, in practice, the highly selective and competitive grading is dominant, and alternative assessment methods fulfilling the above mentioned goals are rare. In secondary education, it is mostly output requirements, such as the secondary school leaving matura examinations, that often determine what is assessed in classes. Hubai (2016) also found that the assessment practices of teachers in Hungary are mostly influenced by the compulsory one to five grading, time management issues, and the educational culture of the given school. In conclusion, resulting from the requirements of the current system in Hungary, "the majority of teachers seem to follow the traditional 'teach-to-the-test and then give a written or oral exam' approach when it comes to feedback and assessment in practice" (Hubai & Lázár, 2018).

2.3.2 Assessment that Supports Students' Development

The second component of my definition of alternative assessment is that it is carried out with the purpose of supporting students' development. This section collects empirical research aiming to identify how assessment can fulfill this purpose. Wiliam (2011) in the meta-analysis of almost 800 studies arrives to the conclusion that "for assessment to support learning, it must

provide guidance about the next steps in instruction and must be provided in a way that encourages the learner to direct energy towards growth, rather than well-being” (p. 13).

Nicol and Macfarlane-Dick (2006) believe that the goal of assessment is to “help students take control of their own learning” (p. 199), and in order to achieve that goal through good feedback practice, they synthesized research on feedback in higher education. They established the following seven principles supporting each of them with empirical research from the literature.

Good feedback practice:

1. helps clarify what good performance is (goals, criteria, expected standards);
2. facilitates the development of self-assessment (reflection) in learning;
3. delivers high quality information to students about their learning;
4. encourages teacher and peer dialogue around learning;
5. encourages positive motivational beliefs and self-esteem;
6. provides opportunities to close the gap between current and desired performance;
7. provides information to teachers that can be used to help shape teaching. (p. 205)

Research conducted by Wiliam et al. (2004) is relevant from two aspects. The first aspect is that they address the following criticism of alternative assessment: that although “increased use of formative assessment (or assessment for learning) leads to higher quality learning, it is often claimed that the pressure in schools to improve the results achieved by students in externally-set tests and examinations precludes its use” (p. 49). For this reason, their study aims to investigate the test results of students who are taught by teachers developing alternative assessment practices. Thus, the second aspect why this empirical research is relevant is that it describes the process of developing alternative assessment methods from the teachers’ point of view.

In the research project carried out by Wiliam et al. (2004), a total of 24 teachers took part in a half-a-year training period in formative assessment developing their own plans, which were then put into action with their classes. The training of the teachers had two main components. There were sessions during which teachers were introduced to the principles of formative assessment and developed their own plans for applying it, as well as class observations of teachers already using formative assessment practices, after which they had the chance to discuss their ideas. As Leahy et al. (2005) refer to Wiliam et al. (2004) as an example of implementing their non-negotiable principles, it can be assumed that teachers based their action plans on the following principles: “Clarifying and sharing learning intentions and criteria for success. Engineering

effective classroom discussions, questions, and learning tasks. Providing feedback that moves learners forward. Activating students as the owners of their own learning. Activating students as instructional resources for one another” (p.20). Leahy et al. (2005) details this type of feedback – that moves learners forward – by saying that “to be effective, feedback needs to cause thinking. Grades don't do that. Scores don't do that. And comments like ‘Good job’ don't do that either. What does cause thinking is a comment that addresses what the student needs to do to improve, linked to rubrics where appropriate” (p. 20).

In order to measure students’ results, comparison groups were established for each class: “either an equivalent class taught in the previous year by the same teacher, or a parallel class taught by another teacher. The mean effect size in favour of the intervention was 0.32” (Wiliam et al., 2004, p. 49). The researchers describe the limitations of the study, for example, that the “comparisons are not equally robust” (p. 62) or that positive effects in comparing classes taught by different teachers might be caused by the personality of the teachers. Despite these limitations, they believe that they have proved that teachers do not “have to choose between teaching well and getting good results” (Wiliam et al., 2004, p. 64).

Teacher autonomy played an important role in the research of Wiliam et al. (2004). Teachers had to develop their own action plans on how to implement the principles of formative assessment. Wiliam et al. (2004) conclude that “for the vast majority of our teachers, involvement in the project has not just spread to all their classes, but has fundamentally altered their views of themselves as professionals” (Wiliam et al., 2004, p. 62). This change in connection with teachers and the positive effect sizes in connection with students’ results imply that supporting teachers in developing their own alternative assessment practices seems to be effective in this context.

Gamlem and Smith (2013) carried out a study to gain more insight into students’ perceptions of when and how students find classroom feedback useful. They conducted class observations and individual interviews with 11 students, aged 13–15, from four secondary schools in Norway. Their main research question was: “What are adolescent perceptions of useful classroom feedback?” (Gamlem & Smith, 2013, p. 150). They identified feedback types and created a feedback typology to provide a framework which can be used to reflect on useful classroom feedback based on students’ perceptions. In connection with the context of the study, it is worth mentioning that assessment *for* learning (AfL) “has been defined as a national goal [in Norway], and a strategy for the years 2010–2014 has been developed [...] to ensure that teachers

gain knowledge about AfL and develop necessary skills. [...] AfL [has been defined] as any assessment that is given during educational situations which helps to promote learning” (p. 156). As the study was carried out during the school year of 2009–2010, it can be seen as preliminary research supporting the need for introducing assessment *for* learning and including more effective feedback types. Although in the background section of the paper they refer to researchers pointing out that “feedback leads to learning gains only when it includes guidance about how to improve, when students have opportunities to apply the feedback, understand how to use it and are willing to dedicate effort” (p. 152), Gamlem and Smith do not analyze their data from this perspective. As my focus is what supports students’ development, I present their feedback typology from this point of view.

Gamlem and Smith (2013) identified four types of feedback denoted by letters A, B, C, and D. Type A feedback means rewarding, grade giving, and punishing. In connection with grade giving, they refer to Black et al. (2004) stating that if there are no comments, students find grades useless. This statement is supported by the opinion of a student who received the top grade in the Norwegian system but was still unsatisfied: “I somehow got a six and got very little feedback even though I knew there was a lot I needed to improve – but I don’t understand how” (Gamlem & Smith, 2013, p. 162). This opinion supports that even besides getting the best grade, students would like to receive guidance on how to improve.

Type B feedback means approving, controlling, and disapproving. According to the study, this is the most frequently experienced feedback type. Gamlem and Smith (2013) conclude that “students find controlling feedback important for motivation and essential for establishing good relations between teacher and student, but not very helpful for improving work” (p. 163). The need for the actions that could help improvement is expressed in this type of feedback as well.

Type C feedback means specifying attainment, reporting, and specifying improvement. According to the data of Gamlem and Smith (2013), the last part of this feedback type, specifying improvement, is found to be the most useful by students. They appreciate it when they are given time to work with the feedback and revise their assignments. This type of feedback also appears in peer assessment, when peers are giving feedback specifying improvement while working together and “this feedback is used to further develop and improve work since it is part of the classroom activity” (Gamlem & Smith, 2013, p. 164). It seems that students appreciate feedback the most when it helps them to improve their work. Reporting in this typology refers to self-

assessment, which students also find useful. However, it “loses its usefulness when opportunities, mostly in terms of time, are not provided to work with the feedback” (Gamlem & Smith, 2013, p. 164). So students are missing the opportunities for actions that would lead to improvement in the case of self-assessment, too.

Type D feedback means constructing achievement, dialogic feedback interaction, and constructing the way forward. “It generates learning, provides information about achievement, gives targeted individual information to proceed and develop understanding, and is used as an interactive dialogue between the teacher and the student(s) or among the students” (Gamlem & Smith, 2013, p. 164) and hardly ever occurs in class. Researchers could only identify a few cases, and students also reported that it is very rare; however, it enhances their learning process greatly. “This was very good! I believe this was the lesson I learned most from during the whole week” (Gamlem & Smith, 2013, p. 165) one student reported after the teacher gave feedback on their homework and discussed the questions they did not understand in connection with it. This type of feedback was identified in peer assessment as well: “I feel I can in a way give help to that person. Like trying to sort of understand how he thinks and then maybe try to sort of understand what he believes. I can in a way say what I mean, so we understand each other so that one feels safe” (Gamlem & Smith, 2013, p. 165).

Gamlem and Smith (2013) conclude that in their study “students were not given time and opportunities to work with feedback from teachers, which causes feedback [...] to be perceived as negative and abandoned. Feedback becomes just a message of little use in the learning process since there are no formal opportunities to apply the feedback. Feedback which is not understood as fruitful seems to be abandoned, or replaced by affective actions such as anger or frustration” (p. 160). All in all, the study supports Sadler’s (1989) claim to consider feedback effective only if the “information about the gap between actual and reference levels [...] is used to alter the gap” (p. 121).

Yeh (2010) looks at the connection between students’ development and school assessment from a psychological point of view. First, Yeh (2010)

synthesizes findings regarding the development of competence and learned helplessness and factors influencing persistence and intrinsic motivation and [...] the process through which small differences in early achievement are magnified by the current structure of

schools, [... then suggests] a specific type of individualized instruction and assessment system [that] may be especially suited to remediate these differences. (p. 170)

Analyzing research regarding the development of self-efficacy, Yeh (2010) concludes that students, as all humans, have the basic need to feel competent, which is connected to beliefs about perceived control, which influences engagement, which, in the end, influences academic performance (p. 171). So if students experience competency through having control over their performance, they are more engaged, and the increased effort and further improvements in achievement result in a virtuous cycle (p. 175). According to several research projects cited by Yeh (2010), students experience lack of contingency between their efforts and academic outcomes “if the dominant form of student evaluation involves comparisons with other students” (p. 172), so on the one hand, students should be assessed in relation to their previous progress, and on the other hand, tasks should be individualized, so that their difficulty matches students’ ability. Yeh (2010) also emphasizes that this individualization of task difficulty does not only affect instruction or the rate of task completion but the actual difficulty of tasks (supported by numerous empirical research studies showing that providing more time in itself is not enough). Connecting assessment and task difficulty, Yeh (2010) suggests informational performance feedback that can be automated utilizing technology, while teachers can supplement it with strategies for improvement. Two specific computer programs are presented using “rapid assessment” that are systems providing “autonomy in task execution, an accelerating standard of performance, and formative testing feedback [information about correct and incorrect answers...], while individualizing task difficulty and performance expectations so that students achieve success on a daily basis” (Yeh, 2010, p. 175). It is also noted that

computer video games typically provide individualization of task difficulty, performance feedback, autonomy in task execution, and an accelerating standard of performance, where students advance to higher levels of difficulty after successfully completing less difficult tasks. This may explain why many students who are disengaged in school are highly engaged when playing video games (Yeh, 2010, p. 174).

Assessment based on gamification that will be detailed in section 2.4.3 also aims to utilize these characteristics of computer video games described by Yeh (2010).

According to the empirical research reviewed in this section, assessment that supports students’ development should “provide guidance about the next steps in instruction” (Wiliam,

2011, p. 13), which can happen through good feedback practice (Nicol & Macfarlane-Dick, 2006) building on teachers' autonomy (Wiliam et al., 2004) and manifested in dialogic interactions between students and teachers (Gamlem & Smith, 2013). Further factors, such as nurturing intrinsic motivation (that will be discussed in section 2.4.1) through individualized performance feedback (Yeh, 2010) should also be taken into consideration in accordance with the context in which assessment takes place. For this reason, the next section reviews research specific to the context of the dissertation: assessment in English as a foreign language and English medium content classes.

2.3.3 Assessment in English as a Foreign Language and English Medium Content Classes

The context of the present research is English as a foreign language (EFL) and English medium content (EMC) classes, so this section looks at the characteristics of assessment in EFL and then in EMC classes. As it is mentioned in connection with definitions of alternative assessment, in many cases it stemmed from foreign language classes due to discrepancies between language testing and real life language use.

Vogt and Tsagari (2014) found, based on large-scale international questionnaire and interview studies, that foreign language teachers' language testing and assessment literacy is not very well developed. Participants, from seven countries across Europe, expressed the need for training especially "alternative assessment formats, such as portfolio assessment and self- or peer assessment, have been named as priorities for in-service teacher training" (p. 391). Their interview study also showed that foreign language teachers mainly model assessment methods of mentors and colleagues, which according to Vogt and Tsagari (2014), might result in a "testing as you were tested [...] perpetuation [that] leaves little space for innovative assessment methods. Indeed, the interview findings confirm that teachers have somewhat fuzzy concepts of more recent assessment procedures and their practical implementation" (p. 391).

In the public secondary school context in Hungary, output requirements of the secondary school leaving matura examinations often determine what is assessed in classes. According to Government Decree 100/1997 (VI. 13.), foreign language is a compulsory part of the matura examinations in Hungary, which consists of the assessment (with a grade from one to five) of communicative competences: reading and listening comprehension, speaking and writing skills, and other strategies necessary to plan, execute, and evaluate tasks. Although there are more detailed descriptions of these competences, it depends on teachers how communicative

competences are developed and assessed in class. Byram (1997), in connection with the assessment of intercultural communicative competence (ICC), argues that complex competences should be assessed in complex ways because the simplification of competences to what can be “objectively” tested has a detrimental effect:

the learning of trivial facts, the reduction of subtle understanding to generalisations and stereotypes, the lack of attention to interaction and engagement because these are not tested. When assessment recognises all aspects of ICC, even if they cannot be quantified and reduced to a single score, then the learner can see their efforts rewarded, and the teacher and curriculum planner can give full attention to the whole phenomenon of ICC rather than only that which can be represented statistically. (p. 111)

As a possible solution, Byram (1997) suggests portfolio assessment. Reviewing policy documents and empirical research focusing on language teaching, Hubai and Lázár (2018) found that EFL teachers in Hungary also tend to use traditional assessment practices such as exam preparation through written and oral tests; however, there is “empirical research about language teachers experimenting with alternative approaches to assessment in order to provide support to the fully engaged, empowered, motivated and autonomous learners that teachers, education experts and policy makers expect to see in our educational establishments” (p. 92). Some of these experiments include using percentages instead of grades, applying game elements, assessing group work, involving self- and peer assessment practices, and having optional homework assignments (Hubai, 2016).

Assessment in English medium content (EMC) classes is even more complex as it encompasses the assessment of both content and language related performances. According to Massler (2011), if teachers come from a foreign language teaching background and are not trained in the respective content subjects’ assessment, teaching and assessing language competences will be dominant and students’ content learning might be neglected. This is more often the case in primary education, while in secondary education, teachers are usually content teachers with sufficient language skills to teach their subjects in a foreign language. In the Hungarian context, the Government Decree 100/1997 (VI. 13.) specifies that in the secondary school leaving matura examinations only content requirements should be assessed, even if the subject is taught in a foreign language, so the assessment of language competences is neglected in class. Research carried out by Kovács (2015) suggests that public schools with EMC classes should clarify that

assessment in such classes differs from assessment in either foreign language or content classes; in addition, teachers should acquire the methodology of content and language integrated learning (CLIL). Kovács (2015) emphasizes the advantages of EMC classes, where assessment could inherently be part of an authentic learning process: “learning by construction, not by instruction” (p. 20). As Kovács (2015) carried out research in primary schools, the study ends with a call for similar research in secondary education in Hungary.

In connection with the Hungarian public secondary school context, it is important to note that content and language integrated learning (CLIL) can appear in two ways. On the one hand, it can be part of EFL classes either by covering special content during EFL lessons or labeling the subjects separately as, for example, business English or IT English. On the other hand, in bilingual secondary schools some content subjects are taught in English. In the first case, teachers are language teachers, and the requirements in connection with assessment are the same as in any EFL classes, so the focus is on language learning. In the second case, teachers are content subject teachers, and as mentioned above, the output requirements are the same as for the subject in Hungarian, so the focus is exclusively on the content. In order to avoid confusing these two vastly different situations, I will not use CLIL (content and language integrated learning), but EFL (English as a foreign language) even for the first case, and EMC (English medium content) for the second type of classes. In both cases, it is up to the individual teacher how to involve the other aspect (content or language) in assessment.

Coyle et al. (2010) discuss issues for assessment in CLIL in detail, most of which are relevant for the Hungarian context as well. They answer the following questions:

What do we mean by assessment in CLIL?

Do we assess language or content first?

Do we sometimes assess one and not the other? If so, which and when (and, more crucially, why and how)?

What about cognition and culture?

Who assesses?

When do we assess?

How do we assess?

What is the role of standard examination systems?

Is there a role for the Common European Framework? (Coyle et al., 2010, pp. 114-115)

Short (1993) acknowledges the need for alternative assessment (“the many varieties of alternative assessment include performance-based tests, portfolios, journals, projects, and observation checklists [...] allow[ing] better demonstration of student knowledge” p. 633); however, Short (1993) raises similar problems in connection with CLIL: teachers have to decide what they assess (content-wise and language-wise) “and decide if one is interfering with the demonstration of the other” (p. 633). Genesee and Upshur (1996) make it clear that in the case of EMC classes, it is the content that should be assessed based on the same criteria for all students (native or non-native speakers alike). Pappamihiel and Mihai (2006) argue that reasons behind rejecting accommodations might be because they would alter the testing environment and would not be fair to students with better language abilities; however, in a linguistically diverse class, “any test given in English automatically constitutes an unequal testing environment” (p. 35). For this reason, they provide specific tools with which teachers can assess their students’ English proficiency level for diagnostic purposes, based on which they can make curricular adaptations and accommodations for assessment purposes. Although they represent the US context, the following pieces of advice can be adapted to EMC classes in other countries as well. According to them, teachers should always go through the following questions: (ELL means English language learner)

1. Do I know my students’ English language proficiencies?
2. Have I designed a test that mirrors classroom objectives, strategies, and activities?
3. Have I made use of all relevant and available visuals and graphics?
4. Have I incorporated true accommodations to level the playing field for my ELLs?
5. Have I created a clear scoring rubric that will allow me to provide culturally sensitive and useful feedback? (Pappamihiel & Mihai, 2006, p. 35)

As the questions show, clear goals, the use of visuals, graphics, accommodations, and rubrics are crucial parts of the approach. In terms of methodology, they suggest specific task types that suit students’ language levels based on Ernst-Slavin et al. (2002), who discuss all the stages of language development and cultural adaptation recommending effective activities and even questioning techniques adjusted to the stages. Using these accommodations can help students respond to content questions within their linguistic boundaries. The last question of Pappamihiel and Mihai (2006) necessitates the use of rubrics. They quote Shepard (2000) who claims that students should always know the criteria based on which their work is assessed, and in EMC classes it has a special importance for all the issues mentioned above. The transparency of rubrics can not only reduce

test anxiety, but it can also act as a study guide for students. In addition, as Brown and Abeywickrama (2010) also state, rubrics (in accordance with preset purposes) can help maintain consistency through the whole teaching-learning-assessment process. Coyle et al. (2010) agree that objectives and success criteria must be clear (first for the teacher) and shared with the students. They also note that these principles and their examples should be transferable to assessment of other classes, levels of work, or ages, and not only in CLIL (or EMC) classes.

As the above reviewed pieces of research show, the context of EFL and EMC classes result in special circumstances for assessment. Teachers should take into consideration output requirements such as the secondary school leaving examinations or language exams that already contain the assessment of complex competences. At the same time, assessment should support “engaged, empowered, motivated and autonomous learners” (Hubai & Lázár, 2018, p. 92) of the language and the content material that students should be able to use outside of school as well. The complexity of EFL and EMC classes might provide the context in which alternative assessment methods can be developed and tested to reach a complex set of goals, some of which are further detailed in the upcoming chapters.

2.4 Further Factors Influencing Alternative Assessment

Nurturing students’ intrinsic motivation and developing complex skills and competences emerged as goals that traditional assessment might not be able to fulfill. The following sections review some of the available literature on alternative assessment in connection with these aims, including assessment based on gamification. Applying gamified assessment appeared as a solution bridging the gap between traditional and alternative assessment by enabling teachers to reach complex goals while still complying with the requirements of the public education system.

2.4.1 Nurturing Students’ Intrinsic Motivation

Reviewing international trends, the OECD (2013) report arrives at the conclusion that “summative classroom assessment activities are a substantial part of education across OECD countries [and...] summative assessment dominates what students are oriented towards in their learning” (p. 145) such as reaching certain grades, transcripts, diplomas, and so on. The report (OECD, 2013) raises concerns that if learning is “targeted to those domains that are rewarded, [...] effort may decrease or disappear when the reward is no longer provided”, in addition, such assessment might “encourage surface learning approaches, generate ego-related priorities, reduce enjoyment of learning and decrease student focus on long-term goals”; moreover, “studies

repeatedly indicated that students with strong initial motivation might be negatively affected by attempts to stimulate their learning by external rewards” (p. 146). As a result, on the one hand, assessment that facilitates students’ intrinsic motivation is encouraged. On the other hand, a combination and balance of intrinsic and extrinsic rewards is suggested, based on the research by Hidi and Harackiewicz (2000), which will be reviewed later in this section.

Intrinsic motivation is “typically defined as the motivation to engage in activities for their own sake” (Hidi & Harackiewicz, 2000, p. 157). In order to nurture students’ intrinsic motivation through assessment, it is important to review what might facilitate intrinsic motivation. According to Deci and Ryan’s (1985) self-determination theory, intrinsic motivation stems from the fulfillment of the basic human needs for autonomy, competence, and relatedness. As a result, assessment that fosters learner autonomy, reflects students’ competences, and provides opportunities for students to experience relatedness might facilitate intrinsic motivation. According to Pink (2009), it is autonomy, mastery, and purpose that drive intrinsic motivation. Beside the common element of autonomy, the desire to be self-directed, Pink (2009) believes that purpose, the desire to do something that has meaning and is important, and mastery, the urge to become more skilled, motivate people to engage in activities for their own sake. It follows that beside supporting learner autonomy, assessment that students find meaningful and important, can relate to its purpose, and believe that it helps them improve might also nurture their intrinsic motivation. Learning for mastery (Bloom, 1968) has long been part of educational theory. In Hungarian education, Csapó (1978) believed that mastery learning could be the future of education by providing step by step individual feedback, so that “students are convinced of the effectiveness of their work every day, which encourages them to make new efforts” (p. 71). Thus, assessment that provides immediate and continuous feedback might also motivate students intrinsically.

Hidi and Harackiewicz (2000) question dichotomous approaches such as dividing motivation into extrinsic and intrinsic elements. Instead, they emphasize the role of interest and goals for achievement. Although they “acknowledge the positive effects of individual interest, intrinsic motivation, and the adoption of mastery goals”, as described above, they also believe that “externally triggered situational interest, extrinsic motivation, and performance goals” should also be considered by teachers (Hidi & Harackiewicz, 2000, p. 151). They recommend combining intrinsic rewards through activities that are inherently interesting to students with external rewards that provide performance feedback. According to Hidi and Harackiewicz (2000), the balance of

such assessment results in optimal and sustained learning efforts over time. The OECD (2013) report concludes that assessment frameworks need to be developed “where a range of formative and summative assessment approaches complement each other to provide the adequate level of challenge and support to each student” (p. 146).

2.4.2 Developing Complex Skills and Competences

The OECD (2013) report also raises the issue that “for assessment to be meaningful, it must be well-aligned to the type of learning that is valued” (p. 147). For instance, tests that require the memorization of factual knowledge “are well-suited to assess the outcomes of traditional teaching approaches based on rote learning and knowledge transfer [...but] are less adequate when it comes to assessing complex competencies” (OECD, 2013, p. 147). Bloom and Krathwohl created a hierarchy of educational goals in 1956. Memorizing factual knowledge is on the bottom of the hierarchy, followed by comprehension, and then application (Bloom & Krathwohl, 1956). Only the next three levels: analysis, synthesis, and evaluation (Bloom & Krathwohl, 1956), are considered higher order thinking skills meaning that students should not only remember, understand, and apply, but also analyze, evaluate, and create (Bloom et al., 2001). Assessing higher order thinking skills requires a more complex approach as well.

The Council of Europe (2016) developed “a conceptual model of the competences which citizens require to participate effectively in a culture of democracy” (p. 3). Later the descriptors of the framework were also created to enable the assessment of what a person is able to do if they have mastered the various competences that are specified by the model (Council of Europe, 2018). The Reference Framework is

built on principles that are common to our democratic societies. It specifies the tools and critical understanding that learners at all levels of education should acquire, [...and] it offers education systems a common focus for their action while respecting a diversity of pedagogical approaches. (Council of Europe, 2018, p. 5)

The following 20 competences are included in the model grouped into the four categories of values, attitudes, skills, and knowledge and critical understanding (Council of Europe, 2016). The category of values contains valuing human dignity and human rights, valuing cultural diversity, and valuing democracy (Council of Europe, 2016). In the category of attitudes, there are justice, fairness, equality and the rule of law, openness to cultural otherness and to other beliefs, world views and practices, respect, civic-mindedness, responsibility, self-efficacy, and tolerance

of ambiguity (Council of Europe, 2016). Skills include autonomous learning skills, analytical and critical thinking skills, skills of listening and observing, empathy, flexibility and adaptability, linguistic, communicative and plurilingual skills, cooperation skills, and conflict-resolution skills (Council of Europe, 2016). The last category of knowledge and critical understanding entails knowledge and critical understanding of the self, knowledge and critical understanding of language and communication, knowledge and critical understanding of the world: politics, law, human rights, culture, cultures, religions, history, media, economies, environment, and sustainability (Council of Europe, 2016).

In order “to support the assessment of the current level of proficiency with regard to each of the competences [...and] to serve as a reference and a toolbox for educators in designing, implementing and evaluating educational interventions” (Council of Europe, 2018, p. 11) descriptors were developed and tested. For instance, when assessing autonomous learning skills, a basic level of proficiency means that the person “shows ability to identify resources for learning (e.g. people, books, internet); seeks clarification of new information from other people when needed”, an intermediate level of proficiency means that the person “can learn about new topics with minimal supervision; can assess the quality of his/her own work”, while an advanced level of proficiency means that the person “can select the most reliable sources of information or advice from the range available; shows ability to monitor, define, prioritise and complete tasks without direct oversight” (Council of Europe, 2018, p. 19) The assessment of such competences necessitates the use of alternative assessment such as various types of observations, reflections, self- and peer assessment meeting several criteria detailed in the framework (Council of Europe, 2018). The question arises whether teachers themselves possess the competences for democratic culture and what other circumstances are necessary for the execution of such assessments.

The OECD (2013) report also mentions the European Commission’s Recommendation on Key Competences for Lifelong Learning, the United Nations Educational, Scientific and Cultural Organisation’s “competencies for life” (Santiago et al., 2012), and the development of 21st century skills as the most commonly used expression in the United States among other examples from all over the world. Trilling and Fadel (2009) categorize 21st century skills into learning and innovation skills such as critical thinking and problem solving, communication and collaboration, creativity and innovation; digital literacy skills such as information, media, and ICT literacy; career and life skills such as flexibility and adaptability, initiative and self-direction, social and cross-

cultural interaction, productivity and accountability, and leadership and responsibility. In the Hungarian educational context, Prievara and Nádori (2018) use the following categorization for 21st century skills: collaboration, knowledge construction, self-regulation, real life problem solving and innovation, ICT use, and skilled communication. Connecting the development of 21st century skills to assessment, Prievara and Nádori (2018) find assessment based on gamification a possible solution, which is detailed in the next section as this approach to assessment plays a vital role in the research conducted for the present dissertation.

2.4.3 Assessment Based on Gamification

Deterding et al. (2011) define gamification as “the use of game design elements in [a] non-game context” (p. 9). They emphasize that gamification does not mean the use of games or the design of full-fledged games for different purposes such as education, but incorporating elements taken from game design in order to reach different goals that are more than entertainment. As a result, gamification is an umbrella term encompassing a wide range of possibilities depending on the applied game design elements and the non-game contexts. Due to the scope of the present paper, only a few game design elements that can be used for supporting assessment in education are mentioned. This is followed by a description of how assessment based on gamification entered Hungarian public education.

As collected in Barbarics et al. (2019), the following game design elements can be used in the non-game context of education. *Achievements* virtually or physically represent something that has been accomplished. In games, they are often locked and when a player achieves them, they become unlocked, which is the game design element of *content unlocking*. It also means that players are “unable to access content without meeting certain criteria” (Kocadere & Çağlar, 2015, p. 89). Games also use the concept of *Cascading Information Theory* meaning that information is released in the minimum possible portions to players (Rab, 2013). These first three game design elements can be linked to mastery learning (Bloom, 1968) as they enable individual players to autonomously progress step by step at their own pace towards achieving mastery, which is one of the building blocks of intrinsic motivation according to Pink (2009). At the same time, this process can also result in the experience of competence, which might also be intrinsically motivating according to Deci and Ryan’s (1985) self-determination theory. The use of *levels* is also closely related as they represent the zones characterized through available interaction elements and difficulty, so they become more and more different and difficult than the previous ones. Kocadere

and Çağlar (2015) design levels in their gamified assessment system according to the revised taxonomy of Bloom (Bloom et al., 2001). In order to support the progression, the game design element of a *narrative* can be applied that can either mean a verbal or visual representation to help players immerse themselves better in a task and experience meaning and purpose that, according to Pink (2009), can also nurture intrinsic motivation. A simpler visual representation can be the game design element of a *progress bar* that provides feedback displaying how far a player is from reaching or achieving an end goal. The game design element of *avatars* represent players in the world of the game. Players often have the chance to develop their avatars as a reward for progressing in the game such as leveling up. In many games individuals cannot achieve the goals, they have to create teams because tasks can only be done together, for example, one is a fighter, the other is a healer, and so on. Thus *cooperation* is a frequently applied game design element. More closely related to assessment, *points*, which are often experience points (XPs), form the basis of most gamified projects (Rab, 2013) as they represent continuous feedback on how well a player is doing in the game. They are often accompanied by *badges* and titles that can be displayed on a *leaderboard* as *status*. There are no direct punishments in games, for example, if players do not collect points, they lose positions in the leaderboard, so games try to activate players through *loss aversion* (Rab, 2013). Overall, the governing idea is *Blissful Productivity* (Rab, 2013), which assumes that people prefer working if they have fun, even to relaxing.

Assessment based on gamification has gained substantial interest in Hungarian public education for making it possible to avoid some of the negative effects of grading while still complying with the law to register the required number of grades. In Hungarian, Rab (2013) wrote first about the theory, related definitions, and examples of gamification in education. Later, Fromann (2017) wrote a book based on his own dissertation about the world of gamification. Both of them described the theoretical background, while Nádori and Prievara (2011) first mentioned some applications of classroom gamification in education, then they introduced how gamification could be used for assessment purposes. Prievara started an experiment and wrote about his findings first in blog posts then in a book entitled *The 21st century teacher* (Prievara, 2015), which was followed by Prievara and Nádori's (2018) book *The 21st century school*.

Designing an assessment system based on gamification, Prievara's (2015) goals of school assessment are the following: provide control to students over their own development and freedom of choice of what they are assessed for, take into account individual differences while not

punishing weaker students (compare students to themselves and not to each other), motivate students intrinsically instead of using extrinsic motivation through good or bad grades; however, be compatible with the official system of public education (still give the required number of grades). According to him, traditional grading is not suitable for supporting individual development or fulfilling any of the above mentioned goals, so the compulsory grades (that students still have to receive) should be kept apart from assessment as much as possible. Applying gamification principles, it means that students collect health points (HPs) in a given time period to complete a level. At the end of each level, these HPs are converted into grades to move on to a next level. For each level, the teacher provides students with opportunities to gather HPs (these can be traditional tasks such as a vocabulary quiz, a writing assignment etc.). Students also have the freedom to choose different tasks and collect HPs with them (for example, summary of a book they have read, translation of a song they like, extra grammar exercises they did, new words they had learnt from a sitcom etc.). They have to negotiate with the teacher how many HPs the chosen tasks are worth.

Part of Prievara's (2015) research is to ask his students' opinion on the assessment system based on gamification in anonymous questionnaires consisting of Likert scale statements and open-ended questions. In connection with grades, students find this point-collecting system more fair as they feel that their grades depend on their effort, which makes the learning process more motivating and less stressful (p. 132). They also feel that this assessment system gives them more opportunities to develop, a better picture of their performance, and continuous feedback, so their grades do not depend on a single test (p. 133). Students also word some criticisms, mainly those who are satisfied with traditional grading practices. As students are used to being compared to each other based on their knowledge, the fact that these grades reflect more their effort changes this hierarchy and forces students who could get good grades without much work to work more (p. 140). Prievara (2015) also notes that in the beginning most of the students' comments were about the relation between assessment and grades, which later changed into other elements of the system, and students became less interested in their grades, so he felt that he had achieved the goal to separate assessment from grades.

The main criticism of gamification is that it fosters extrinsic motivation by rewarding the required behavior through points, badges, and leaderboards (Fitz-Walter, 2019). According to van Roy and Zaman (2017), gamification in education often fails because "adding game elements as

external, meaningless regulations is likely to cause detrimental effects on learners' intrinsic motivation" (p. 485). They argue that instead of which game elements to add, teachers should focus on how gamification works. After a theoretical overview, they postulate the following nine "Gamification Heuristics that account for (the interplay between) design, context and user characteristics" (van Roy & Zaman, 2017, p. 485):

- #1 Avoid obligatory uses
- #2 Provide a moderate amount of meaningful options
- #3 Set challenging, but manageable goals
- #4 Provide positive, competence-related feedback
- #5 Facilitate social interaction
- #6 When supporting a particular psychological need, wary to not thwart the other needs
- #7 Align gamification with the goal of the activity in question
- #8 Create a need-supporting context
- #9 Make the system flexible (van Roy & Zaman, 2017, p. 504)

In summary, they emphasize that these heuristics wish to afford autonomous (intrinsic) as opposed to controlled (extrinsic) types of motivations. They meant to be understood holistically, adapted to the context, and designed *for* the learners' experience instead of designing the experience itself. Van Roy and Zaman (2017) also acknowledge that their heuristics are theory-based and "may benefit from empirical validation and refinement," so they "call upon future researchers to put them into practice and further extend our knowledge of gamification" (p. 503).

2.5 Assessment in Emergency Remote Delivery

The Covid-19 pandemic resulting in periods of remote delivery of education worldwide had a great impact on assessment as well. The last three months of the 2019/2020 school year had to take place online due to the first wave of the pandemic, followed by further remote and in person periods in the upcoming years. As the present research ended in 2020, only first impressions and assessment-related questions are detailed. First, terminological issues, then the reactions of the world's and Hungary's education systems and their effect on school assessment are discussed.

2.5.1 Terminological Issues

At the beginning of the pandemic, the possible need for distance education was mentioned. Although the expression "distance education" existed long before the current pandemic, it describes a different situation. It originally started with correspondence courses for students who

could not attend school physically, and with the advancement of technology the forms of distance education have also evolved. For example, in the Eurydice glossary of the European Commission, the definition of distance education in the context of higher education emphasizes the use of information and communication technologies (ICT): “Distance education [...] means a particular form of training, involving the use of ICT teaching aids, teaching-learning methods and digital study materials, based on the interactive relationship between lecturer and student and the student’s individual work” (European Commission, 2022). Distance education has its own methodology including the assessment of online teaching (Tobin et al., 2015). What is relevant in the context of the present dissertation is that in distance education one of the main roles of the teacher is assessment and not, for example, creating digital content. According to Tobin et al. (2015) “the assumption that the instructor of an online course is also its designer is a common belief that must be challenged” and instead the evaluation of online teaching should look at the quality of “content expertise, instructional design skills, delivery skills, assessment skills, and course management skills” (p 9). Although it is useful to learn from the existing methodology, during the Covid-19 pandemic circumstances were very different, and for this reason, it is better to use different words describing the situation (instead of distance education or online teaching).

In Hungarian, the words “online” and “digital” were frequently used in connection with education during the pandemic as they express the medium or the form that was used. Research articles also refer to “online education” (e.g., Molnár et al., 2020), and country reports, such as for the European Observatory on Health Systems and Policies, as well: “education continues in an online format using a digital curriculum”, “education will continue in digital form”, “out-of-classroom digital education will be maintained”, “education remains digital”, “online education for high schools also remain in place”, “5-8 graders will stay at home and continue with online classes”, “education has started again using online distance learning” (Gaal et al., 2020, November 10). As the examples show, there are plenty of expressions used to describe the situation. In order to choose and consistently use only one, I will look at the translations of the official government decisions and decrees issued in connection with the Covid-19 pandemic.

The European Centre for the Development of Vocational Training (2020) collected legislation responding to Covid-19 in connection with Hungarian national vocational education and training. The Government Decision issued March 14, 2020, officially declared the educational changes in effect from the following Monday (March 16). The Hungarian title of the decision

applies a “new working mode”, which is detailed in the text as a “digital working mode” (1102/2020. (III. 14.) Korm. határozat a koronavírus miatt a köznevelési és szakképzési intézményekben új munkarend bevezetéséről); however, its translation uses “remote learning” both in the title and in the text:

Government Decision No 1102/2020 (14 March) implementing remote learning outside the classroom in general (public) education and VET institutions due to Covid-19. [...] A series of government and ministerial decisions were published between March and May 2020: remote delivery in general (public) education [...]; [...] and school operation till the end of the academic year (2 to 15 June) continues in remote mode. (European Centre for the Development of Vocational Training, 2020, para. 1)

So here the expressions “remote delivery” and “remote mode” are used.

After the first wave of the Covid-19 pandemic, several others followed, resulting in the variations of in person and remote delivery in the following school years. In order to differentiate between the first wave (the first three months of remote delivery from March 16, 2020) and the periods of remote delivery from the 2020/2021 school year, the first three months is referred to as “emergency remote delivery”. The nature of the situation in Hungary is well illustrated by the fact that March 13, Friday morning, prime minister Viktor Orbán said that public schools would not close because the virus does not infect children, and if schools were to be closed, it would mean the end of the school year, sending teachers on unpaid leave. This declaration was published in writing at 15:47 on the official website of the government (Viktor Orban in the “Good morning, Hungary” broadcast by Radio Kossuth, 2020, para. 8). Six hours later at 21.44 (Friday evening), the prime minister announced that children cannot go to school from Monday (Viktor Orbán's emergency announcement about government measures related to the coronavirus, 2020, para. 1). The following day (March 14, Saturday) the above-mentioned Government Decision was issued, so that teachers should start remote delivery on their next work day (March 16, Monday).

After the 2020/2021 school year, research papers also started to refer to the first months as “emergency remote teaching”, for example stating that “work-based learning has not only been reshaping continuing education beyond emergency remote teaching, but also challenges the mindset and the protocol of adult learning” (Szederkényi et al., 2021, p. 405). The European Commission’s report also describes “how families handled emergency remote schooling during

the Covid-19 lockdown in spring 2020” (Vuorikari et al., 2020). The next sections detail some of the effects of the pandemic on education, focusing on assessment.

2.5.2 Emergency Remote Delivery in the World

As mentioned above, emergency remote delivery refers to the first months (the first wave) of the Covid-19 pandemic, which meant most of the second term of the 2019/2020 school year (March-June, 2020). Vuorikari et al. (2020) describe the remote schooling experiences of “parents and their children at the end of primary education and in secondary education (10-18 years old) from 9 EU countries (Austria, France, Germany, Ireland, Italy, Portugal, Romania, Slovenia and Spain) in addition to Switzerland and Norway” (p. 2). In these countries almost all students could participate in school-related activities using digital technologies, and in many cases, schools provided digital communication and learning platforms. However, “the readiness of schools and families to support remote instruction through digital technologies was uneven” (p. 3) as there were varied accessibility and availability of parental support and devices at homes in addition to teacher competences. Findings also show worries of all stakeholders, and it is interesting to note that assessment is among the worries of all stakeholders although for different reasons. Children were worried about getting bad grades, and parents were also concerned that their children would fall behind with school work and fail exams. Parents also expressed their need for guidelines on how to support their children’s distance education activities (Vuorikari et al., 2020).

The Council of the European Union (2020, June 26) published their conclusions on countering the COVID-19 crisis in education and training, in which they highlighted that “one of the greatest challenges has been how to manage assessment and grading, with some Member States encouraging a focus on formative assessment and consideration of different learning conditions” (14th point). They also discuss issues with the end of the school year due to the importance of school leaving examinations and how students would receive certificates and final degrees that might influence students’ enrolment in further levels of their education. They encourage the “exchange of information, cooperation and dialogue between the relevant education and training stakeholders, including quality assurance and recognition authorities and social partners, [to] provide a basis for a useful coordinated approach” (14th point). They also report on how Member States rapidly introduced measures including “guidelines and instructions on how to take part in distance learning and deliver distance teaching, for example, on safe use of digital tools and online assessment” (18th point). However, they also suggest taking into consideration the lessons of the

emergency remote delivery period and “support further development of teachers’ and trainers’ digital skills and competences, in order to facilitate teaching and assessment in digital learning environments” (24th point). It is beyond the scope of the present paper to analyze the further waves of the pandemic, resulting in further periods of remote delivery in education worldwide.

2.5.3 Emergency Remote Delivery in Hungary

As Molnár et al. (2020) describe the situation in Hungary, the Government and the Educational Authority declared that “in the state of emergency [...] the educational system has to be continuously operated, students must be provided with the opportunity to learn and schools and pedagogues have to adapt to the unexpected situations” (p. 606). Monostori (2021) details that it meant that “teachers had to find the proper online education platform, make digital curriculum, ensure the digital equipment and revise the forms of assessment and evaluation within days. Due to the unpreparedness of the education system, including schools and teachers, the transition to remote education has been accompanied by many problems” (p. 4). Monostori’s (2021) main finding (taking into consideration the next school year as well) is that remote delivery spotlighted a lot of issues that the Hungarian education system had been struggling with for a long time. One of the greatest problems is the highly selective nature of the education system resulting in several problems such as great differences between schools, ethnic segregation, and so on (Radó, 2020). According to Monostori (2021), this meant that those schools that had already used digital tools and teaching materials and focused on developing students’ creativity, innovation, individual skills and competences could see the new situation as a positive challenge, while in other schools where teachers did not have much (or any) experience with digital education in the past, it was a great achievement if teachers could hold online classes; however, it usually meant that they “tried to replicate classroom instruction to an online environment, and they basically used the traditional frontal instructions” (p. 2). In addition, teachers’ and students’ digital competences also greatly varied even within a school, not to mention the differences in infrastructural background.

The other major problem that remote delivery highlighted is that “the content of the curriculum is predominantly knowledge-based, and it leaves little room for the development of independent learning, creativity, group work, playful learning as well as for the development of individual abilities. Closely related to this is the fact that [Hungarian] student performance lags behind the European average” (Monostori, 2021, p. 2). This leads to the question of assessment

practices. Emergency remote delivery made the situation apparent as Monostori (2021) also describes:

Most teachers identified assessment and grading as the most problematic issue during remote education. During school closure in 2019/2020, hardly any schools and teachers had an opinion of how to evaluate in the given circumstances. The most important issue was considered to be how the grades would be given. Teachers tried to write exams using different digital platforms, but they were also aware that students often did not solve the tasks by themselves, making it hard to assess their actual performance. Teachers used different practices to avoid fraud: they set serious time limits for solving the tests, trying to make tests with different sets of questions for each student. Others simply did not give grades and did not give exams, but closed the year as of March 2020. In the 2019/2020 school year, very few teachers came to the realization that grading was not the most important element of assessment. (p. 11)

Láncos and Christián (2021) look at the situation from a legal perspective: “the government decided to declare a state of danger on 11 March 2020, a form of special legal order under the Fundamental Law” (p. 77). Only due to these special legislative powers was it possible to institute a lockdown and take further actions, such as closing schools. These actions were “complemented by administrative authorities, bodies and agencies playing an important role in both informing the public and issuing non-binding measures” (p. 78). The Government Decision declaring the closure of schools (March 14, 2020) did not name any responsible institution, but stated that methodological recommendation was issued to implement remote delivery (nor did it include anything on the content of the recommendation).

It was the National Education Authority that finally adopted these methodological recommendations on 9 April 2020, comprising the following instruments: the ‘Methodological recommendation for out-of-class digital work schedule’; the ‘Methodological recommendations for kindergarten, arts and other special education situations’; and the ‘IT security recommendation for teaching and learning in a digital work schedule outside the classroom’ and published them on its website (Láncos & Christián, 2021, pp. 86-87).

In connection with the implementation of the first version of the recommendations, Láncos and Christián (2021) interviewed four principals from public, private, and church-run schools. All

of them answered that they had not implemented the recommendations, on the one hand because they “came with considerable delay [...] since the schools were bound to transition to digital teaching mid-March, the recommendations issued mid-April were no longer considered relevant by the respondent schools” (p. 90), and on the other hand, “the recommendations were too generally framed and were difficult to adapt to the diverse situations in which the different (public, private or church-run) schools and the affected students and families found themselves when the closures were announced” (p. 90). Nevertheless, Láncoš and Christián (2021) believe that “external soft law has the potential to become an important regulatory instrument in Hungarian administrative governance, ensuring useful professional guidance to domestic stakeholders” (p. 92).

This first version of the recommendations (published 9 April 2020) was already irretrievable when Láncoš and Christián (2021) wrote their case study about it. For this reason, only later versions of the Educational Authority’s recommendations will be discussed in the next section.

2.5.4 Recommendations of the Educational Authority Focusing on Assessment

As Monostori (2021) reports, the “Educational Authority (Oktatási Hivatal) published its recommendations for the 2020/2021 school year on 25 August, a few days before the start of the school year” (p. 6). It was about three pages long and contained about one page in the following topics each: “IT security recommendations”, “Additional useful materials to support learning”, and “Monitoring the learning process and assessment”. Almost a year later, the Educational Authority (Oktatási Hivatal, 2021, June 29) added further sections that can also be seen under the same link. Later in 2021, the Educational Authority published a 277-page-long volume (Farkas et al., 2021) of digital pedagogical methodological recommendations. As it is highly relevant for the present dissertation, I detail the first available recommendations of the Educational Authority about monitoring the learning process and assessment (Oktatási Hivatal, 2020, August 25) and the relevant chapters of the 277-page-long volume (Farkas et al., 2021).

In the introduction, the authors of the recommendation (Oktatási Hivatal, 2020, August 25) encourage schools to motivate students by regular reporting through online tasks, mainly for self-monitoring and self-assessment, and they also acknowledge the existence of numerous opportunities to monitor the learning process and the importance of their varied use and

assessment. The main part presents a theoretical overview of assessment, and the recommendations end with questions to consider when choosing the forms of assessment.

The theoretical overview begins with a warning that teachers should keep the amount of tasks manageable for students and their assessment achievable, as well as “provide feedback that is understandable, reasonable, and inspires students’ development” (Oktatási Hivatal, 2020, August 25, para. 3). Assessment methods are categorized into live (real time) reporting, delayed remote reporting, and in-process assessment. There are examples for all of them, such as live (real time) reporting can happen through video conferences, phone calls, chat programs, etc. while delayed remote reporting can be supported by forums, shared documents, online questionnaires, tests, and so on. Further examples of what teachers could use are listed: homework, presentation, multimedia tasks such as “photo, video, montage, collage, moodboard, mindmap, problem tree” (Oktatási Hivatal, 2020, August 25, para. 6), blogs, wikis, quiz games, questionnaires instead of tests, offline tasks (posters, creative or artwork) submitted through pictures or videos. According to the recommendations, in-process assessment can be supported by different testing, voting or polling programs. Moreover, educators experienced in digital pedagogy can use digital project labs, project rooms, webquests, or gamified frameworks. What is highlighted in connection with gamification is that in gamified frameworks students set goals for themselves and select from a variety of tasks offered by their teachers to achieve those goals. The recommendations state that self-assessment opportunities for students can also be added to the assessment process. Listed self-reflection methods include checklists, surveys about different elements of the learning process (motivation, attitude, interest, time management, self-efficacy, and so on), journals, forum posts, moodboards, or emoticons reflecting steps of the learning process, tasks, or students’ performance. The ninth paragraph ends with the conclusion that using these methods has the substantial benefit of improving students’ self-knowledge and reflective approaches “that are crucially important among the so-called 21st century skills” (Oktatási Hivatal, 2020, August 25, para. 9).

The recommendations (Oktatási Hivatal, 2020, August 25) use the categorization of assessment based on purpose: diagnostic, formative, and summative, which are described in the following way. Diagnostic assessment identifies students’ attitude to a topic (or learning itself), their level of motivation, and knowledge. Summative assessment is the comprehensive assessment of the learning process at the end of a given topic. It aims to compare student performance to requirements thereby grading student performance. Formative assessment is an assessment

procedure that aims to determine learners' development and adapt their learning to it. It compares (student performance) to the learning pathway in which students are able to achieve the set goal (for example, set by the curriculum or the task). Formative assessment focuses on the process of development, providing an opportunity to traverse individual developmental paths and to continuously adjust, correct, and shape the learning process.

According to the recommendations (Oktatási Hivatal, 2020, August 25), the role of diagnostic assessment is not that important in remote, digital delivery as teachers have already gathered information about their students' knowledge, attitude, and motivation beforehand; however, the Educational Authority in its recommendation also notes that experts suggest using diagnostic assessment not only at the beginning but also for in-process assessment, for example, through mini tests or quizzes. One method is recommended in detail for summative assessment: student portfolio. Tests are also mentioned but recommended only if it can be made sure that they reflect students' individual work. From the three categories, the importance of formative assessment is emphasized. Authors of the recommendation state that during formative assessment, teachers can help students progress in two main ways: first, providing constructive and supporting feedback on students' progress, knowledge acquisition, independent questions, research, experience, and problem solving; and the second is feedback positioning students to make them capable of solving tasks and thinking creatively. What feedback positioning students mean is not detailed any further. The recommendations continue with the role of formative assessment in remote delivery. It is also claimed to be important because metacommunicative signals, which are usually well-decoded by teachers in the classroom, are typically not accurately perceived during online work. Therefore, it is suggested planning pedagogical processes in a way that formative assessment is embedded in the processes. For formative assessment, teachers can be assisted by self-assessment tasks and tests, which support students' progress and independent knowledge acquisition in an adaptive way.

The recommendations (Oktatási Hivatal, 2020, August 25) end with the following list of questions to consider when choosing the forms of assessment: What is the purpose of assessment? Who is it for? Does it facilitate the learning process, individual development, or cooperation in the group? Does it provide feedback on the result of a learning process or on previously acquired knowledge? What are the special features of the given age group? What do they understand? What can they use? In connection with a given platform: does it have age restrictions? Is it accessible to

everyone? Can it be customized for individual solutions (too)? Considering infrastructural possibilities: Does everyone have access to the internet and digital equipment suitable for learning? What is a channel that everyone involved can use? How can students access the platform used for assessment? Considering the class, the subject, and teacher's preferences and traditions: What assessment methods do they know? What are they used to? What can they easily learn to use? What provides feedback that is understandable to both students and parents? Considering aspects of data protection (especially in the case of peer assessment): Is sensitive data generated? Who has access to the assessment? Who has access to the platform used? Do students (or parents) need to create a profile on the platform, and if so, what information does it contain? Are there any privacy settings that can be used to ensure privacy? These questions lead to the other sections: "IT security recommendations" and "Additional useful materials to support learning" that will not be detailed here.

It is interesting to note that, as mentioned earlier (e.g., by Monostori, 2021), grading was the most problematic issue for teachers during remote delivery, and in the recommendations (Oktatási Hivatal, 2020, August 25) grading is only mentioned by saying that the results of summative assessment influence students' final grades at the end of the year. So the recommendations do not provide any answers to the question of many teachers of how to give grades, while they strongly support "alternative assessment" (based on my definition: assessment that contains different elements different from grades in order to support students' development).

Later in 2021, a collection of digital pedagogical methodological recommendations was added (Farkas et al., 2021). The 277-page-long volume deals with

the technological tools of digital pedagogy, learning organization solutions, digital curricula, forms of assessment, the problems of disadvantaged children and children with learning difficulties. The volume also includes a list of additional websites where teachers can find digital learning materials, methodological support, and good practices by grade and subject. (Monostori, 2021, p. 6)

Before detailing the chapter on assessment, it is important to note that the volume uses the term "traditional" to refer to teachers' in class practices in contrast to the new situation caused by the pandemic resulting in hybrid or fully remote, digital delivery. Moreover, the recommendations repeatedly emphasize that the current situation only accelerated the necessary paradigm shift caused by the advancements of the 21st century, and the following principles and methods are also

referred to as “traditional” that needs to be changed: the source of information is the teacher, the focus is on transmitting pieces of information from the teacher to the students, the students are passive recipients of this information transmission, and assessment is mainly evaluating acquired knowledge in paper based written or traditional oral tests (Farkas et al., 2021). As it is detailed in section 2.1.3, The Pedagogical Programme of the School of the Action Research Project, the main forms of assessment are indeed written and oral tests. In connection with “traditional oral tests” the pedagogical programme only contains that they do not have to be announced in advance; however, they traditionally happen by choosing students randomly who have to answer the teachers’ questions about the material in front of the class. Farkas et al. (2021) recommend instead that teachers renew their approaches and practices by, “for example, providing interactivity, personalization, support for learning instead of or in addition to teaching, opportunities for alternative learning and knowledge acquisition, acceptance of knowledge gained elsewhere, supportive assessment approaches, and encouraging community learning” (p. 87).

The eleventh chapter of the volume is entitled “Digital Technology in Assessment for Hybrid and Online Delivery - Recommendation for Educators on Forms and Measurement Tools for Student Assessment Supported by Digital Technology” (Farkas et al., 2021, p. 131). It has five sub-chapters: after an introduction, it goes on with “Pedagogical Principles of Digitally Supported Student Assessment”, then “Considerations for Planning Assessment”, followed by “Assessment and Evaluation Tools in Practice - Some Tips”, and ends with “Assessment and Support with Digital Tools” (Farkas et al., 2021, pp. 131-144).

The chapter on pedagogical principles (Farkas et al., 2021, pp. 133-134) begins with the hope that the assessment systems teachers use in class do not only mean giving the number of monthly grades prescribed in the pedagogical programme of the institution. The authors also call attention to the fact that assessment methods working in class cannot be automatically transferred to processes used during remote delivery. For these reasons, they suggest that teachers should not try simply modernizing traditional classroom assessment practices, but using the opportunities offered by digital tools and applications, it is necessary to renew their methodology including assessment: by answering the question whether a particular assessment method supports the learning process in remote learning, too, and change their methods accordingly. As the authors state, most methods that work in remote delivery can also be used in hybrid or in person teaching, while it is not true the other way around as in remote delivery, assessment is based on

differentiation. Teachers should also note that there are several factors influencing students' performance, so they should choose assessment methods that minimize the effects of factors threatening the reliability of assessment. If teachers have the chance to meet students in person, the time should not be spent on realizing traditional assessment practices, but on processes that support student learning, such as, developing skills necessary for digital communication, tasks, and assessment.

The recommendations continue with the applications of the above mentioned pedagogical principles (Farkas et al., 2021, pp. 134-135) meaning that traditional assessment methods should be replaced by complex assessment methods based on formative assessment, different tools for self- and peer assessment, checklists and rubrics, oral and written feedback, and online discussions. They ask traditional achievement tests, pop quizzes, and oral tests, the three most common, and sometime only, ways of assessment, not to be used unless it is absolutely necessary to meet requirements or curriculum expectations. They suggest the application of the following instead: project work, student portfolio, cooperative work utilizing self- and peer assessment, online debate, and gamification. According to the authors, these examples show that it is possible to provide supportive assessment to students while fulfilling the obligation of giving grades. However, they do not specify how to give grades when using the above mentioned alternative assessment methods. They also suggest alternatives for oral tests: discussions, debates, and different types of student presentations, and explain the advantages of using these methods: it is possible to avoid the stress that almost always accompanies evaluation, and teachers do not have to worry about students using unauthorized tools or help during oral or written tests. Finally, as in remote delivery teachers cannot see students' immediate reactions, the authors also advise asking regular feedback from students and even from parents to further develop methods used during online delivery.

There are the following eleven aspects listed to consider when planning assessment (Farkas et al., 2021, pp. 135-137). First, collecting and sharing individual measurement and assessment methods offers an opportunity to expand and update assessment at the institutional and even school district level. Second, it is important for the teacher to consider students as partners during assessment. Third, strive for the use of diagnostic and formative measurement and assessment methods as the results of summative assessment might be misleading. Fourth, a toolkit of diagnostic assessment forms helps in collecting information during the school year (such as the

level of knowledge required for progress in the given subject, the extent of missing knowledge affecting the subject, the degree of progress expected based on individual abilities). Fifth, it is worth striving for the widest possible use of formative assessment methods, and adding self- and peer assessment to the different forms of assessment provided by the teacher. “Let’s not forget that we can best support students in learning and development through formative assessment!” (Farkas et al., 2021 p. 136) as they highlight. Sixth, in the case of summative assessment as well, interpret the concept of assessment as broadly as possible and use complex tasks. Seventh, in the case of all types of assessment, personalization, adaptability, and the development of individual development and learning paths differentiated on the basis of results should be sought. Eighth, the use of tools developed for other assessment purposes can also be used (e.g. standardized entrance exams, assessment tools developed for research purposes, and so on). Ninth, avoid traditional forms of evaluation: oral or written testing. Tenth, continuously collect information about missing students, and plan how to help dropouts. Eleventh, ask students, parents, and colleagues how they feel in the online education space. Evaluate their responses and incorporate the lessons learned into system-wide development planning. It is interesting to note that the recommendations (Farkas et al., 2021) interchangeably use the expressions “formative assessment” and “developmental assessment” or “assessment for students’ development”. It is not clear if there are differences between them; however, their importance is continuously emphasized as opposed to traditional assessment. In addition, Farkas et al. (2021) do not address the question of grading in the 277-page-long volume either.

The chapter entitled “Assessment and Evaluation Tools in Practice - Some Tips” (Farkas et al., 2021, p. 137) provides a short overview of the types of tools that are detailed and expanded with exact examples in the last, Assessment and support with digital tools, chapter. Both chapters are structured based on the three types of assessment: diagnostic, formative, and summative. It is emphasized that all types of assessment should aim for making (virtual) classroom processes interesting and engaging through interactivity. The “Assessment and Support with Digital Tools” chapter (Farkas et al., 2021, pp. 138-144) contains the specific tools presented in a chart, in which the first column defines the tool, the second includes the exact links to their digital sites, and the third includes further links to videos or documents explaining how to use them. In the next paragraphs, I detail these definitions of the different types of assessment. The exact links of the recommended digital tools and examples can be found in Appendix B.

Farkas et al. (2021, p. 138) describe diagnostic assessment as follows. It is used to assess students' prior knowledge and level of knowledge in order to make the pedagogical process more effective. This feedback is primarily for the teacher (or the school). Interactivity between teachers and students during (virtual) classroom processes also makes it possible for the teacher to gather information that later can be used for supporting the learning process of the students. Diagnostic assessment can happen through surveys or national and international instruments for assessing students' different skills and competences. Moreover, less complex methods (such as brainstorming or creating a word cloud) can also serve the development of students and provide information about their prior knowledge of a particular topic, and while working together, the learning process can be supplemented by questions or short quizzes that require immediate answers with a variety of looks and compositions, too, and these tools can be the starting point for formative assessment.

According to Farkas et al. (2021), formative assessment provides information to students during learning based on clear assessment criteria as to whether they meet output requirements and work with the right method and pace. In addition, it provides meaningful assistance in determining the possibilities of correction and possible directions for development. Formative assessment can come from the teacher, but self- and peer assessment should also be used.

In connection with formative assessment coming from the teacher, the following recommendations are listed (Farkas et al., 2021, pp. 138-139). Comments, advice: short written or oral feedback on student performance. Supportive, developmental ideas and advice can be delivered to students in text or audio notes using digital technology and can be easily recalled later, thus supporting the process of entirely remote or blended learning. Quizzes: a form of assessment using short, limited number of question types. They can be used for warm up, revision, or motivation. They can be easily applied in all types of working modes using digital technology. Questions, criteria, hints: brief oral or written feedback on student performance. Supportive, constructive criticism, questions, and help can be provided to learners through digital technology (such as giving criteria in the form of a checklist, or an overview table) and can be easily recalled later. Portfolio: (digital) portfolios compiled from students' work are not only collections, but also show students' development. In addition, teachers can provide feedback on learning outcomes and methods. A portfolio can contain a variety of easily retrievable data and information through digital technology. Feedback can happen in the form of texts, voice memos, or even video messages.

Rubrics: a tabular tool that helps students categorize their performance showing the teacher's planned assessment. This table shared through digital tools guides students in what to look for while completing a given task. Rubrics or any other recommended tools can serve as a basis for consistent implementation of formative assessment.

In the category of self-assessment, the following methods and tools are recommended (Farkas et al., 2021, pp. 139-140). Checklists: a list of student tasks broken down into elements, which can be used as a basis for self-assessment. The checklist should be shared with the students in advance, so that they can keep track of what they have already done from the tasks, and with the help of digital technology, the teacher can also follow the completion of the tasks. Self-assessment, practice tasks: while practicing students can keep track of what they have acquired, how confidently they know the material, and whether they have met the requirements. With the help of digital technology, teachers can also follow the progress of an individual student or even a whole class. Rubrics and quizzes are also mentioned here implying that they can also be used as self-assessment tools. Games: when completing playful tasks, whether individual or group, carrying out the activity itself contains information for self-assessment. A self-assessment criteria system provided in advance can further support this otherwise spontaneous process. Exit tickets: at the end of the lesson, teachers can ask students a question about the lesson, which is an important piece of feedback for the teacher and the student themselves about what they understood well, what they liked, what caused them problems, what they would like to get help with, and so on.

For peer assessment the recommendations contain the following methods and tools (Farkas et al., 2021, pp. 140-141). As described in connection with self-assessment, checklists and rubrics can also be used for peer assessment. Oral feedback: feedback on student performance in a live or online channel. Supportive, constructive, longer verbal feedback can also be recorded using digital technology or transcribed from it, so that the ideas and pieces of advice can be easily recalled later, too. Discussions, debates: forms of group assessment for which the main points and guidelines must be clarified in advance. Questions and aspects that arise can be recorded digitally for later discussion or assessment. Brainstorming: a creative way for groups to find new, alternative ideas on a given topic. Implemented in digital group work (e.g., working together in a shared document), it can be an effective tool for introducing a new topic, task, or approach to a problem. Mind map: a tool for visualizing connections between knowledge elements and concepts of new knowledge materials available in textual form (written or oral). Tools available online provide an excellent

opportunity to assess reading comprehension individually or in small groups. Once created, digital mind maps become effective learning support tools. Games: it is worth taking the time during or at the end of a game for group assessment, in which the teacher also participates as an equal partner. The efficiency of the assessment can be increased with a pre-issued system of criteria, which also directs opinions to the desired channel.

According to Farkas et al. (2021), summative assessment shows how well a student is performing relative to a given desired level. Assessment happens at the end of the learning stages and categorizes the effectiveness of learning (e.g., into grades), but the result does not influence teaching strategies, nor does it provide sufficient support and assistance in the learning process. The recommended tools are categorized into oral, written, and complex forms (Farkas et al., 2021, pp. 142-144).

The recommended oral forms of summative assessment are the following. Oral tests: in the case of online education, it is advisable to focus more on student presentations and include topics and questions whose answers cannot be found immediately in the textbook or on the World Wide Web. However, it can also be good if students search because it is also part of the learning process and a kind of feedback if students know what to look for and where to find it. To support oral tests and record results, spreadsheets or learning support systems can be used. Student presentations: they are recommended in all work modes (in person, hybrid, or remote). It is advisable to specify the assessment criteria in advance in the form of a checklist for the preparation and presentation stages. Presentation, demonstration, artistic performance: an individual or group performance, which (if necessary) can be recorded for subsequent analysis and assessment. Discussions: a type of group assessment, for which assessment criteria must be clarified in advance and provided in the form of a checklist. The questions, aspects, decisions that arise can be recorded digitally for later discussion or assessment, such as using a word processor, mind map, or collaboration tool.

Written forms of summative assessment during remote delivery can be the following. Different types of written tests (achievement tests, pop quizzes, exams): depending on their purpose and timing, they vary in scope (consisting of one or more questions) and duration. Under fixed conditions, the individual performance of several students can be assessed simultaneously with the help of these pre-compiled questions and tasks. Implementing them in remote delivery is difficult, especially guaranteeing that students' work individually. The same definition and examples are given for quizzes here as under formative assessment coming from the teacher or as

a form of self-assessment; however, in Hungarian different words are used (in the formative section “kvíz”, while in the summative section “teszt”). Authors even state that “they can be used for practicing, in-depth learning, and self-assessment both in class and at home, and collaborative, shared materials from colleagues can also be easily adapted” (Farkas et al., 2021, p. 143) implies that authors mean the same tools used for different purposes. However, there is no further explanation of how these different purposes would manifest in practice. Authors continue with worksheets: a variety of task types can be used both offline and online. Essays, home assignments: creative writing in a given or possibly freely chosen topic according to goals and expectations set in advance. With carefully designed assessment criteria that are well known to students, it can become a versatile form of assessment. It is worth considering sharing and assessing the writings together with students. Assessment criteria (in the form of a checklist or overview table) or comments by the teacher (in the form of recorded voice notes) can also be used.

The last group of recommended tools consists of the following complex summative assessment methods. Project tasks: during pedagogical projects, activities are continuously assessed primarily in a formative way. It is advisable to consider assessment tools when planning the project. Checklists, overview tables, which are also useful in self- and peer assessment, as well as scoring tables, oral feedback, and discussions are very common. Escape Rooms: a complex system of tasks characterizes escape room type activities with a linear, branched, or even star-like gameplay. In these activities, the feedback is usually textual and included as an element of the game. Portfolios are mentioned for both summative and formative assessment. Similarly to other examples, the same definitions, explanations, and digital tools are given without specifying what their use for different purposes mean in practice. Gamification is also listed as a tool for summative assessment with the following explanation. Activities assessed with the means of gamification are for a longer period of time. The feature of the method is that processes contain game elements. In many cases, students complete levels and earn points and rewards online. Lastly, contests and competitions are detailed: competitive forms of assessment in which students receive quantifiable results (e.g., scores, time scores, rankings) using digital tools.

In connection with terminology, it is also interesting to note that there are some similar expressions used, and based on their definitions, the difference among them is not clear, and in most cases, the same digital tools are recommended, implying that they are synonymous: rubrics,

assessment tables (in Hungarian “értékelő táblázat”), overview tables (in Hungarian “áttekintő táblázat”), and scoring tables (in Hungarian “pontozó táblázat”).

As mentioned in connection with the recommendations issued August 25 (Oktatási Hivatal, 2020), the question of grading was not discussed in those three pages. The 277-page-long recommendations from the following year (Farkas et al., 2021) also have only a few mentions of grades. As detailed above, the chapter about assessment begins with the hope that the assessment systems teachers use in class do not only mean giving the number of monthly grades prescribed in the pedagogical programme of the institution (Farkas et al., 2021, p. 133). In addition, when providing examples of alternative ways for summative assessment, it is also noted that these methods suitable for remote delivery (e.g., project work, portfolio, group work, online debate, gamification etc.) can also result in grades. The other contexts in which grades appear in the recommendations are the following. The fifth chapter describes the national compulsory e-gradebook system (e-KRÉTA) (Farkas et al., 2021, pp. 49-72) that schools must use for communication with parents and students, register each lesson, grade, and optionally add notes such as homework. The chapter reminds teachers of submitting grades as it remained the only compulsory output requirement, so the pandemic did not induce any changes in that regard. The other chapter where grades are mentioned is chapter six: “Online Learning Support Systems and Virtual Classroom Applications” (Farkas et al., 2021, pp. 73-82), suggesting that students’ work should be assessed textually, with graphic elements (e.g., badges) or numerically (e.g., grades, points, percentages) depending on whether it is formative or summative assessment. Chapter twelve focuses on intervention for groups at risk of falling behind (Farkas et al., 2021, pp. 145-170). According to the authors, in remote delivery, grades should be treated as “signs” to identify students at risk (Farkas et al., 2021, p. 149). Moreover, the methods used for giving grades, including grades for behavior and diligence, should be revised (Farkas et al., 2021, p. 150). Although there are recommendations on how to modify teachers’ perceptions of behavior, diligence, and attendance in remote delivery as opposed to in class practices, there are no recommendations for grading, while it still remains the only compulsory output requirement.

In summary, as the Recommendations of the Educational Authority (Farkas et al., 2021) also emphasize, the Covid-19 pandemic only accelerated the necessary paradigm shift caused by the advancements of the 21st century. Farkas et al. (2021) agree that traditional assessment in which the teacher is the sole source of information and the focus is on testing the transmitted

factual knowledge should be renewed, and teachers should provide interactivity, personalisation, and support for learning in alternative ways as well as different types of assessment facilitating these processes. International policy documents, such as the OECD's (2013) report: Synergies for better learning (as part of the Reviews of Evaluation and Assessment in Education series) also suggest policy directions that establish a good balance between formative and summative practices, support effective formative assessment processes and a variety of assessment types, put the learner at the center and involve them in their own assessment. The report (OECD, 2013) also adds that it should happen through maintaining the centrality of teacher-based assessment and promoting teacher professionalism. These directions are in line with the goals of alternative assessment that, in this study, mean different elements in addition to or instead of grades with the purpose of supporting students' development.

3. Research Niche and Research Questions

On an international level, governments and education policy makers are increasingly focused on assessment to understand better how well students are learning and to provide information to parents and society about educational performance and to improve school leadership and teaching practices (OECD, 2013). As presented in the literature review, although international trends in student assessment show that policy priorities should take a holistic approach, align assessment with educational goals, focus on improving classroom practices and build on teacher professionalism, avoid distortions such as teaching to the test, put students at the center, and so on, "information collected in the OECD Review indicates that the use of innovative assessment approaches remains quite limited within the national assessment frameworks of OECD countries" (OECD, 2013, p. 149). Besides presenting some promising alternative assessment practices in different countries, as Hungary did not opt to take part in the OECD's country reviews (2013), data could only be collected from official legislation and policy papers in connection with Hungary. Therefore, research is needed to determine what kinds of assessment, if any, takes place in Hungarian public secondary education apart from what is legally required.

Hungarian national legislation (Act CXC on National Public Education, 2011) only requires summative assessment in the form of one to five grades in public primary and secondary education and centralized secondary school leaving examinations, while there are no central frameworks for formative assessment on any levels of education (OECD, 2013). Official policy

papers include ideals on different purposes of assessment such as providing learning support, developing students, and carrying out ongoing evaluation and analysis of educational processes (Act CXC, 2011, Section 64), or applying differentiation as a basic principle, taking into consideration the development of talented students, and incorporating different types of assessment (National Core Curriculum, 2012). Although in theory, schools could apply different assessment systems as long as they can also provide grades if requested (Act CXC, 2011, Section 54, paragraph 4), pedagogical programmes show a similar duality to official legislation. On the one hand, as output requirements refer to grading exclusively, grading procedures are detailed extensively. On the other hand, further purposes and assessment methods are only mentioned, such as “we use guided self-assessment tools”, “we implement open, well-established, predictable, justified and fair assessment”, or “we adhere to the principles of regularity and methodological diversity in assessment” (Gál-Berey, 2022, pp. 88-89) without any further specifics, practical implications, or guidelines for how teachers should carry them out. As a result, it depends entirely on individual teachers if and how they fulfill the above mentioned purposes. Only few research studies have been conducted in connection with assessment in Hungary focusing on ways that aim to fulfill other roles than grade giving (Hubai & Lázár, 2018). Consequently, research should map how teachers apply alternative assessment methods.

The advancements of the 21st century bring further challenges, for instance, the need for developing and assessing more complex skills and competences. Both international (Vogt & Tsagari, 2014) and Hungarian (Radnóti, 2006) studies show that teachers express the need for further training in assessment of complex competences in English as a foreign language (EFL) and English medium content (EMC) classes. The Covid-19 pandemic resulting in emergency remote delivery worldwide also highlighted the inadequacy of some traditional assessment methods (Farkas et al., 2021). Hence, research needs to explore ways of assessment fulfilling these roles and meeting the challenges of the 21st century.

Thus the present research aims to gain insights into the use of alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary. In order to map alternative assessment methods in this context, the following research questions are asked: (the first group of questions starting with number 1 refer to teachers’ perspectives, and the second group of questions starting with number 2 refer to students’ perspectives)

- 1.1 What do teachers mean by assessment and alternative assessment in particular in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 1.2 What alternative assessment methods do teachers claim to use in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 1.3 What are teachers' views of using alternative assessment in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 1.4 What are teachers' motivations and purposes for using alternative assessment in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 1.5 What are teachers' experiences with using alternative assessment methods in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 1.6 How can alternative assessment methods in English as a foreign language and in English medium content classes in public secondary education in Hungary be adapted to remote delivery?
- 2.1 What are students' experiences in connection with assessment in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 2.2 What are students' perceptions of alternative assessment in English as a foreign language and in English medium content classes in public secondary education in Hungary?
- 2.3 What are students' experiences in connection with assessment during emergency remote delivery in public secondary education in Hungary?
- 2.4 What are students' perceptions of the adaptation of alternative assessment methods in English as a foreign language and in English medium content classes in public secondary education in Hungary to remote delivery?

4. Methods

The research consists of three main phases. In order to answer the first five research questions (What do teachers mean by assessment and alternative assessment, what alternative assessment methods do they claim to use, what are their views of using alternative assessment, what are their motivations and purposes for using alternative assessment, and what are their experiences with using alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary?), an interview study was conducted with Hungarian public secondary school teachers already using alternative assessment methods.

In the second phase, to gain more insight into the use of alternative assessment methods (research question 1.5) and to understand students' experiences with school assessment in general (research question 2.1) and alternative assessment in EFL and EMC classes in public secondary education (research question 2.2), an action research project built on the results of the first phase was carried out in the school year 2019/2020.

The third phase of the research was not planned originally; however, as the second, action research phase was interrupted by the Covid-19 pandemic resulting in emergency remote delivery, further research questions emerged: 1.6 How can alternative assessment methods in EFL and EMC classes in public secondary education in Hungary be adapted to remote delivery? And 2.3 What are students' experiences in connection with assessment during emergency remote delivery in public secondary education in Hungary? 2.4 What are students' perceptions of the adaptation of alternative assessment methods in EFL and EMC classes in public secondary education in Hungary to remote delivery? In order to answer these questions, the last three months of the action research project were modified in ways that will be discussed in the upcoming sections, and a follow-up interview study was organized with the participants of the first interview study to look into their experiences with emergency remote delivery.

In the Methods section, first the research paradigm is presented, followed by participants, instruments, procedures, data analysis, and the chapter ends with ethical and legal considerations.

4.1 Research Paradigm

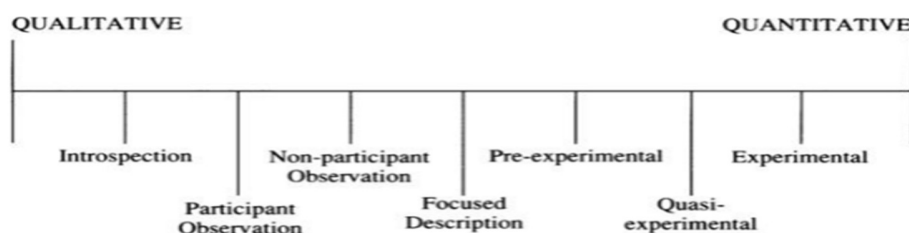
The research paradigm “provides the largest framework within which research takes place. It is the world view within which researchers work, [...] a set of overarching and interconnected assumptions about the nature of reality” (Maykut & Morehouse, 1994, p. 5). Holliday (2007) compares the quantitative and qualitative research paradigms as different ways of thinking about

the world. “The belief in quantitative research that reality can be mastered by the right research instruments is *normative*. It maintains that there is a normality that we can fathom and understand, and master by statistics and experiment,” while “the qualitative belief that the realities of the research setting and the people in it are mysterious and can only be superficially touched by research which tries to make sense is *interpretive*” (Holliday, 2007, pp. 5-6). According to Guba (1981), “the issue here is not which assumptions are ‘true’ but which offer the best fit to the phenomenon under study” (p. 77). As “social/behavioral phenomena exist chiefly in the minds of people, and there are as many realities as persons [...] if one attempts to focus attention on certain portions of reality, the whole falls apart,” moreover, “human behavior is rarely if ever context-free; hence knowledge of human behavior individually or in social groups is necessarily idiographic, and differences are at least as important as similarities to an understanding of what is happening” (Guba, 1981, pp. 77-78). It follows that research questions aiming to interpret social, behavioral phenomena in a given context should be explored within the qualitative research paradigm. The purpose of the present research is to gain insight into the use of alternative assessment methods in the context of EFL and EMC classes in public secondary education in Hungary. As a result, the research questions are exploratory in nature and thus follow the qualitative paradigm.

Maykut and Morehouse (1994) “take a radical position on the nature of qualitative research[, ... which] has important implications [...]. It means, among other things, that one must carefully match research questions with methods of collecting and analyzing data” (p. 5). As Larsen-Freeman and Long (2014) describe, qualitative data collection is recommended when the goal is to understand processes and human behavior without previous hypotheses to be tested. However, they also argue that research methodologies are on a qualitative-quantitative continuum portrayed in the following figure.

Figure 2

The Qualitative-Quantitative Continuum of Research Methodologies (Based on Larsen-Freeman & Long, 2014, p. 59)



As research questions ask about teachers' views (1.3), motivations and purposes (1.4), experiences (1.5), and students' experiences and perceptions (2.1 – 2.4), introspection is one of the most important “methodologies” (Larsen-Freeman & Long, 2014) of the research. Participant observation is also significant as a researcher, I also take part in the activities I observe during the action research phase, which is also supplemented by non-participant observations. Elements of the action research use quantitative data collection, too, part of a pre-experimental design as there are no control groups and participants are not randomly assigned to the groups taking part in the action research (Larsen-Freeman & Long, 2014).

There are several approaches to ensure the quality of qualitative research. Here I mention the theoretical background that I will refer to when justifying the different aspects of my research. Guba (1981) renames the aspects of trustworthiness (truth value, applicability, consistency, and neutrality) and provides means to achieve them. According to Guba (1981), the naturalistic term equivalent of internal validity is credibility, of external validity or generalizability is transferability, of reliability is dependability, and of objectivity is confirmability. Similarly, Holliday (2007) identifies the sources of validity and states that “the qualitative researcher needs to justify every move - demonstrating particularly how the overall strategy is appropriate to the social setting and the researcher-subject relationships within it, and the steps taken for thorough engagement” (p 8). According to Holliday (2007), when quantitative research details the population, the survey questions, the statistics, the variables, and groups exposed to in the experiments, qualitative research should provide the rationale for the choice of the social setting, the research activities, the themes and focuses, “the dedication to and thoroughness of fieldwork”, and “a judicious balance between opportunism and principle” (p. 9), respectively. For these reasons, the choice of participants, instruments, procedures, data analysis, and ethical considerations are detailed in the following sections.

4.2 Participants

In connection with qualitative research, Maykut and Morehouse (1994) refer to the “human-as-instrument” concept, which “simply means that it is the person with all of her or his skills, experience, background, and knowledge as well as biases which is the primary, if not the exclusive, source of all data collection and analysis” (p. 24). It also results in the key role of the researcher as “the qualitative researcher has the added responsibility of being both the collector of relevant data - data whose relevance changes as the study proceeds - and the culler of meaning

from that data, which most often is in the form of people's words and actions" (Maykut & Morehouse, 1994, p. 41). For these reasons, in this chapter I first introduce myself, then the participants of the interviews and the action research.

4.2.1 The Researcher

In connection with the researcher, Holliday (2010) suggests that it should be explicitly stated how "her own ideology is defined, directs her research methodology and thus has a certain type of impact on the research setting and the people involved, in terms of all aspects of how she sees, interacts with and treats it" (p. 48). For this reason, I will introduce myself through my experiences with school assessment and reflect on the influences that shaped me as a researcher.

I grew up in the Hungarian public education system. I was an exemplary student, which meant that I always got grade fives in all subjects. Although I attended special math classes both in primary and secondary school and took the higher level matura exam from both mathematics and English with grade fives, I did not feel challenged enough. In 9th grade, I took a B2 level language exam so that I did not have to attend the EFL classes in school. As a student, I felt that it was extremely unfair that I could meet requirements without much effort while I saw others investing a lot of time and energy and still not get good grades.

For these reasons, when I studied to become a mathematics and English teacher, I was already looking for ways of alternative assessment. As a teacher trainee in 2013, I attended a pedagogical conference and listened to a presentation by Tibor Prievara on the findings of his action research experimenting with assessment based on gamification. It raised my interest and I chose it as the topic of my master's thesis entitled Possible Uses of the Assessment System of Gamification in ELT (Barbarics, 2014). I started conducting interviews with teachers using gamification and launched a Facebook group to connect them in 2014. My Facebook group, Gamification in Hungarian Education, had more than 2800 members of teachers interested in the use of gamification by the beginning of 2019, and more than 6000 by 2023. In 2015, Prievara asked me and six other teachers who were experimenting with assessment based on gamification to write about our experiences in his book (Prievara, 2015).

In 2015, I graduated from Eötvös Loránd University (ELTE) as a teacher of English language and culture and teacher of mathematics and continued teaching at the bilingual vocational school of my teaching practice. As I studied the methodology of teaching EFL and mathematics separately, but I also had to teach mathematics in English, I organized an Erasmus+ mobility

program for myself and eleven other colleagues of my school to study content and language integrated learning (CLIL) methodology in England. In 2017, I attended a summer school in Germany organized by the Council of Europe's Pestalozzi Programme, entitled Competences for a democratic culture: from onlookers to engaged and active citizens. As a result, I volunteered to be a member of the trainer team of Pestalozzi Fridays, the Hungarian professional development community of the international Pestalozzi Programme (Lázár, 2015), and I have held workshops for teachers and teacher trainees on various topics such as alternative assessment, gamification, motivation, democratic schools, global competences, and so on.

As a practicing teacher, I have always given grades to my students based on different point-collecting systems inspired by gamification since my teaching practice as a teacher trainee. I have never used traditional grading and testing methods. From the beginning of my career, I have regularly asked my students' opinions in anonymous questionnaires and face to face discussions to enhance my teaching and assessment practices, and I have analyzed that data from different perspectives. For example, in 2018 after the Pestalozzi summer school, I looked at my students' answers from the point of view of whether my assessment helped in developing democratic competences (Barbarics, 2018). I could detect most of the components of competences for democratic culture (Council of Europe, 2016) in my students' responses. I also came to the conclusion that the assessment methods teachers use tend to be congruent with their own attitudes, skills, knowledge and values. Thus, reflecting on Holliday's (2010) suggestion, my intentions to develop attitudes, skills, knowledge and values for a democratic school culture primarily in myself and then in my students have an impact on my research and on my behavior as a researcher.

4.2.2 Participants of the Interviews

As Guba (1981) explains, social and behavioral phenomena are context bound; therefore, instead of generalizability, the research aims for transferability through describing and interpreting a given context with purposive sampling, collecting "thick" descriptive data, and developing thick descriptions (p. 86). For these reasons, the participants of my interviews were chosen through purposive sampling in order to interview teachers already using alternative assessment methods. Piloting the interview schedule (as suggested by Cohen et al., 2007), in 2018 I reached out to four teachers who were known for using alternative assessment methods based on their conference presentations and publications on the topic. I advertised the research in the above mentioned Gamification in Hungarian Education Facebook group and in other online communities of

innovative teachers. Using snowball sampling (Cohen et al., 2007), I asked the teachers whom they had heard the methods from and whether they knew any other teachers using alternative assessment methods. Then, using case sampling (Flick, 2009), based on the suggestions I chose the group of participants to ensure that it is as diverse as possible (male and female, novice and experienced teachers, EFL and EMC teachers, vocational and grammar school teachers, teaching in the capital and in the countryside). In April 2019, I conducted seven more interviews. In order to answer research question 1.6 (How can alternative assessment methods in EFL and EMC classes in public secondary education in Hungary be adapted to remote delivery?) I reached out to all the participants of the interview study and carried out follow-up interviews with all eleven of them in June and July 2020. I carried out one more interview in July 2020 as I wanted to include another science teacher who had EMC classes. The data of the twelve interviewees can be seen in the following table.

Table 1

Data of the Participants of the Interviews

Pseudonym	Age (years) in 2019	Subject(s) taught	Type of school and location
1. Anna	35	Science	Secondary general grammar school in the capital
2. Mari	38	Hungarian grammar and literature, learning methodology	Vocational secondary school in the capital
3. Lali	30	English and Hungarian as a foreign language	Secondary general grammar school in the capital
4. Eszter	30	English as a foreign language	Primary and secondary alternative school in the capital
5. Ilona	53	English as a foreign language	Primary and secondary general grammar school in the countryside
6. Júlia	49	English as a foreign language, drama in English	Secondary general grammar school in the capital
7. Bea	50	English as a foreign language	Primary and secondary general grammar school in the countryside

8. Ella	50	English as a foreign language	Vocational secondary school in the countryside
9. Csilla	45	English as a foreign language	Secondary general grammar school in the capital
10. Edit	51	English as a foreign language	Secondary general grammar school in the capital
11. Dóra	35	English and French as a foreign language, IT English	Vocational secondary school in the capital
12. Fanni	35	Biology in English	International primary and secondary school in the capital

4.2.3 Participants of the Action Research

Holliday (2007) describing research paradigms, strategies, and methods, categorizes “participatory action research” into the postmodern qualitative research paradigm, in which “researchers are part of research settings” (p. 16). I have introduced myself in section 4.2.1 The Researcher, and my participatory action research project included three groups of students I taught in the 2019/2020 school year in a public bilingual vocational secondary school in Budapest, Hungary; thus, applying convenience sampling (Cohen et al., 2007). The data of the groups can be found in the following table.

Table 2

Data of the Groups Participating in the Action Research

Group	Grade	Subject	Abbreviation	Number of students	Number of lessons in person	Number of lessons online
1	9th	IT English	9thITE	17	45	21
2	11th	Mathematics in English	11thME	16	89	44
3	11th	Elective advanced mathematics in English	11thAME	11	42	24
Total:				44	176	89

Members of the 9th grade class (9thITE) coming from different primary schools had just started their secondary school studies when the project began. At the beginning of the school year, they had written a diagnostic test to measure their language knowledge, and based on the result of the test, the class was divided into a beginner and an advanced English group. The participants of the action research were in the beginner group. As the class was in the information technology (IT) track of the vocational school, they had both general English and IT English lessons taught by different teachers. The action research took place in the IT English lessons. Thus we had beginner IT English lessons twice a week (on Mondays 11:40-12:25, which is the 5th lesson of the day, and on Fridays 8:55-9:40, which is the 2nd lesson of the day) adding up to a total of 66 lessons in the 2019/2020 school year. Initially, the group had 17 students: 1 girl and 16 boys. The girl had already been the student of the school in the previous school year as she had failed some of her subjects, she had to repeat 9th grade, which she eventually continued at a different school from December, 2019 (she was present in 13 lessons of the action research). As a result, the majority of the action research (the remaining 53 lessons) included 16 students in the 9th grade group.

I had been teaching the 11th grade mathematics in English group (11thME) using alternative assessment methods for two years at the start of the action research, so in the Results and discussion section, I will separate the results of the period before the action research and after it. The group attended the bilingual chemistry track of the vocational school. As a result, they had a preparatory language year before 9th grade and started studying mathematics, chemistry, and civilization in English (in addition to their general EFL lessons) from 9th grade. I had been teaching the 11thME group as part of a four-year-long mathematics experiment of the Content Pedagogy Research Program of the Hungarian Academy of Sciences since 9th grade, so the class was not divided into two groups based on their performance but based on their choice if they wanted to participate in the mathematics experiment or not. This resulted in a highly heterogeneous group both in terms of mathematical and language skills. The group had four mathematics in English lessons a week (on Mondays 8:55-10:35 double: 90-minute-long lesson, on Wednesdays 13:40-14:25, which is the 7th lesson of the day, and on Fridays 8:00-8:45 the first lesson of the day) adding up to 133 lessons in the 2019/2020 school year.

The 11th grade elective advanced mathematics in English group (11thAME) consisted of students who chose to take advanced level mathematics lessons in addition to their general mathematics classes. If it is possible, the school offers these elective courses separately in

Hungarian and in English for the bilingual classes; however, in this case students from all the groups volunteered to attend my lessons in English. As a result, members of the elective group came from four different mathematics groups (taught by four different teachers) in the school, resulting in the fact that six students out of the eleven had previously learned mathematics only in Hungarian. The double, 90-minute-long lessons were on Wednesdays 14:30-16:00 (after their seven other school lessons) adding up to 66 lessons in the 2019/2020 school year. Due to time management issues, one student left the group after the first half of the school year.

4.3 Research Instruments

As mentioned above, based on the main aim of the research (gaining insight into the use of alternative assessment methods in EFL and EMC classes in public secondary education in Hungary), the inquiry is of exploratory nature and thus follows the qualitative research paradigm using mainly verbal data. In addition, in order to compare students' experiences in connection with assessment in general (research question 2.1) versus alternative assessment (research question 2.2) in EFL and EMC classes in public secondary education in Hungary and also their adaptation in remote delivery (research questions 2.3 and 2.4), some quantitative data collection and analysis took place as well. For the data collecting processes, in the first and last phase of the research, interview schedules were the main instruments, and in the action research phase, a teaching journal, class observation instruments, questionnaires, and focus group interview schedules are presented in the next sections.

4.3.1 Interview Schedules

Semi-structured in-depth interviews are suggested when there is a list of topics that are common to all respondents, and the goal is to identify commonalities and differences across individual respondents on these topics, while also giving space for emerging themes (Lapan et al., 2012, p. 94). For these reasons, semi-structured in-depth interviews were conducted in the first and last phase of the study.

Although Dörnyei (2007) states that piloting the instruments is more important in quantitative studies as they "rely on the psychometric properties of the research instruments" (p. 66), Cohen et al. (2007) suggest that piloting can enhance the reliability of interviews as well. Thus, the first four interviews served the purpose of piloting the instrument. The instrument was first based on a list of preliminary questions and was later developed into an interview guide. The first version of the interview guide for teachers was subjected to expert opinion, after which the

questions were grouped, reordered, and some of them rephrased. The four interviews were transcribed, and after each one of them, following expert opinion, smaller modifications were made to the interview guide. Ensuring the credibility of the research, such “peer debriefings” made it possible to detach from the data and interact with other professionals who could act as “jury of peers and deal with whatever questions they may pose” as recommended by Guba (1981, p. 85).

Thus, the final interview guide contained an introduction, questions on background data, and the main questions with possible sub-questions in brackets (see Appendix C1). The questions asked the interviewed teachers to define assessment, particularly what traditional and alternative assessment meant for them; to list and describe in detail the types of assessment they use in class including their advantages and disadvantages; to give reasons why they introduced those forms of assessment and what their students’ reactions were; and whether they noticed any changes in their students’ behavior. The final interview guide proved to be suitable for the main study (Appendix C1).

In the follow-up interviews, participants were the same interviewees as in the main interview study. The main aim of the follow-up interviews was to gain insight to the experiences of the interviewed teachers in connection with emergency remote delivery and especially assessment in that context. As one to two years had passed between the two interviews, three out of the twelve interviewed teachers did not teach in public education anymore (namely: Anna, Eszter, and Csilla). Therefore, developing two interview schedules became necessary (see Appendix C2 and C3). Those who did not teach anymore were asked to summarize what had happened since the previous interview, when and why they had left their workplace, what their retrospective views on their assessment practices were, and whether they intended to go back to teaching and if so, under what conditions (Appendix C3). I also asked individual follow-up questions based on each participant’s first interview. Those who were teaching during the Covid-19 pandemic were asked the following different questions (see Appendix C2). They had to describe their groups who had to participate in emergency remote delivery, the advantages and disadvantages of the situation, and their assessment practices. All the assessment practices each participant mentioned in the main interview study were discussed one by one, whether they continued using them online, how they had modified them, what their goals were, what the reactions of their students were, what the advantages and disadvantages of each were, how effective they seemed to be, and if the interviewees noticed any changes in their students’ behavior.

They were also asked to detail any other assessment methods they used during the emergency remote delivery. Finally, I asked about their future plans: which assessment methods they would like to continue using and why. All the interview schedules can be found in Appendix C.

4.3.2 Instruments of the Action Research

Based on the results of the interview study and my former experiences, I designed the second phase of the research: the action research project. As Guba (1981) describes the neutrality aspect of trustworthiness, instead of objectivity, “naturalists shift the burden of neutrality from the investigator to the data, requiring evidence not of the certifiability of the investigator or his or her methods but of the confirmability of the data produced” (pp. 81-82), which entails that researchers “use themselves as instruments” (p. 81).

In order to ensure this confirmability, Guba (1981) suggests triangulation, practicing self-reflexivity, and arranging for confirmability through an audit trail. Triangulation, which also ensures credibility (Guba, 1981), means “collecting data from a variety of perspectives, using a variety of methods, and drawing upon a variety of sources” (p. 87). For this reason, the action research phase was documented by a teaching journal including field notes, lesson plans, and self-reflections. In addition, for investigator triangulation (Maxwell, 1992) observers’ reports, proformas (Burns, 1999), class recordings, discussions, student questionnaires, and focus group interviews were further instruments applied. These represent the variety of perspectives: mine as researcher and participant, my student participants, and different observers. The variety of methods and sources are also presented in the following sections. According to Guba (1981), “one indispensable technique in support of practicing reflexivity is to keep a continuing journal in which introspections are recorded on a daily basis” (p. 87), which is detailed in the next section. Arranging for confirmability through an audit trail means “certifying that data exist in support of every interpretation and that the interpretations have been made in ways consistent with the available data” (Guba, 1981, p. 88), will be presented in the Results and Discussion chapter.

4.3.2.1 The Teaching Journal. As Burns (2015) puts it, a “classic” instrument of action research is the “researcher journal or diary,” that is, a self-reflective tool used for different purposes such as using “grids or tables with relevant headings, for instance dates/times, issues arising, actions taken, changes made, reflections, comments, reactions, literature references” (p. 196). I also created a template for my teaching journal that I could quickly fill in after each lesson. It contained the date, class, number of the lesson, absent students, and the description of the lesson.

The primary reason for including these pieces of information was that these had to be registered in the national compulsory e-gradebook system (e-KRÉTA), too. For the description of the lesson initially I had three parts: I included my lesson plan (before the lesson), my field notes of what happened in class (during the lesson), and my reflections (after the lesson).

At the beginning of the action research project, I was not sure what pieces of information would be useful for my research purposes, so when describing each lesson, the amount of happenings in class that could be analyzed seemed overwhelming. Moreover, it soon became apparent that I had to focus my reflections and find a sustainable structure that I could maintain throughout the whole year. As a result, during the year, the description of the lessons organically changed: it became a bullet-point list of what happened during a lesson including my reflections (merging the three: before-during-after the lesson parts) and ended with a new section: “tips” for myself. This structure was sustainable, and the tips became the basis of the “cycles of (i) planning, (ii) action, (iii) observation and (iv) reflection” (Burns, 2015, p. 189) of the action research resulting in a focus on specific goals for each lesson and valuable sources of data for the research. A sample screenshot can be found in Appendix D1.

4.3.2.2 Class Observations. As Larsen-Freeman and Long (2014) categorize class observations: participant observations (carried out by me as the teacher) are documented in the teaching journal and non-participant observations (carried out by others who observe activities without engaging in them directly) are based on different observation criteria. Burns (2015) emphasizes the importance of different kinds of observations for data triangulation in action research to verify the researcher’s own observations. During the action research, nine instances of class observations took place by seven individuals and two groups of teacher trainees. These class observations had several different roles, and for this reason, the observation instruments also varied, which will be detailed in the following paragraphs.

In the case of an EFL teacher colleague from the same school, the goal of the observation was to create a report for the leadership of the school, so he used the observation template and criteria required by the school. In addition to the general data of the lesson, the template contained the following observation criteria: lesson plan and execution, assessment, revision and practice, and student activity and discipline. In the case of an EMC teacher colleague from the same school, I asked him to use a proforma (Burns, 1999). It also contained the general data of the lesson and the following three columns: time, happenings, remarks (see Appendix D2). The aim of the

proforma was to differentiate between the observed happenings and their interpretations. As both my colleagues were teaching the same students, they could also share insights about student behavior in their lessons compared to mine. In the case of teacher trainees, they all had their own observation criteria required by the teacher training course they attended. Some of their tasks were to describe the circumstances and atmosphere of the lesson, compare the lesson to ones in the USA (in the case of American teacher trainees), compare the lesson to other observed lessons in Hungary, note ideas that they can use as future teachers, mention points for improvement, analyze how instruction is accommodated to online settings (in the case of online lessons), and list the online techniques used.

As emergency remote delivery became necessary during the action research, online class observations had to be organized. They happened in the following two ways. Observers either joined the live lesson, or they watched the recording of the online lesson. In both cases they had the chance to discuss the observed lessons afterwards in a separate online meeting both with me and with students who volunteered to answer the observers' questions, which were also recorded. These discussions served as member checking (Dörnyei, 2007) both in person and in emergency remote delivery. Observers also included these reflections in their observation reports. Further ensuring the credibility of the research, these recorded online lessons and discussions acted as "referential adequacy materials" as Guba (1981) explains: "if the inquiry deals with the behavior of classroom teachers, videotapes of actual classrooms can be made and stored. Later, when it is asserted that teachers exhibit such and such behavior, that assertion can be tested by reference to the archives" (p. 85).

In addition to recorded lessons, the online part of the action research resulted in vast amounts of digital data. As the school started using Microsoft Teams, the following features of the virtual classroom provided data: the Calendar included the data of the online lessons held; students' work was either submitted through Assignments, or they worked in the Class Notebook; and class materials were created in the forms of short video recordings, pictures, and links shared in Posts in addition to general discussions (a sample can be found in Appendix D3). The online classroom was also suitable for recording students' results and provided different options for assessment. Student questionnaires contained detailed questions about the use of these features.

4.3.2.3 Student Questionnaires. My MA thesis paper, entitled Possible uses of the assessment system of gamification in ELT (Barbarics, 2014), contains a quasi-experiment of

introducing an alternative assessment method based on gamification. Since then, I have been experimenting with alternative assessment methods and collecting data from field notes and student questionnaires. For this reason, the basis of the student questionnaires used in this study was the questionnaire I designed for my MA thesis in 2014 and used until 2017 with all the groups I taught in that period. In 2014, I consulted an expert and piloted the questionnaire with one person from the same age group as the participants (Barbarics, 2014). For predictive validity (Cohen et al., 2007, p. 140) I used the same questions in future questionnaires as well. However, for the action research, I considered the 2014-2017 period as a pilot study, and revised the questionnaire according to the aims of the action research and quality control. In Appendix E, I marked in blue the parts I kept the same and in red that I changed for different reasons. In the following paragraphs, I elaborate on the instruments and these changes based on the groups participating in the action research.

In the case of the 9th grade IT English (9thITE) group, whom I started teaching at the beginning of the action research, I designed an entry questionnaire (see Appendix E1). I named it “entry” questionnaire in order to differentiate it from the numbered questionnaires that students filled in after each unit. Compared to the questionnaire I used between 2014 and 2017, I made the following changes. I kept the question if they play computer games and if so, what kinds. I added, however, a question concerning the language of the games they play. I also repeated the task to write down the first three words or expressions that come to mind when they hear “school assessment”. In case of questions containing Likert scales, I changed the six-point scales (0 to 5 or 1 to 6) to five-point scales (1 to 5) as suggested by Cohen et al. (2007). I also modified the question “How much do you like English?” because the piloting showed that it was not “exhaustive and discrete” (Cohen et al., 2007, p. 328) and might be affected by the following questions, so I created three different questions: “In your opinion, how important is it to learn English? (1= not important, 5 = very important)”; “How much do you like the English language? (1 = I hate it, 5 = I love it)”; and in the upcoming questionnaires “Based on your experiences so far, how much do you like the IT English lessons? (1 = I hate it, 5 = it’s one of my favourite subjects)”. As it was a beginner group, I also added a question about how they had learned English before (they had to check all that applied): primary school lessons, private teacher, by themselves (e.g., from movies, games, etc.), other:... or they could also check if they had not learned English before. In connection with their previous studies, I asked about their foreign language lessons indicating that foreign

language refers to English if they had English lessons in 8th grade, but if they did not, they had to answer in connection with the foreign language they learned (Appendix E1).

In connection with grades, I asked the same set of questions in all the questionnaires: what grade they have just got (at the end of 8th grade and at the end of each unit during the action research), how well that grade reflects their knowledge and how well that grade reflects the energy they put in studying (from 1 = “it doesn’t reflect it at all” to 5 = “it reflects it perfectly), how satisfied they are with the ways the teacher had assessed their work (from 1 = “I am not satisfied at all” to 5 = “I am absolutely satisfied”), and they had to give their reasons, too, because as Cohen, et al. (2007) state that although “rating scales are more sensitive instruments than dichotomous scales [...], they are limited in their usefulness to researchers by their fixity of response caused by the need to select from a given choice” (p. 328), so they suggest adding open-ended questions that allow respondents to detail or explain their answers. In addition to the questionnaires I used between 2014 and 2017, I asked about the students’ satisfaction at the end of 8th grade in connection with other subjects as well: mathematics, Hungarian grammar and literature, history, and a subject chosen by them. The entry questionnaire ended with the same two open-ended questions: if they had the chance, how they would change assessment in school, and they were encouraged to detail any answers or share an example or story of assessment in school if they wanted to (Appendix E1).

As the 11th grade mathematics in English (11thME) group filled in the entry questionnaire in 2016, they did not have a new entry questionnaire. During the action research, all of the groups had to fill in a questionnaire after each topic, ending with a grade that they received based on alternative assessment methods (Appendices E2, E5, E7, E9). For predictive validity (Cohen et al., 2007, p. 140), these questionnaires had the following recurring questions: how much they like English/mathematics (in general), how much they like our lessons, if they would like to continue learning using the same point system in the next period (yes, no, other:...), their reasons for their previous answer, what they would change in this point system, and the above mentioned questions about grades (what grade they have just got, how satisfied they are with the way their work has been assessed, why, how well this grade reflects their knowledge and the energy they put in studying, and what they would change in connection with assessment). Furthermore, they had to say if they had been graded in the traditional way, the grade would have been: better / worse / the same / they can’t decide whether the grade would have been better or worse / other:... Introducing

self-assessment in the questionnaires, I also asked their opinion on their own performance: what went well, not so well, and what they would like to do to make it better. The last question was an opportunity to detail any of their answers or share anything else that they liked or did not like during the given period (Appendices E5, E7, E9).

In addition to the recurring questions, each questionnaire asked students' opinion on specific issues regarding the given period. Based on the interview study, I tried out 32 different methods connected to assessment in group 9thITE (Appendix E5). Whenever one appeared, the questionnaire asked if in the next period they would like to have it with answer options: yes, no, other... , and they also had to give their reasons. In connection with the midterm and year-end evaluation, I repeated the questions in connection with grades: what their midterm/year-end grade was from IT English/mathematics, how satisfied they were with the ways I assessed their work, why, how well this mark reflects their knowledge and the energy they put in studying the subject, and what changes they recommend for the next period in connection with assessment, or the lessons, or the tasks, and so on (Appendices E6, E8). I also used the questionnaire to introduce students' self-assessment and conscious planning of their studies in the next learning period (Appendices E6, E8). The last two questionnaires were administered during the pandemic, so they also asked for feedback in connection with students' experiences with online instruction, such as, how much they liked their online lessons and why, how they prefer working in this new environment, and how I can help them (Appendices E6, E8, E9). All in all, there were seven questionnaires administered in group 9thITE (one entry questionnaire, and six after each topic grade).

In group 11thME, I did not try out that many new methods as I had been teaching them for three years already with long established working modes. They were also used to filling in my questionnaires, and they were more willing to answer open-ended questions. For these reasons, in addition to the above mentioned recurring questions (about the lessons, grades, etc.), I asked what their opinions were about eight assessment related issues, and several questions aimed to develop effective teaching and learning methods during the pandemic (Appendices E7, E8). All in all, there were six questionnaires administered in group 11thME (one after each topic grade).

In the 11th grade elective advanced mathematics in English (11thAME) group there were three questionnaires administered (Appendix E9). Although they received four grades, at the beginning of the semester it was not clarified by the leadership of the school whether they should

receive any grades, so for this reason, I merged the first questionnaires into one when it became clear that they could also be asked about grading. The first questionnaire thus contained elements of the entry questionnaire (such as their experiences with computer games and their associations about “school assessment”), the recurring questions about grades and alternative assessment (what their last grade was, how satisfied they were with the assessment, why, how well their grade reflected their knowledge and the energy they put in studying the subject, what they thought of our assessment system, if they wanted to continue learning in this point system, why, and what they would change in this point system). In each of the questionnaires, their opinions of the different assessment methods were also asked, as well as self-assessment and pandemic related questions. In the last questionnaire I asked what they liked the most and least in general and also in connection with the online lessons (Appendix E9).

A summary table of the information about the groups and the questionnaires can be found in Appendix E10 (which group filled in which questionnaire when). In addition to the regular questionnaires, data on students’ perspectives was collected through focus group interviews as well.

4.3.2.4 Focus Group Interviews. The last research instrument during the action research was focus group interviews. According to Kamberelis and Dimitriadis (2005) “focus groups facilitate the exploration of collective memories and shared stocks of knowledge that might seem trivial and unimportant to individuals [...] and help researchers to work against premature consolidation of their understanding and explanation” (p. 903). They also note that focus group interviews are practical as they generate a lot of data from a relatively large number of participants in a relatively short time. “In addition, because of their synergic potentials focus groups often produce data that are seldom produced through individual interviewing and observation and that result in especially powerful interpretive insights” (Kamberelis & Dimitriadis, 2005, p. 903). Answering the last two research questions, I was interested in students’ collective experiences with assessment in emergency remote delivery in general (research question 2.3) and their perceptions of the adaptation of alternative assessment methods in our EFL and EMC classes (research question 2.4).

For these reasons, during the focus group interviews students were asked the following questions: 1. What did you like in connection with emergency remote delivery? And specifically about our lessons? 2. What didn’t you like in connection with emergency remote delivery? And

specifically about our lessons? (I added the following prompts in case students needed them: What were the most memorable positive / negative happenings?) 3. What do you think are the advantages of our assessment system? 4. What do you think are the disadvantages of our assessment system? I was also interested in students' self awareness, so if we had time at the end of the focus group interview, I intended to ask the following additional questions: What do you think would help you in learning (from me / from a teacher in general)? What motivates you (e.g., in other subjects)?

The first two questions did not focus on assessment as I was interested in what topic areas emerge in connection with emergency remote delivery in general, and if assessment was among them. As Dörnyei (2007) puts it, the “strength of this format is the discussion that emerges about a broad topic” (p. 129).

The variety of instruments used during the action research project aims to ensure dependability, the consistency aspect of trustworthiness, through overlap methods (Guba, 1981), “whereby different methods are used in tandem [...] in such a way that the weakness of one is compensated by the strengths of another” (p. 86). It is one form of triangulation, while triangulation aiming to collect data from a variety of perspectives is ensured by including the perspective of the researcher, that of other teachers, and the students “in the hope these actions will lead to credibility” (Guba, 1981, p. 83). The following sections elaborate on these procedures of data collection followed by the methods of data analysis.

4.4 Procedures

This chapter presents the procedures of how the actual data collection happened and the quality control measures that were taken. These sections follow a chronological order, and as the student questionnaires have a history reaching back to 2014, I start with them. Then the pilot interviews and the main interview study follow, after which the action research (first in class, then in emergency remote delivery) comes, ending with the follow-up interview study.

4.4.1 Student Questionnaires Between 2014 and 2017

In 2014, I started working on my Language Pedagogy MA thesis entitled Possible uses of the assessment system of gamification in ELT (Barbarics, 2014). As I started my teaching practice, in the same bilingual secondary school where I conducted my action research, I designed a questionnaire to inquire about students' opinions on school assessment in general and then the gamification based assessment I introduced. In all of the new classes I taught, I conducted the same experiment. Each time I started teaching a new group, I asked students' opinion on school

assessment (before they had known anything about the new assessment system I was about to introduce), then they got acquainted with the assessment system based on gamification and could tell their opinions in the form of anonymous questionnaires as well as face to face discussions (Barbarics, 2015, 2016, 2017). At the start of my PhD studies in 2017, I had asked the opinion of 133 students, among whom I taught English (IT English or English as a foreign language) to 49 students and mathematics in English to 84 students (EMC classes). The 133 students filled in one to six questionnaires depending on how long I taught them. As described above, it can be considered a pilot study for the questionnaires used during the action research. As Dörnyei (2007) notes “qualitative piloting differs from quantitative piloting in that that after these ‘trial runs’ we do not have to discard the obtained data but can use them for the final analysis” (p. 67), so most of the data obtained in this period can be used for the final analysis. In addition, this “prolonged engagement” can also enhance the credibility, the truth value aspect of trustworthiness, of the research (Guba, 1981, p. 83).

In terms of administering the questionnaires, all of them are in Hungarian as it is the mother tongue of the participants, and as Cohen et al. (2007) highlight, for “cultural validity” (p. 139), language issues must be considered and the researcher should make sure that participants clearly understand the questions. As my mother tongue is also Hungarian and I grew up in the Hungarian public education system, creating the questionnaires in Hungarian was not a problem; however, translation issues did arise in connection with students’ responses containing terminology specific to the Hungarian public school system. This is one of the reasons for creating the glossary in Appendix A. Students always filled in the questionnaires during their regular lessons, which made it possible that they could ask if something was not clear (which proved to be a great source of information for piloting the questionnaires), and they had ample time for taking it seriously and answering the questions at length. In the beginning, questionnaires were printed, and I digitized the data afterwards. Later, students filled in the questionnaires in Google Forms, which was beneficial for two reasons. First, I did not have to digitize data as it could be exported into an excel sheet. Second, students’ anonymity could be preserved as opposed to handwritten answers because I could recognize their handwriting after a while. The only complications were that although I emphasized that they should remember the pseudonym they chose for themselves, some of them either did not remember it or misunderstood the task and wrote new pseudonyms in subsequent questionnaires, which reduced the sample size of those whom I could follow in a longitudinal

manner meaning that I could not look at the same student's answers in several questionnaires. Although I had been collecting a lot of data on the students' views, I was also interested in teachers' points of view. The next section describes the interview study with teachers between 2018 and 2019.

4.4.2 Interview Study in 2018 and 2019

As mentioned in the Participants chapter, four pilot interviews took place between April and June 2018; when I reached out to teachers who were known for using alternative assessment methods. In order to get diverse feedback, I chose participants teaching different subjects in different secondary schools in Hungary. The interviews were carried out in Hungarian as it is the mother tongue of each participant. After the consent of the participants, the interviews were recorded, transcribed, and I consulted expert opinion after each.

After analyzing the results of the pilot interviews, using the final interview guide (see Appendix C1) seven more interviews happened in April 2019 and one more in 2020. All of them were administered in Hungarian. The lengths of the interviews were between 34 and 87 minutes (62 minutes on average), and the lengths of the transcripts were between 2379 and 8659 words (6265 words on average, so it is interesting to note that the average rate was about 100 words a minute). One interview happened online as the interviewee lived further away from the capital. All the other interviews took place at home or in cafés in Budapest (when other interviewees from the countryside happened to be there). There were no complications or unexpected events in data collection.

In terms of quality control, there can be several threats to the validity of qualitative research. Dörnyei (2007, pp. 46–52) lists three of them: insipid data, the quality of the researcher, and anecdotalism; then Dörnyei (2007) details further categorizations of quality control in the qualitative research paradigm and strategies to ensure them. Insipid data in this context was soon refuted as in the pilot interviews, the four participants listed 119 ways of alternative assessment that they claimed to use (each one of them between 25 and 35 ways). In the twelve interviews, 217 assessment methods emerged. Interview by interview, the assessment methods described kept repeating themselves and fewer and fewer new ones were added. Other emerging themes also contained quality data answering the research questions, which proved the adequacy of the interview guide as well. In addition, to ensure researcher integrity, I intend to leave a detailed audit trail with contextualization and thick descriptions of the data. To ensure the interpretive validity

of the interview data (Maxwell, 1992), the follow-up interviews, in addition to the newly emerged research questions about remote delivery, acted as validation interviews to member check my interpretations with the interviewees (Dörnyei, 2007, p. 49). In connection with anecdotalism, it is important to be aware of researcher bias. According to Guba (1981), it should be avoided by triangulation, practicing reflexivity, and leaving a detailed audit trail to ensure confirmability, the neutrality aspect of trustworthiness. Considering transferability, the applicability aspect of trustworthiness (Guba, 1981), I interviewed teachers in public secondary schools in order to analyze how the different methods could be used by other teachers in other public secondary schools in Hungary.

4.4.3 Action Research in the School Year of 2019 and 2020

Before the start of the school year, planning the action research project was guided by data from two main sources. On the one hand, based on the results of the interview study, I selected alternative assessment methods to try out and emerged themes to pay attention to. On the other hand, based on the literature, I planned instruments for data collection. In addition, specific tasks from Smith and Rebolledo's (2018) handbook for exploratory action research made it possible to formulate aims, analyze practices, find sources of information, and keep track of all of them.

The action research was conducted in the 2019/2020 school year with the participants and instruments described in sections 4.2 and 4.3, respectively. In terms of administering the instruments, the teaching journal (see Appendix D1) was filled out after each of the 265 lessons (89 of which were online). The unexpected introduction of emergency remote delivery meant an unforeseen interruption and made me change the original plan. However, I decided not to end the research, but expand it with further research questions about adaptation possibilities to remote delivery. For this reason, I continued the documentation of the action research resulting in the detailed description of the transition to remote delivery. The teaching journal contains day-to-day accounts from my point of view, the questionnaires and focus group interviews show the participating students' perspective, and the follow-up interviews present other teachers' experiences of the emergency remote delivery manifesting investigator triangulation as suggested by Maxwell (1992).

In connection with the trustworthiness of action research, Burns (2015) recommends the following measures to enhance it: different types of triangulations, "member-checks", "perspectives comparison", "cyclical iteration", "testing through practical application",

“compatibility with educational aims and democratic values”, and “compatibility with teaching demands” (pp. 193-194). Triangulation also happened through the use of various data-collection methods. As described in the previous sections, in addition to the teaching journal, questionnaires, and interviews; class observations, recordings, and digital data recorded in the online classroom (Microsoft Teams) were also included.

During the action research (including both the in class and the emergency remote delivery phase in chronological order), the following class observations took place (Table 3). The abbreviation for the classes are the following: 9thITE stands for the 9th grade IT English group and 11thME stands for the 11th grade mathematics in English group.

Table 3

Class Observations of the Action Research

Observer(s)	Observed class	Date of the observed lesson(s)
1. EFL teacher colleague from the same school	11thME	September 28th
2. EMC teacher colleague from the same school	11thME	November 11th
3. EFL teacher trainee from ELTE	9thITE	January 27th and 31st
4. Math teacher trainee group from the USA	11thME	February 21st
5. Math teacher trainee from the USA	9thITE	February 21st
6. Mathematician interested in teaching	11thME	February 24th
7. Math teacher trainee group from the USA	11thME - online	May 4th
8. Math teacher trainee from ELTE	11thME - online	June 5th

The class observations were always discussed with all participants in advance. I asked for the observers’ consent for their reports to be used in this dissertation, and students also knew about the observations and gave their consent, which was especially important when they had to perform (such as in the 9thITE beginner group, when students gave short presentations on February 21st while a native English speaker was observing). These were also instances of triangulation when

my field notes in the teaching journal, observers' reports, and students' experiences of the same events could be compared. As observations were followed by informal discussions with the observers, it provided the opportunity for observers to ask me and the students questions. In addition, after receiving the observers' reports, I could also apply member checking.

These are examples of triangulations and member-checks. In connection with “perspectives comparison” (Burns, 2015, p. 193), the follow-up interviews make it possible to compare my experiences during the action research to that of other colleagues'. Burns (2015) also mentions looking at “rival explanations that show alternative data explanations, or negative cases” (p. 193), which will be presented in the Results and Discussion chapter. During the planning phase of the action research I was not sure how long it would last, and the reason why I continued it for the whole school year was that I could carry out “cyclical iterations” in which I could compare with and test the trustworthiness of my findings and interpretations “against previous iterations of the [action research] cycle in order to build on previous evidence, expand the scope, purpose and central questions of the study, further triangulate the data and guard against researcher bias” (Burns, 2015, p. 193). Expanding the scope and purpose was especially true when previous findings and interpretations of the action research could be tested in a new, online environment during the emergency remote delivery phase. One of the aims of action research is to “generate improvement in practice” and thus quality control also means “testing through practical application” to see whether the applied intervention strategies actually produce improvement in practice (Burns, 2015, p. 193). Proof of that is the introspective accounts documented in the teaching journal presented in chapter 5.4, *Alternative Assessment Methods and Their Consequences* (together with the findings of the interview study). The last two quality criteria mentioned by Burns (2015): “compatibility with educational aims and democratic values” and “compatibility with teaching demands” are important because “action research is deliberately interventionist aimed at disturbing and unsettling the status quo”, so considering the educational context in which it happens is crucial. This is one of the reasons why I defined alternative assessment in relation to current Hungarian rules and regulations of assessment and why the context of my research is public secondary schools in Hungary. These compatibility issues will be further discussed in the Results and Discussion chapters.

The procedures of the use of two more instruments of the action research need to be discussed: student questionnaires and focus group interviews. Administering student

questionnaires during the action research happened similarly to the ones collected between 2014 and 2017: in Hungarian, during their lessons, in Google Forms (and with some students still forgetting their chosen pseudonyms). There was a questionnaire after each grade registered in the e-gradebook (which meant six grades each for 9thITE and 11thME, and three for the elective 11thAME group as this is the compulsory annual number of grades that the respective classes should get). Two of the grades had to be administered during the emergency remote period. In the Literature review, the section describing Emergency remote delivery in Hungary shows how problematic this was in general, so the action research presents an alternative view on it both from the teacher's perspective and students' reactions through the questionnaires and focus group interviews. It is worth mentioning that when questionnaires were administered in class (in person), 78% of students filled in the questionnaires on average, while during the emergency remote delivery, it dropped to 59%, so focus group interviews were also necessary to ensure that students who did not fill in the questionnaires could also express their opinion.

The last data collection phase for the action research was the focus group interviews. It was also discussed with students that their last double (90-minute) lessons were dedicated to a recorded discussion about their experiences. (As according to the timetable, I did not have a 90-minute-long lesson with group 9thITE, I had to switch a lesson with a colleague to have enough time for the focus group interview.) At the end of the school year, it was still not allowed to meet in person, so the focus group interviews were organized online. The advantage of the online interviews was that they could be recorded easily in perfect quality, while the disadvantage was that they could not have the natural flow of in person discussions. One of the greatest challenges of online group discussions is to determine who speaks when, and if there is no moderator, it can cause a lot of confusion and take a lot of time. In focus group interviews, the researcher acts as a moderator, and as Dörnyei (2007) explains it involves tasks such as “making sure that nobody dominates the floor and even the shier participants have a chance to express their views. In addition, moderators need to prevent any dominating and inhibiting group opinion [...] from emerging by actively encouraging group members to think critically” (p. 129). I believe that our practice of online lessons including group discussions during the months of emergency remote delivery not only made it possible, but made it even more effective to carry out focus group interviews this way. We could maintain our friendly atmosphere. I asked students one by one calling their names to tell their opinions, and I was surprised how otherwise shy students elaborated on their experiences at

length. The interviews were conducted in Hungarian and were recorded and later transcribed. In the 11th grade groups everybody was present. In the 9thITE group, there were 11 students present, so I organized another date for the missing six students, but only three participated (so all in all, there were only three students out of 44 who did not participate, making it a 93% attendance rate). The advantage of the interview with three students was that I could record their opinions, too, and from a methodological point of view, I could compare focus group interviews with different numbers of participants.

4.4.4 Follow-up Interview Study in 2020

The emergency remote delivery during the action research created a unique situation that resulted in further research questions on students' experiences and perceptions of the adaptations of alternative assessment methods to remote delivery. I documented my experiences and that of my students through the above mentioned processes; however, I was also interested in the experiences of my interviewees. As the school year ended, I organized follow-up interviews with all of them (in June and July 2020). In addition to member checking the interpretations of the main interview study, the follow-up interviews provided prolonged engagement as well (Dörnyei, 2007, p. 52). They enabled me to inquire about the use of specific alternative assessment methods in the perspective of years, and also their adaptation possibilities in remote delivery.

During the summer it was possible to meet in person; however, eight interviewees chose to have the interviews online and only four in person. Similarly to the main interview study, all of the interviews were in Hungarian; they were recorded and transcribed afterwards. The fact that two-thirds of the interviews were recorded online in a much better quality, made the transcription process easier. There were no complications or unexpected events in data collection.

4.5 Methods of Data Analysis

For the interview studies, I recorded, transcribed, and analyzed the data using the constant comparative method first described by Glaser and Strauss (1967). Initial coding happened parallel with transcribing the interviews to find emerging themes and see if the questions elicited sufficiently rich data. The constant comparative method was used to compare data with codes, and codes with other codes, so that it would result in sorting of the initial codes into more elaborate ones (Lapan et al., 2012). For the pilot interviews, data were copied to the atlas.ti program for focused coding. Categories that summarized groups of open codes were created. These categories linked with the initial research questions of the pilot interviews are presented in Table 4.

Table 4*Research Questions Linked With Emerged Categories From the Pilot Interviews*

Research questions of the pilot interview study	Emerged categories	Number of labels
1.) How do teachers define assessment and alternative assessment?	definitions of assessment (in general, traditional, and alternative)	63
2.) What alternative assessment methods do teachers claim to use in public secondary education in Hungary?	forms of assessment (specific examples, practices)	119
3.) Why do teachers use alternative assessment methods in public secondary education in Hungary?	motivations and purposes (why they do something as they do it)	131
4.) What are the views of teachers using alternative methods on assessment?	beliefs about assessment	348
	consequences (what results they see due to using alternative assessment methods)	194
	teacher characteristics (personality or behavioral traits that could affect their views on assessment)	302

There are several advantages of coding in a program. It is easy to label words, expressions, or even sentences with a code, which then can be compared to other codes and categorized accordingly. For example, in the pilot interviews one category was “definitions of assessment” with 63 labels (Table 4), which was further divided into smaller categories of definitions of “assessment in general”, “traditional assessment”, and “alternative assessment”. The program can display all quotations coded into a category, and the researcher can quickly switch between codes, search for words, create new categories, reorder or group labels, and add comments. The program can also count the labels that belong to a category (Table 4). In addition, the program codes data to its source by adding ID numbers to the quotations. Figure 3 shows a screenshot of some quotations for definitions of alternative assessment in the atlas.ti program with their ID numbers: the first number represents the interview, the second number shows the line of the transcript. Examples of the second category, forms of assessment, can be found in Figure H in Appendix H.

However, this program could only be used to analyze the pilot interviews, so further data analysis procedures are detailed in the next paragraphs.

Figure 3

Screenshot of Some Quotations for Definitions of Alternative Assessment in the Atlas.ti Program

Show quotations coded with <i>alternatív é. def.</i>		
ID		Name
2:11		Ha szűken veszem, akkor minden, ami nem jegyadás
2:12		a jegyadást is lehet alternatívan csinálni
3:7		minden, ami más
3:8		ha az valamiben többet vagy mást nyújt, mint a hagyományos értékelés
5:9		kiveszik az egymáshoz való viszonyítás

As the free version of the atlas.ti program allowed only a limited number of labels, and there were no financial resources for the paid version, interviews 5-11 had to be analyzed in a different way. Interviews were transcribed into separate google documents where open coding happened by highlighting words, phrases, sentences, and adding the codes and notes in the form of comments. Then codes of the pilot interviews were exported into an excel sheet, and the new open codes were either categorized into one of the already existing categories, or new categories were created. The already existing categories of the pilot interviews can be seen in Table 4 (definitions, forms of assessment, motivations and purposes, beliefs, consequences, and teacher characteristics). In the new categories, I divided the forms of alternative assessment into oral, written, numerical, and further forms, and placed them on a continuum from simple to complex. This categorization can be seen in my earlier publication on innovative ways in assessment (Barbarics, 2019b). The category of “motivations and purposes” was renamed “teachers’ goals for using alternative assessment methods”, and “teacher characteristics” was split into “personal characteristics”, “assessment related views”, and “struggles”. The new categories were also checked in the pilot interviews and vice versa. Due to the iterative nature of qualitative research, there was a constant movement between data collection, analysis, and interpretation. The analysis of the interview study had to be finalized, on the one hand, due to a publication (Barbarics, 2019b), on the other hand, due to the beginning of the action research in the 2019/2020 school year as I planned the action research project based on the results of the interview study.

Before detailing the analysis of the action research, the follow-up interviews were analyzed similarly to the main interview study using the constant comparative method (Dörnyei, 2007).

However, as it happened after the online teaching period, I did not want to use any digital tools, so I printed out the transcripts of the follow-up interviews (142 pages) and applied color coding by using eight different colors for highlighting. I added pencil notes at the end of the lines where information that I could not categorize into the existing labels emerged. So as advised by Flick (2009), I merged the different coding approaches in order to proceed with the analysis based on my research questions. First, I looked for the categories that emerged from the main interview study. New definitions of assessment did not appear in the follow-up interviews. Exact alternative assessment methods were highlighted in gray. Teachers' goals for using alternative assessment methods were blue. Teachers' experiences of using alternative assessment methods were green. In addition, there was a miscellaneous category in yellow, which included beliefs, personal characteristics, views, and struggles that I knew I had to further analyze in the next round of coding. However, I was more interested in the aspects of the emergency remote delivery, so I continued with open coding in relation to questions about the pandemic. Four main categories emerged: advantages of the emergency remote delivery that were highlighted in orange, disadvantages were red, how the interviewed teachers carried out assessment was pink, and their experiences with others (assessment by their colleagues, issues with the system, families etc.) became purple. I labeled the remaining pieces of data, which resulted in further miscellaneous categories. I followed the iterative nature of comparing data with new codes, and codes with other codes in order to sort them into more elaborate ones (Lapan et al., 2012).

As most of the analysis happened in Hungarian, after the analogue work with highlighters, pencils, and sticky notes on paper in Hungarian, I typed the codes into a word document and added English labels with the page numbers of the printed transcripts for the quotations to prepare for the writing phase. Figure 4 shows the screenshot of the colored headings in the document with the main categories, under which further subcategories, labels, and page numbers are hidden. In addition, Figure 5 shows an excerpt when the labels, short versions of the quotations, and the page numbers of the printed transcript are also shown. All in all, these notes resulted in a 35-page-long document. This document provided the basis for the writing phase.

Figure 4

Screenshot of the Emerged Categories of the Follow-up Interviews

Methods	Views / beliefs / characteristics
<ul style="list-style-type: none"> ▶ How did they give grades? ▶ Alternative assessment methods ▶ Assessed products, activities ▶ Digital tools, platforms 	<ul style="list-style-type: none"> ▶ ● Teaching methodology ▶ ● Beliefs ▶ ● Tasks ▶ ● Materials ▶ ● Conflict management ▶ ● Cheating ▶ ● Teacher sets an example ▶ ● Teachers' self-reflection / criticism ▶ ● What they say about themselves ▶ ● What I noticed about them ▶ ● Assessment-related <ul style="list-style-type: none"> ▶ ○ Feedback (from the T to the Sts) ▶ ○ Value -> points reflect: ▶ ● Problems with the system
<ul style="list-style-type: none"> ▶ Emergency remote delivery solutions 	
<ul style="list-style-type: none"> ▶ Goals ▶ Subject/school/learning related ▶ Online learning related ▶ Assessment related ▶ "Life" related 	<ul style="list-style-type: none"> ▶ CLIL or rather EMC ▶ Why did they leave?
<ul style="list-style-type: none"> ▶ Experiences 	
<ul style="list-style-type: none"> ▶ Advantages (of remote delivery) 	
<ul style="list-style-type: none"> ▶ Disadvantages (of remote delivery) 	
<ul style="list-style-type: none"> ▶ Assessment by others / issues with the system 	

Figure 5

Screenshot of the Emerged Categories Including Labels, Quotations and Page Numbers

Views / beliefs / characteristics

- Teaching methodology
 - minimizing frontal teaching p9, 29
 - learning by doing p11
 - Learn effectively -> through experience p11
 - Experiment p11
 - Project week pp 57-62
 - No leaderboards (Sts only see their own results) p28, only team leaderboards p32,
 - Differentiation:
 - heterogeneous groups -> point is to develop compared to themselves, can't (shouldn't) compare Sts to each other (especially in language) p28, 110...
 - develop the skills they need p33
 - cater for special needs Sts (with learning difficulties) p89
 - different teachers start doing things differently with the different groups p110
 - student with selective mutism... p111
 - students with different types of autism, dyslexia, other learning difficulties + levels of English knowledge -> all different goals p114
 - T recommends a schedule (what to do when), but there is deadline and Sts have the freedom/flexibility within it p42, 51, 53
 - criticism about testing online: short time for tests, disadvantages for those who are not that good in reading etc. p73

During and after the action research, analyzing the teacher's perspective through self-reflections, field notes, and observers' reports were also subject to content analysis using the constant comparative method (Dörnyei, 2007). Action research has even more of an iterative nature; thus, constant movement between data collection, analysis, and interpretation characterized this phase of the research. Based on Kemmis and McTaggart's (1982) model of action research, I followed the steps of focusing, planning, acting, observing, reflecting, revising, and refocusing, while maintaining a thick description of the data. As I was not sure which data would be useful for my dissertation, during the time of the action research, I focused on enhancing the teaching and learning processes and analyzed the data accordingly. As described in section 4.3.2.1, The Teaching Journal, I highlighted tips for myself that I learned from a lesson and wanted to focus on the next one. It became such a useful tool for my teaching practice that I have kept this habit even after the end of the action research.

In addition to the teacher's point of view, when analyzing students' perceptions for data triangulation during the action research (Burns, 2015, p. 195), I used questionnaires, personal discussions, and focus group interviews. Analyzing the focus group interviews, I also applied the constant comparative method, but I applied a third variation of the process. I divided the transcript of the focus group interviews into two parts (henceforth halves; however, they were not equal in length). For personal reasons, I could analyze the focus group interviews only a year later, so using this time as an advantage, in the first half of the transcript, I started open coding without any preconceptions as suggested by Glaser and Strauss (1967). As I was not working with the data for a year, I felt that I had enough distance from the data to see what themes emerge. However, I was also interested in comparing students' experiences of the emergency remote delivery to teachers', and for this reason in the second half of the data, I used the same codes as for the teacher interviews as suggested by Strauss and Corbin (1990). Finally, I compared the codes of the two halves and created the following categories: advantages and disadvantages of the emergency remote delivery; experiences with assessment in general and in my lessons, both of which had the subcategories of positive and negative experiences; exact assessment methods mentioned by the students; and further factors influencing students' learning experiences. I printed the transcripts and color coded these categories in Hungarian and translated only the quotations when writing the dissertation.

The student questionnaires contained both qualitative and quantitative data. In addition, there is data from questionnaires administered between 2014 and 2017 and from the action

research. In connection with qualitative data, students' answers from 2014 can also be used in accordance with Dörnyei's (2007) claims that qualitative data collection results from the pilot study can be used in the final analysis. Moreover, this prolonged engagement can carry more "face validity" (Dörnyei, 2007, p. 52) looking at students' experience throughout seven years and a larger sample of 152 students in total. When analyzing textual data, depending on the questions, I used different types of content analysis. For example, looking at students' associations in connection with school assessment, I categorized the words into negative, positive, and neutral categories. Looking at open-ended questions with longer textual answers, such as their suggestions of what they would change in school assessment or their reactions to different alternative assessment methods, I copied their answers into a separate document and applied the coding process of the constant comparative method (Dörnyei, 2007). I started with open coding, then created categories and compared the codes to the data and to other codes coming up with more elaborate ones.

Analyzing quantitative data, I used descriptive statistics reporting frequencies, averages (arithmetic means), standard deviations, ratios, and so on. In connection with Likert scale statements, Cohen et al. (2007) have a very strong opinion that rating scale questions are ordinal data, so researchers should use modal scores and non-parametric data analysis, "though one can find very many examples where this rule has been violated, and non-parametric data have been treated as parametric data. This is unacceptable" (p. 328). Dörnyei (2007) also reports that "data yielded by Likert-scale items have typically been treated as interval data, even though theoretically we cannot be certain about the 'equal-intervalness' of the categories (e.g. strongly disagree, disagree, neither agree nor disagree, etc.)" (p. 202); however, he also states that "researchers have found that if such 'quasi-interval' variables are sufficiently continuous, that is, have at least five values, they can be analysed by parametric procedures" (p. 202). Based on the latter claim, analyzing Likert scale items between 2014 and 2017, I used parametric procedures such as ANOVA, independent-samples and paired-samples T-tests (Barbarics, 2015, 2016, 2017). As Dörnyei (2007) provides their non-parametric equivalents, for the present dissertation I also checked the "Mann-Whitney U test [...as] the non-parametric alternative to the independent-samples t-test", the "Wilcoxon signed-rank test [...as] the non-parametric alternative to the paired-samples t-test", and the "Kruskal-Wallis test [...as] the non-parametric alternative to one-way ANOVA" (p. 203).

The final step of the data analysis was to compare students' and teachers' perspectives and organize the results based on the emerged themes. As Holliday (2007) suggests, "moving from data collection and analysis to the writing of the data analysis part or chapter of the written study involves organizing the data into themes which then act as headings in the written study" (p. 113), I describe how the headings of the Results and Discussion chapter were formed.

As alternative assessment was defined in relation to traditional assessment, first I collected all the data in relation to traditional assessment both from the students' and the teachers' experiences. This resulted in the chapter entitled Students' and Teachers' Experiences With Traditional Assessment (5.2). As it emerged from the data, teachers based their alternative assessment practices on their various goals, so before presenting the exact alternative assessment methods, I collected teachers' goals for using alternative assessment methods (chapter 5.3, Teachers' Goals for Using Alternative Assessment Methods). Based on the emerged themes, there were goals related to assessment, to learning and school subjects, to students' life beyond school, and to the emergency remote delivery, in addition to some personal goals for using alternative assessment.

Then I categorized the emerged alternative assessment methods and decided to add the consequences that the interviewed teachers experienced in connection with them, (which was one of the subcategories under the Experiences label) resulting in chapter 5.4, Alternative Assessment Methods and Their Consequences. As mentioned earlier, the exact alternative assessment methods had several categorizations during the different stages of the data analysis (e.g., based on mode: written, oral, numerical, or based on complexity: from simple to complex, and so on). Due to their great number, it was difficult to present them in a systematic way. Finally, I turned to my definition of alternative assessment and organized the methods in relation to the traditional, resulting in the following two sections: 5.4.1 Solutions for Grading and 5.4.2 Alternatives for Testing. After that the following two categories emerged from the remaining methods: one consisted of different activities that teachers used to help process the material (5.4.3 Assessment Activities That Help Process the Material) and the other consisted of further methods including diagnostic, peer, self-assessment, and feedback from the teacher to the students (5.4.4 Further Alternative Assessment Methods). The last category of alternative assessment methods was connected to the emergency remote delivery (5.4.5 Alternative Assessment Methods Used During the Emergency Remote

Delivery). For each section, I gathered data from the interviews and from the action research and presented them in separate sections.

In connection with teachers, I still had the miscellaneous category, which included teachers' beliefs, personal characteristics, views, and struggles with several further subcategories, which I felt were pivotal in my research. Not only did I have a research question on teachers' views, I also considered it crucial to present who these teachers were. I struggled a lot with how to organize their beliefs and personal characteristics. Then I came across an earlier publication of mine (Barbarics, 2018) in which I analyzed whether my assessment practices helped in developing democratic competences (Council of Europe, 2016), and I realized that there was a great overlap between the descriptors of the competences and my labels of the interviewed teachers' characteristics. On the one hand, I became interested if I could find instances of all the components of competencies for democratic culture in my data, so I analyzed the data based on this external category. On the other hand, it gave a structure for the first part of the Results and Discussion chapter presenting who these teachers are and how their pedagogical views are connected to their use of alternative assessment.

There was one more instance when I analyzed my data based on external categories. In connection with teachers' goals related to students' life beyond school (section 5.3.3), the interviewed teachers frequently mentioned that they would like to develop their students' 21st century skills. They only gave a few examples, so I was curious if there were instances of further 21st century skills. For this reason, I analyzed the data based on the categorization of Trilling and Fadel (2009) to find examples of 21st century skills that the interviewed teachers wished to develop.

Finally, students' experiences were analyzed and structured. I moved their perceptions of traditional assessment next to teachers in chapter 5.2, Students' and Teachers' Experiences With Traditional Assessment. Then in chapter 5.5, Students' Perceptions of Alternative Assessment Methods, I present their reactions to all the alternative assessment methods used during the action research in the same order as it is presented in the previous chapter (5.4 Alternative Assessment Methods and Their Consequences).

A lack of systematically unifying the coding of the data to its source was a weakness of the data analysis process. As the constant comparative method was applied on different platforms: first in the atlas.ti program, then in excel sheets, and finally on printed paper, the coding of the

data to its source also varied. As I did not unify them, I can find where each quotation comes from, but I failed to establish an “audit trail that will make it possible for an external auditor to examine” (Guba, 1981, p. 87). In future research, I will make sure to use a unified coding procedure, so that any piece of data can be traced back to its source by any external auditor, too.

4.6 Summary Table of Research Questions and Methods

Table 5 summarizes the methods of data collection and analysis linked to the research questions. As most of the questions are of an exploratory nature, the gathered qualitative data is analyzed with the constant comparative method (Dörnyei, 2007).

Table 5

Summary of Research Questions and Methods of Data Collection and Analysis

Research questions	Methods of data collection	Methods of data analysis
1.1 What do teachers mean by assessment and alternative assessment in particular in EFL and in EMC classes in public secondary education in Hungary?		
1.2 What alternative assessment methods do teachers claim to use in English as a foreign language and in EFL and in EMC classes in public secondary education in Hungary?	semi-structured interviews with twelve innovative teachers already using alternative assessment methods	content analysis with the constant comparative method
1.3 What are teachers' views of using alternative assessment in EFL and in EMC classes in public secondary education in Hungary?		
1.4 What are teachers' motivations and purposes for using alternative assessment in EFL and in EMC classes in public secondary education in Hungary?		

1.5 What are teachers' experiences with using alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary?	in addition to the interviews:	content analysis
1.6 How can alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary be adapted to remote delivery?	field notes, teaching journals, proformas, self-reflections, discussions, observers' reports, and class recordings	with the constant comparative method
2.1 What are students' experiences in connection with assessment in EFL and in EMC classes in public secondary education in Hungary?		
2.2 What are students' perceptions of alternative assessment in EFL and in EMC classes in public secondary education in Hungary?	field notes, focus group interviews, monthly	content analysis with the constant comparative method of textual data, inferential and
2.3 What are students' experiences in connection with assessment during emergency remote delivery in public secondary education in Hungary?	questionnaires for students of the action research project	descriptive statistical analysis of questionnaire results
2.4 What are students' perceptions of the adaptation of alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary to remote delivery?		

4.7 Ethical and Legal Considerations

Permission and consent were asked from all the participants. At the beginning of the interviews participants were informed about the aims and the process. The statement about the purpose of the research explained that I intend to investigate the different methods used for assessment purposes in public secondary schools in Hungary, so it is an exploratory study and that is why there are no correct or incorrect answers. My intention was to make it clear that I am interested in the honest opinion of the participants. Confidentiality and anonymity were granted. Pseudonyms were to be used so that participants would not be identifiable. Their participation was voluntary, and they could withdraw or refuse to participate at any time. The data would be kept confidential and used only for research purposes. I asked for their consent to record the interview, started the recording, and repeated the introduction (including their consent) on record as well.

In connection with ethical considerations of action research, Burns (2015) asks the following three questions: “Whose permission or consent is needed for the research?”, “Who will be affected by the research?”, and “Who should be told about the research when it is completed?” (pp. 198-199). Answering the first question, permission needs to be asked from all stakeholders: the leadership of the school in which the action research takes place, the students and their parents, and colleagues if they are involved (e.g., through class observations, discussions, etc.).

My 11thME students were participants of an ongoing experiment of the Content Pedagogy Research Program of the Hungarian Academy of Sciences. The stakeholders (the school, the students, and their parents) agreed to take part in the research conducted by me (including questionnaires, class observations, photo and video recordings of lessons) at the beginning of the experiment in September 2017. This consent was extended to incorporate students’ answers to questions on aspects of alternative assessment as well. The supported Content Pedagogy Research Program in question is about the methodology of mathematics teaching. It has no components in connection with assessment or English medium content teaching; however, the output requirements (given number of grades and exams) must be met; therefore, it is an ideal context to carry out the action research project on developing ways of alternative assessment.

Originally, I did not have any English as a foreign language courses planned for the 2019/2020 school year, so in order to include such a group in my action research, I asked the leadership of the school for the possibility hence I could teach the above mentioned 9th grade beginner IT English group (9thITE). During my first lesson with the 9th grade group, I included

the introduction of my research and asked both students and parents for their consent. The actual form to be signed by parents is in Appendix F, which all participating students brought back with their parents' signatures in the upcoming lessons.

Although consent forms were signed, it was still important to remind students of the data collection, that confidentiality and anonymity were granted through pseudonyms (Appendix E), and participation was always voluntary, which meant, for example, filling in the questionnaires, performing in front of observers, or participating in the focus group interviews. As Burns (2015) answers the second question ("Who will be affected by the research?"), "no harm, risk or disadvantage should ensue, [...] explanation and communication about the purpose of the research should be foregrounded" (p. 198). Finally, the last question ("Who should be told about the research when it is completed?") can be interpreted in different ways. On the one hand, participants should be informed about the research process. On the other hand, results should be disseminated. The latter is yet to come; however, the former was also relevant as in the beginning of the school year I communicated that my action research would take place during the school year, but I was not sure when it would end. Moreover, the unexpected introduction of emergency remote delivery further complicated the situation, so it was necessary to inform participants that the action research continued and expanded involving questions about the adaptation possibilities in remote delivery. As the focus group interviews were on the last lessons with the groups, I announced that it would be the last data collection event of the action research. In connection with the interview study, after I wrote the Results and Discussion chapter, I sent it to the interviewed teachers for "member reflections[, which] allow for sharing and dialoguing with participants about the study's findings, and providing opportunities for questions, critique, feedback, affirmation, and even collaboration" (Tracy, 2010, p. 844). Five interviewees (Anna, Mari, Eszter, Lali, and Fanni) answered that they would read it when they have the chance, and later, one of them (Eszter) replied the following: "I read it through. Congratulations, very nice work, brutally thorough. I hope and wish that those who read it will understand and appreciate it. I think it's great that everything we talked about is in it and in a way exactly how we talked about it". So informing participants about the research process was fulfilled, and I also intend to disseminate my findings in professional conferences, workshops, publications, and so on.

In terms of pronouns, reporting results of the anonymous participants, when the gender of the participant is not stated or unknown, I will use the gender neutral, singular "they" as advised

by the Associated Press (Easton, 2017) or the Chicago Manual of Style (Chicago Manual, 2017). Holliday (2007) suggests the use of the first-person singular pronoun as “a major device for separating the researcher’s agenda from the other voices in the text, thus increasing transparency and accountability” (p. 136). For this reason, I use the first-person singular pronoun “I” as a tool for self-reflexivity.

5. Results and Discussion

Results and their discussion are presented according to the themes that emerged during the data analysis including answers to the research questions as well. As a result, the first part (5.1 How Teachers’ Pedagogical Views are Connected to Their Use of Alternative Assessment) answers research question 1.3 (What are teachers’ views of using alternative assessment in EFL and EMC classes in public secondary education in Hungary?) referring to teachers’ views in connection with alternative assessment, in addition to their pedagogical views in general, too. The analysis of the interviews showed that teachers’ pedagogical views are related to their assessment practices; moreover, their competences for democratic culture also emerged, so the section starts with teachers’ competences for democratic culture, followed by their views on conflict and sensitive issues, teaching and learning, their subjects, school materials, and tasks, ending with their views on assessment. The second part describes traditional assessment, starting with answers to the first research question: (1.1 What do teachers mean by assessment and alternative assessment in particular in EFL and EMC classes in public secondary education in Hungary?) teachers’ definitions of assessment and their views of traditional assessment in schools. In connection with traditional assessment, students’ views are also presented when answering research question 2.1 (What are students’ experiences in connection with assessment in EFL and EMC classes in public secondary education in Hungary?). In connection with traditional assessment, issues in the emergency remote delivery are also discussed to answer research question 2.3 (What are students’ experiences in connection with assessment during emergency remote delivery in public secondary education in Hungary?). Based on questions about traditional assessment, teachers’ goals for using alternative assessment methods are described to answer research question 1.4 (What are teachers’ motivations and purposes for using alternative assessment in EFL and EMC classes in public secondary education in Hungary?). After that, the specific alternative assessment methods are categorized and detailed, answering research questions 1.2, 1.5, and 1.6 (What alternative

assessment methods do teachers claim to use, and what are their experiences with using alternative assessment methods in EFL and EMC classes in public secondary education in Hungary? And how can alternative assessment methods in EFL and EMC classes in public secondary education in Hungary be adapted to remote delivery?). Chapter 5.5 elaborates on students' experiences in connection with alternative assessment methods answering research questions 2.2 (What are students' perceptions of alternative assessment in EFL and EMC classes in public secondary education in Hungary?) and 2.4 (What are students' perceptions of the adaptation of alternative assessment methods in EFL and EMC classes in public secondary education in Hungary to remote delivery?). In the last part, further issues raised by the emergency remote delivery are discussed. Thus the Results and Discussion chapters are organized around the emerged themes and research questions thematically.

5.1 How Teachers' Pedagogical Views are Connected to Their Use of Alternative Assessment

The findings of the pilot interviews (the first four interviews) were published in the *Journal of Adult Learning, Knowledge, and Innovation* (Barbarics, 2019a), and parts of the results and discussion of the main interview study were published in Hungarian (Barbarics, 2019b) as a chapter of the book *Innovation in education* (Polonyi & Abari, 2017). The following sections present the results and discussion of the final in depth analysis of the interview study including the first, main interview study and the follow-up interview study after the emergency remote delivery as well. Note that the pseudonyms of the interviewed teachers in the publication in English: Ann, Mary, Luke, and Ivy (Barbarics, 2019a) were replaced by Hungarian pseudonyms: Anna, Mari, Lali, and Eszter (Barbarics, 2019b), respectively, for the *Innovation in education* book (Polonyi & Abari, 2017) in Hungarian, and due to the fact that all the interviewed teachers are native Hungarians, the Hungarian pseudonyms appear in the present dissertation as well. Interviewees were purposively selected because they all made use of alternative assessment.

Analyzing the interviews, teachers' pedagogical views appeared to be closely related to their assessment practices. In addition to teachers' views on conflict and sensitive issues, teaching and learning, their subjects, school materials, tasks, and assessment, another theme emerged. During the coding phase, the category of "teacher characteristics (personality or behavioral traits that could affect their views on assessment)" emerged from the pilot interviews as presented in Table 4 in section 4.5, *Methods of Data Analysis*. Both the main interview study and the follow-

up interview study enriched labels in this category such as empathy, curiosity, student-centeredness, autonomy, critical thinking, self-reflection, righteousness, cooperation, communication, and so on. I realized that the Council of Europe (2016) framework of competences for democratic culture lends itself well to analyze the category, so first teachers' competences for democratic culture are presented.

5.1.1 Teachers' Competences for Democratic Culture

The framework created by the Council of Europe (2016) describes a conceptual model of the competences which citizens require to participate effectively in a culture of democracy, living together as equals in culturally diverse democratic societies as the subtitle of the document displays. The model (Council of Europe, 2016) identifies 20 competences grouped into the four categories of values, attitudes, skills, and knowledge and critical understanding. Teachers' competences for democratic culture are also grouped into these categories providing the titles of the following sections. Descriptors of competences for democratic culture (Council of Europe, 2018) were also devised and tested. There are "statements referring to concrete observable behavior of a person with a certain level of competence" (p. 11): basic, intermediate, and advanced. In most cases, there are two descriptors per level of proficiency. It is also assumed that when a person displays behavior assigned to a higher level of proficiency, that person can display behaviors for the lower levels as well, for instance, "it is highly probable that a person who displays the behaviour in the descriptors associated with the intermediate level of proficiency will also be able to display the behaviours corresponding to the basic level of proficiency" (Council of Europe, 2018, p. 12).

5.1.1.1 Attitudes.

The first attitude listed in the Council of Europe's (2016) competences for democratic culture is openness. Although in the framework, openness refers "to cultural otherness and to other beliefs, world views and practices" (Council of Europe, 2016, p. 12), which is seemingly not directly connected to the topics of the interviews, teachers display a high degree of openness towards their students (who can represent other beliefs or world views, too), and this openness also affects their assessment practices. For instance, when Fanni's students give feedback to each other in a shared google document, one of her reasons for introducing this assessment is that she is "genuinely interested in students' thoughts". In addition to openness to students' thoughts, there

is a general student-centered attitude that has a great influence on interviewees' practices. Anna explains this connection in the following way:

I realized that those people inspire me who are student-centered, so it doesn't matter what exact methods you use, it all boils down to the question whether you are interested in the students or just yourself, your power, your performance, etc. So if you want to serve these meaningful connections, you will come up with innovations because if traditional methods don't work, you will change, not because you want to create something great, but because you are constantly thinking about ways that work for you and your students.

Student-centeredness appears in all interviews and seems to be the overarching attitude that affects other competences as well. Supporting this argument, all the attitudes of the framework (Council of Europe, 2016) are also interpreted in the teacher-student relation.

The second attitude is respect. Thus, in connection with the teacher-student relation, Ella when talking about alternative assessment methods says:

The essence of this [...] is that teachers recognize that students are not below or behind them, but only come after them in time, and this does not entitle teachers to anything else, just to give students everything they can. It, in no way, entitles them to treat students as subordinates. For example, this gamified assessment system, or anything with which we help without oppressing them, is all to promote that we are equals. I think there is indeed an ideology behind this, even if we don't always articulate it.

According to the descriptors of the competences, respect means - on its basic level - "giv[ing] space to others to express themselves, [and] express[ing] respect for other people as equal human beings" (Council of Europe, 2018, p. 17).

All of the interviewed teachers ask for feedback from their students giving them opportunities to express themselves and respecting their opinions even if it is criticism towards the teacher. For instance, Fanni shares a story of very negative feedback from a student including, among others, that she teaches "in a shitty way". She talked to the student afterwards and told him the following: "I respect you and your opinion, and I appreciate honest criticism over dishonest praise, but I would express such thoughts differently", and she gave him a note in which she rephrased his criticism. (Later this student reflected on how he did not like her assessment practices at the beginning but then realized their merits.) When Edit reports that her students tell her that they have found a particular assessment system fair, she adds that "I will survey this more

thoroughly with an anonymous google form at the beginning of next term, too, because it is one thing what they tell you in person and anonymous feedback is another”. Whenever she tries something new, she always asks students’ opinions, but she also says that she does “not want to push this feedback thing all the time because if it’s too much for students, they might not answer honestly”. So she gives space to students to express their opinions, but also respects them if they do not want to because the goal is to get honest feedback. Anna admits that she has changed a lot in this respect because when she started teaching “it was irrelevant whether it was good for the students or not because I was the teacher, I knew the material, I knew what they had to learn etc.”, so she has become more student-centered giving more and more space to students, which is reflected in her assessment practices as well. In connection with this change, she also mentions that “I used to be very self-centered, but it has become a task: I have a talent for drawing people’s attention, so it’s my responsibility if I turn this attention to myself or to things, causes, opportunities beyond me etc.”, which raises the theme of responsibility, which is the next attitude to be discussed.

“Responsibility is an attitude towards one’s own actions. It involves being reflective about one’s actions, forming intentions about how to act in a morally appropriate way, conscientiously performing those actions and holding oneself accountable for the outcomes of those actions” (Council of Europe, 2016, p. 13). A portrayal of this attitude is when Eszter introduced a new assessment system that she had doubts about. “After the first month, I was so unsure whether it would work that I held a vote in every group if they wanted to continue in this system, and all of the groups voted for it”. She was ready to face her students’ decision, and after the students’ authorization, she continued improving her system for the students’ benefit. Taking responsibility for one’s mistakes can be seen in a story that Anna shared. After a student’s disrespectful behavior, she got so angry that she tore the student’s test paper in pieces. After calming down, she admitted her mistake in front of the whole class, apologized for her behavior, and allowed the student to write the test. She was surprised that after the class, the student came to her and apologized for the disrespectful behavior (without her asking the student to do so). Moreover, when she asked for students’ feedback at the end of the school year, they referred to this incident appreciating that she was reflective about her actions and apologized for her mistakes. Not only “forming intentions about how to act in a morally appropriate way” (Council of Europe, 2016, p. 13), but also acting in a morally appropriate way are shown in Fanni’s bribing story. Once a student asked her in an

email how much she had to pay to get a better grade. Fanni immediately forwarded the email to the school director, and together they discussed the criteria of assessment and what she did and did not have to pay for in the school.

The next attitude is self-efficacy, which is also an attitude towards oneself. It includes “a positive belief in one’s own ability to undertake the actions that are required to achieve particular goals, confidence that one can understand issues, select appropriate methods for accomplishing tasks, navigate obstacles successfully and make a difference in the world” (Council of Europe, 2016, p. 13). The last thought of the definition is voiced by Anna and Ella almost word by word. Anna reveals that “I have some kind of mission awareness that I serve something, that I’m a gearwheel in a bigger machine that works for making the world better”. Ella’s belief is that “if you like teaching, you step into the classroom, and even without words somehow you change the world”.

On a more practical level, interviewed teachers show a high level of professionalism. As Mari notes “professional knowledge cannot be replaced by alternative methods, so first you have to know your profession and why you are doing what you are doing and only then can you innovate”. Anna agrees that “in connection with assessment, you have to be conscious, really thinking through why you’re doing what you’re doing, what your goals are, etc.”. They are aware that it is time-consuming, but this is not something they would like to save time on. For instance, Csilla acknowledges that “even after twenty-something years of teaching, I can’t prepare less for the lessons [...] I don’t feel well if I’m not prepared enough: I have to think through what I will do and why I do something”. Bea, giving her reasons for using alternative assessment methods, goes through her classes one by one and tells the different goals why she introduced them (such as catering for the needs of a heterogeneous group including beginners and B2-level English speakers, or involving students’ interests in another group, or exploiting students’ creativity in a third one). Moreover, there was a group in which Bea did not change anything in connection with traditional assessment as she did not see the need to. The same attitude is present when selecting the tools to reach those goals (sometimes in contrast to other possible goals). For example, in connection with gamification, Bea explains that the “self-management, goal setting, systematic thinking aspects attracted me, not the playfulness aspect”, and it’s “the positive effects it can have on students’ lives, not the video game graphics that made me interested”. A last example is Ella describing assessment in a 12th grade group during the pandemic. They were solving matura exam

tasks of previous years, which is, according to Ella, “nothing too revolutionary [...], nor too creative”, but students needed “an exam preparation course, so I was correcting lots of matura exam essays”.

The penultimate attitude is tolerance of ambiguity. “Tolerance of ambiguity is an attitude towards situations which are uncertain and subject to multiple conflicting interpretations. It involves evaluating these kinds of situations positively and dealing with them constructively” (Council of Europe, 2016, p. 13). The Covid-19 pandemic and the resulting emergency remote delivery definitely tested people’s tolerance of ambiguity. Interviewed teachers, on the one hand, retailed how difficult the situation was for everyone. On the other hand, they could acknowledge advantages and “express enjoyment of tackling situations that are complicated”, which is the advanced level descriptor of tolerance of ambiguity (Council of Europe, 2018, p. 18). For instance, Bea says that figuring out what to do in emergency remote delivery

had the advantage that I had to understand a new system, create a new plan because I saw that I couldn’t carry out my original plan, and putting things into a new perspective is always good, so from the point of view of self-reflection it was very beneficial.

However, even before the pandemic, Dóra exhibited a similar attitude in connection with groups that were regarded “problematic and not too motivated to study” in the school, but she found this situation challenging and said “you know colleagues don’t vie for getting these groups, but I’m very happy to teach them”. In connection with alternative assessment, Ella said that “I love the complexity of assessment based on gamification”.

The last attitude is civic-mindedness (Council of Europe, 2016). According to the descriptors, its intermediate level can be recognised through the expressions of “commitment to not being a bystander when the dignity and rights of others are violated, [and the discussions of] what can be done to help make the community a better place” (Council of Europe, 2018, p. 17). Starting with the latter, the interviewed teachers often express how they communicate their views and goals with alternative assessment to the greater community. For example, when teaching a new group, both Fanni and Anna send out letters to students and their parents introducing themselves, their methods, and alternative assessment practices explaining their goals with them and how they can be incorporated into public education. In the school of Ella and Dóra, the pedagogical programmes of the institutions were extended to include alternative assessment following an official vote of a staff meeting. Both of them aim to work with other teachers (at least

the ones who teach the same classes) representing the same attitudes to help make the community a better place. Mari has also noticed a change in this regard: “[seven years ago] when I started using assessment based on gamification, nobody had heard about it. [...] A new, young colleague came last year who said that we should introduce assessment based on gamification... that was cute... so nowadays you can’t avoid these ideas anymore”.

Unfortunately, there are also negative experiences in connection with the violation of the dignity and rights of others. Eszter found out that some teachers applied corporal punishment (e.g., hitting a student) in her school. She was outraged and discussed the issue with colleagues, leadership, and eventually the whole staff in official meetings, just to find out that the majority of teachers accepted corporal punishment as a possible educational tool despite the fact that it is illegal in Hungary. As she could not convince them of the opposite, she decided to leave the school and education as a whole. She misses students and teaching enormously, but in order to teach again she formulates a set of criteria on top of which now stands the community’s attitude to abuse: not only openly condemning physical violence but also any kind of verbal abuse as she finds that just as harmful.

5.1.1.2 Skills. Following attitudes, the second set of competences for democratic culture include the following skills: autonomous learning; analytical and critical thinking; listening and observing; empathy; flexibility and adaptability; linguistic, communicative and plurilingual skills; co-operation; and conflict-resolution skills (Council of Europe, 2016).

The results of the pilot interviews are written up from the point of view of teachers’ autonomous learning in developing alternative assessment methods (Barbarics, 2019a). However, not only do the first four interviewed teachers exhibit autonomous learning skills, but the following eight teachers do as well. These skills enable teachers “to pursue, organise and evaluate one’s own learning in accordance with one’s own needs, in a self-directed manner, without being prompted by others” (Council of Europe, 2016, p. 13). As a prerequisite, it is necessary that teachers are committed to lifelong learning for themselves. Fanni says “you always have to renew because the world is changing, science is changing, and your students are changing”. Dóra, presenting new online applications, adds that “you always have to update your knowledge because students get bored with everything after a while”. Mari shares a more personal point of view. Based on the feedback her school asks from students, she reports: “I started as a 58% teacher, now I got 94%. It’s good to see my own progress, and this feedback is important for me because I always know

what to pay more attention to”. (The exact statements of the school’s feedback form can be found at the end of section 5.4.4.1, Diagnostic Assessment.) So there should be an openness to professional development.

On the advanced level of autonomous learning skills, teachers should be able to “select the most reliable sources of information or advice from the range available, [and] show ability to monitor, define, prioritise and complete tasks without direct oversight” (Council of Europe, 2018, p. 19). Bea, in the first interview, reflects on the introduction of a new alternative assessment system in the following way: “you just have to carry it through to see how you can improve the system”. Then two years later, in the follow-up interview, she details the improvements she made year by year analyzing the process. Eszter, when asked what she would have done differently, answers that “I don’t think that I have done anything wrong, but you can always fine tune your method”, and then lists exact steps for improvement. Similarly Ella, elaborating on her alternative assessment system, stops at certain elements and says “I don’t think this restriction was there at the beginning, I introduced it when I realized that I needed to hold the reins a bit more in this area”.

Thus teachers are monitoring, defining, prioritizing, and completing tasks without direct oversight in order to improve their assessment practices. In terms of selecting reliable sources of information and advice, they refer to professional development opportunities learning from courses, conferences, workshops, discussions with colleagues or even students. For example, Dóra attended an eTwinning professional development course where they worked on a certain online platform that she later used with her students. Lali mentions “stealing the idea” of video recording his formative feedback from a methodology workshop. Bea describes a peer assessment method heard at a conference that she would like to implement in the future. As mentioned earlier, all the interviewed teachers ask for feedback from their students, from which they can also learn. As a last example, both Fanni and Edit say how much they can learn from their students’ peer assessment, too.

Teachers do not only display autonomous learning skills, but they also express the need for autonomy in their professional development. For instance, Csilla tells the story of her first attempts with alternative assessment:

In the very beginning when I started experimenting with alternative assessment, the leadership didn’t know about it. I didn’t keep it a secret, or anything, but I thought that if

I wanted to try something innovative I wouldn't need to ask for permission because it's all about the students and how they can learn the best.

Fanni explains how she finds freedom and autonomy extremely important and adds that “I feel that I always stick out from the crowd with my ideas”. Mari formulates something similar: “I'm always the eccentric one who does her things”. Or Ella expresses in connection with colleagues: “they know about me that in order to stay calm I must be allowed to innovate”.

The second skill set is analytical and critical thinking skills that “are the skills required to analyse, evaluate and make judgments about materials of any kind (e.g. texts, arguments, interpretations, issues, events, experiences, etc.) in a systematic and logical manner” (Council of Europe, 2016, p. 13). As mentioned above, in order to improve, interviewed teachers continuously analyze and evaluate their assessment practices. One of their main sources is the feedback they ask from students. As Dóra states, it is extremely important for her to receive feedback on her work, so that she knows what to keep and what to change. She shares an example when students denoted specific outdoor activities as their favorites in a project, which really surprised Dóra as she had previously worried that students would find those activities childish, but now she is planning to incorporate similar ones in the future, too, asking for feedback about them as well. At the end of emergency remote delivery, Ilona created a very detailed feedback form to assess the online tools they used based on a list of criteria. She had her own experiences and thus opinion about the tools, and in some aspects, her students' perspectives were the same, but in others she was also surprised by students' answers. One such example is when students had to choose the tools that they perceived the most helpful for their learning processes. Students chose the tools that had better graphics, were more colorful, and more interactive, and not the ones with more explanations (as Ilona anticipated).

Júlia also assessed her practices in emergency remote delivery and has the following critical attitude to her findings: “I don't want to say that I have solutions because what works now in emergency remote delivery in the short run, might not work in the long run”. So it is also important to take the context into consideration. (The methods of alternative assessment mentioned in this section are further detailed in chapter 5.4, Alternative Assessment Methods and Their Consequences.)

The next skills are listening and observing, which “are the skills required to notice and understand what is being said and how it is being said, and to notice and understand other people's

non-verbal behaviour” (Council of Europe, 2016, p. 13). As interviews are indirect sources, teachers’ skills of listening and observing towards their students can only be implied based on their accounts. Student-centeredness is already discussed in connection with teachers’ attitude of openness, so we know that it is crucial for all of the interviewed teachers to have a good relationship with their students. For Mari, this is one of the most important things: “during the remote delivery I had some kind of contact with all of my students, so that’s why my conscience is clear”. Lali also mentions, in connection with emergency remote delivery, that although he has a good relationship with his students, not meeting in person was extremely difficult. For example, students log into his English class half asleep and can only say “I’m good” during a supposedly advanced vocabulary lesson, for which his reaction is that he can understand them completely because “this is the cognitive capacity one can perform five minutes after waking up”. So he listens to his students and can understand not only their words, but also what is being implied by them.

In connection with noticing and understanding non-verbal behavior, the emergency remote delivery contains further indications. All the interviewed teachers who were teaching in emergency remote delivery were suffering from not seeing students’ nonverbal behavior if students did not or could not use cameras during live online lessons. Teachers were missing eye contact, students’ non verbal reactions, seeing if the “message got through” (Fanni). Ilona compares it to in person classes, where “you can see how students are looking at you if they understand or if you need to ask them, not to embarrass them but for clarification”. Mari says that “speaking to a dark screen just made me wanna die”. For Ella, it is also connected to her alternative assessment practices: “I think that assessment based on gamification requires personal contact because it is based on trust. For me, [it works only with] looking into each other’s eyes, knowing about each other’s feelings, noticing who is having a bad day, and such non-verbal clues”.

A last indication of teachers’ skills of listening and observing is their very detailed descriptions of their students. All the interviewed teachers share cases and their highly elaborated presentations indicate skills of advanced levels of listening and observing. Ella summarizes it with the following words: “in my class, we know each other so well that if a student doesn’t turn up, I’ll just call their mother”.

Student-centeredness not only results in skills of listening and observing, but also empathy towards students, which is “the set of skills required to understand and relate to other people’s thoughts, beliefs and feelings, and to see the world from other people’s perspectives” (Council of

Europe, 2016, p. 13). Some examples of empathetic behavior from the interviews are the following. At the end of the school year, Júlia was assessing her own work in emergency remote delivery in connection with which she wanted to ask her students' feedback as well; however, she saw how overwhelmed students were, so she said she would "not want to burden them more", so they postponed the assessment session to the beginning of the next term. Ella reported the following in connection with students during the emergency remote delivery:

I didn't feel that it was easy for them. I was so sorry when I saw what some of my students looked like. Some of them even started crying, so it really wasn't easy for them. [...] Some of them were not allowed to leave the house because of an ill grandparent, which was extremely nerve wracking. So I just tried to pour some life into them.

The next skills are flexibility and adaptability (Council of Europe, 2016). According to Mari, they are necessary skills for alternative assessment, so her advice for teachers is to "have a goal in mind, think it through from as many different perspectives as possible, but don't be scared if life then raises completely different aspects. I think it requires a great deal of flexibility and courage or rather self-confidence". Ilona also describes applying alternative assessment while continuously monitoring "which learning method or way of interaction works, which does not, and the same with groups, how they can work together, and so on". When Fanni started teaching as a novice teacher, she was told how many grades she should give and the deadlines for each of them, but it was up to her how to do it. So she started with traditional written tests, but soon looked for alternative assessment methods, which she - similarly to all other interviewed teachers - constantly adapts to changing circumstances. For example, students had to give presentations, and "based on their performance, I could clarify the instructions and detail the rubrics more in depth each time" (Fanni). Edit mentions that texts for reading and listening comprehension are often outdated or boring for students, so she tries to adapt materials to students' interests as well. Csilla also spends a lot of time searching for materials because "for example a video might work with one group but not with another". Lali lists examples of things that he changed based on the feedback of his students, such as deadlines, or in a point-collecting system, how many points students can get for which task, and so on. He also adapted "perks", rewards that students could work for, when he saw that they were not motivating enough for students. He says that he is open to trying out anything: "I'd throw my assessment system out the window and start something completely new if that helped students more". Fanni concludes that this is a continuous discussion because she

always reflects on students' suggestions telling them what can or cannot be implemented and why. In connection with assessment based on gamification, Mari gives the following summary:

Its nature is that it varies group by group. [...] In the past 7-8 years, it has become more individual for me. I have realized that I can, moreover, I should let go of strict frameworks created by others, and adapt it to each and every subject, group, or even student. [...] It often fails when others just copy the rules that will either work for them or not. You have to get used to the fact that it is a constantly changing system with unlimited ways and possibilities depending on your groups. [...] So if you'd like to introduce it, give it enough time to clarify it for yourself and for your students.

Linguistic, communicative and plurilingual skills are defined as “the skills required to communicate effectively and appropriately with people who speak the same or another language, and to act as a mediator between speakers of different languages” (Council of Europe, 2016, p. 14). As interviewed teachers are mostly teachers of English as a foreign language or other content subjects in English, these skills are their strengths. In addition, all of them spent time abroad either working (Júlia in the US, Anna in Austria, Fanni in Belgium), or studying for at least a semester (Eszter in the US, Bea in the UK, Csilla in Kuwait, Ella in Russia), or participating in workshops (Ilona, Dóra, Lali, and Edit) in a foreign language, which was mainly English, but German (Anna) and Russian (Ella) also appeared.

The penultimate skill in the list of competences for democratic culture is cooperation: “Cooperation skills are the skills required to participate successfully with others in shared activities, tasks and ventures and to encourage others to co-operate so that group goals may be achieved” (Council of Europe, 2016, p. 14). In the interviewed teachers' case, cooperation can be looked at among colleagues or between teachers and students or among students. First, working towards the same goals with colleagues is pivotal for the interviewed teachers. Anna says that wherever she has taught, she would not start without the support of the leadership. Dóra had to change schools because the leadership did not agree with her alternative methods (and she rather changed schools than methods). Ella is a member of the leadership, so on the one hand, she actively encourages cooperation, but on the other hand, it has its disadvantages, too. For example, during the emergency remote delivery, Ella wanted to cooperate in using shared google classrooms; however, colleagues preferred creating separate online platforms for each subject and group. Ella assumes that instead of cooperation, teachers interpreted it as a violation of their autonomy, and they did not want the

leadership to see everything they do. The first weeks of emergency remote delivery looked very similar in all of the interviewed teachers' schools. Mari's narrative is the following:

The announcement was on Friday, 10 pm. Saturday morning, we already sent our letters to the leadership saying 'we are here, how can we help, let's do this'. [...] On Monday, we held a staff meeting, and it was decided that the whole school would use Microsoft Teams. [...] On Tuesday, students were registered in the system, [...] and some of us learned how to use it, so on Wednesday, we could hold a three-hour workshop teaching it to our colleagues. [...] We laid the foundations, created ground rules, discussed roles and responsibilities, and so on. [...] There were four of us managing this. We scheduled the week, so one of us was always available to help colleagues. We wrote a Teams 'manual' with FAQ and continually mentored each other and the students. It was a good story.

As in Monostori's (2021) report on the difficulties of the emergency remote delivery, detailed in section 2.5.3, *Emergency Remote Delivery in Hungary*, the above described cooperation and efficiency in dealing with the emergency was probably the exception. The reason why the interviewed teachers share similar experiences might be because of their skills enabling them to handle the situation effectively and support their colleagues, too. However, beside the emergency situation, the interviewed teachers' also experienced similar tendencies of their colleagues' attitudes to their alternative assessment methods.

The interviewed teachers usually group their colleagues into three categories based on their reactions to alternative assessment. First, those who are interested in their alternative methods and might even try some of them. Second, those who are indifferent. Third, those who are "skeptical" (Eszter), "critical behind your back" (Anna), "not too supportive" (Dóra), and so on. All of them agree that the majority of colleagues are in the second category as most of them do not show any interest. According to the interviews, the number of teachers in the first, supportive category ranges from one to five. As Mari describes her school: "it's really positive here. Although nobody does it like me, at least they don't hurt me, and I don't expect them to do it my way. They support me, and we often have discussions, giving ideas to each other".

The other context for cooperation is the teacher-student relationship. Mari says that "it should be a healthy partnership based on mutual respect and trust. They trust you that as a teacher, you know what you are doing professionally". In addition to respect discussed above, trust is just as important for cooperation. Ella aims to build "a level of trust, on which they accept any

innovation”. When asked about students’ reactions to specific alternative assessment methods, Ella answers that students usually say “if Ms says so, we’ll give it a try”, “but of course you don’t have such a good relationship with all classes,” Ella notes.

There are also examples that reflect teachers’ cooperation skills. For example, in Dóra’s cooperative projects everybody has their roles, including her, and when students present their projects, she always starts with her own reflections on what she has learned working in these projects. In connection with Lali’s alternative assessment system, he reports: “We talk a lot about the [alternative assessment] system. We discuss the reasons behind its elements, for example, why I’m pushing some things etc.”, and as mentioned above, based on students’ feedback, they modify the system together. When Bea started experimenting with new applications and online tools (before the pandemic), she says that her students were very nice and understanding: “they knew that I was 50 years old, so they were smiling at my ICT skills, but we were taking baby steps together helping each other along the way”. Both Ella and Júlia give points to students based on students’ self-report of executing some tasks. They say that this trust contributes to building a good relationship with their students.

The last set of skills are conflict-resolution skills, which “are the skills required to address, manage and resolve conflicts in a peaceful way by guiding conflicting parties towards optimal solutions that are acceptable to all parties” (Council of Europe, 2016, p. 14). Section 5.1.2, Teachers’ Views on Conflicts and Sensitive Issues presents several cases underscoring the conflict-resolution skills of the interviewed teachers, for this reason they are not detailed here. Beside attitudes and skills, values and knowledge and critical understanding contain further competences for democratic culture as described in the next section.

5.1.1.3 Knowledge and Values. The next dimension of competences for democratic culture to be examined is knowledge and critical understanding. First, knowledge and critical understanding of the self, then language and communication, and the world (“including politics, law, human rights, culture, cultures, religions, history, media, economies, the environment and sustainability” Council of Europe, 2018, p. 23).

Teachers find knowledge and critical understanding of themselves important for different reasons. Anna connects it to work: “what we believe about ourselves determines how we work, [...] and how we define our role and see ourselves as teachers, will also affect what others think about us”. A more specific reason, according to Eszter, is that teachers should be aware of their

own competences, responsibilities, and limits. For example, when students had psychiatric problems (such as selective mutism or presumable mental illness), it was crucial that she knew how much she could help, what her responsibility was, and what to do, who to turn to when events were beyond her limit. According to the descriptors of the framework, the advanced level knowledge and critical understanding of the self means that one “can reflect critically on his/her own emotions and feelings in a wide range of situations” (Council of Europe, 2018, p. 22).

This critical self-reflection is often portrayed in the interviews. The interviewed teachers sometimes seem to be too critical towards themselves. For example, Mari questions her students’ positive feedback about her lessons because they are just being too nice and concludes that her interpretation “might be just my usual self-flagellation”. Bea admits that “I always feel that I’m not that skillful when I read about all these [...] incredible methods, but even with my hybrid [traditional and alternative assessment methods mixed] ways, I can see the good effects”. Another example is Anna saying: “I’d like to identify as an innovative teacher, but it might be an exaggeration”. They know that they are not perfect, and they do not hide it from their students. When Lali records his feedback on students’ work, he is aware that the recordings contain hesitations, filling sounds, even mistakes, but he says “after the first few, you let go of these”.

As an advanced level of knowledge and critical understanding of the self, teachers also feel comfortable describing and reflecting on their emotions in a variety of cases. For example, Ella felt disappointed when a student applied for the higher level matura exam and failed (missing one point for passing) as Ella did know about the application, and she felt that she could not form a realistic self-assessment for the four years they worked together, so Ella regarded it as a failure. Dóra felt shame when she eventually gave grade ones to students who did not attend online lessons dedicated for test writing. Eszter admitted that maintaining continuous administration of points in her alternative assessment system was very challenging for her. She also described difficult feelings when different groups of hers started comparing themselves to each other: “it was difficult how to convey that different classes don’t collect the same number of points, and it doesn’t mean anything, what matters is that they all collect more and more points each month”. More details about her system are provided in chapter 5.4, Alternative Assessment Methods and Their Consequences. Bea mentioned difficult feelings, too:

I had this image of myself that I was good at planning and using ICT, but after the first week [of the emergency remote delivery] it was all ruined. I was paralyzed by the situation, and I felt that I didn't have the strength to figure out one more thing.

According to Júlia, it is also important not to take things personally such as if students do not come to offered online classes. She says that “it doesn't hurt me”, but she still tries to reach everybody through multiple channels (e.g., online learning management system, e-gradebook, the classes' form teachers, and so on).

Knowledge and critical understanding of language and communication “includes knowledge and critical understanding of the socially appropriate verbal and non-verbal communicative conventions that operate in the language(s) which one speaks, of the effects that different communication styles can have on other people, and of how every language expresses culturally shared meanings in a unique way” (Council of Europe, 2016, p. 14). In connection with non-verbal communication, the interviewed teachers' skills are discussed under skills of listening and observing. In addition, linguistic, communicative and plurilingual skills have also been mentioned. The descriptors of the framework (Council of Europe, 2018) also include knowledge and critical understanding of language and communication with other cultures, which is not the topic of the interviews and has not been referred to by interviewees. Similarly, knowledge and critical understanding of the world “includes a large and complex body of knowledge and critical understanding in a variety of areas including politics, law, human rights, culture, cultures, religions, history, media, economies, the environment and sustainability” (Council of Europe, 2016, p. 14), which was not touched upon in the interviews either.

The last category is values, which comprise valuing “human dignity and human rights”, “cultural diversity”, “democracy, justice, fairness, equality and the rule of law” (Council of Europe, 2016, p. 12). It is interesting to note that although these themes did not emerge in the discussions beyond what has already been elaborated on (for example, none of the forms of the word “democracy” were mentioned by any of the teachers), based on the above detailed attitudes, skills, and knowledge, the interviewed teachers seem to demonstrate almost all of the components of competences for democratic culture.

5.1.2 Teachers' Views on Conflicts and Sensitive Issues

Teachers' views on conflicts and sensitive issues is a separate section because although cases shared by the interviewed teachers show their conflict-resolution skills mentioned above, the

emerging themes paint a more detailed picture on teachers' views on conflicts and sensitive issues. As the focus of the interviews was assessment, there were no direct questions about conflicts, so the examples below emerged in assessment related situations. It is interesting to note that the interviewed teachers display the definition of the Council of Europe (2016) in connection with conflict-resolution skills, too, as they try to “manage and resolve conflicts in a peaceful way by guiding conflicting parties towards optimal solutions that are acceptable to all parties” (p. 14) through immediate and open discussions. The following examples are detailed in the order of conflicts and sensitive issues among students, between teachers and students, and of teachers in relation to the education system.

5.1.2.1 Teachers' Views on Conflicts Among Students. Before conflicts emerge among students, it is important to acknowledge the importance of prevention. Eszter knew that bullying, mobbing, and other behavioral problems were a constant issue in the school, so to prevent conflicts among students in her lessons, she decided not to use anything that would promote competition. Although she created an assessment system based on gamification, she did not use leaderboards, badges, or titles (see game design elements in section 2.4.3 Assessment Based on Gamification) with which students could compare themselves to each other. Instead, students collected points with highly personalized tasks, and only the teacher and themselves knew how many points they had, so they could not compare themselves to their classmates. There were two occasions when it still caused conflict among students. First, after the first point-collecting period, all students collected the set amount of points and received a grade five (so everybody got the best grade). Second, when they realized that other groups collect different totals of points. In both cases, Eszter held a discussion with her students about it. She asked them to look through their work, see what they collected the points with, and based on that if they thought the grade reflected their performance. As it was objectively there that they worked for every single point, they could not say anything. So she concluded that as long as everybody did the work for the set amount of points, everybody would get a grade five. The reason why her different groups collected different sums of points was that after each point-collecting period, she increased the number of points to be collected in the next period. The amount of the increase was also discussed with the students, which resulted in the different sums for the different groups. Discussing this process with her students, she wanted to convey the message that they might do different things, but what matters

is if they do everything they can at any given moment. She hopes that the fact that her students have first-hand experiences of it makes them accept this variety.

Dóra's handling of a sensitive issue among students also contains elements of prevention and resolution. Preparing for a project week, when Dóra created the teams, one of her students said that he would not participate in the project if he had to be in the same team with a particular student, who happened to be autistic. Dóra tried to find another student to participate instead of the uncooperative one, but as she could not, she eventually put them in different teams. She had a discussion with the uncooperative student telling him that such behavior would result in his dismissal at a real-life workplace. She assigned a mentor student to work with him and have discussions about inclusion with him as she believed that he would accept his peer's opinion more than a teacher's preaching. She put the autistic student into a very supportive team (the members of which actually told Dóra that they would not want to be in a team with the other guy), and she gave the autistic student an extremely challenging task (being the communication manager of his team). On the one hand, Dóra had some bad feelings that the uncooperative student had his way by getting into a separate team; on the other hand, the autistic student could thrive in his group, and his achievements were highlighted in front of everyone. Moreover, at the closing ceremony of the project week, there was a surprise VIP guest, who was a former student of the school, who happened to also be autistic, and he assessed the teams' projects from the point of view of an autistic person. So Dóra hopes that these events might have had an impact on the uncooperative student as well.

The last case to illustrate handling conflicts among students happened in Fanni's group during the emergency remote delivery. Students had to collect information into a shared google document that everybody could edit. They had to add their names next to the parts they wrote. Some students changed the names and claimed parts that were not written by them. Those students whose writings were "stolen" reported it to Fanni. Fortunately, she could easily check the editing history of the document and saw the changes, so she wrote to the students individually. She explained what the problem was and why it was unethical to do something like this. She did not punish the students, but clarified the rules and said that if they broke them, they would get no points for the task and lose the right to choose that kind of task in the future. Fanni started teaching this group shortly before the pandemic, so they did not know each other well enough. For this

reason, she believes that once students get to know her and her ways of assessment, such letters will not be necessary.

5.1.2.2 Teachers' Views on Conflicts Between Teachers and Students. The emergency remote delivery resulted in various conflicts mainly in connection with traditional assessment practices and the question of cheating. Mari acted as a mediator between her students and other colleagues. As she puts it: "I was corresponding with everyone writing nice, long, friendly letters all the time". For example, if she saw that a student wrote an impolite message to a colleague on the news feed of the class, she immediately sent the student a message about how to express criticism in a polite way, "then came the student's explanation of the situation, so we discussed the problem with the colleague, then the rude message was deleted, so the situation was settled, but multiply thirty-something stories like this by I don't even know how many messages". She kept her patience with conflict resolution, but still feels sorry: "for me the worst is that some relationships between students and teachers deteriorated. Some teachers lost face. I might be among them, I don't know, some of my students might think that I haven't dealt with the situation well enough". In the next chapter on traditional assessment there are more details under the section entitled Issues with traditional assessment in the emergency remote delivery. The following paragraphs elaborate on teachers' views on students cheating.

As during the emergency remote delivery students were sitting at home in front of the computer using the internet, all the interviewed teachers expressed that it was not possible to create classroom circumstances, for example, for test writing. As Edit says: "we can trick each other, I can deceive myself, you can fool yourself, but what's the point?! So there were no tests in emergency remote delivery during my lessons". Mari formulates it similarly: "testing lexical knowledge? Who are we making a fool of?" So Mari's students had to write individual essays using the information they found online, which took a long time to correct and write individual feedback on, but it still made more sense for Mari than traditional tests. Fanni created tests for which students could get a passing grade if they were able to find, copy and paste the correct information, but for better grades they had to critically think, evaluate, and apply the information. Lali, knowing that students work together, automatically assigned tests (and other tasks) to be solved in study groups. All in all, they did not prohibit students from using different sources, but they created tasks in which students had to use higher order thinking skills (meaning not only

knowledge and comprehension, but also application, analysis, synthesis, and evaluation as described by Bloom and Krathwohl, 1956).

The interviewed teachers' further views on cheating contain two more aspects. They believe that students cheat if the stakes are high or if they do not see the point of a task. It follows that if they can make tasks low stakes and meaningful for students, in addition to having a good relationship including mutual trust, students might not cheat. In connection with tests, Edit says that "I might be old fashioned, but I think individual tests are necessary. But if they make up only one element from the many ways how students are assessed, they don't matter that much". This idea that tests are only one from many elements, which makes them low stakes, appears in all interviews. The exact methods the interviewed teachers use for testing are detailed in chapter 5.4, *Alternative Assessment Methods and Their Consequences*. Both Dóra and Ilona say that plagiarism occurred about as much in the emergency remote delivery as otherwise, so they dealt with it the same way: discussed it with students and those affected had to rewrite their essays, and according to their experiences, it usually solves the issue. Ilona, Júlia, and Lali all uploaded the keys to workbook exercises, and students had to solve and correct tasks for themselves, and they discussed questions during live lessons. The fact that these were again only one element from many point-collecting opportunities and that students could choose to do it only if they felt the need for it (so hopefully if they found it meaningful), made teachers hopeful that students would not cheat. Júlia says that

I don't want to play Cops and Robbers because it's unnecessary. [...] I have given a lot of tasks from which I can see what they understand, what not, what else we should work on, etc., so the tasks have served this feedback function for me and for the students, too, after that, I don't care about their grades.

Ilona's experience is similar: "Students like it when the number of points reflects how much they work. [...] Otherwise, I just had to trust them that they told me the truth".

5.1.2.3 Teachers' Views on Conflicts in Relation to the Education System. The last category of conflicts and sensitive issues is teachers' problems with the education system. Although it was not the focus of the interviews, as carrying out alternative assessment happens in the context of Hungarian secondary public schools, problems with the education system were frequently mentioned by the interviewed teachers. Out of the twelve interviewed teachers, three left public education before the follow-up interview was conducted with them, so this chapter ends

with their reasons. Ella was also on the verge of leaving: “I was offered some non-teaching jobs that I seriously considered because the working conditions are not ideal to say the least. But [during the emergency remote delivery I realized that] I needed teaching to feel healthy [...] I’d suffocate in an office, I want to be with my students”.

So what are these working conditions that are not ideal? One is the lack of equipment. During the emergency remote delivery, several companies offered educational tools such as websites, applications, and so on for free; however, when talking about sustainability Júlia asks: “How much should I sacrifice for public education? It’s a constant dilemma: Shall I buy the chalk, the pen, the paper, pay for the online tools etc.? I honestly don’t know”. Almost all of the interviewed teachers pay for some online tools that they use with their classes, and their explanation for buying them themselves is that they use these tools with their private students as well. Having private students (or other jobs) is also the case for everyone as the salary is so little that none of them can afford having only that income.

A further problem is that there are lots of uncertainties due to constant structural changes, especially in vocational education. Dóra says that in the previous five years the school had five different names due to the frequent changes in which educational authority the school belonged to. Ella as a member of the leadership of her school, describes this situation in the following way:

It feels like being after a war when they draw new borders. We don’t know where the school will belong to and what will happen to our classes. We would like to continue our work based on the vision we have about ourselves. We were promised that all the classes in the school now can go on with us, but you never know.

In connection with the introduction of the emergency remote delivery, Bea admits the following: “to be honest, I didn’t do anything during the first weekend as I was absolutely sure that we would get a week to adjust, to figure out how to do this”, so she was shocked that teachers were expected to start teaching online on Monday morning after a Friday evening announcement without any help. As mentioned earlier, the interviewed teachers had better ICT skills than average, and could help their colleagues, but they also expressed that they could not imagine how more disadvantaged schools dealt with the situation.

The interviewed teachers also complain about bureaucracy and the amount of administration, which they are used to, but find it time consuming and unnecessary because what needs to be registered in the e-gradebook does not reflect either students’ or teachers’ real work.

As Ella says: “meeting the assessment requirements of the pedagogical programme - giving enough grades - is not a problem, you learn how to manage bureaucracy so that the management is happy, but the real question is how to help students achieve their goals”. Still in connection with time, the number of lessons and the structure of timetables are also problematic. Csilla complains that “these 0th and 8th lessons are sick. We are literally making ourselves and the students sick”. The 0th lesson is a lesson before 8 a.m. (or in some schools even before 7.30 a.m.) and the start of the 8th lesson depends on the lengths of breaks between the 45-minute lessons but around 2.30 p.m. or later. Teachers feel that the number of lessons both students and teachers have is too many. Students have 6 to 8 lessons each day. The interviewed teachers teach 18-25 contact lessons a week for 4-9 groups, which means about 60-150 students. (Note that the full-time salary in public schools does not depend on the number of lessons or groups taught, but on how long someone has been in public education.) This affects teachers in different ways. For example, it affects carrying out differentiation as Ella says: “when you’re trying to handle more than 90 students individually that might be a big challenge”. It also affects innovation explained by Bea: “I’d like to get to know these LMS [learning management systems] and the possibilities they offer, but with this workload I don’t know if I have any energy left for it”. The last piece of criticism of the Hungarian public education system affecting teachers’ assessment practices is the curriculum and the corresponding output requirements. Further criticism is detailed in connection with traditional assessment in section 5.2.2, Teachers’ Views on Traditional Assessment.

The above mentioned problems played a part in some of the teachers (Csilla, Anna, and Eszter) leaving public education. Csilla was offered a non-teaching job (still in the field of education), which she took for the following reasons (listed in the order as they appeared in the interview). First of all, she felt that she could not motivate students anymore. Then she mentioned that she found problematic the recent legislative changes in the education system, for example, the new National Core Curriculum (2012), the new compulsory coursebooks, and the whole tendency of limiting the autonomy of teachers and schools. In addition, the salary accompanying the workload resulted in the fact that after more than 20 years of teaching she still had absolutely no time for herself. Anna solved some of these problems by never teaching full time in any of the schools. However, after a break due to health issues resulting in hospitalization and surgery, she reconsidered her priorities, and although she stayed in contact with education in general, she taught in public education only through extracurricular activities. Eszter’s reasons for leaving education

completely are already discussed earlier at the end of section 5.1.1.1 about teachers' attitudes. Although she started working in an entirely different field, at the end of the follow-up interview she says "this might be the first time that I say it out loud, but 'I am a teacher', and I miss it greatly, I'd love to teach again". However, she does not think it is likely to happen and lists similar reasons as above: workload, number of lessons and groups taught, salary, restriction of autonomy, and as described in her story, on the top of her list is the necessity that the community should openly condemn any form of verbal, emotional, or physical abuse.

5.1.3 Teachers' Views on Teaching and Learning

In addition to teachers' competences detailed above, there are some further thoughts on how the interviewed teachers see their roles as teachers. For instance, Anna emphasizes teachers' self-acceptance: "you should have a good relationship with yourself [... because] even if others trust you but you can't trust yourself, nothing will change, this is my version of Gandhi's be the change you wish to see in the world". Lali sees his role not only in teaching the material but also educating for life, or as Fanni adds it: "even achieving attitude change in students". Ella says that "I feel like I'm a coach who has to motivate, push, encourage, and give realistic feedback". Júlia had a very explicit discussion with her students about her views of the role of a teacher:

I had this group who asked me to 'beat the work out of them'. I just spread my arms and said: 'Who am I? Why should I do that? I think that my job as your teacher is to teach you autonomy and self-direction, [...] so it would go against my values and my holistic picture of my role as a teacher'. [...] Of course, when I see that they don't reach their full potential, it hurts, but I still don't think that my role would be to 'whip them to do their best'. [...] Moreover, my previous experience is that even if I give grade ones, it doesn't necessarily make them do the tasks.

All of the interviewed teachers would like to represent their views authentically. This includes being an example of what they require from students. Such as Mari saying, in connection with trying out new assessment methods, that "if you tell students that it is OK to make mistakes, and failure is part of the learning process then you should also apply it to yourself". Anna expresses similar thoughts in connection with the fact that learning sometimes involves changing your previous opinion about something, and teachers should not feel ashamed if they change their minds about something. Bea also talks about difficult times and connects it to assessment saying that because everybody has bad days, if it happens to be a day when students have to perform, they

should get the chance to make up for it. The emergency remote delivery was an especially difficult time for everyone, and Ella, as a mentor teacher, kept her spirits up to show an example: “I had two teacher trainees, so it was very important for me that they’d have the experience of what it’s like to be a teacher. So [...] you have to stay authentic in your role as a teacher”. Showing good examples can also mean small, practical things such as Fanni presenting her own poster to students, or looking up a word online that does not come to mind immediately, or telling her student how to phrase the criticism towards her in a constructive way instead of using hurtful words.

The interviewed teachers’ views on teaching are connected to their student-centered approaches. They would like to teach how their students can successfully learn. They do not find one-way (from the teacher to the student) information transmission effective, which results in minimizing frontal teaching and instead involving students as much as possible. They formulate it in different ways such as “learning by doing” (Fanni), “learning through experience” (Lali), or as Dóra explains “they are not going to remember fantastic vocabulary tests on Monday mornings, but what an experience that project week was”. Anna explicitly states that “methods that are about memorizing lexical knowledge are not close to me, I really don’t like them, and I can’t identify with such output requirements either”. These views directly affect their assessment practices as well. As Bea connects the two, what she likes about the alternative assessment system she uses is “the philosophy that students should take their fate into their own hands and not wait for others to push them from one station to another”. The specific goals that these teachers would like to achieve through using alternative assessment methods are detailed in chapter 5.3, Teachers’ Goals for Using Alternative Assessment Methods.

The participating teachers’ different views on what makes learning successful also appear. According to Fanni, learning should be a fun and enjoyable process. As a general tendency, all the interviewed teachers would like to create a safe environment for learning and minimize stress. However, all of them note that as this is usually not the case in other lessons, it can have varied consequences. According to Mari, some students are really grateful and thrive in a peaceful environment, while there are always some students who abuse it, so they have to learn how to work even if “there is no imminent threat”. She also adds that less stress might make her subject less important for them. Ilona also experiences that students prioritize subjects in which they are afraid of punishment before hers. Ella agrees that “it really works against any student-friendly system if

for other subjects they have to memorize large amounts of texts for the next day's tests". She gives an example from the emergency remote delivery:

Most students didn't do my tasks because they had to make 30 pages of history notes, and they rather didn't do the English homework because they weren't afraid of me. If we look at the degrees of dread, I stumble in the back - thank god.

So they do not mind that students are not afraid of them; moreover, they would like their students to experience what learning means according to their views.

In Eszter's very disadvantaged classes, students had no motivation and were not willing to participate in anything, so first she wanted to raise their interest to a level that they would choose to do anything English related. Then she created a lesson structure in which everybody could work on their own level and was not compared to anybody else, only to themselves. She wanted her students to see positive consequences of every effort they made, to experience that they are capable of achieving good results, and to connect their own performance to learning. Her alternative assessment system was designed to accomplish these goals.

Eszter's views on teaching and learning that are behind her structure can be discovered in other interviews, too. For example, teaching in heterogeneous groups in a way that students can perform on their own levels is the case for almost all the interviewed teachers. Ilona designs cooperative tasks in which students can teach each other. Dóra uses projects in which students work from different roles. Edit rotates responsibilities in group work during her lessons. In connection with making students' results visible to each other, Eszter held a vote, and although students voted for an open leaderboard, she felt that it might have been influenced by peer pressure and was afraid that it would be destructive (seeing each other's points), so she decided not to make it public. Lali also believes that if results show development comparing students to themselves, there is no point in making it visible to each other. Mari agrees and adds that it would also violate students' data protection rights. They do not want to promote students comparing themselves to each other because according to their views of learning, it is individual development that counts. They believe in giving every student the chance for improvement to develop the skills they need. Furthermore, there are students with different levels of learning difficulties in all of the interviewed teachers' groups, so catering for their needs and providing them with the experiences of success is only possible through differentiated and adaptive instruction, which teachers believe students should receive.

5.1.4 Teachers' Views Related to School Subjects, Materials, and Tasks

Teachers' assessment methods are also influenced by their views of the subject they teach and the materials and tasks they use. In connection with English as a foreign language (EFL) and English medium content (EMC) classes, the notion of English as a lingua franca often appears. Ilona says that "English has become a mediating language used mostly by non-native speakers", which makes effective communication of utmost importance. Placing effective communication to the center of teaching appears in almost all interviews. Teachers' views on teaching grammar in EFL classes are also reflected in their assessment practices. According to Lali's views "grammar is not something you can memorize, it just clicks, sooner or later, which might not happen for a grammar test, so it's not fair to grade it". He includes it in achievement tests, but before an achievement test, students only write grammar tests without grades in order to see what they need to practice more. Other interviewed teachers express similar thoughts that they find it unfair to give bad grades for grammar tests. Ella mentions that some students just "cannot do better than a grade three in a grammar test however hard they try". In Ilona's school, grammar is a separate subject, so students get separate grades for grammar within EFL. One of the reasons she introduced grades based on group work was the joy she saw on students' faces getting good grades in grammar that otherwise they would never achieve individually. These examples show that they regard communication more important than grammar knowledge. Júlia says that "I'm not worried about their English knowledge as everything around them is in English, and they use English so much in their free time that they practically live in English". So teachers would like to bridge this gap between students' personal lives and lessons in school. Those who teach EMC enjoy that the focus is automatically on the professional content, so they believe it resembles the role of English in the world and in students' lives better.

The above mentioned views imply using authentic materials. Materials that are not necessarily in the compulsory curriculum (Fanni), that can be connected to students' interests (Mari), that resemble real life situations (Dóra), and that students choose for themselves. The latter category is present in some ways in all the interviews. For example, Ilona's students should teach each other new vocabulary that they have found for their individual essays. Júlia's students can choose any book to read, create their own list of new words, and test themselves on that vocabulary. Bea's students can choose the focus of their presentations in a given broad topic. She also supports students who are above secondary school level to work on whatever English related material they

would like to as far as it fits their level. And sometimes students' lives become the material, such as during the emergency remote delivery:

It didn't matter what we were talking about, [the goal was] just [to] talk in English, so we were talking about how they moved, how their sibling wrote poetry, how their mother was never around etc. [...] I realized that it was useless to do tasks in the book, [...] so I tried to find all kinds of interesting articles and videos and gave tasks in connection with them (Ella).

Other teachers also mention that they like to work with interesting materials and enjoy coming up with varied tasks, which is connected to the next category: teachers' views on tasks.

Students' freedom of choice is not only present in relation to materials but also tasks. Anna's students can select from a variety of means how they would like to deal with a certain topic (such as a presentation, essay, test, individual creative product, cooperative project work, etc.). In addition, students can also offer tasks (Bea). Dóra holds votes in which students can choose what to do (e.g., which exam topic to discuss in more detail). In Eszter's groups the last month of the year was completely built upon students' choice of topics and task types based on their experiences of the school year. The interviewed teachers all believe that it is important to introduce a diversity of task types that help in developing lots of different skills (which is discussed in the chapter about teachers' goals in more detail). This diversity can mean the skills they develop, but also the levels of difficulty or the practice that students need.

In connection with practicing, the interviewed teachers also share some views. They find it an opportunity for differentiation, so there are no compulsory practice tasks that everybody should do. In Eszter's lessons, practicing is built into class work, which, together with the following presentation phase, creates focus that is intensive enough, so she does not see the need for giving any extra homework. Bea makes it compulsory to practice something out of class, too, but it is up to students to choose what they would like to (or need to) practice. Similarly to Ella, who says "I don't care what you do, just use the language". Lali, after experimenting with different versions for giving homework, found that "the best homework is no homework". One of Júlia's motivations for introducing alternative assessment was also in connection with homework:

I had enough of starting every lesson with registering missing homework. I wrote down that half of the group had no homework, and whatever I did, they still had no homework.

It was really frustrating. [...] I also read research questioning the efficacy of homework, which resonated with my inner feeling that there was no point in pushing it.

As a general tendency, interviewed teachers believe in giving a lot of freedom of choice to students within a set framework. For example, there is a given number of points to collect, set deadlines, recommendations of tasks, and individual plans, which are realized in their alternative assessment practices.

5.1.5 Teachers' Views on Assessment

The greatest impact on teachers' assessment practices is caused by their views on assessment. These views are also part of the reasons for applying alternative assessment methods. In addition, they affect their teaching and vice versa. Mari is wondering whether it is true in general or not, but for her, the fact that she has changed her assessment practices based on her views has an effect on her teaching, too. It has given her more opportunities and allowed her to experiment more. She has realized that she can assess, for example, if students are reading in their free time, and it has resulted in a change in her attitude to teaching as well. All of the interviewed teachers are using some kind of point system. What they give points for reflects their views through showing what they value; in addition to what they believe is necessary in a given situation. One of the main characteristics is a conscious goal-driven attitude displayed by the interviewed teachers. When asked what her reasons for introducing alternative assessment methods are, Bea answers that it always depends on the pedagogical goal she has in a particular group. The specific goals are detailed in chapter 5.3, Teachers' Goals for Using Alternative Assessment Methods.

This section first details teachers' views through what they value by giving points for it, then their views on what assessment should be like (fair, transparent, individualized, supportive, and motivating), ending with their views on self- and peer assessment. Eszter, similarly to Bea above, also says that what she gives points for depends on the situation and it must change over time:

In a very unmotivated group, in the beginning, you give points for even lifting the pen [to start working on a task], but after a while it must be abandoned [and different things should be rewarded]. So if two years later somebody refuses to lift a pen and work on a task, it should have different consequences than in the beginning.

So points can be given for starting a task, or for finishing it, sometimes they reflect only the effort, in other cases the quality of the work, always depending on teachers' goals. Fanni had an argument

with her students when they thought they did not get enough points for something they spent a lot of time with. Her explanation was that she allocated the points based on how much students could learn from the tasks. That is why copying texts from the book might take a lot of time but is not worth many points because students cannot learn much from it. So through the number of points students can collect with different tasks, Fanni tries to motivate them to choose task types that she thinks enhance learning more. Lali tells of an instance when he gave more points for new task types (e.g., video editing) in order to drive students towards trying out something they have not done before. During the emergency remote delivery, he registered bonus points for small daily tasks, so that students would do something English related every day. Ilona also gave points for practicing based on the number of tasks solved regardless of their quality (students had to correct the tasks for themselves) as the goal was only to regularly deal with English. Giving extra points for keeping deadlines (or handing in assignments way earlier) or losing points for late assignments can also affect students' time management skills, such as in Júlia's practice when the different phases of essay writing are awarded with points if done within deadlines (e.g., choosing the topic, writing an outline, first draft, corrections based on feedback, and so on). In communicative tasks, Júlia awards the points for achieving the communicative goal, so, for example, students do not lose points for grammar mistakes (if they do not affect the communicative goal). Another task type Júlia uses is a vocabulary test in which students get as many points as many correct answers they have; however, they can do the test as many times as they want to (within a reasonable time limit), so it depends on students how well they would like to perform. This view is connected to fairness.

All of the interviewed teachers would like their assessment practices to be fair. What they mean by fairness varies. In connection with Júlia's examples above, one meaning is that students' results depend on them. It involves objectivity, so "nothing depends on the fact whether the teacher likes or dislikes somebody as the amount of students' work is objectively reflected in their points" (Bea). During the last lesson of Eszter's first year with a group, they held a Q&A session in which students could ask her anything. One student asked who her favorite student was. She asked back who they thought it was. "Students put their heads together and concluded that with such a fair assessment system, it was impossible to have a favorite", so Eszter was happy that she could change students' view that assessment reflects teachers' subjective opinion of them because for her fairness means that everybody who is working for it can achieve something. Lali also connects fairness to students' efforts. He says that it would not be fair if students could get good grades just

because they already have a higher level language knowledge compared to those who work a lot in order to improve their level, so he would like his assessment to reflect the amount of work students invest in studying the language. For Fanni, fairness means that whatever students do is reflected in the assessment, too, so nobody works in vain. She also tries to reinforce what students do well, which is important for the other interviewed teachers, too. When Bea gives feedback, she always tries to “find positive things in students’ work even if they are difficult to find”. Dóra creates categories so that “everybody can win in something”. Edit believes that her assessment system is fair because students know in advance what will be assessed and exactly how, which is linked to the idea of transparency.

The interviewed teachers find it crucial that assessment is transparent for several reasons. One is fairness. The other is creating a safe learning environment. As Eszter describes the process:

I told them the number of points they had to collect in a monthly period [...] they knew in advance what was worth how much, what the topics were, and at the beginning of each lesson what was going to happen on that specific lesson, and that gave them a sense of security. [...] I think it also enhanced their motivation that they always knew what was going to happen.

The view that students should know in advance how they are assessed appears in every interview. Fanni simply says “they should know the criteria” and “if something is not clear I always explain the reasons for the points”. In Ella’s system students also know in advance “what the compulsory tasks are and which ones they can choose from together with their point values”. Ilona always gives 10-point tasks. “I tell them in advance the success criteria, what they have to do for 9-8-7... points, what I will look at, so, for example, if it’s an essay I might look at the correct usage of one structure”. Ilona envies those colleagues who can plan the whole month ahead, so her students see their individual progress bars always compared to the maximum number of points that they could have collected by that time, to which she keeps adding new tasks until the end of a topic. Bea’s students have a list of what they have to pay attention to during a specific oral test (the criteria of assessment). Dóra also uses different versions of showing assessment criteria to students. There are organizing tables, mental checklists, rubrics, and so on (they are detailed in chapter 5.4, *Alternative Assessment Methods and Their Consequences*). “Whatever they do, they know what kind of product I expect and the criteria based on which it will be assessed” (Dóra). Anna says that “obviously, I have provided them with the criteria for all the grades because it is extremely

important that they know the expectations how they will be assessed”, later she also notes that “the more transparent assessment is, the more humane it can be, which makes students grateful resulting in a better relationship between the teacher and the students”. So transparency can also strengthen teacher-student relationships. Mari emphasizes the aspect of predictability: “my students say that it is better that in my subject, they can schedule their time, they know what to expect, there are no surprise tests, etc. and it matters a lot for them”. Mari uses the rubrics of the matura exam, so students can see how they would perform at the exam, but without the high stakes. Beside the points, she also adds “what was missing, how to make it better [...] so I write a lot, and I assess every single piece of writing like that”. Moreover, she also uses a point-collecting system

in which students can continuously see their own progress, the types of tasks they choose, the types they don't, and the skills these tasks develop. [...] So any time we can sit down, have a look at it, [...] and see their strengths, the things that don't go that well, what to develop, what they like doing, what they want to improve, and maybe what cannot be improved that much, so it's OK to know that, too, we should have this self-knowledge.

These examples show that the transparency of assessment is important for many reasons, one of which is to make students' improvement visible. As Eszter says “a result might not be visible from the outside because people would compare it to something, but I compare it to the student's starting point (and where they have arrived during the semester)”, which can be connected to the next theme that emerged in the interviews: facilitating students' development.

In addition, all the interviewed teachers share the idea that assessment should serve the development of students. As it is one of their main goals, it is elaborated on under chapter 5.3, Teachers' Goals for Using Alternative Assessment Methods. However, it is interesting to look at their views on what development is and how assessment can reflect it. Mari says that she tries to “manipulate students' egotism and self-esteem” telling them that “let's not applaud when you can do something that your six-year old sibling can”, and according to Mari students “usually realize that development means solving increasingly more difficult or new types of tasks, and what was enough yesterday might not be enough today. This might go with some sadness, tears, and discipline issues, but we put up with it”. The fact that Eszter has been raising the number of monthly points to be collected reflects the same view. However, she says that “students were excited for the next month's stakes”, so according to Eszter, students saw it as an objective, numerical representation of their development. In Lali's groups the “exchange rate of grade five”

is also changing. He believes that “if you explain your reasons, students accept the changes”. Ilona argues that

the point is never what grades students get but the feedback they receive. They should get the chance to improve and see their own development. Learning a given unit only makes sense if they see the goal, where they started and where they have arrived, and if they can correct if they are not doing something well [...] so you can give them feedback, for example, ‘pay attention to word order, or punctuation, or you know the first person singular’ etc. [...] then if they can do it again, they will pay attention to it.

For Ilona, the key is the feedback teachers give to students together with opportunities to use that feedback. Fanni uses similar strategies, for example, “next to an incorrect answer, I wrote a helping question or suggestion where to find the correct answer”.

In connection with the emergency remote delivery, these views became even more apparent. Both Bea and Lali video recorded their feedback while correcting students’ work. Mari could handle only five assignments a lesson, but for those five assignments she gave very detailed feedback. It was either her who chose the assignments or students could ask for feedback on their work. Ella also says that “I give a lot of feedback. I evaluate the situation and how to go on, so I communicate a lot” and in connection with remote delivery she adds that “luckily I’m a graphomaniac”. Júlia agrees that “because of the pandemic, I spend even more time on giving individual feedback”, but as a result “a dialogue emerged, based on which students could improve their work, hand it in again, and so on”. Bea reports similar thoughts that “I feel that students can only step forward if I discuss with them individually what and how they should change”.

According to Dóra, students seeing their improvements has a positive effect on their personality development as well. For this reason, she favors methods that can reflect development by making visible a period of time and the results of students’ efforts. (More details about these methods are in chapter 5.4, *Alternative Assessment Methods and Their Consequences*.) Appreciating students’ efforts appears in other interviews as well, and it can also be connected to teachers’ views on assessment being supportive and motivating. Detailing what supportive assessment means, Eszter says that it “should promote autonomy and not dependence on points, the teacher, the school, etc. but that students should work for themselves, taking responsibility, consequences for their actions, and so on”. As she teaches adults, too, her observation is that school assessment can have lifelong effects. Among adults, she cannot differentiate as adults find “easier

tasks degrading”. Eszter derives this from school assessment that “always compared them to a given standard [...] and everybody had to conform to that standard”, it also results in the attitude that “they want you to pour knowledge into their heads”. Eszter’s theory is that if students are forced into a unified assessment where there is only one correct way, and if students cannot perform that, the message is that they are not good enough, so they stop making efforts after a while, which stops their development. She attributes today’s advertisements of “you will speak English on a C2 level in two days” to this attitude that you can get knowledge from someone else, and it is not due to your own efforts. So her view of alternative assessment is that it supports students’ autonomy, conveying that development is a result of students’ efforts, which “they can take, and they can also choose their tools for it” (Eszter). Ilona formulates a very similar view. Through alternative assessment she would like to make her students “aware that only they can learn and that there are several ways to do so”, for this reason Ilona introduces “lots of tools with one goal in mind: so students can have autonomy over what they do, so they can control and influence what results they achieve”. Teachers’ goals for using alternative assessment methods is further elaborated on in chapter 5.3 with the same title.

In connection with autonomy, the interviewed teachers also believe that students should be involved in their own assessment processes. Students are usually surprised by such requests, which also shows that it is not a traditional practice. As Lali describes the process:

I asked for feedback after each unit. [...] In the beginning, only a few students filled them in, but the numbers grew when they realized that what they had written about really changed. [...] I also emphasized that I changed those things precisely because they asked.

Further detailing the process, Lali adds: “in these online questionnaires there are given criteria [...] based on which] they have to assess themselves, their group, and their group members, and then I also tell them what I have seen, [...] so basically we agree on the assessment together”. In Mari’s accounts, students’ reactions look like the following: “Students love to be involved. They love that they can be part of making the rules. They love that they can create tasks. It is really cute when a student thanks you for a task”. As detailed earlier, students of the interviewed teachers have the freedom in choosing task types, including materials, bringing topics, and also affecting their assessment. Eszter consults with her groups about how much they can raise the number of points to be collected in the next period. Ilona claims that “students definitely have influence on how their grades turn out”. Moreover, all the interviewed teachers give their students the chance to opt

out from their alternative assessment systems completely (and be assessed in the traditional way). Mari says that it has never happened even with those students for whom alternative assessment does not seem to yield results. Her view is that probably traditional assessment would not work for them either, so she continues to look for individual solutions. Edit had a group who voted against introducing alternative assessment as they were in their final year and preferred keeping assessment practices as before.

How teachers view motivation through assessment emerged in the interviews as well. Providing the above mentioned autonomy is one factor. Another one is including students' interests and goals. Júlia says that students' personal lives are so separated from school that even if there is a chance to connect them, it might be difficult for students. She tells two cases as examples. When she first introduced that students could collect points with any English related activity they did outside of school, one student said that she would fail the class, and in fact, she barely passed. She viewed it as homework that she did not want to do, and it took a while until it turned out that she regularly watched series in English, read fun-fiction novels, and did a lot of English related things outside of school, all of which could have been incorporated into her English school grade. Similarly, only after two years did it come to Júlia's attention that another student of hers tested drones and wrote professional articles about them in English, which "is very hard work involving real learning, but it took a while before he realized that this could also be included". Júlia was very surprised by that as she reports: "I really think that I have described, told, projected, handed out in a thousand different ways that anything can be done for the points, but still it was really strange how long this realization took for some students". Ella uses a very similar system, and according to her, "knowing all the lyrics of Dua Lipa" is absolutely worth it if this is how she can motivate a student. Ilona gives another example that "there are always those students who enjoy creative tasks such as making a movie, [...] and they are so happy if they can use their skills, [...] and if they enjoy dealing with a task, they will spend more time on it". Eszter has also found that if she takes into consideration what students like, "it creates such motivation that they will do things that they wouldn't in the beginning". Finally, Lali brings up students' goals such as preparing for a language exam:

being at an English lesson can be a must or it can be a useful thing where you can reach a goal [...and] this goal shouldn't be what I say as a teacher, but it should be your goal, [...] for example, [some of my students] did only language exam tasks and got their grades for

that, [...] and they were really happy that they didn't have to do unnecessary tasks that didn't help them reach their goal.

So the interviewed teachers believe that assessment should support students becoming autonomous learners and motivate them to find their interests and pursue their goals.

The last category of views is about self- and peer assessment. Self-assessment is closely related to students becoming autonomous learners, so all the interviewed teachers find its development important. Ilona's experience is that introducing self-assessment practices often meets students' incomprehension because it is missing from traditional assessment practices. She has lots of unsuccessful experiences, such as including "I can..." statements at the end of a unit, which students found completely unnecessary. Ilona reports: "if I see that students take it just as another compulsory task to do, it drives me crazy. It only makes sense if they see its point". She keeps trying in different ways, and in some cases, students seem to be more involved, for example, when reflecting on which task types helped their language learning. One reason might be that it works when students see a direct connection between the self-assessment and what happens in class. This is not only the case with Lali's anonymous questionnaires, but other interviewed teachers describe similar experiences of student engagement, too. All of them ask for feedback from their students, and they shape assessment practices together; however, students reflecting on their own learning processes for themselves (and not for the sake of the teacher) is much more difficult to develop.

Peer assessment practices also appear in all interviews reflecting different views of the teachers. Almost all of them mention that it is important that students get feedback from different sources through which they can look at their and others' work from multiple perspectives. The interviewed teachers would also like to harness the importance of peers for adolescents. It might happen through mentoring each other, teaching each other, showing their own learning methods, correcting each other's work, learning in study groups, role playing exams, and so on. They are also aware of the dangers of peer assessment. Lali says that it would be "self-deception to say that nine-graders are too young and childish for peer assessment because they can't do it properly as they have never learned it". So he believes that peer assessment should be "taught and slowly built up just as self-assessment practices". Júlia warns to be careful with introducing peer assessment as "there could be a lot of tension between students, and I might not know what I tap into, so I rather not do it, but I'm really searching for possibilities and boundaries on how to do it well".

Eszter also pays close attention to students' needs, and because they "really hated to be in the spotlight, stand on a stage, or perform in front of each other, I never forced them". They played the game Activity at the end of the school year, in which students had to explain words in English and "that was the first time they dared to say anything in English in front of each other. They were still very shy, but by that time, they had reached a level". So students assessing each other was out of the question in that group. All in all, all the interviewed teachers focus on students' needs and develop their assessment methods to facilitate their improvement.

To conclude how teachers' pedagogical views are connected to their use of alternative assessment, the following themes emerged. The interviewed teachers display competences for democratic culture. For instance, they try resolving conflicts "in a peaceful way by guiding conflicting parties towards optimal solutions that are acceptable to all parties" (Council of Europe, 2016, p. 14) through immediate, open discussions. Their student-centered approaches influence their teaching and assessment practices. For example, they choose their methods based on the pedagogical goals they would like to achieve. Their views on learning also affect how they teach and assess. They believe in creating a safe environment for learning by minimizing stress and making the learning process enjoyable. In connection with their subjects, their view of English as a lingua franca results in focusing more on developing communication. As a general tendency, interviewed teachers believe in giving a lot of freedom of choice to students within a set framework, such as selecting from varied tasks and materials involving students' interests. Answering research question 1.3 (What are teachers' views of using alternative assessment in EFL and EMC classes in public secondary education in Hungary?), interviewed teachers use alternative assessment as a result of their views mentioned above, in addition to the further views more directly linked to assessment. Teachers believe that through alternative assessment, they have wider possibilities in facilitating the learning process. They express their views on what assessment should be like: fair, transparent, individualized, supportive, and motivating, in addition to focusing on students' needs and developing assessment methods to facilitate their improvement.

5.2 Students' and Teachers' Experiences With Traditional Assessment

First, based on the interview study, teachers' definitions of assessment are discussed, differentiating between what traditional and alternative assessment mean for them, and detailing their views on traditional assessment. After that, students' perceptions of traditional assessment are presented expressed in their questionnaire answers and data from the action research. The last

part of this chapter links both teachers' and students' experiences with traditional assessment to the emergency remote delivery period.

5.2.1 Teachers' Definitions of Assessment

Defining assessment in general, the interviewed teachers use either the word "information" or "feedback" (or both) for assessment and emphasize different aspects of it. Júlia defines assessment as any kind of feedback that teachers give to students "from a facial expression to a word, it can be a lot of things". While Edit defines it as different types of information: on the one hand, it is information for the teacher "whether students have learned what you have taught them or to what extent they have acquired it". On the other hand, it is also information for the students about their progression, which, according to Edit, is necessary because through "tests and challenges, we can see our development and also face where we have to do more". Many of them mention these two directions of information: feedback to students about their work and also to teachers about theirs. Anna adds that assessment can also be a form of measurement that provides information for parents and schools, too, creating comparisons among students, classes, schools, and so on. This reflects the norm-referenced nature of assessment (Gipps, 1994). In contrast, Ilona emphasizes that the feedback should refer to comparisons to preset expectations, which implies criterion-referenced assessment (Gipps, 1994). Ella includes the case when assessment models a situation outside of school (such as a language exam or other life challenges) making it part of the learning process, which reflects assessment *as learning* (Earl, 2006). According to Anna, what we assess should also be discussed when defining assessment.

All teachers formulate similar views in connection with what traditional and alternative assessment mean for them. All of them define traditional assessment as the compulsory grade giving defined by the law (and elaborate on the practices how it traditionally happens in schools), and alternative assessment as anything that is different from that (with a great variety of their different practices). For example, Ilona adds that traditional assessment happens through grading written or oral tests (and many of them emphasize the prevalence of testing in connection with traditional assessment). Anna says that alternative assessment is "anything other than grade giving". As Lali explains, alternative assessment is what "provides more than just giving a grade, or for example you don't give a grade at all". Ella defines it as "breaking free from the one to five scale". However, many of them express that it is difficult to differentiate between the two because even if you would like to be innovative, as it is compulsory to give a certain number of grades,

“most alternative assessment sooner or later results in the traditional [...], ways to reach it might differ, but eventually you must place the kids on the one to five scale” (Mari). As Edit puts it, “grades are a very important extrinsic motivation, which we fight against, but it is part of our lives”. Ella gives an example of the transition between the two: “when a student writes an essay and I give points and grades for it [...] that’s traditional, but if I give feedback on the thoughts in the essay, that’s not [traditional assessment] anymore”. It indicates the summative-formative distinction meaning that summative assessment is categorized as traditional and formative assessment as alternative. For example, Bea explicitly states that she would like to continue giving individual video feedback to students although it is not summative because it does not result in a grade. According to Anna, teachers have to be creative in giving grades in alternative ways. For Eszter the main difference lies in the fact that traditional assessment compares students to each other creating categorization and competition among them, while assessment should help the development of the individual student. “For me assessment means that the teacher evaluates students’ performance compared to themselves and not to somebody else’s performance” (Eszter). Ilona also highlights that traditional assessment is final, while alternative assessment always provides chances to improve and reflects students’ own development. Dóra expressed a similar caveat: “my greatest problem is that I have to give grades because I want to assess students’ own individual performance”.

It is interesting to note that although the law claims that school assessment should ensure the development of students (Act CXC, 2011, Section 1, paragraph 1) and teachers should provide learning support and individual treatment (Government Decree 326/2013, Section 7), as only grade giving is mandatory and there are no other output requirements, ensuring individual development and providing learning support are categorized as alternative assessment by the interviewed teachers often in contrast with the traditional and compulsory grade giving. However, when talking about their own assessment practices, the interviewed teachers do not differentiate between traditional or alternative assessment; they rather express all the different functions that school assessment can mean.

In summary, many of the definitions of assessment reviewed in section 2.2, Categorizations and Definitions of Assessment Types are mentioned by the interviewed teachers, and there seems to be a common understanding of what traditional assessment means in public secondary education in Hungary. It is also interesting to note what the interviewed teachers categorize as alternative

assessment and their relation to the traditional. In the next sections traditional assessment is further detailed through the views of teachers and students ending with the special case of the emergency remote delivery.

5.2.2 Teachers' Views on Traditional Assessment

When describing assessment, in particular traditional assessment, all of the interviewed teachers express concerns about compulsory grade giving: finding it inadequate or even impedimental for reaching different pedagogical goals, hence their need for alternative assessment. As Lali states: “my greatest struggle at the moment is to provide assessment that is fair and complies with all the rules and regulations of the school at the same time, which I find really difficult”.

As all definitions above include, assessment should convey information; however, according to the interviewed teachers, grades in themselves cannot provide adequate feedback either to students on their work, or to teachers on theirs. As Lali asks: “What should grades represent?” and shares the following dilemma:

Let's say you assess students' knowledge. There comes a student who can perform perfectly with zero extra effort, and you have to give a grade five even if they don't do anything at all in class. On the other end of the range, let's assess students' performance in class. There can be students who do their best, work really-really hard, but still cannot reach a passing grade in their knowledge according to the curriculum, but if you look at their work, they have put in the work, so based on that, you should give them a grade five.

They also all emphasize the summative nature of grade giving in contrast to the need for supporting students' development. Ella gives the example of students who cannot get a grade five on language tests because of the compulsory grammar parts, and as a consequence, their grades do not express how they are progressing or how well they are communicating. Bea also believes that the complexity of a student's language knowledge cannot be expressed in a grade. Fanni says that if the range were wider than a five-point-scale, students would also have more opportunities to develop. So another issue is how to acknowledge development. According to Fanni, meaningful feedback that explains the reasons for a point or a grade and how students can change or develop is essential. Ilona suggests that instead of the one to five scale, giving separate percentages would be better to assess different components of language proficiency that a cumulative grade cannot

express. “For example, if there were separate results for vocabulary, grammar, and the four skills [reading, listening, speaking, writing], you could see what is behind a result, and you might get 73% as an average, but you would know that you are really good at writing for example” (Ilona).

Lali and Edit both use percentages, but they still have problems with the system. Lali says that it is great that it shows the difference between an 80% grade four and an 89% grade four, but if a student does not hand in something, registering 0% affects the student’s average much more than just a grade one would. The other problem is that anything above 100% cannot be registered, which he would also need in order to show when students perform way above the expected level. Edit also immensely enjoys that it is possible to officially register percentages in the e-gradebook because out of an 88% it is much easier to create a 90% than from a grade four a grade five. However, her problem is that percentages are eventually converted into grades in a way that students cannot see. As she reports: “we have been corresponding with the national compulsory e-gradebook system for a year now, so that students could follow this conversion, but it hasn’t happened, yet” (Edit). It is interesting to note that a student in their school wrote a program and made it public for everyone, so that students can plug in their percentages and the program converts them into their final grade. In summary, Edit just wishes that they did not have to convert everything to grades.

All the interviewed teachers would prefer not to have grades at all. As Júlia puts it: “if I could, I would let go of this whole grading thing”. Ella says that “if my school allowed me not to give grades just textual feedback on the progress of the students [...] I would be happy to assess like this without grades for all the four years of high school”; however, Ella admits that not everybody thinks like that, for example, “students are so fixated on going for the grades that it would be just as difficult for them to change as for some colleagues”. Ilona also says that she simply does not see the point in grading and perceives it “as a greater and greater obstacle”. She admits that when students need the grade for an entrance exam, she would give a grade five of course, “so they wouldn’t be disadvantaged compared to students from other schools, where I have no idea how grading happens, so there is absolutely no point in comparing students like this, and yet, this is what matters and not their knowledge”. Edit also mentions the issue with end-term grades counting into further studies from the point of view that it can extrinsically motivate students, so in these cases, students might be more willing to do some extra work; however, it is clearly not intrinsic motivation. Ilona’s other problem with grading is similar: she does not want

students to do something just because they get a grade in return. Bea argues that it is an artificial coercive force that teachers must give a certain amount of grades that can be registered in the e-gradebook in different categories reflecting the traditional forms of assessment: “vocabulary test, this test, that test, etc. I think it is extremely one-dimensional”.

According to Mari, the National Core Curriculum (2012) is a great example of the duality of public education in Hungary. The introduction of the curriculum describes values and attitudes that Mari can wholeheartedly agree with. However, according to her, the description of the compulsory materials makes teachers wonder

What? Where? When? How? Why?... there are two opposing messages in it, which I believe represents the two groups of teachers with a great gap between them: one says ‘performance, achievement, lexical knowledge, Prussian methods’, while the other says ‘be humane instead, apply the knowledge, use individual learning paths, assess the development of students compared to themselves’... (Mari)

Anna also mentions these values, attitudes, and skills that are in the National Core Curriculum (2012), but she concludes that “this system doesn’t care about the student, it is not the student that is in the focus, but the material, the lexical knowledge... and as long as the output requirements only focus on these, there is no use for all the skills and competence development in the curriculum, [...] as nobody cares about them”. Fanni refers to the National Core Curriculum (2012) in the context that it creates a great performance constraint, which she finds really frustrating as well. Ella also comments on the effects of the compulsory higher level matura exam: “if rules stay like this that it is compulsory to have one higher level exam, we will have a lot of failed exams because when you correct the higher level matura exam papers, you can clearly tell [...] who is there just because of this law and thinks that English is the easiest”.

Lastly, the interviewed teachers find that traditional assessment also affects students’ learning strategies. As they are required to memorize lexical knowledge, they apply rote learning of materials, which is really difficult to change. For example, Ilona asks angrily “how is it possible that in 9th grade the students’ only way of learning is rote memorization, and how can you change that when it has been a successful strategy so far?” Preparing for the English matura exam and discussing learning and exam strategies, Ella expresses very similar feelings: “I was furious that they were still copying and memorizing the history book, they shouldn’t have done it in the first place, but if that’s what the history teacher told them, I couldn’t do much”. In connection with

introducing alternative assessment methods, all the interviewed teachers share stories of surprised students, such as “when the achievement test was a creative task, they were staring at it for days because they couldn’t believe that that was really the achievement test” (Mari). According to Ilona’s experience, students have rarely met alternative forms of assessment, such as self-assessment, peer assessment, creating a group product and being assessed as a group, or assessment based on gamification, so for her it is difficult to do something completely different in five lessons out of the 36 that students usually have in a secondary school week. To illustrate the effects of traditional assessment, she talks of her experiment using “interactive notebooks” for three years. Students should take notes only on the left pages of their notebook during the lesson, and they should add anything that is connected to it on the right (such as practice, thoughts, quotes, illustrations, examples they see in a movie, song, online, etc.). The goal would have been to connect students’ life outside the school to the school material and to notice where what they learn in school appears in real life. However, the students’ only question was whether it was graded or not, and if it was not, they did not do it, and when it was, they quickly drew some illustrations, so Ilona’s primary goal was not reached. (Although it had the advantage that students learned to take systematic notes and had very nice notebooks.) Eszter also notes that on the one hand, it is great that English lessons can be “the islands of peace compared to other subjects”, but on the other hand, she thinks that much greater progress could be achieved “if other colleagues would have similar attitudes, [...] and for real change more time and similar input from others would have been necessary”. Further achievements and consequences of using alternative assessment methods are detailed in chapter 5.4, *Alternative Assessment Methods and Their Consequences*. After teachers’ views, in the next part students’ perceptions of traditional assessment are presented.

5.2.3 Students’ Perceptions of Traditional Assessment

From 2014 onwards, I used an entry questionnaire inquiring about students’ experiences with school assessment in general (see Appendix E). 152 students have filled in these questionnaires altogether. They are from ten different groups, seven of which were filling in the questionnaire at the beginning of their secondary school studies, meaning that their answers reflect their primary school experiences with assessment, and one tenth- and two eleventh-grade groups with two and three years of secondary school experience, respectively, including their extra preparatory language year in the bilingual school.

One of the questions of the entry questionnaire is what the first three words or expressions are that come to mind when students hear the expression “school assessment”. Based on 150 students’ 415 expressions that they associate with school assessment, more than two-thirds of the students (67%) mention grades (with 103 instances, which is almost a quarter of all the associations). It shows how deeply grades are connected to school assessment for students as well. The second most frequent association is tests (written or oral) including 12% of the words, reflecting the most frequently used forms of assessment. They are followed by learning (6.5%) and teacher (4%). These words are categorized as neutral together with (e-)gradebook, school, knowledge, work, half or end of year, average, school report, behavior, diligence, system, performance (all of them are mentioned at least three times), and some more words mentioned once or twice. 69% of words are in this neutral category, while 22% are clearly negative and 9% positive. It is also worth mentioning that 19 students only have negative associations with assessment, while only three students have only positive ones.

The negative associations can be categorized into adjectives or expressions describing (or rather criticizing) assessment and nouns expressing either specific forms of assessment or accompanying phenomena. (After the words or expressions the number of instances they are mentioned are included in brackets if they are mentioned more than once.) Adjectives are the following: bad (11), rigid (5), difficult (3), sucking (3), unfair (3), subjective (2), unnecessary(2), disgusting, insignificant, boring, stupid, overrated, old fashioned, disappointing, chaotic, confusing, unsteady, not comparable, not exact, not about knowledge, not reflecting real knowledge. Nouns are the following: grade one or failing (9), fear (8), stress (5), anxiety (4), re-take exam (2), lots of expectations, unmatchable expectations, too high expectations, favoritism, depression, compliance constraint, suffering, no free time, too much learning material, nausea, feeling unwell, longing. Finally, one student’s three associations are the following: written warning by a teacher, I don’t read it, my parents shouldn’t read it.

In connection with positive associations, there are words or expressions that appear only once, suggesting that these are rather describing the experiences of individual students and are less of a collective experience. I categorized these words into adjectives possibly describing assessment, the students, and nouns possibly referring to their experiences. Adjectives possibly describing assessment are the following: fair, interesting, good, organized, serious, high level, informative; adjectives describing possibly students themselves are: excellent, exemplary, all 5s,

good grades, great, top, best, first, brilliant, distinguished, outstanding, informed, satisfied; and nouns possibly referring to students' experience are: success, recognition, reward, praise, hope, development, improvement, happiness, excitement, curiosity, good news, honesty.

After the associations, the entry questionnaire asks about students' grades, how satisfied they are with them, how much they reflect their knowledge, and the energy they put in studying the subject. As the same questions reappear in questionnaires after the introduction of alternative assessment methods, these answers will be discussed in comparison with them in chapter 5.5, Students' Perceptions of Alternative Assessment Methods. The entry questionnaire also contains open-ended questions: asking for reasons after each Likert scale statement, what students would change in connection with assessment in school, and they can detail any of their answers or share an example or story of assessment in school. These open-ended questions are all optional for students to fill in. Table 6 shows how many students wrote suggestions on how they would change assessment in school. 51% of students (78 students) who filled in the entry questionnaire wrote some exact suggestions. 49% either left this question empty (22 students), or wrote that they would not change anything (48 students), or did not know what to change (4 students). It is important to note that all of these students filled in this entry questionnaire when I started teaching them, so they did not know anything about me back then. In Table 6, the last three groups are the groups of the action research, and "pre-9th" grade refers to the preparatory language year of the bilingual school.

Table 6

Number of Answers to the Question: "If You Had the Chance, How Would You Change Assessment in School?"

Year	Grade	Subject	No. of students filled in the questionnaire	No. of students who wrote suggestions	Their ratio
2014	11th	Mathematics in English	14	9	64 %
2014	9th	English as a foreign language	17	10	59 %
2014	pre-9th	Mathematics in English	16	7	44 %
2015	10th	Mathematics in English	16	10	62.5 %

2015	9th	Mathematics in English	16	9	56 %
2015	9th	English as a foreign language	16	2	12.5 %
2016	9th	IT-English	16	8	50 %
2016	pre-9th	Mathematics in English	17	3	18 %
2019	9th	IT-English	17	14	82 %
2019	11th	Elective advanced mathematics in English	7	6	86 %
		Sum:	152	78	51 %

Categorizing these 78 answers, there are remarks about grades and grading, tests and ways of testing, and assessment in general (e.g., what should be assessed and how, what assessment should be based on etc.). In connection with grades, at least four students say something along the lines that everybody should get good grades (e.g., “only good grades should exist” or “I’d give everyone 5s” etc.), which might have a humorous tone; however, I believe it can also express the desire for positive feedback. Instead of the one to five scale, different ranges are recommended: “I’d use only three grades: unsatisfactory, average, excellent”, “wider range of grades, like one to seven or eight”, “grading could be from 1 to 10 because it would better differentiate if somebody is for example at the top or the bottom of a grade 5”, “I’d use a wider scale, maybe 1-10, 1-20, or even 1-100 for more accurate assessment”. Some students refer to “small grade” systems when three or five small grades make up the grade that gets registered in the e-gradebook, and they suggest that this should be applied in all subjects. It is connected to the criticism that each teacher assesses differently. Wishing that assessment would be more unified is expressed mostly in connection with tests, such as, “the level of difficulty and scoring should be unified for tests”. However, subjectivity also appears in the statements that teachers should not give grades based on how much they (dis)like a student (one suggestion is constant anonymity for tests), and teachers should not punish students with bad grades for behavioral problems (such as leaving PE clothes or books etc. at home). Still in connection with tests, they should be fair in a way that they contain what has been taught and discussed. In relation to frequency, students have different preferences. Some students would like to have more time to prepare for tests (“at least 1.5 weeks”) or just have fewer tests in general that are “more to the point and really about the material”. Others would like

to have more frequent tests of shorter materials (e.g., “maintaining the required level, regular, about 10-minute-long quizzes would be more expedient than tests at the end of the month that cover a lot at once”); however, another student would “abolish pop quizzes”. The most extreme note in connection with frequencies is the following: “it wouldn’t only be at mid-term or at the end of the year, but at least monthly because it would give a more accurate picture of our knowledge or the lack of it”. I assume that “it” refers to the official school report (that students get at mid-term and at the end of the year), so this student would also like to receive more regular feedback on their work. In connection with tests, the last group of students’ criticism is about the scoring of tests. At least thirteen students mention grade boundaries (that it should be different, lower, more unified, similar to exams, and possibilities to get a better grade when only one or two points are missing).

These exact recommendations about grades and tests imply that students have either positive or negative experiences with them; however, one can only infer the specific reasons behind these suggestions. The last category of student answers might explain more why they like or dislike these practices and what would be important for them in connection with assessment. First, what assessment should be based on or take into account: “diligence”, “class participation”, “homework”, “the fair amount of work put into studying”, “the material covered in class (not something else and more difficult)”, “our real knowledge”, “not crammed lexical knowledge, but skills”, “continuous work and not only at one given moment”. The last one refers to the fact that tests measure knowledge only at one given moment, which together with averaging (“that one bad day can ruin the whole year’s work”), creates high stakes for tests resulting in a lot of stress for students. One student says that “for example, if your mother is in hospital, you cannot concentrate even if you know the material, so in these situations, tests could be postponed”, or another student asks for “taking into consideration students’ out of school activities, too”. Many students ask for different opportunities, to have a chance to improve their grades. Some of them would even abolish grades and suggest instead: “textual feedback”, “corrections with explanations”, “personal discussions with the students and their parents”, and above all “individualized assessment”. The need for learning and then expressing themselves in varied ways that suit them (and be assessed accordingly) appears in some ways in almost every answer. Some students would like to have more oral tasks instead of “writing tests all the time”. Others would prefer tasks that include creativity instead of testing only their memory. Projects and cooperative tasks also appear together

with tasks that can be done at home taking as much time as the student needs. Many students write in general about the need for flexibility, individualization, and lots of different opportunities for students to choose from (some mention individualized assessment for students with anxiety or learning difficulties). All in all, as one student puts it “BE FAIR”.

The above mentioned results are from all the entry questionnaires between 2014 and 2019; however, I made some modifications to the entry questionnaires for the action research (for September 2019) to fine tune some of the questions. One such question is how much they like the subject on a one to five scale (from 1 = I hate it, to 5 = I love it). Students pointed out that it depends whether I mean the subject in general or the lessons in school. So the answers of the 9thITE group (n=17) for the modified three questions can be seen in Table 7.

Table 7

Students' Answers in Connection with English (n=17)

Question	Arithmetic mean	Standard deviation
In your opinion, how important is it to learn English? (from 1= not important, to 5 = very important)	4.76	0.42
How much do you like the English language? (from 1 = I hate it, to 5 = I love it)	3.80	0.54
In primary school (8th grade) how much did you like foreign language lessons? (from 1 = I hated them, to 5 = I loved them)	3.47	0.98

It is interesting to note the differences between the averages (arithmetic means). The results of a Wilcoxon matched pairs signed rank test indicate that students find learning English significantly more important than how much they like the language, $z = -3.29$, $p = .00048$, and how much they liked their foreign language lessons, $z = -3.18$, $p = .00074$. There is no statistically significant difference between how much they like the language and the lessons.

The other modification in the entry questionnaire is the inclusion of other subjects as well. Students had to rate how satisfied they were with the ways their teachers assessed their work in 8th grade in the following subjects, (in addition to foreign language,) mathematics, Hungarian grammar and literature, history, and a chosen subject (from 1 = “I was not satisfied at all”, to 5 = “I was absolutely satisfied”). They also had to give their reasons (see Appendices E1 and E2). Although there were only 17 students in the group, as each of them answered in connection with

five subjects (and they came from different primary schools), there are a lot of interesting results. One is the averages (arithmetic means) of students' satisfaction with the assessment of the different subjects (Table 8).

Table 8

Students' Satisfaction With the Assessment of the Different Subjects (n=17)

Subject	Arithmetic mean	Standard deviation
Foreign language	4.18	1.10
Mathematics	4.47	1.14
Hungarian grammar and literature	4.18	0.98
History	4.29	1.07
Chosen subject (PE, geography, IT, arts, physics, biology, chemistry)	4.41	1.33

It is interesting to note that foreign language and Hungarian have the two lowest averages, while mathematics has the highest. In connection with the chosen subject, the 4.41 average resulted from 14 students choosing "positive" examples (giving 5s for satisfaction) and three students choosing "negative" examples (twice geography with 1s and once physics with a 3 for satisfaction) out of the 17 students. It is also worth looking at the explanations for the different numbers: why students are satisfied or dissatisfied with the assessment methods of specific teachers (teaching different subjects in different primary schools). There is only one student with a 5.0 average meaning that he or she is absolutely satisfied with the assessment of every subject with the explanation (written next to all of them): "I think I got the grade I deserved". All the other students have varied numbers and explanations, which seems to indicate that everyone took the task seriously and hopefully answered the questions honestly.

The reasons for their satisfaction or the lack of it are the following (starting from the 1s - why they are not satisfied at all, up to the 5s - why they are absolutely satisfied). Explanations accompanying 1s: "we had tests every lesson", "my average was 3.8 and I didn't get a 4", "the teacher hated me", "the teacher wanted to screw me up". Reasons next to 2s: "test corrections weren't thorough", "my teacher taught us very badly, so I don't think the grade I received reflects my own abilities", "the teacher was mean with us, e.g., many tests, unrealistic scoring system, and

didn't give a 4 for a 3.5 average". Explanations for 3s: "there were always different teachers, so they didn't know us, and we didn't have the chance to get to know each other", "the teacher didn't teach much", "the teacher was too strict", "I didn't have a good relationship with the teacher", "the teacher didn't like me", "the teacher hated me from 5th grade", "I messed it up a lot". This last one is the only one where the student blames themselves. The rest of the answers show that the students mainly have problems with their teachers' attitudes, grading, and the frequency of tests.

As the averages above indicate, the majority of the answers are 4s and 5s with lots of reasons for satisfaction. When choosing 4s, students' responsibility appear more, for example, "I missed a lot of lessons and didn't make up for them", "I could've been more diligent", "it was what I could do", "when I put myself there (e.g. in written and oral tests) then I could do quite well, and the teacher was also good". As in the latter example, it is clear that the student is not entirely satisfied, but almost, such as when "the teacher helped a lot but also talked about other things through one third of the lessons". In other cases, it is not obvious why the student chooses 4: "my teacher was strict enough as they kept the order in class quite well, it was possible to follow things well in class, so I achieved a good result", "the teacher gave a 5 for everything", "it wasn't on the level of my group", "we wrote five-word vocabulary tests and the teacher never came to class", "we didn't learn anything", "there were too many oral tests".

When students are absolutely satisfied with the ways their teachers assessed their work in 8th grade in different subjects, they give reasons that can be grouped into four different categories: answers in connection with the subject, the teacher, the student, and assessment. The first category contains all the answers that reflect that they like that particular subject, e.g., "math is good", "I love history", or "I like drawing". In connection with teachers, students mention that the teacher "was kind", "fair", "thorough", "nice and funny", "used perfect methods", "taught well", "explained the material well", "held very good classes", "asked only what they taught", and connecting to the next category: "we had a good relationship", "they liked me", "praised me a lot". It is interesting to note that in connection with students' achievements there is one instance when a student says "I studied for it", but in all other cases it is in connection with grades: "we never did anything, but I was excellent (got only 5s)", "I always understood math and got 5s", "my average was 4.9", "I became [*sic*] a 5 from it too (just like from every other subject)", "I studied a lot for English, and sometimes it was hard, but eventually I became [*sic*] a 5". The Hungarian statements that literally translate as "I was a 5" or "I became a 5" appear in the answers of eight students (out

of the 17), which also indicates how much their satisfaction is tied to grades. There is another instance when a student's chosen subject is physical education (PE) and writes that "despite my stature, I got a 5 because I struggled and tried, I didn't give up". In connection with assessment, students appreciate it when "the assessment was realistic and fair" (three students mention it), "there were no random oral tests" (three students mention it), "the teacher was aware of the students' knowledge and graded accordingly" (grading reflecting their knowledge also appears three times), "I was assessed based on my performance". One answer includes all the categories, statements about the subject, the teacher, the student, and assessment: "I love playing sports, my teacher was also nice, and I was good at it, so my results were also good".

Another source of information from the action research is observers' reports and recordings where students' views on assessment and their earlier memories in connection with traditional assessment appear. One student said that he had a very traditional education, and the American visitors asked him to elaborate on what he meant by traditional. He answered: "Nothing special. We had like one topic, and we had a book, and we went through the tasks in the book, and then we had a revision of the topic, a test, and then started a new topic and did the whole process all over again". Another student added that they only had to memorize things, which she found very "boring and dry". A third student highlighted the fact that he had to solve all the tasks exactly as the teacher wanted, which caused a lot of stress: "I hated that so much because I cannot really work under pressure".

The appearance of grades in students' stories recorded by observers is also worth noticing. One student describes her different primary school teachers: the ones that she liked because they were nice, understanding, had interesting lessons that made her pay attention, and one whom she did not like because the teacher gave a lot of homework. Although it was her favorite subject (math), as she did not always do the homework, she "got some bad grades, which made [her] demotivated". Describing herself, another student mentions that "maybe I am hard-working, but I am getting a bit burnt-out from it because I get worse grades now with more work". A third student, someone who had to move to the school dormitory due to the severity of the problems in his family also talks about his performance, and it is interesting how he describes it for an outsider observer: "there are some circumstances why I wouldn't say that I'm lazy, but my grades show that I'm lazy". A fourth student, when reflecting on her primary school experiences, evaluates everything through grades. For example, talking about their placement test at the beginning of a year: "We

wrote a test, and there were three different levels: the highest, the medium - I was in the medium because I got a four for my test - and there was a group for people who were not the best". She explains that first she was disappointed because she did not get into the highest group "because I always got fives for everything, so I didn't like it", but later she realized that it was better this way as the group represented her level: "I felt I was on my level, and I had the chance to get good grades". Coming into secondary school she was in my math group, and this is how she describes the experience to one of the observers: "I was really afraid of math. Maybe it looked like the worst subject. [...] I was scared because I didn't want to have bad grades. [... Later,] it was still difficult, but I don't feel that [scared anymore] because I can get good grades". And the same student in connection with the vocational subjects says that "I pay a lot of attention, and I do a lot of practice tasks, but I don't feel I have great grades from them. I get angry when I study a lot for a test, but then I get a bad grade for it."

Students' perceptions of traditional assessment seem to match teachers' views in multiple aspects. Data collected from both groups show how traditional assessment is tied to grading through different kinds of tests, and the need for changing this tradition also appears from both teachers' and students' perspectives. The introduction of emergency remote delivery further escalated this desire for change detailed in the next section.

5.2.4 Issues With Traditional Assessment in the Emergency Remote Delivery

Although I did not directly ask anything about traditional assessment in relation to the emergency remote delivery, both in teachers' follow-up interviews and in students' focus group interviews several related issues were raised. Those interviewed teachers who were still teaching during the pandemic were usually the ones who helped their colleagues and listened to their students' accounts, so during the follow-up interviews they shared how they saw the execution of emergency remote delivery in their schools. In the focus group interviews, my students also reported on their experiences in connection with other subjects, too. First, themes that appear from both teachers' and students' perspectives are presented, followed by further issues separately only from teachers' then from students' points of view.

The main lesson that both teachers and students voice is that the emergency remote delivery highlighted lots of problems shedding light on things that do not work in schools. As a tendency, teachers tried to replicate online what they do in person. So those who usually have frontal classes did the same online. In the interviews, for example, Mari mentions that some of her colleagues

were happy that there were no discipline problems (students did not disturb the lessons). My students describe similar situations from their perspective like this: “when I knew that we didn’t have to do anything during the lesson, I was listening to the teacher through the speakers while working out”, or “I was listening to the first lessons of the day still lying in my bed”, or “sitting in front of the computer is in itself extremely disturbing because when you have to listen to a boring lesson, you rather watch a random video, so I usually ended up muting the lesson and watching the video”, others played games, chatted with their classmates, and so on. Other teachers did not even attempt to deliver live online lessons; they only sent materials to study, or held a few online lessons, but it was not enough for students. Students complained that from pictures, textual descriptions, or page numbers from a book, it was really difficult to learn the material without further explanations. They also mentioned a case when the teacher recorded explanations, but it was in such bad quality that they still could not learn from it. According to the students, problems appeared when there were very difficult tests from these materials.

The complaints described above raise the question of what is assessed. As tests traditionally contain lexical materials that students have to memorize, how this can be tested remotely is an interesting question. Both teachers and students mention examples when using a camera during live test writing was compulsory (Mari mentions a colleague who required two cameras: one showing the student’s face, the other their hands). Edit mentions parents boasting about how they could trick cameras and still help their children “as it was not about anything in the long run, the only point was what grade the child would get from English or any other subject”. Other testing solutions were strict time limits and the use of randomized questions in mixed orders in which students could not move between questions. According to Júlia, this is a really bad direction as knowledge is not always reflected in speed, so these kinds of tests are measuring speed instead of knowledge, and those students who are quick can still look up solutions, so it does not eliminate cheating either. Some teachers were not concerned about cheating at all. For example, Edit asked a colleague about assessment who said that they wrote tests the same way as they would in person. In connection with cheating, the colleague directly stated that students surely would not cheat. Edit (and Júlia and myself) also heard students arranging how they would solve such tests in conference calls together during upcoming lessons. Finally, some teachers gave more tasks with stricter grading that, for example, Mari’s students found extremely unfair.

Most students felt overwhelmed with school-related online tasks. On the one hand, they had to study by themselves. Bea mentions students complaining that they had to study three-four pages from a book for a lesson, while according to Bea, teachers often covered 10-15 pages a lesson. On the other hand, students either had to prepare for tests, which was the only assessment method most teachers used; or if teachers could let go of tests, students had to do lots of other tasks (essays, projects etc.). As one of the students says: “In the school, it was different. We didn’t get so many tasks. I think it’s because teachers don’t really have any ideas on how to give us grades”. Teachers also confirm that most of the discussions, correspondence, and online meetings were about grading. Lali says that in his school many teachers complained that the emergency remote delivery ruined evaluation. Both Mari and Ella, as form teachers, had to deal with lots of conflicts between their students and other colleagues, most of which were about assessment. For example, Ella tells the story of a parent who called in despair because an otherwise excellent student, who had only grade fives, wrote a bad biology test and got a grade three, so according to the average could not get a final five at the end of the year, which would count into the student’s further studies. Ella was extremely disappointed that they had to discuss this with the colleague because it was not obvious to her colleague in this situation that the student should get a chance to improve the grade. According to Mari, she has never had to manage such a high number of conflicts in her life. In several schools, the leadership had to interfere and ask teachers not to give grade ones. For example, in Ella’s school “we agreed that when assigning a task with a deadline and if a student does not do it, we should send out a written warning about it, and we give a grade one only after three warnings about the same task”. In Bea’s school the leadership sent out a similar letter asking teachers for several warnings and finding out information why a student did not do a task instead of registering grade ones “to try to be as humane as possible”. In Fanni’s case, the school prohibited failing students, but other than that, it was up to the teachers how they gave grades. Teachers also mentioned debates about how to register absences and the grades for diligence and behavior.

One explanation for these debates might be the use of the compulsory national e-gradebook system. As it is the only official source monitored by authorities, all teachers are very careful with the administration. Before the pandemic the e-gradebook system was used to register all the lessons with absences and grades of students. During the emergency remote delivery, these functions were not changed, so on the one hand, teachers were frustrated not knowing how to continue administration. On the other hand, there was a dire need for a platform that could be used for

necessary functions of remote delivery such as communication among the school, teachers, students, and parents, sharing materials of various kinds, providing live lessons, and so on. As all of the interviewed teachers are experienced using information and communications technologies (ICT), they quickly recognised that the e-gradebook system could not be used for such purposes. So for many teachers it resulted in a “double administration” as Ella says “everybody paid close attention to the requirements: what needed to be written in the e-gradebook, then there was the normal communication with students in the Google Classroom”. Ella suggested using Google Classroom together, but some colleagues did not want to use other platforms than the compulsory e-gradebook system, and others did not want to learn new platforms, so they used the ones that they were already familiar with. This resulted in a very chaotic situation. According to Ella, “students spent the first 90 minutes of their day searching for tasks on different platforms”. My students also complained that it was impossible to follow all the different platforms teachers started using, and they got bad grades for assignments they did not know existed or got notified only shortly before submission deadlines. This chaos characterized only about the first month of emergency remote delivery as schools recognized these problems and tried to limit and unify the platforms used. Our school, for example, made it mandatory to use only Microsoft Teams in addition to the compulsory national e-gradebook system. For me it meant that for each task I assigned during a live online lesson, I registered it in the e-gradebook, posted it in the News Feed of the online classroom, and created an Assignment with a deadline for it in Teams, hoping that at least one of the four methods would reach each student.

As the examples above show, there are several reasons why traditional in person school practices of assessment cannot be successfully replicated in remote delivery. Another point to consider when all classes and assessments take place online is the role of the family. The fact that students were at home resulted in various situations. In some cases, the whole family was working on the student’s tasks, so it is questionable whose work the teacher assesses. In other families, school was not that important, and students went to work with family members neglecting school assignments completely. In addition to these two extremes, there were large families where students could not concentrate because of the noise or because they had to look after younger siblings, not to mention differences in the availability of equipment and internet connection (computers, laptops, smart phones, microphones, speakers, headphones, web cameras, and so on) in a family. Even though none of the participants of this research comes from a very disadvantaged

background, teachers reported that from one or two students to a third of the group disappeared in every single class. Teachers tried to find the reasons why students did not join the emergency remote education, and some reasons seemed to be the above mentioned family problems, internet connection or equipment issues, students taking jobs and going to work, but there were also mentions of addictions, lack of time management skills or motivation, and mental health problems.

Although these issues with traditional assessment in general and especially in the emergency remote delivery period present plenty of reasons for applying alternative assessment methods, there are several further goals that teachers would like to achieve. The following chapter details teachers' goals for using alternative assessment methods followed by a presentation of the data describing the specific methods collected in the interview study.

5.3 Teachers' Goals for Using Alternative Assessment Methods

The present chapter aims to present and discuss the data collected about teachers' aims with the use of a variety of alternative assessment methods and tools. Teachers' goals for using alternative assessment methods are categorized into five domains: 1) goals that are directly related to assessment; 2) goals that are related to learning and subjects in school in general; 3) goals that are about students' lives beyond school; 4) goals that are connected to the teachers themselves; and 5) goals that emerged during emergency remote delivery. The results will be presented in this order.

5.3.1 Teachers' Assessment Related Goals

Although the interviewed teachers wish they did not have to give one to five grades at all, as it is compulsory, at least they would like to give grades that in the end reflect students' effort and energy put into studying, their development compared to themselves, in addition to their performance and knowledge of a certain section of the material as compared to standards described in the curriculum. It is also important for them that students understand what is behind a certain grade and find it fair. Furthermore, in connection with grades, their goal is also to relieve the stress caused by traditional grading and testing practices and change the direction of assessment to something meaningful. Sorting teachers' goals related to assessment, the categories based on the purposes of assessment - summative, formative, and diagnostic - appeared.

In addition to the above mentioned goals referring to what a certain grade should reflect, participants believe that summative assessment should also provide realistic feedback on a final product. What realistic feedback means depends on the criteria applied. It can be compared to a

given material, a certain language level, or exam requirements (more about these in the next section discussing learning and school subjects related goals). It can also model performance evaluation of a real-life work situation (more details in the Life related goals section). Lali, for example, aims to create “objective statistics on students’ different skills,” so it is not only the “intuitive knowledge of the teacher how well each student performs”, but students can also see it broken down to skills. Mari and Bea also mention showing achievements by skills, and this kind of transparency (making it visible for students) appears in all interviews, which can be connected to diagnostic goals.

Diagnostic purposes such as obtaining information about what students already know, what they still need to practice, and so on, also appear in all interviews together with transparency, so that this information is not only accessible for the teacher but also for the students. It occurs in different situations and forms. For example, before starting a project, Dóra always creates a diagnostic test to see students’ attitudes to projects and different roles, and she designs the groups and students’ tasks accordingly. In addition, transparency frequently appears among teachers’ goals in emergency remote delivery, too (see also section 5.3.5, Teachers’ Goals Related to Assessment During the Emergency Remote Delivery). To conclude, assessment for diagnostic purposes is definitely used by the interviewed teachers more than just for creating as homogeneous groups as possible at the beginning of students’ secondary school studies. They all seem to collect information on students’ needs, experiences and expectations in order to make informed decisions about instruction and assessment.

Looking at summative, formative, and diagnostic purposes of assessment, most goals described by the participating teachers are in connection with formative assessment. All the interviewed teachers want to reflect on the learning process (and not just the outcome) and also facilitate it through assessment. One of their goals is that students also learn how to reflect on their own learning process. As it is not a one-time event, assessment should reflect on processes, so the interviewed teachers would like to provide students with continuous feedback. This feedback should not only show students’ current level, but also the goal, and how to get there. The first two steps can be present in traditional assessment practices (even in summative assessment showing if an answer is correct or incorrect and what would have been correct); however, as Sadler (1989) highlights, what makes the difference is whether students get help in how to close the gap between their current level and the goal. There are examples in all interviews of teachers aiming to provide this help. For Mari it means that when students develop and learn something that they have not

known before, teachers should “strengthen the belief that wherever they stand is temporary, so the assessment should never impose permanent categories, [...] it should show a present state that they have already achieved together with the way for development,” and then “appreciate the development compared to the student’s own previous state”.

Alternative assessment as defined in this dissertation contains the assessment of a student’s work or performance including elements in addition to or instead of grades, and that it is carried out with the purpose of supporting students’ development. Supporting students’ learning process and their development as the goal of assessment also appears in all interviews. For example, when Bea video records personal feedback for every student on their mini project work, she adds that her goal with this assessment is for students to see that creating the project had a point: what they can learn from it, how they can apply the material, meet the criteria, and so on. As development is a process, another goal that all the interviewed teachers find important is to give specific opportunities for students to improve. For example, based on feedback, they can improve their work, or their grade does not depend on one performance but several components from which they can choose and freely experiment without the high stakes of getting bad grades on rare tests or assignments. Not only do the students have one opportunity to improve their grade but there are many examples when students can do a task as many times as they want to, so they can improve it until they feel it is ready. A good example of this is Júlia’s task of students recording their presentation at home and then uploading it.

In connection with assessment related purposes, when applying peer assessment, the interviewed teachers also formulate specific goals. For instance, Lali says the following:

The 1-5 scale cannot provide constructive, useful feedback on their performance, how they should develop, what their strengths and weaknesses are, so on the one hand, I’d like to provide them with this information, but on the other hand, I’d also like them to be able to form such statements, give, and get this kind of feedback from their peers, too.

Fanni would also like her students to learn and practice giving constructive feedback to each other. Dóra would like students to be able to express their appreciation of each other’s work. Edit asks all the students to assess all the others’ work (in specific tasks based on given criteria), with which she aims to generate enough amount of data, so that on the one hand, students get a more objective picture of their work, and on the other hand, they gain enough practice to assess their own work as well. Ultimately all of the interviewed teachers express that they would like their students to be

able to objectively assess their own work, so they can become self-directed learners, which overlaps with the next category of goals related to learning and specific school subjects in general.

5.3.2 Teachers' Goals in Connection With Learning and School Subjects

Students being able to assess their own work and become self-monitoring learners are very important goals within developing students' autonomy, which is the goal the interviewed teachers mentioned the most times. Developing students' autonomy has several further aspects. Lali formulates it in the following way:

I have realized that learner autonomy is not something that you can give to students [...] and then they immediately become autonomous learners, [...] but it is something that they have to learn, and in the beginning, they don't know how to deal with this freedom, [...] so my goal is to develop autonomous learners who can take responsibility for their own learning processes.

Thus one such goal is that students take responsibility for their own learning processes. It can be manifested in many ways. Ilona's goal is that students can consciously plan their learning activities that they need in order to reach their goals. Bea also mentions teaching students how to set goals for themselves and then reflect on their achievements. According to Edit, it also includes learning time management and keeping deadlines. Júlia, describing tasks in which students have a lot of freedom of choice, says that it is about developing autonomy, students' independence to take their learning into their own hands. According to Anna, not having compulsory tasks but offering opportunities can involve students much more. Fanni also wants her students to experience freedom and learn how to take advantage of it instead of getting lost in it. Similarly to Lali, who wants to motivate his students to "use the [alternative assessment] system and its advantages".

The second most frequently mentioned goal is in connection with motivation. Traditional assessment is often criticized for strengthening extrinsic motivation as opposed to the goal of the interviewed teachers to spark and maintain intrinsic motivation. How to motivate students is discussed from several angles. According to Deci and Ryan (1985), one way is the above detailed autonomy. So when nurturing students' autonomy, teachers' goal can also be to develop intrinsic motivation. Csilla mentions the opportunity for students to bring their own interests to school and have them incorporated to the material with the purpose of motivation. Eszter also declares that the most important thing for her is the appearance of intrinsic motivation and that she can support it. In her groups with cumulative disadvantage, her goal is that students experience that learning

can be something they like and even want to do. It is closely related to the second component of intrinsic motivation by Deci and Ryan (1985): competence. The interviewed teachers also mention the goal to design circumstances in which students can experience success, that they can effectively carry out tasks, reach mastery, and as a result feel competent. As a result of their attempts to ensure such circumstances, the interviewed teachers list further goals, such as creating an environment “to feel safe gaining the skills to get to know the subject, me, my methods” (Mari). Supporting students to do so, teachers can help them find learning strategies that effectively work for them. Ilona’s goal is for her students to see that “they are able to learn and it can be done in several ways”. As students vary greatly, this approach requires individual solutions through differentiation.

Differentiation is also a reappearing goal in the interviews. Dóra, when giving feedback to students, aims for finding different aspects, so that “everyone can stand out in something”. As mentioned above in connection with students’ development, teachers would like to assess students compared to themselves, which also requires differentiation and creating a learning environment that facilitates individual development. For instance, when Mari creates “individual learning plans [... or] contracts” with her students, she wants them not to be discouraged by the results of others and compare themselves to others but set goals for themselves and choose tasks accordingly. Differentiation can happen in different domains. Students can learn at their own pace, for example, Ilona constructs lessons using the online tool Sutori (see link in Appendix B) with which students can learn individually at home taking as much time as they need to. Students can learn on their own level for example, Eszter has very heterogeneous groups in terms of language level, so students can do completely different tasks that match their level. Students can learn according to their skills, Lali assigns tasks according to the skills students need more practice in. Finally, students can learn through the medium they prefer as in Anna’s case when in connection with a topic she offers the possibility to give a live or recorded presentation, write an essay, create a poster, or solve online quizzes when learning the same material. The participating teachers claim that such individualized possibilities can all enhance students’ motivation.

The last element of intrinsic motivation, according to Deci and Ryan (1985), is relatedness. Humans are social beings, so meaningful connections can also motivate us. In terms of goals, this is demonstrated by the interviewed teachers aiming to develop students’ social competences. Since one of the criticisms of traditional assessment is that it punishes students’ cooperation, for

example, Lali would like to not only legalize solving a test together, but actually have students acquire real life collaboration skills. Mari would also like her students to learn how to work together. Dóra aims to develop students' social skills in projects through working with each other, mentors, and even real-life customers. Ilona wants her students to experience cooperation in a way that "there is always somebody I can count on and the others can also count on me". She sets up situations in which she hopes that students teach each other, learn strategies from each other, and so on. Eszter designs role playing tasks in which the explicit goal is to help each other, so students can work on their weaknesses and develop their skills in a safe environment. There are further related skills that teachers aim to develop described in more detail in section 5.3.3, Teachers' Goals Related to Students' Life Beyond School.

In connection with their school subjects in general, the interviewed teachers also enumerate various goals. They would like to raise students' interest in the subject (Eszter), have varied and diverse lessons (Ella) with interesting and humorous materials, too (Bea). They wish students to be actively involved (Mari), like the subject (Ilona), have an exciting class (Edit), enjoy the learning process (Lali), understand connections in the material (Fanni), have fun even during an oral test (Anna), be able to share their experiences (Dóra), and try out new things (Csilla). They also all mention complying with compulsory output requirements as well. Their students' success is important for them, so preparing them for standardized tests and exams, especially the secondary school leaving matura exams or language exams, occurs in all interviews.

Lastly, specific goals connected to English are the following. In addition to the matura exam, preparation for language exams is something that teachers would also like to include in their lessons (if it fits students' goals). In relation to different language skills, communication is emphasized by all teachers. For example, Lali says "I never give grades on oral tests, so that they don't feel stressed because for me the point of oral tests is to develop their speaking skills, which is only possible if they're not stressed". Teachers also aim to teach different sub-skills of communication, for instance, how to give presentations (Bea), carry out small talk (Dóra), or interview someone (Anna). Ilona emphasizes the lingua franca nature of English, so that the goal is "to understand others and make yourself understood because there are so many non-native speakers of English, so you have to adapt to a lot of things". Júlia points out the importance of being aware of the aim of a task: if there is a specific grammar focus then only that should be assessed, similarly to an exact vocabulary focus, or in a communicative situation Júlia says that "I

don't care if it's grammatically correct or not because what matters is if it's understandable what the student tries to convey". Those who teach English medium content classes also lay emphasis on assessing content such as Dóra saying "we have to assess professional and not language knowledge", or Fanni who gives feedback on language use but affirms that grades should only reflect content knowledge. However, she shares a situation in which the two were mixed: "I told a student that I could not accept their assignment because it did not make any sense as the sentences did not have subjects or verbs", so she claims that in English medium content classes it can occur that the language use affects the assessment of content. As a result, it is the goal in both EFL and EMC classes to develop language skills including reading, writing, listening, and speaking skills. Improving speaking and writing skills is part of developing communicative competences, which, together with further competences, are mentioned in the interviews and categorized into the next section, Life related goals.

5.3.3 Teachers' Goals Related to Students' Life Beyond School

All the interviewed teachers stress the importance of students' life beyond school and formulate goals to develop competences for future studies, the world of work, or just coping with the challenges of life in the 21st century. Future studies do not only mean possible tertiary education but lifelong learning including the world of work. As Júlia articulates it, preparing for life outside school she would like to provide circumstances for her students in which they can practice their skills in a safe environment, where they can experiment without consequences, for example, how to stand in front of each other, express their opinions, debate, and so on. This goal of preparing students for situations in a workplace, for example, makes them move assessment practices into this direction as well. For instance, instead of traditional oral tests reciting the material, students have to give presentations including interactivity (Anna), comprehension checks and feedback from the audience (Edit), using notes and visual aids because the goal is to show what students know instead of trying to find what they do not know (Júlia). When assessing these kinds of products (not only presentations, but essays, posters, project work etc.), the teachers' goal is to provide realistic feedback and show what the value of these products would be in a real life work situation. Although Eszter is not teaching anymore, if she taught again, she would further fine tune her assessment system to more closely resemble real life situations. For Dóra, this goal is so important that she even involves "customers" and professionals to assess students' project work based on preset criteria.

Coping with the challenges of the 21st century is often associated with the “so-called 21st century skills” as the Educational Authority’s recommendations refer to it (Oktatási Hivatal, 2020, August 25, para. 9). Based on the categorization of Trilling and Fadel (2009), interviewees refer to the following 21st century skills that they wish to develop. All the learning and innovation skills appear: critical thinking, problem solving, communication, collaboration, creativity, and innovation. They either say directly that “I’d like my students to become independently and critically thinking adults”, or it appears in goals related to specific tasks such as “instead of cramming lexical knowledge, I’d like them to be able to filter the appropriate and necessary information, see what is trustworthy, and find the connections” (Fanni), or as Mari says: “of course I use lexical knowledge, but students have to search for it, find it, word it, paraphrase it etc. so they have to work with the data they find”. Dóra and Anna both “want [their] students to solve real life problems”.

Developing communication skills is already detailed in connection with EFL and EMC subject related goals. However, Anna also mentions it in relation to science subjects that she would like “to develop their cooperation, their communication skills with each other, and it is also a very important part of the subject knowledge that they can communicate using the jargon, which is the first step in applying knowledge”. Cooperation is also elaborated on above in connection with catering for the intrinsic motivation of relatedness. Anna explains what she means by “applying knowledge:” she would like her students to learn how to construct and apply knowledge for themselves. She refers to the categorization of 21st century skills Prievara and Nádori (2018) apply in which “knowledge construction” is one of the 21st century skills. It can also be connected to creativity and innovation. Creativity is also mentioned by almost all the interviewed teachers, and just as it is put together with innovation in the categorization of Trilling and Fadel (2009), it appears with similar ideas in the interviews, too. For instance, Fanni, after mentioning the goal of students being able to critically deal with information, continues with the goal of creating something new out of it. In Dóra’s projects the assessment criteria contain aspects of “presentation style, professional content, [English] language use, creativity, and cooperation”. Lali would also like to assess something new that students have learned, created, and not something that they already possessed.

After learning and innovation skills, Trilling and Fadel (2009) continue with digital literacy skills as a separate category. It is important to note that almost all the interviewed teachers included

developing digital literacy as their goal even before the pandemic (and it became the center of attention during the pandemic, which is detailed in section 5.3.5, Teachers' Goals Related to Assessment During the Emergency Remote Delivery). One example is Anna stating that "I want them to be able to use ICT [information and communications technology]". Lali constructed his assessment system around three main goals: developing learner autonomy, digital literacy, and differentiation. Autonomy and differentiation have already been discussed above. Developing digital literacy was realized by giving students a lot of tasks to choose from and most of them required the use of ICT. During the emergency remote delivery, Bea did not want to teach her students any new ICT tools, but her goal was to use the well-known ones comfortably to carry out tasks from home, and of course if students wanted to explore something new, this was also welcome, but it was not compulsory.

The last group of skills by Trilling and Fadel (2009) is career and life skills from which there are also examples of all categories (initiative and self-direction, social and cross-cultural interaction, flexibility and adaptability, leadership and responsibility, and productivity and accountability skills) appearing in the interviews. In connection with initiative and self-direction skills, Edit says that her goal is that students would take initiative and "not only do what is compulsory but things that would broaden their horizon". According to Dóra, "connections to real life begin by being able to assess our own work". Although self-assessment is touched upon in the Assessment related goals as all the interviewed teachers would like to develop it, teachers' goals in connection with students' self-regulation (Anna) and self-image (Mari) can also be categorized into the career and life skill of self-direction (just as their goals to develop students' autonomy detailed above). The goal of developing students' personalities in some ways appears in every interview. Ilona and Ella would like their students to gain self-confidence. Mari uses the expression to "facilitate students building up themselves little by little", while Eszter's goal is "for the students to come to terms with themselves and to be able to say at least something about themselves". So developing self-knowledge is frequently mentioned as a goal. Fanni's aim is for her students "to think about what is good for them, what makes them happy, how they can take control of their lives, and affect things around them".

Developing social skills is partly discussed in connection with collaboration; however, it has other aspects as well. For example, it is interesting to note that when Dóra organizes mock exams for her students, her goals are not only school related, but she would also like to prepare

students for real life situations such as a job interview mentioning social skills in relation to behavior in such circumstances. It is no surprise that for English as a foreign language teacher improving cross-cultural interactions is also an aim.

In connection with flexibility and adaptability skills, Trilling and Fadel (2009) define the goal as “work[ing] effectively in a climate of ambiguity and changing priorities” (p. 77). These skills were tested during the pandemic, so more details about them are in the Emergency remote delivery related goals section. Other components include incorporating feedback effectively, which is one of the main goals thoroughly discussed in the Assessment related goals section. In addition, “deal[ing] positively with praise, setbacks and criticism” (Trilling & Fadel, 2009, p. 77) is portrayed in Dóra’s practice in which students have to give and receive feedback in the following categories: things they liked, things they would change, questions they have, and ideas they got. Her goal is to look at a product from all these different angles and deal with feedback in a constructive way. Developing adaptability skills is also an aim when teachers use cooperative group work with changing roles, responsibilities, and tasks (used by almost all the interviewed teachers). It is closely connected to leadership and responsibility skills.

Developing leadership and responsibility skills as a goal is present in tasks where students get roles including leadership responsibilities. Dóra carries out diagnostic assessment in order to find who seems to have leadership skills or who would like to be a leader (which includes self-assessment), and then students carry out the project by taking separate roles (CEO, marketing manager, social media communication manager, web designer, and content developer) with the corresponding responsibilities. During the project there is formative assessment aiming to support students in their roles and summative assessment evaluating the end result. All stages include self-assessment when students can reflect on these roles and responsibilities. Similar leadership skills can appear in smaller scale projects or even in single tasks involving cooperative group or pair work that appear in almost all interviews. Students mentoring each other also requires leadership and responsibility skills, so interviewees organizing such opportunities (Dóra, Ilona, Ella) aim to develop these skills as well.

The last category is productivity and accountability skills, which means managing projects with various attributes necessary for producing high-quality results such as “set and meet goals, [...] prioritize, plan and manage work, [...] manage time and projects effectively, multitask, participate actively, [...] be reliable and punctual, present oneself professionally [...], collaborate

and cooperate effectively [...], respect and appreciate team diversity, [and] be accountable for results” (Trilling & Fadel, 2009, p. 83). Csilla’s motivation for introducing alternative assessment methods was to shift accountability and make students responsible for their own learning processes. So the goals discussed in connection with learner autonomy in the Learning and school subject related goals section are all in connection with productivity and accountability skills as they include students taking responsibility for their learning and accountability for their results. Developing further sub-skills also appear such as setting goals (mentioned by all the interviewed teachers), choosing appropriate tools and using them because they are suitable for the task (Fanni), systematizing (Edit), thinking in a system (Bea), managing time effectively (it arises in connection with emergency remote delivery frequently from both teachers’ and students’ perspectives), multitasking (Anna), participating actively (all interviews, especially in connection with remote lessons), and presenting professionally (again in almost all interviews).

5.3.4 Teachers’ Personal Goals for Using Alternative Assessment

There are some goals that stem from the teachers’ perspectives such as their experiences when they were students or now when they are teachers. Teachers’ negative experiences as former students encouraged them to innovate as teachers. When Júlia went to school, she developed a stomach ulcer and migraines caused by the stress in the school. As a result, “it has become [her] motto to reduce stress wherever [she] can”. Anna admits the following: “when I introduce something new, I have [...] more like gut feelings, for example, I hated oral tests [as a student], so I knew I should do them differently [as a teacher], and then I find the pedagogical reasons what is good for what”. Eszter’s assessment practices were also influenced by her school memories: “what happened with me in secondary school was that you got a grade from 1 to 5, and together with that grade, you got a category in the class, a label about how you could perform, [...] you were constantly compared to others”. So one of Eszter’s main goals of assessment is to compare students to their own previous state and not to someone else.

Teachers’ goals for using alternative assessment methods are also influenced by their own personal reasons. Such as to be liked by the students (Anna); to have opportunities to experiment as a teacher (Mari); to be perceived as fair (Edit); to have some fun, cheerfulness, joy during the lessons (Lali); to get feedback on their own work as a teacher in order to know what to do differently (Dóra). Fanni was searching for alternative assessment methods because she wanted to teach according to her personality, so her goal is also to give students enough space and freedom

to develop their personality. It often appears in the interviews that teachers would also like to enjoy the assessment process and what makes it enjoyable depends on the personality of the teacher. For Fanni, it was extremely monotonous to correct the same tests over and over again. Although it takes much more time, she enjoys assessing the variety of students' creative work, so her goal is to involve as much creativity as possible. There are some teachers in all of the interviewed teachers' schools who also experiment with alternative assessment methods, and all of them change it to some extent to fit their own personality.

Csilla was the only interviewee who categorized her experimentation with alternative assessment as a failure. In the first interview, she attributed this to her personality. According to her, alternative assessment methods require teachers to have a "lucky balance of being able to keep boundaries, be consistent but somehow still easy-going, [...] be attuned to students, [...] and definitely not be a control freak". Although she could not pinpoint which traits she did not have, she concluded that she had lost her enthusiasm. Two years later at the time of the follow-up interviews, she was not teaching anymore. In retrospect, she thought that experimenting with alternative assessment methods would have helped, but she listed several reasons for not continuing teaching, none of which could be changed by using different assessment methods. In connection with alternative assessment, she believes that these methods can only work if teachers feel authentic in representing them.

5.3.5 Teachers' Goals Related to Assessment During the Emergency Remote Delivery

In the follow-up interviews, teachers expressed some specific goals that refer to the unique situation of the emergency remote delivery period. Anna, Csilla, and Eszter were not teaching anymore during the pandemic, so this section does not include them.

As mentioned in the literature review, according to the Educational Authority's recommendations (Oktatási Hivatal, 2020, August 25), the role of diagnostic assessment is not that important in remote, digital delivery as teachers have already gathered information about their students' knowledge, attitude, and motivation beforehand. The interviewed teachers, on the other hand, express that due to the special circumstances caused by the pandemic, for them it became extremely important to carry out diagnostic assessment precisely because the situation might have changed students' knowledge acquisition, attitude, and especially motivation. For example, Mari initiated one to one calls as "oral tests" for those students whose activity she did not see otherwise, so she wanted to gather information on "where those students were". Júlia created lots of tasks in

the Canvas learning management system to be able to continuously follow students' results and make decisions accordingly ("to see what students understand, what we still have to work on, etc.").

As a general tendency, all the interviewed teachers were very concerned about how their students could cope with the pandemic, which fundamentally influenced their goals. As Mari puts it:

For me the most important was that none of the children should feel lost in this situation, and everything else was subordinate to this goal. Other colleagues might find this unprincipled or not good because I saw that many were worried about how to grade, how to write tests, how to avoid cheating, and so on, but I didn't care about these.

Fanni says that "I just want them to be there [in the live online lesson] and talk". Bea's goal was also to encourage participation in online classes. Ella felt tremendously sorry seeing students' difficulties and said "I was trying to entertain them, to keep them alive, and to just stay in touch with them, that was more or less the goal".

Teachers also tried to maintain familiar practices naming different goals. Edit, for instance, kept her alternative testing practices, so that students could feel that work outside of school was just as important as if they were sitting in class. Dóra also conducted her mock exams to preserve students' motivation for studying. Júlia's goal was to proceed with the chapters of the book for the same reason. Mari selected which practices were suitable for remote delivery and kept only those that developed students. On the other hand, some methods were not sustainable due to the circumstances. For example, both Mari and Ella said that their point-collecting systems in person could provide continuous feedback and lots of opportunities to collect points, while during the emergency remote delivery they rather gave chances for students to receive good grades doing single activities. Their goal was to save students from receiving bad grades just because of the pandemic situation.

The final category of goals are the ones that take advantage of the situation. The fact that students were at home separate from each other gave opportunities for more individualized treatment. Students had the chance to solve tasks in their own rhythm and timing, so the interviewed teachers laid greater emphasis on supporting individual learning paths. Both Ilona and Mari started using online tools, such as Sutori (see link in Appendix B), that are designed for studying independently, so their goal was to develop their students' autonomous learning. As

mentioned in connection with Life related goals, Lali's assessment system was constructed around developing learner autonomy, digital literacy, and differentiation, which remained his goals during remote delivery, too. In addition, he wanted his students to use this opportunity to get to know new tools, to try out things they have not done before, learn new skills (such as how to edit a video), and deal with English - even if only very little - every day. Edit also saw possibilities in finding ways that can be used later in the classroom as well, for instance, exploring new smartphone applications. Ilona also formulated a long-term goal to make students practice self-assessment that can be perpetuated even after the pandemic. Júlia summarizes the emergency remote delivery's relationship to assessment as follows:

Now the whole Hungarian teacher society is thinking about assessment. I think that if we could find out what works in remote education, that would be a good direction in the long run, too. We have to think through whether we want to assess what we told students, meaning cramming, memorizing, and repeating... because that obviously does not work in remote education (and my question is what its point is anyway), or if we want students to think independently, critically, then we can assess them online as well [...] I've always hated tests based on memorized information, and now it has become more visible how restricted their value is.

In conclusion, the interviewed teachers seem to be guided by their goals in adapting their alternative assessment methods to emergency remote delivery as well.

To summarize teachers' goals for using alternative assessment methods and answer research question 1.4 (What are teachers' motivations and purposes for using alternative assessment in EFL and EMC classes in public secondary education in Hungary?), teachers seem to be motivated by their dissatisfaction with traditional grading practices, in addition to expressing numerous purposes for using alternative assessment. The interviewed teachers aim to achieve goals related to assessment, learning and school subjects in general, and students' lives outside of school. Moreover, their aims are connected to the teachers themselves, in addition to goals that emerged during emergency remote delivery. In connection with assessment, participants would like to give grades to students that reflect their effort and energy put into studying, their development compared to themselves, in addition to their performance and knowledge of a certain section of the material as compared to standards described in the curriculum. The interviewed teachers also aim to utilize assessment that students find fair and understand what is behind a certain grade.

Furthermore, all diagnostic, summative, and formative purposes of assessment appear in the interviews. Through feedback, teachers would like to support students' learning process and development by not only reflecting on students' current level, but also the goal, and how to get there. Students being able to assess their own work and become self-monitoring learners are very important goals within developing students' autonomy, which is the goal interviewed teachers mentioned the most times. Fostering intrinsic motivation, not only through autonomy, but also competence and relatedness, appear as goals for using alternative assessment. Preparing students for further education, the world of work, or lifelong learning in general are further goals, just as developing complex competences and students' 21st century skills. These skills are, for instance, critical thinking, problem solving, communication, collaboration, creativity, innovation, ICT use, initiative and self-direction, social and cross-cultural interaction, flexibility and adaptability, leadership and responsibility, and productivity and accountability skills. Although the introduction of the emergency remote delivery shifted the emphasis on the importance of different goals, supporting students' well-being and their development remained in the center. Details of the specific alternative assessment methods both before and during the emergency remote delivery and their consequences are presented in the next chapter.

5.4 Alternative Assessment Methods and Their Consequences

This chapter details the specific alternative assessment methods described in the interview study together with the consequences interviewed teachers experience in connection with them. Based on the interview study, alternative assessment methods used during the action research are also presented, followed by students' perceptions of these methods in the next chapter (5.5). After some general thoughts on alternative assessment methods as expressed by the participants, solutions for the perceived problems caused by the need to award grades are listed, including the point-collecting systems of the action research and further methods and tools mentioned by the interviewed teachers. Then alternatives for testing are followed by assessment activities that help process the material. In addition, further alternative methods, such as diagnostic assessment, peer assessment, self-assessment, and different types of feedback are elaborated on. Finally, assessment methods involving ICT both before and during the emergency remote delivery are discussed. Each section also presents the results concerning the consequences of using these alternative assessment methods as seen by the participating teachers.

When describing alternative assessment in general, interviewees say that “it is not for everyone, only for those who are student-centered” (Ella). According to Anna: “it is based on a good relationship between the teacher and the students, or at least a basic openness from the teacher toward students”. In Mari’s words, “it is a completely different state of being than the traditional” and “it is difficult to always swim against the current”. Lali also admits that “not everyone has the energy to innovate all the time, it’s expedient to do things the same way”. All in all, the interviewed teachers are aware that these methods need extra time and energy to introduce; however, their application yields several benefits presented in the following sections.

5.4.1 Solutions for Grading

As the context of the research is public secondary education in Hungary, teachers must comply with the rules and give the required amount of grades to students. For this reason, the first section deals with teachers’ alternative methods for grading. All of the interviewed teachers use some kind of point-collecting system originating from assessment based on gamification (see section 2.4.3 in the literature review). They share the basic idea that instead of grades, students should collect points with their activities that are converted into a grade only after a period of time. First, the exact point-collecting systems used during the action research are presented. Then further point-collecting methods are discussed based on the interview study. Using these point-collecting systems has already yielded results in the participants’ practice, so in the last section, their perceived achievements and disadvantages are listed.

5.4.1.1 The Point-Collecting Systems of the Action Research. In order to fulfill the obligation of giving the required amount of grades, during the action research, all three groups of my students received grades based on the following point-collecting systems. Table 9 shows the conversion of points into grades in the different groups. Next to the grades, there are the slightly modified verbal descriptions of the grades (officially it would be excellent (5), good (4), average (3), satisfactory (2), and unsatisfactory (1) according to the Act on National Public Education, Act CXC, 2011). As these descriptions were only visible for my students (only grades appear in the official e-gradebook system), I wanted to express that grade two only means passing (and should not be “satisfactory”) compared to failing, while grade three should not be “average” because to my mind, it refers to the normative nature of grading that I wanted to avoid even on a semantic level. I agree that grade four is good and grade five is excellent; however, for grade five I wanted

to include the idea of mastering a topic and also including a higher category for overachievers to be able to become “Super Masters” of a topic.

Table 9

Converting Points Into Grades in the Groups of the Action Research

11thME	11thAME	9thITE	Grades
from 60 points	from 150 points	from 2000 points	5* - Super Master
46 - 59 points	120 - 149 points	1600 - 1999 points	5 - Master
33 - 45 points	90 - 119 points	1100 - 1599 points	4 - Good
27 - 32 points	60 - 89 points	900 - 1099 points	3 - So-so
21 - 26 points	30 - 59 points	700 - 899 points	2 - Pass
0 - 20 points	0 - 29 points	0 - 699 points	1 - Fail

In the 11th grade mathematics in English (11thME) group, I started using point-collecting methods in 9th grade and made small changes in the following semesters. The points in Table 9 were used because we finished the previous year with these points, and students voted for keeping these grade boundaries in 11th grade, too. In the 11th grade elective advanced mathematics in English (11thAME) group, it was not clear at the beginning of the school year if I had to give them grades or not, and at the beginning of October, it turned out that students had to receive four grades during the school year. Both in the 11thME and 9th grade IT English (9thITE) groups, students had to have at least six grades during the school year. As some students attended both of my 11th grade classes, I used almost the same number of points for activities in both groups, but in the elective group more points could be gathered due to the longer time period for a grade. In the 9thITE group, I used a slightly modified version of the point-collecting system presented by Prievara and Nádori (2018, p. 139). It is slightly modified because there might have been a mistake in the book as grade one goes until 700 points, but grade two starts from 1000 points; in addition, there are only 100 point differences in between the grades from two to four (Prievara & Nádori, 2018, p. 139), so I kept the 700 points boundary for grade two and extended the ranges of grades two to four to include 200 points. In order for students to receive a grade five they needed to collect more points because I wanted to express that mastering a topic requires greater effort than advancing in the previous stages.

All student activities counted into the point-collecting system, so there were no other ways for students to receive grades during the school year. The 11thME students had already known my assessment system, but for the 9th grade group I explained this point-collecting system both in class and in writing (see Appendix G). They could collect points with several activities that are detailed in section 5.4.3.2, Assessment Activities in the Action Research Project. Above the maximum number of points, students could collect extra super master points for which they could either receive grades registered in the e-gradebook (with the same grade boundaries), or they could buy privileges such as a yes-no question in a test (for 200 points), an extra day to hand in work after a deadline (for 400 points), a 5-minute-long chosen in-class activity (for 800 points), a game lesson for that one student or a grade 5 (for 1600 points), a chosen in class activity for the whole group such as a lesson consisting of games only or ordering pizza (for 5000 points), or an out of class activity for the whole group such as going for ice cream during the lesson or an excursion in the afternoon (for 10000 points). These categories and their price were based on my previous experiences.

As the last two categories required so many points that one student could not collect enough, group members had to add their points together, so that the whole group could enjoy eating pizza in class, going for ice cream, or participating in an escape room activity out of class. All of these were thoroughly enjoyed by groups of mine before the action research. Unfortunately, due to the emergency remote delivery, neither in class nor out of class activities could be organized during the action research, which might have made collecting these points less interesting for students. In the 9th grade group six students collected some extra super master points, and they exclusively used them for extending deadlines. The reason for this might be the fact that all point-collecting periods ended with the achievement test of the topic, and all other assignments were due on the date of the test, too. Some students had already collected enough points, so that the achievement test did not have high stakes, but when it did not have the results someone hoped for and only a few points were missing for a better grade, students could extend the deadline and submit something more for the missing points.

5.4.1.2 Further Point-Collecting Methods. The interviewed teachers use similar point-collecting methods to the ones presented in connection with the action research; however, there are differences not only in the conversion of points to grades, but in several other aspects, too. One of the main differences is whether students only get grades through collecting points or not. Bea,

Edit, and Lali introduced their system in a way that students get traditional grades for their achievement tests at the end of each topic, and separately there are additional grades that students can get through collecting points.

In Lali's case, they are called XP-grades (based on the XP abbreviation of experience points from computer games) and the point-collecting periods are called levels. He also uses unit points (UPs) that students collect for compulsory tasks connected to the unit, based on the given material, in a given time limit. Students know in advance how many points they have to collect to receive a given percentage. According to Lali, its disadvantage is that students can only reach 100%, so over achievers cannot go above the maximum points. Experience points (XPs) must also be collected through tasks chosen by the students.

Although XPs and UPs are separately collected, for the year-end grade, students must reach a certain level from both. For example, Lali tells a story of a student whose UP average would have resulted in a grade four, but because of the XP-level, the student only got a grade two. According to Lali, it was a difficult decision to be that strict; however, in retrospect he believes it was worth it. On the one hand, the rules were discussed and available for students and their parents in advance, and there were no complaints about it. On the contrary, students expressed in their feedback forms that they saw the point of this system. On the other hand, there were disagreements from colleagues, which, according to Lali, proved to be unnecessary as at the end of the year all students (except the one mentioned above) performed well. Students also confirmed that this system made them work harder, and in the traditional system they might have been satisfied with a grade three based on tests. The compulsory individual work increased their grades, which provided a great experience of success.

The third element of Lali's assessment system is the credit point system through which students can receive "perks", which are fun elements such as "bonuses, extra credits, rewards, e.g., opportunities to rewrite a checkpoint [equivalent to a pop-up quiz], to double your XPs, to watch a film together, and so on" (Lali). Lali also created cards representing the different elements (see some examples in Appendix H1).

Bea uses a hybrid method based on where she would like to diverge from the traditional. For example, there is a "homework grade", in which students have to collect a set amount of points within a deadline practicing anything they would like to or need to. Bea always provides suggestions, but they are not compulsory, so students have the freedom to practice in different

ways. In another group, she uses point collecting for differentiation, so advanced level students work on materials that fit their level and collect points with them in class (so they get a “class work grade” after a period of time). It also results in advanced students staying in class and still developing (while these students are often exempt from attending language classes). In a third group, students collect points with everything they do, but Bea sets compulsory and optional elements, which is very similar to the point-collecting systems used by Ella and Mari.

Ella’s students have to collect ten points from the set of compulsory tasks and ten points from the set of optional ones. For example, students cannot collect twenty points by only translating lyrics for a month (only the optional ten points), so they also have to collect ten points with an achievement test or a set of language exam tasks (or something complex enough, negotiated with the teacher). Ella calls the point-collecting periods “missions”. In Mari’s system, there are three ways to collect points: class work, tests, and home assignments. She calls the compulsory elements “core materials” that everybody must acquire, otherwise they cannot pass the “level”, and there are additional tasks. First, she also started with 10-20 points, but soon realized that she prefers giving points on an itemized basis, so she offers around 400 points to be collected in a month, about 300 of which should be collected from the core materials. However, she also notes that the maximum number of points to be collected also changes based on the length of a topic (which is known to students in advance). In one of her classes, students asked for a separate grade for the achievement test as they wanted to see their results in the traditional way, too. Interestingly, it is the case for Júlia and Csilla as well. So they also give traditional grades for achievement tests following students’ requests. Júlia calls the point-collecting periods “rounds” and her assessment system the “Freedom of Learning”.

Dóra uses separate point-collecting systems for assessing content and behavior. This way students’ grades in the e-gradebook system only reflect their knowledge, but she still gives feedback on students’ behavior. It is interesting to note that although points for behavior do not result in grades, students are still invested in collecting them, according to Dóra, due to its gamified structure. Students have avatars that they can customize, and this ownership element motivates them (e.g., they take selfies with their avatars and post them on social media). Students’ points can be seen in an open leaderboard, so it creates competition among students. Dóra emphasizes that she uses this only when she assesses their behavior and attitude, but never for assessing products or content knowledge. Another difference is that in this category, she uses negative points as well

(which she does not do in connection with content, either), such as taking away points for “not listening to my peers” or “being off task” and so on. However, she still believes in positive reinforcement, so she gives extra points for “a genius idea”, “being a teamplayer”, “doing my share of the tasks”, and so on. Dóra likes this system because she can customize what behavior she awards, so she can motivate students to act in desired ways. Júlia has also experimented with making a difference between HPs meaning health points (for content) and APs meaning action points (for behavior). She considered it as a failure because as a teacher she did not feel invested enough in keeping track of action points and felt it to be too subjective, so she eventually abandoned collecting action points and kept only health points for content related activities.

In the assessment systems of Anna and Fanni, students also collect points with everything they do. There are no other ways of receiving grades, and there are no compulsory tasks either. A plethora of tasks are offered (all connected to the given topic) with detailed descriptions of criteria including their point values, deadlines, and final grade boundaries, which are all known to the students in advance. As they teach content subjects, students have less freedom in connection with the material (that is the reason for all tasks being connected to the given topic), but students can choose from several ways of dealing with the material. These task types are detailed in section 5.4.3.1, Assessment Activities in Interviewees’ Practice. Both of them admit that designing each point-collecting period in advance with all the different tasks connected to the topics means a substantial amount of work.

Ilona used to do the same as Anna and Fanni (see the very detailed example in Appendix H2); however, due to its complexity, she could not maintain it, so she simplified her system by making all the tasks worth 10 points and not telling students the maximum number of points in advance, only the maximum points that can be collected up until that time. In students’ progress bars, Ilona created “sample students” depicting the current point values for grades 2, 3, 4, and 5 that she updates after each task, so students can compare their points to that of the sample students and continuously monitor their current state. So on the one hand, it is easy for the students to follow their scores as they can see their results in a progress bar always compared to the percentage of the maximum points that can be collected by that time. On the other hand, Ilona allows herself the flexibility to see how much she would like to cover in a given topic or how much fits in the given time limit.

Eszter's point-collecting system differs from all of the ones above in the sense that she only gives students one point limit that everybody should reach while working on their own individualized tasks in a given topic. In addition, this limit is increasing topic by topic as students are able to carry out more and more tasks. As a result, if students work through the month, which is the time frame of a unit, they all receive grade fives. As it is discussed in section 5.1.2, Teachers' Views on Conflicts and Sensitive Issues, the unusual outcome that everybody received the best grade caused some confusion the first time it appeared. One student even started crying as they had never got a grade five in English before. Eszter discussed with them the tasks which they collected the points with showing that it was a well-deserved grade they really worked for. Later students accepted it, and in all of Eszter's groups students did not receive grades worse than four (excluding dropouts who did not receive any grades due to the fact that they left the school).

There is one exception where there is no point collection involved in grading. Fanni teaches biology in English. In an 11th grade group, as this was the students' last semester for learning biology, they did not need the grade for future studies and were not preparing for any exams, Fanni agreed with the students that they would all get grade fives if they attended and actively participated in class. They distributed topics that students were interested in among themselves, created presentations and held interesting conversations about them, which according to Fanni worked well for this group. Lali also entertains the idea of giving everybody grade fives by default and concentrating only on formative assessment, but he has never tried it in practice.

All in all, there are several differences in the point-collecting systems presented, but there are some common elements, too. Students have a variety of tasks to choose from (although the degree of freedom varies), they have several chances to develop, correct their work, improve their grades, and so on. Teachers can also cater for students' different needs, for example, these systems can be customized to provide success for slower students at the same time satisfying overachievers, for example, by collecting more points than the set maximum. The next section details the consequences that teachers experience using these point-collecting methods including the benefits and the drawbacks they encounter.

5.4.1.3 Consequences of Using Point-Collecting Methods. In section 5.3.1 Assessment Related Goals, the interviewed teachers expressed that they would like to give compulsory grades that reflect students' effort and energy put into studying and their development compared to themselves, as well as their performance and knowledge of a certain section of the material.

Students receiving grades through point-collecting methods can achieve these goals. According to the interviewed teachers, anybody who is hard-working has the chance for good grades through point collecting. They all report that the results of their students got better after introducing point-collecting methods. As Ilona claims, “those who work accurately and regularly will get a better grade, not necessarily because of their knowledge, but because of their work and knowledge together”. However, she also admits checking students’ achievement tests to see that if they were graded, would the grade be different from what students get after the end of the point-collecting period, and if she sees big differences, she modifies the rules of point collecting, so that grades more accurately reflect students’ knowledge as well. There are similar reasons behind the mandatory components of the point-collecting systems, so that students could not get good grades unless they acquire a given minimum. According to Lali, students realize that it is a complex learning management system, in which they have to work in order to succeed. All the interviewed teachers agree that one of the greatest challenges in creating such a system is determining the point values of tasks and the rules of the system, so that it achieves the above mentioned goals.

When asked about the effectiveness of her point-collecting system, in the first interview Mari responded that the objective measurement would be students’ results of their upcoming secondary-school leaving matura exam. So in the follow-up interview, three exam periods later, I asked about these results, and Mari confirmed that all of her classes performed well in the exams according to her expectations, not better, nor worse than other classes. Eszter also mentioned the matura exam in connection with a group that had the skills but lacked the motivation to do anything such as reading through a longer text because they lost their patience. According to Eszter, her gamified assessment system could motivate her students, so eventually all of them applied for the matura exam in advance, a year earlier than finishing secondary school and passed it with good results. In addition, all of Eszter’s dyslexic students collected points and thus received grades even though they were officially exempt from grading.

Creating a fair assessment system has also been mentioned as a goal in all interviews. Teachers ask feedback from their students and modify the system accordingly until they all find it fair. This flexibility also helps in fine tuning the system and reducing stress for teachers as well by providing opportunities for experimentation and development. In addition, it also encourages students to take initiative: “My students have become more proactive. They have ideas on how to collect points, what else they can do, and so on” (Ilona). According to Ella, for those students who

already use English confidently “any decent language education would lead here, but what is really conspicuous is that students who lack self-confidence, and would otherwise just wait for the lesson to end, are beginning to be active and become more autonomous” which Ella considers the greatest achievement. Similarly, Bea says that “I could see how their attitude changed from ‘tell me what to do’ to ‘oh, I’ve got an idea, I’ll do this’ and this is the genius of this point-collecting system”. Lali also adds that his students “tried out new things that otherwise they wouldn’t dare to because here there are no punishments if they are not doing so well, so they are not afraid to experiment”.

Using these point-collecting methods also results in the development of learner autonomy that became evident during the pandemic. All the interviewed teachers who had groups learning through these point-collecting systems for various time periods encountered the same phenomenon. The longer a group had been studying in this system, the more autonomous they became, which made it easier for them to face the challenges of the remote emergency education. The interviewed teachers also experienced a difference in the level of students’ engagement. They concluded that “it’s easier to motivate students with this system” (Fanni). Fanni illustrates this with one of her students’ story: “I had conflicts with her in the beginning, but eventually I could motivate her, which was a really good feeling, especially because my colleagues were shocked that she received a grade five and asked if she really did the tasks” (Fanni). Eszter brings up a dropout student who quit school and probably will never finish secondary education, but in connection with this student, she expresses the following:

I could still see the development in how much they wanted to do something, and I think assessment based on gamification helped in it, which they can use in other areas of life. So they might never learn English, but the experience that learning can be liked might motivate them in the future.

Eszter believes that this system keeps students motivated because it addresses both their short- and long-term attention, which can be observed by how engaged students are during the lessons (short-term) and how excited they are about the monthly changes (long-term). In connection with Ilona’s leaderboard where students can continuously compare their points to that of the “sample students”, she also experienced increased engagement. Students followed it, and when they fell behind, there could be immediate intervention, for example, they asked what tasks they were missing or what else they could do to make up for the missing points.

Some of the participating teachers' students also shared some of their reservations and negative opinions when asked to evaluate the point-collecting assessment systems. On average, there is about one student in each class who prefers traditional assessment and they have different reasons for their preferences. One of Fanni's students writes that "it's too easy to get a grade five". Fanni speculates that as it is written by an eminent student, others also receiving good grades might have their pride hurt. One of Lali's students comments that "this system is too complicated". Others also report students who do not understand the point of the system and get lost in its complexity. First, Eszter dedicated two weeks to the introduction of her gamified assessment system; however, she experienced that students felt flooded and overwhelmed by information, so she had to slow down and dispense information in smaller doses over a longer period of time. One of Edit's students is dissatisfied due to the following: "if I'd had to write only tests, I would have a 5 now, but I got a 4 because I didn't do all the tasks". One of Csilla's students expresses a similar thought: "if I do everything well, why do I have to work more?" These criticisms also appeared in Prievara's (2015) findings. In conclusion, all the interviewed teachers agree that most students like the point-collecting systems, which is also proven by the fact that all of their groups voted for continuing their studies in these assessment systems, and they address these criticisms individually.

The interviewed teachers also experience larger scale consequences among colleagues, in the school, or even on the level of the school district. All the interviewed teachers mention that some of their colleagues are inspired to experiment with alternative assessment methods as well. As Mari frequently requested absences in order to give presentations and workshops about her assessment system in other schools and conferences, her boss asked her to hold one for their own colleagues, too. First, Mari opposed the idea of imposing anything compulsory on their colleagues; however, her boss convinced her by saying that "it is not compulsory to do anything, but it is to listen to it". After her presentation some of her colleagues also adapted elements of alternative assessment. The same happened in Ella's, Eszter's, Dóra's, Júlia's, and Lali's school. In fact, Lali together with Edit and Csilla formed a gamification working group regularly meeting and discussing their experiences. In Dóra's and Ella's school, even the school's pedagogical programme was modified to include a grade category that students receive as a result of point collecting. In Dóra's school the following addition was entered (written by one of Dóra's colleagues who also uses point-collecting methods):

A topic grade is equivalent to the weight of an achievement test awarded at the end of a topic based on smaller written or oral grades, partial credits, or collected points. Teachers must inform students about the specific details of such assessment at the beginning of the teaching period.

The text written by Ella in their pedagogical programme is the following:

Teachers can use different methods for grading registered in the gradebook. In the so-called gamified assessment system, students' performance is assessed through points that are converted into a grade at the end of a learning period. All grades registered in the gradebook this way encompass assessments of diverse tasks aggregating and compensating each other. Assessed tasks and their point values are determined when planning the learning period and registered in the gradebook as a double grade when closing the period. During a term at least three grades should be registered taking into consideration their even distribution. The criteria system and rules of the alternative assessment are thoroughly explained to students at the beginning of the school year.

Teachers provide continuous feedback to students and their parents through messages in the electronic gradebook and by making the point-collecting chart available.

In order to change a school's pedagogical programme, the majority of teachers in an official meeting should vote for the acceptance of the exact changes, so in these schools alternative assessment methods were presented to all the teachers of the school. In Mari's school district, the following official change has taken place: at least one compulsory homework assignment each term must be counted in the grade of the achievement test, which they call a "project". According to Mari, its naming is a huge mistake as projects methodologically mean something completely different. She is still happy that even though it is a small step, it is a step away from the tradition that assessment means giving grades for tests that are mostly achievement tests. The next section presents the interviewed teachers' alternatives for testing.

5.4.2 Alternatives for Testing

Traditionally, students get grades for their test results. As it is detailed in connection with traditional assessment, the interviewed teachers have several problems with traditional testing and try to find alternatives for it. First, these alternatives are presented and discussed. For the interviewed teachers, the primary goal of testing should be for students to show what they know in a particular topic. For this reason, alternatives to testing also appear, so methods used for

summative purposes instead of testing are presented in this section as well; followed by the consequences of using these methods that the interviewed teachers experience. The chapter ends with a description of a selection of these methods that were applied during the action research. Students' perceptions are also gathered in connection with the methods used during the action research and can be found in chapter 5.5, Students' Perceptions of Alternative Assessment Methods.

5.4.2.1 Alternatives for Oral and Written Tests. The interviewed teachers find traditional oral testing problematic for the following reasons. Choosing a random student from whom the teacher asks random questions about the material in front of the whole class causes incredible stress for students. In addition, "after the unlucky student is chosen, the rest of the class is relieved, and the better scenario is that they don't do anything, or the worse case is that they take the classroom apart" (Anna). The first problem (choosing students randomly) can be avoided by students volunteering for the oral test, or they can be informed in advance who will be tested when and can prepare accordingly. The subjectivity of what teachers would ask during an oral test can also be changed if students know the criteria in advance. Anna uses pair interview tests in which students interview each other about the material, and they also have some preparation time, so they can discuss the "interview questions". In order to involve the class, they set it up as a "talk show" where "the audience" can also ask questions (Anna). Edit involves the class through students writing feedback to each other based on a set of criteria. Students can also get tasks (such as taking notes) during the presentation, which leads to alternatives to oral tests, for example, instead of calling them oral tests, they can be students' presentations. Presentations can be pre-scheduled and based on set criteria providing numerous further opportunities. For instance, students can create an outline for their peers to follow and possibly answer comprehension questions as done by Edit. Students can create visual aids supporting their presentations (which element appears in all interviews).

The interviewed teachers mention further summative oral methods that diverge from testing: simulations (of exams or work-related situations such as a job interview), role plays, monologues, dialogues, and drama activities. For students who would not like to perform in front of their peers, there are also solutions mentioned by the interviewed teachers. Eszter listens to her students' presentations one by one at their desks while the others work on their individual tasks. Students can also record their presentations and send them to the teacher (Anna and Ella call them

video tests, while Lali calls them vlogs). These recorded presentations together with live online presentations by students have been used by all the interviewed teachers since the emergency remote delivery was introduced.

Traditional written tests usually have two main types. One is a pop quiz (see translation in Appendix A) that focuses on recently learned material, does not take the whole lesson (can last approximately between five to twenty minutes), and is often unannounced. The other is the achievement test at the end of each unit (see translation in Appendix A), which lasts the whole lesson and has to be announced in advance (see section 2.1.3, The Pedagogical Programme of the School of the Action Research Project). Similarly to random oral tests, instead of unannounced pop quizzes, the interviewed teachers prefer if students know about them in advance and can prepare accordingly. Further decreasing the stress, pop quizzes are always worth points but never individual grades in the interviewed teachers' point-collecting systems even if achievement tests are graded separately. Júlia calls them "point winners". Mari tells the story when her boss visited her lesson and said that "'basically your students were writing a pop quiz', which was funny because neither I, nor the students felt that way, but in fact, you could call it that". Another alternative for pop quizzes that the interviewed teachers applied even before the pandemic was online quizzes in class such as Kahoot!, Quizizz, Quizlet, Socrative, Redmenta, and so on (see their links in Appendix B).

In connection with achievement tests, numerous alternatives appear in all interviews. There can be alternatives for how the tests are administered and how they are assessed. Moreover, there are further summative methods that can replace achievement tests entirely. An alternative for how tests are administered is students writing tests in pairs or groups. For example, Ilona makes her students write a test first individually, then they get the same test again in a group, after which they can go back to their individual test and modify it if they want to. Students can also write a sample test before the "real test", or they can create test items, questions, and tasks. that are included in the achievement test of the class. Furthermore, students can compile the whole test for themselves individually or in pairs or groups. In relation to assessing achievement tests, if the test is created by the student, the task could include creating the answer key for the test, too. The class can also correct tests together (students can correct their own tests or the tests can be randomly distributed among students and they correct each other's). The test can be assessed based on one criterion only, for instance, in the case of essays, assessing the use of a specific grammar structure only. An

achievement test type can be when tasks are in order of increasing difficulty, which also makes differentiation possible (e.g., at the end of the test, there are optional challenging tasks). The interviewed teachers emphasize the importance of feedback, so when correcting tests, not only grades, points, correct or incorrect answers are marked, but they also add praise, encouragement, explanation, advice on how to improve, and highlight things that are good. Further methods that interviewees use for summative purposes are essays, translations, extended reports, reflective journals, project journals, further project work, creative (individual or group) products, posters, and portfolios. An additional category is achievement tests presented in a gamified framework such as a board game (for which students can create task cards, questions, or even design the whole game), treasure hunt, mystery game, quiz show (for example, Jeopardy! or Who wants to be a millionaire?), or an escape room activity (in class or online).

Teachers apply these alternatives in order to minimize the negative impacts of traditional testing while still providing students with opportunities to present what they have learned, as well as fulfilling the goals of summative assessment: identifying overall levels of achievement and measuring students' results against them as suggested by Rea-Dickins (2000). The next section elaborates on the consequences the interviewed teachers experience when using the above mentioned alternative methods.

5.4.2.2 Consequences of Using Alternatives for Testing. As a main consequence of using alternatives for testing, all interviews mention the attenuation of stress in students. As Ilona phrases it:

I can see that the majority of students are relieved, so they are not stressed because they know that even if they make mistakes, things can be done about it, and it doesn't result in a bad grade registered immediately, their parents seeing it, and so on. So they take tests easier without that usual stress.

Bea also emphasizes that if a test is not the end of everything, students still have opportunities to correct their mistakes, most students use these chances, correct their mistakes, and send her corrected versions even if it does not change their final grade. So if the stakes of tests are not high, it can reduce stress.

There are several consequences of the fact that students have different ways of presenting what they know. Fanni's students express that they enjoy the freedom and flexibility. Lali's students write in their feedback that the tasks make them work with the language, use the language,

which they really enjoy, and Lali hopes that students can see that besides having fun they are also developing in the meantime. According to Dóra, students do see how they are developing, and she tells the example of an autistic student:

He was the communication manager of their project. He also has several learning disabilities, so this role was a huge challenge for him. [...] At the end of the project, he wanted to speak with me privately and asked me to involve him in all of my projects because he could see how much he developed and learned through this.

Teachers also have more flexibility in differentiation and providing students with individual learning paths. During the emergency remote delivery, Mari had to talk to some of her students individually. She told one of her students the following: “you see that the others have done 20 assignments and this is your second, so don’t be scared, but let’s discuss what, when, and how you can do”. Mari created individual learning plans together with her students that they called “contracts”, and as a result, students always did what they had agreed on this way. Mari adds that it might not have resulted in a grade five, but in those cases it was not the plan either. Eszter had a similar experience (even before the pandemic) that “there was nobody with whom we wouldn’t have found a way to develop somehow”. According to Ilona, if students experience that they are in control of their results (as their grades do not depend on a single good or bad performance), they become more proactive and use the chances to learn and show what they know in different ways.

All the interviewed teachers celebrate that the bad habit of cramming for tests is replaced by continuous work with the material and an outcome that students are happy to share. Some instances are the following. In connection with using projects, Dóra’s students write feedback on their exit tickets such as “this gave me trust, so I dare to step out of my comfort zone” or “I’m proud of my achievement” or “I’ve learned so much”. Bea’s students enthusiastically bring their own ideas for their individual projects. When Júlia’s students create their Quizlet vocabulary tests for themselves, they appreciate the freedom that they can achieve the result they desire. As Júlia reports, some students create a screen recording of doing a 100% test (probably they do it as many times as needed for a perfect result), while others send a test including mistakes. For Ella’s oral tests in groups, students create study groups for themselves and practice together for their desired outcome. Eszter’s students like different task types, and for example, one group especially enjoys picture descriptions based on their own drawings, which resulted in developing new habits in the school: they dedicated a place to exhibit students’ drawings. Fanni has a similar experience that

students creating posters for summative assessment developed a lot of skills (such as their creativity, language, organizing skills, and so on). She can see how the quality of students' posters have been increasing unit by unit (based on the feedback she gives on students' work).

Differentiation in summative assessment allows students with different skills to all experience success and can also cater for the needs of overachievers. There are one or two students in each group who enjoy performing above the required maximum. One of Lali's students already collected the maximum number of experience points and still submitted further homework assignments reaching a 140% total. Similarly, Edit's student did all the five compulsory tasks and handed in three more so that the total goes above the 100% maximum. These students also enjoy competition among themselves. Some of Bea's students place bets on how many points above the maximum they can reach. Júlia can see how many times her students' fill in a WordWall test (or other online tests), and some students do the same tasks several times over and over again to either make it perfect or to compete with the others on the leaderboard. Even though grade fives have been registered, some students still send corrected versions of their work to Bea. These are clear indications of students not doing something for the grade.

A last consequence from the teachers' point of view is that assessing diverse students' work is more enjoyable than correcting the same tests. According to Fanni, the fact that there are no two assignments that would be the same protects her from burnout and makes her job more fun as well. Dóra also describes how much she enjoys listening to students' presentations of their projects, and similar enthusiasm can be detected in all the interviewed teachers' description of their students' performances.

5.4.2.3 Alternatives for Testing During the Action Research. During the action research, as mentioned in connection with the point-collecting systems, all of the following listed activities were worth points. Going in the same order as the methods presented above (in section 5.4.2.1, Alternatives for Oral and Written Tests), in relation to oral tests there were voluntary oral tests, scheduled student presentations both in class and, after the introduction of the emergency remote delivery, online, while the not presenting students could also collect points by taking notes on the presentation, and live online and recorded presentations, which, similarly to Lali, I also called vlogs.

There were no unannounced traditional oral tests, so students could either volunteer or prepare knowing the assessment criteria in advance. In the beginner 9thITE class, first we threw

Wendy, a plush toy, around, and the student who held Wendy had to speak, so the group collected words, expressions, then sentences together. In the beginning, there were only voluntary oral tests in which students received 10 points for each sentence they said in English in connection with the material. It was up to them how much they wanted to speak, so they could say even just one sentence. Grammar mistakes were not penalized, but afterwards we discussed the correct forms. According to my teaching journal, students gradually spoke more and more. In the second term, we introduced presentations. Students also volunteered and chose the date and the topic in connection with the unit that they wanted to present about. They could also present in pairs or groups of three. We discussed what makes a good presentation and introduced a peer assessment sheet (see Appendix I8). If students prepared visual aids and talked for at least five minutes, they received all the points, so the number of points did not depend on the quality of the presentation. In addition, they also collected their peers' and my formative feedback sheets. On average, half of the students volunteered during each unit to give presentations.

After the introduction of the emergency remote delivery, students could either present during our live online lessons, in which powerpoint visual aids could be replaced by illustrations through screen sharing, or they could record themselves and send me their vlogs. One of the greatest surprises of the action research for me was when students gave presentations about their favorite video games. They shared their screens, launched the games, explained how they worked, and narrated gameplays. Even though they were a beginner group, students talked fluently for 20 to 30 minutes each time. Others had to take notes and collect useful expressions. After each presentation we discussed how different vocabulary items could be used in other contexts outside of the video game. For example, they used the word "skin" as a graphic element they could download to change the appearance of a character, through this, they could learn its everyday usage, too. Instead of the peer assessment sheets on paper, we gave formative feedback to the presenter going through the criteria in speaking.

In the category of written tests, there were "mini point-collecting opportunities" that were simply called minis, meaning preannounced pop quizzes, vocabulary quizzes, online quizzes such as Kahoot! both before and during the emergency remote delivery, sample tests, tests with increasingly more difficult tasks, students writing test questions that are incorporated into everybody's tests, and students compiling achievement tests for themselves (see samples in Appendix I). Minis usually lasted for about 10-15 minutes and included two to three questions in

connection with the material of the previous lesson(s). The aim of the minis was twofold. On the one hand, they gave feedback to students on how well they knew that material. On the other hand, they had diagnostic purposes showing me how much I could build on that knowledge, what elements were missing, what had been misunderstood, and so on. Samples of minis from the 11thME group can be seen in Appendix I. Vocabulary quizzes in the 9thITE group had similar functions: motivating students to learn expressions that were especially important in relation to IT. As the American teacher trainee also noticed during a class observation:

The vocabulary quiz consisted of several translation tasks, giving definitions for English words in English, and recalling information from the two presentations that day. I was surprised by the difficulty of the vocabulary being assessed. All of the terms were Internet-related and highly domain specific, e.g. modem, ISP, etc.

The mentioned vocabulary quiz can also be seen in Appendix I. In addition to offline vocabulary quizzes, students played Kahoot! (see link in Appendix B) during lessons. In class, it looked like the following. Multiple choice questions were projected, and students had to choose the correct answer on their own device within a time limit. After each question, we discussed the solution, and the program showed a leaderboard with only the first five places. Students could individually see their own place on their own device. Regardless of the leaderboard, I gave points for students based on their number of correct answers. As students seemed to really enjoy playing Kahoot! in class, I created one occasion each unit for which students could also submit questions to be included in the game (more details in section 5.4.3.2, Assessment Activities in the Action Research Project).

In the 11thAME group, students could also collect points with a sample test the lesson before the achievement test. I introduced sample tests for two reasons. On the one hand, it was the first time for the students to meet advanced level math problems, so I wanted them to practice higher level matura exam tasks with low stakes first. On the other hand, it was the first time for some students to learn math in English. As I did not want to create an achievement test in two languages, I wanted to familiarize them with the language of an advanced test in time. In addition, I also offered them to answer any language related questions during the achievement test as well. According to my teaching journal, all students seemed to cope with the language, and there were only a few comprehension related questions clarifying if they understood everything correctly. In the 9thITE class, in person Jeopardy! quiz shows had a similar function: practicing tasks similar

to the ones in the achievement test the following lesson. I wrote titles and point values on the blackboard; more difficult questions were worth more points. Students played in groups of three or four. Taking turns, each group could choose a topic and a point value corresponding to the difficulty of the question. If they answered correctly, each member of the group received the number of points. If they did not answer correctly, other groups could “steal” their points. During a 45-minute lesson, we covered and thus summarized the whole material of the given unit.

Increasingly more difficult questions also appeared in the achievement tests of both the 11thME group and the 11thAME group (see sample from the 11thAME group in Appendix I3). My main goal was differentiation. As both groups contained extremely diverse students, I wanted to offer challenges to everyone. In addition, I discussed with students that the first set of problems were the compulsory, core material according to the curriculum that everybody should be able to solve, and “bonus” problems labeled with plus signs were optional, more challenging tasks. According to my experiences, the strategy achieved its goal as most students only solved the core tasks, one or two students each time solved only the challenging problems, and some students solved from both categories. Students’ opinions are discussed in section 5.5.2.1, Reactions to Alternatives for Testing.

In the 9thITE group students had the chance to compile achievement tests for themselves (see sample in Appendix I4). Each unit, we agreed on the criteria, such as in the first unit “it had to include at least 20 full sentences and all the collected parts of the computer” (quote from my teaching journal). They had to send it to me, including a key how they meant to solve it, by a deadline we agreed on in advance, so that I could check, correct, and send back their achievement tests which then they were writing together with the others who did not choose this option and were writing the achievement test I created. On average half of the group (seven or eight students from the group of 16) chose to compile achievement tests for themselves. I hoped that this way students would spend more time with the material and also have the chance to present their knowledge in a way they preferred. I had the same aims with presentations and posters (see sample in Appendix I9). Before the introduction of the emergency remote delivery, students created posters on paper, while after it, posters also became digital. In addition, instead of students compiling tests for themselves, I created online escape-room-themed achievement tests that will be detailed in section 5.4.5.3, Alternative Assessment Methods of the Action Research During the Emergency Remote Delivery.

5.4.3 Assessment Activities That Help Process the Material

The next category of alternative assessment methods includes activities that help students process the material. They are categorized into alternative assessment because students can collect points with these activities, so on the one hand, they also affect the grade that students receive, but on the other hand, they provide opportunities for teachers to reach further goals they set when using alternative assessment methods. Elaborating on all the activities that interviewees use (even with the restriction on what students can collect points with) is beyond the scope of the present paper. Therefore, this section presents only two sets of activities: 1) the ones that the interviewed teachers mention to illustrate how they reach their goals by using alternative assessment methods and what consequences they experience while using these assessment activities and 2) the ones that were used during the action research (see section 5.4.3.2, Assessment Activities in the Action Research Project) and yielded data on students' perceptions as well (for which, see section 5.5.2.2, Reactions to Activities That Help Process the Material).

5.4.3.1 Assessment Activities in Interviewees' Practice. As described in section 5.4.2.1, Alternatives for Oral and Written Tests, interviewees' students can write test items or compile whole tests for themselves, and they can collect points by creating practice materials, too. On the one hand, the interviewed teachers claim that "students can have fun" while creating these tasks for themselves in order to practice the material. On the other hand, these tasks can be given to other students as well, so solving (and possibly correcting) each other's tasks results in an even more varied learning experience. For this purpose, Mari mentions the website Learningapps (see link in Appendix B) where several templates can be filled with content material. For example, there are grouping, ordering tasks, filling in a text, creating audio or video subtitles, "Who wants to be a millionaire?" themed multiple-choice games, word guessing games, memory games, word grids, crosswords, and so on. Lali adds word finders, flash cards, and matching games. Eszter lists encrypted messages, hangman and activity game versions. Ella makes her students write exam items in pairs and then role play the exams (with the items written by other pairs). According to her, after a while students get so good at writing exam items that they can organize full exam preparation weeks based exclusively on contributions by students.

Teachers offer a variety of tasks types out of class that students can choose from such as solving puzzles, riddles, drawing illustrations, cartoons, creating videos, subtitles, practice with online quizzes, translations, essays, compositions, task sheets, interviews, experiments, reports,

posters, creative project work, and so on. Offering such variety seems to result in the following positive consequences. The interviewed teachers claim that students are more motivated by experiencing autonomy through their freedom of choice. Thus they enjoy the learning process more. In addition, they have the chance to experiment and see which learning methods work for them and learn accordingly. As Fanni says “they can do what they are good at and that makes learning fun for them, so I think that the most I can hope for is that my students enjoy learning”.

Moreover, as students can also suggest task types or include their interests and activities in the point-collecting system, it has further advantages. Bea and Júlia find it much more efficient if it is students who choose which words to learn for themselves. “I can see that they have worked with these words, and they learn them differently from the ones I tell them to learn” (Bea). Ella highlights that she “could immediately see what movies they watch, what they read, what they listen to, and so on”. According to her, students enjoy sharing their lives, and this personal connection also drives them towards autonomy. “I can already see how independently they use the language” (Ella). Júlia notices the advantage that students have a “better momentum to look for English outside of class”, for example, “they get used to watching series and reading in English, which makes language learning sustainable, which is the foundation of lifelong learning”. She experiences it even with beginners that “they start watching films in English and connect that freetime activities can also be used for learning”.

Ilona wanted to achieve something similar with the “interactive notebook”. However, as detailed in connection with traditional assessment, the idea to look for opportunities to learn the material outside of school as well seemed to be an extra burden for her students. Csilla also hoped for more activity in connection with students choosing extra tasks for themselves. Based on the interviewed teachers’ experiences, it seems that students can perceive tasks outside of school as compulsory homework with all its negative connotations, even if the teacher thinks of the tasks as creative or interesting. Interviewees’ students do not seem to mind extra out of school tasks when teachers include such activities in the point collecting that students have already been doing in their freetime and students’ task is to notice possible connections to the material. The reason might be that out of class activities and homework are often too overwhelming for students after spending their whole day in class. Mari believes that

since students have to stay in school for 35-37 hours a week, I can’t give them 40 extra tasks because they can’t do them even if they want to, so I don’t want to further frustrate

them with this. For this reason, I started consciously planning that students could collect most of the points during lessons through class work and tests, and home assignments are only possibilities to request help.

Students' freedom of choice also creates "a more trusting environment in which we get closer to each other or at least the gap between us can get smaller" according to Júlia. Mari also adds that "it builds a better relationship". All the interviewed teachers find it crucial to have a good relationship with students that is manifested in their continuous communication with each other (for example, about task types, likes, dislikes, modifications, etc.). According to Anna, her adjustments of the system, assignments, points values, and so on, based on students' feedback work very well. Eszter mentions an extreme case of a student with selective mutism. The student was speaking in ninth grade, but for unknown reasons stopped talking and has not spoken in the school for more than a year (until the time of the interview). Eszter communicated with them in writing and as they did all the non-speaking tasks, collected the set amount of points, they received grades. Eszter always offered the chance to speak but did not pressure it and accepted alternatives in writing that the student chose. They maintained a good relationship.

Besides alternatives for testing (such as a board game achievement test), for class work Mari incorporates task sheets and "knowledge organizing tasks", such as mind maps or comparison charts that students can collect points with. Using this point-collecting system enables her to include more task types in class as well. Such as small creative tasks (e.g., creating illustrations, writing chat messages between two literary figures, or alternative story endings etc.) that otherwise she would not include because "you can't give a grade five for such a small task, but you can give two points for it, and students will immediately do it" (Mari). As a consequence, teachers also have greater freedom in experimenting with tasks and determining what students can collect points with.

This also leads to a negative aspect that the interviewed teachers detect: collecting points often remains a source of extrinsic motivation for students. According to Ella, "now students don't ask if a task is for grades, but if it's for points". So teachers try to balance what students can collect points with. Sometimes the interviewed teachers offer tasks, upload materials, and so on that are not for points, and according to Edit, only a few students have the intrinsic motivation to deal with them. Lali also confirms that some of his students develop the intrinsic motivation to do tasks for their own sake, but it is definitely not the case for everyone. According to Júlia, it greatly varies among the different groups and students. Mari agrees and supports it with students' feedback that

they are often motivated by completely different things, such as some students writing that “I loved Sutori” while others: “I don’t want to do Sutori ever again” (see link in Appendix B) or “I enjoyed the amount of autonomy we had” versus “I wish there were more explanations from the teacher” and so on. Mari draws the conclusion that “this shows how diverse learning is, and I was happy that students could write about their learning methods maybe for the first time and formulate things that we can work with”.

Mari uses students’ feedback as a source of information, so that she can help students find what works for them. Dóra also mentions that some of her students are crazy about Rebus puzzles, while others are not motivated to solve them at all (see an example in Appendix H3). Júlia talks about similar ideas in connection with extracurricular activities: she had two American students volunteering for after class activities. From one group there were two students who attended, and as they had only been studying English for two years, these weekly discussions with native speakers developed their skills immensely. From another group, there were four or five students attending who were more interested in cultural topics. Júlia believes that the fact that these were not compulsory classes made students with intrinsic motivation join and learn much more than if it had been imposed upon a whole group.

Lali tried to combat points being extrinsic motivators with a project in which he separated the gamification of processes from assessment. He included numerous game elements that made the lessons more playful. On the one hand, students reported having fun, enjoying the lessons, and according to Lali, it was a team building experience. On the other hand, students felt that these elements were not contributing to their language learning, and as it did not help students reach their goals, Lali gave it up and instead developed his gamified assessment system presented in section 5.4.1.2, Further Point Collecting Methods.

All in all, these examples seem to indicate that intrinsic motivation can be fostered by developing student autonomy, thereby achieving this goal of the interviewed teachers (as discussed in section 5.3.2, Teachers’ Goals in Connection With Learning and School Subjects). In addition to autonomy, Deci and Ryan (1985) mention relatedness as another factor developing intrinsic motivation that interviewees also wish to increase through group work as shown by the data above and also in the next paragraphs.

The interviewed teachers apply and assess group work in various ways. On the one hand, all of the interviewed teachers incorporate several communicative tasks to develop their students’

communicative competences. On the other hand, they also find it crucial to develop their students' social skills including cooperation. For these reasons, they experiment with different types of pair and group work. Enumerating them all is beyond the scope of this paper, so there are only some examples listed to illustrate their variety, especially in connection with their ways of assessment. Edit mentions the cooperative group work described by Kagan and Kagan (1994), in which students on different (in this case, language) levels work together in heterogeneous groups. In other cases, for example, in Eszter's groups, students on the same language level work together in homogeneous groups but with different roles. Ilona employs jigsaw learning (such as students forming homogeneous expert groups first, for example, based on their grammar or vocabulary skills, and then heterogeneous groups) in which students teach each other. Role playing is used by almost everybody from simple tasks, such as acting out a dialogue, to complex drama activities in English. Ella uses a mentoring system in which students help each other in an organized setting taking place both in class and after school as well. Teachers also apply study groups in which students automatically work together for a set amount of time (that can last from a given set of tasks to a whole school year). Ilona's students thoroughly enjoy working in these groups, and they are always relieved when they can work on something together instead of individually in class. According to Ilona, they encourage each other and appreciate the freedom of progressing at their own pace. A further advantage she reports is that students who otherwise do not spend time together also get to know each other. In terms of effectiveness, Ella and Ilona both say that students preparing for exams together seems to be highly effective. Dóra organizes project weeks that students always look forward to, and in their feedback they express that they wish they could learn like this all the time.

Assessing group work also happens in diverse ways. During the project weeks, students create collaborative project portfolios and present their work, which is then assessed by an expert board based on preset criteria. In connection with other tasks, there can be formative discussions between the teacher and the students when they talk about what went well, what to do differently, and so on. The interviewed teachers also ask for students' feedback in anonymous questionnaires, so students can express issues that otherwise they might not speak about face to face. When students collect points with group work, it can be administered also in different ways. For instance, Edit gives a set amount of points to a group that members have to allocate among themselves. Sometimes it reflects how much work each student put in, but mostly students distribute the points

evenly among themselves or even give more points to someone who needs it for a better grade, which, according to Edit, shows a kind of “comradeship that is an important skill as well”. Students can also receive the same number of points (or grades) together as a group. Ilona shares how happy students can be getting a good grade that they would not have been able to achieve individually. Finally, there can be communal point collecting when students’ individual points are added into the points of their “clan” (Lali). In Lali’s case, leaderboards only show the clans’ points, so it is not detailed how much individual students contributed. Bea adds that in one of her groups, students created competitions among each other to see which group can collect the most points in the given time period placing bets in chocolate, which Bea regarded as a “collateral positive effect”. Mari also gives an account of students organically creating positive competitions through which they motivate each other.

5.4.3.2 Assessment Activities in the Action Research Project. From the category of alternative assessment methods that help students process the material, this section presents activities that were used during the action research project and yielded data on students’ perceptions as well (for which, see section 5.5.2.2, Reactions to Activities That Help Process the Material). The different activities can be categorized into two groups: what students could get points for and what they could not. Similarly to the interviewed teachers, I also asked students’ opinions about group work in general and also about specific activities during the action research that they either collected points with or I was interested in students’ reactions for different reasons.

Beside alternatives for testing mentioned above (in section 5.4.2.3, Alternatives for Testing During the Action Research), students could collect points by playing Kahoot! (see link in Appendix B) both in class and at home. For homework practice, they could go through the questions infinitely many times, for this reason in order to receive points for it, they had to do it until all their answers were correct. Students could also write Kahoot! questions that are multiple choice questions with one correct and three incorrect answers. First, they received points based on the number of questions submitted. Soon I had to set a limit for the maximum number of questions written by a student so that they would not collect all their points through this one activity that turned out to be extremely popular among some students. I corrected their questions and answers and sent them back saying that next time they could only get points for them if they did not make the same grammatical mistakes. I also selected the ones to be incorporated into the games that the group played together.

In the 11th grade math classes, students could practice the material at home with Khan Academy (see link Appendix B). Khan Academy is a website that includes short video explanations organized into topics based on the American curriculum of different subjects such as mathematics. It also includes supplementary practice exercises and the possibility for teachers to create groups and follow their students' progress. I assigned the practice exercises to my students that were connected to our unit. I gave points only for perfect solutions as students could go through the exercises as many times as they needed to, in addition to asking for built in hints, or watching the corresponding video explanations. The website is also gamified including several game design elements such as points, badges, leaderboards, avatars, progress bars, content unlocking, and so on (see section 2.4.3 Assessment Based on Gamification).

Further individual point-collecting opportunities were solving task sheets, making notes about each other's presentations, listening tasks both in class and for homework, and essay writing tasks only as homework. According to my observations, the fact that they could collect points with these tasks increased students' engagement; however, I was also interested in whether students could see any other benefits of doing these tasks. Their opinion will be presented in section 5.5.2.2, Reactions to Activities That Help Process the Material. Students could also collect points in groups. For example, creating a poster with given roles or a treasure hunt in which students had to go to different "stations" in the classroom and solve tasks there as a group. They had to write their names on their task sheet and everybody in the group got the same number of points based on the group's results. Students also worked in pairs or groups without collecting points with their activities. I was interested in their opinions about the different tasks and also about working in groups in general.

There were a number of further activities that students did not collect points with. Students' opinions on them are detailed in section 5.5.2.2, Reactions to Activities That Help Process the Material. Activities that the whole group did together were brainstorming about a topic by saying words, expressions, or sentences in English while throwing Wendy, a plush toy, around, and different ways of summarizing such as creating a mind map together (see Appendix I5) or working in a shared Google document to collect all the material of the unit. Students worked in pairs when they had to correct their mistakes while getting through a mistakes maze (see Appendix I6), and there was a reading comprehension task in which students had to teach the material to each other after working with the text individually. I also asked students' opinion about the instance they had

to submit their individual answers through the website Socrative (see link in Appendix B) in class when the lesson was in a computer room before the emergency remote delivery. In the mathematics in English group, I also asked students' opinions on the amount of time spent solving problems and discussing problems together in class.

During the action research project, my main goal with these questions was to gather data on students' perspectives, so that by comparing them with my own observations and reflections I could follow the cycles of replanning, action, observation, and reflection of action research as suggested by Burns (2015).

5.4.4 Further Alternative Assessment Methods

There are further alternative assessment methods that the interviewed teachers apply. The following sections detail ways of diagnostic assessment including feedback that teachers ask from students, peer assessment (meaning ways of students assessing each other), students' self-assessment, and feedback that teachers give to students. Consequences of using these methods are also mentioned at the end of each section. From these alternative assessment methods used by the interviewed teachers, the last section describes the ones that were also utilized in the action research project.

5.4.4.1 Diagnostic Assessment. As discussed in the literature review, diagnostic assessment aims to analyze a situation in order to gather detailed information before making a pedagogical decision (Vidákovich, 1990). It is categorized into alternative assessment as, despite its important aim, it is traditionally rarely used in Hungarian public education perhaps because a diagnostic test is not a form of assessment to be graded (Golnhofer, 2003). However, it is exactly for the reasons listed above that diagnostic tests fit the definition of alternative assessment of the present dissertation. The interviewed teachers regularly apply it in accordance with their goals (as mentioned in chapter 5.3, Teachers' Goals for Using Alternative Assessment Methods). They can gather information about students' knowledge, needs, experiences, or opinions to make pedagogical decisions about instruction, materials, tasks, assessments, and so on.

Diagnostic assessment traditionally appears in Hungarian public education in the form of placement tests at the beginning of secondary school when there is a chance to learn a subject in smaller groups. Students are usually placed in groups based on the results of these placement tests. So the interviewed teachers also apply them; however, classes are not always divided, and for this reason, the interviewed teachers experienced highly diverse groups, too (sometimes even students

from A2 to C1 levels of English in the same group). As mentioned earlier, this is one of the reasons why they started using alternative assessment methods.

Beside these placement tests, the interviewed teachers also use diagnostic tests to see students' previous knowledge before starting a unit. For example, Dóra used language tests in Google forms (even before the emergency remote delivery) to determine her students' levels. She also uses the "KWL-chart", which includes three columns labeled with the three letters: K stands for "what I already know", W stands for "what I would like to learn", and L stands for "what I learned". So at the beginning of a unit, students fill in the first two columns in connection with the new topic. Dóra uses it for diagnostic purposes to see what students already know about a topic, but according to her, it is also beneficial for students with low self-esteem to realize how much they already know about something in English. She mentioned another tool that can be used for diagnostic purposes: exit tickets. Exit tickets can contain short questions or sentences to finish, which students can answer quickly at the end of a lesson and hand in while leaving the classroom. Ilona mentions the 3-2-1-card, which can have the same function depending on what the three, two, and one statement is that students have to write. For example, Dóra's students have to finish in connection with the lesson: "I was surprised by...", "The best was...", and "It would be good if next time...". Ilona uses it in connection with a situation: write "3 things you feel", "2 things you do", and "1 thing you don't know". However, instructions can be modified to anything that teachers would like to gather information about.

The other main form of diagnostic assessment that the interviewed teachers use is online questionnaires in which they ask students' opinion about various matters. Their students confirmed that it might happen that they have to provide feedback, but teachers taking it into consideration and changing anything accordingly was new and surprising for students. Lali has an anonymous online questionnaire that he asks students to fill in at the end of each unit (which is about once a month). In the beginning, only a few students filled it in, but after a few months when students realized that he actually modified things they asked for, the answer rate increased to 80-90%. It stayed a source of permanent communication between them. Lali adjusts the questionnaire according to the specific situations of the groups that he would like to gather information about. For example, there was a group in which their relationship deteriorated, and he wanted to find out whether it was because of his attitude, so he included a Likert scale statement in which students had to rate how supportive they find their teacher's attitude. Eventually, the answers came from

discussions with other colleagues teaching the same group, but he kept this statement in the questionnaire to monitor how supportive students find him. Further Likert scale questions (that students have to rate from one to five) in his questionnaire are mostly about assessment: “How satisfied are you with your results? How fair do you think your teacher’s assessment is? How difficult was the achievement test? How much have the tasks brought by the teacher supported your learning?”

Another form of diagnostic assessment Lali uses is that students write grammar tests only to see what they need to practice more (so they do not get grades for it). After introducing this, he adapted his questionnaire to include the question: “How much do you like that we don’t write grammar tests for grades?” and instead of compulsory homework, based on the results of the grammar tests, students can practice workbook exercises and check the answers for themselves, so he also included the questions: “How much do you like that we don’t check the workbook exercises together?” and “Have you used the workbook key to check exercises that you have solved?”. For further diagnostic purposes, he asks the following open-ended questions, as well:

What have you learned very well in this Unit?

What have you not managed to learn completely in this Unit?

Which in class task did you enjoy the most? Why?

Which in class task did you enjoy the least? Why?

If you would like to write about anything else, you can add it here.

Based on students’ feedback, he has got rid of some unpopular task types and a plush ball that they were throwing around and made his students anxious (Lali assumes that it might be some PE related anxiety about catching and throwing balls). He also modifies parts of his assessment system (such as the point values of tasks or new rules etc.) to support students’ learning processes more effectively.

Similarly to Lali, all the interviewed teachers ask for feedback from their students. When Fanni started teaching, she simply asked for feedback without any specific questions, then she asked students to write at least one positive and one negative thing, and now she asks about specific issues she would like to know her students’ viewpoint about (in addition to always providing the opportunity for any anonymous feedback in general, too). This is a great example of how the practice of a novice teacher develops. Edit uses Mentimeter at the end of lessons (similarly to an exit ticket) for feedback; however, she warns that it should stay short and meaningful otherwise

students might not take it seriously. The fact that almost all of them ask if students find their assessment system fair proves how crucial it is for the interviewed teachers. The types of activities, tasks, and tools that students like (or dislike) also reoccur.

Sometimes students surprise teachers with the content of their feedback. Mari mentions an instance of students complaining that the test was not difficult enough. Dóra illustrates the consequence of using polls with the following story. She asked her students to vote for the most difficult topics in order to choose which ones to revise for the upcoming exam, and students voted for completely different topics that she would have considered difficult for them, so she concluded that it was worth asking students' points of view in the future instead of assuming what they might think.

All the interviewed teachers mention group discussions that they have from time to time to discuss issues with their students. These issues might be related to the material; however, it is important to note that opportunities are provided for the students to discuss problems of any kind. Teachers can collect valuable information from these discussions, which can be structured, such as Eszter's scheduled Q&A sessions, or spontaneous, for example, interviewed teachers often say something similar to "we had long discussions about this".

In the beginning of emergency remote delivery before having live online lessons, Bea sent tasks to students and asked them to also add how long it took them to solve the tasks as she did not want to overwhelm students with too many assignments. For similar reasons, later she also asked students how many times they had to fill in the online quizzes to reach the desired points (because students could fill them in as many times as they wanted to). All of the interviewed teachers who were teaching during the emergency remote delivery period continuously gathered information on students' state (learning conditions, knowledge acquisition, attitude, motivation, and so on) on various platforms: through the e-gradebook, the online learning management systems they used, calling the students (or their parents) on the phone, and asking other colleagues. Therefore, they could take into consideration students' state and make pedagogical decisions accordingly, which are detailed in section 5.4.5, Alternative Assessment Methods Used During the Emergency Remote Delivery.

At the end of the school year, all of them discussed their experiences with their students about the emergency remote delivery. For example, Ilona made her students assess all the online tools they used asking if they were easily manageable, interactive, helped them in learning, and if

they would recommend it to others. The questions about what to keep for the future and what to change appeared in all verbal or written assessments. Dóra created an online board (see its screenshot in Appendix H4) together with her students about the emergency remote delivery in Conceptboard (see its link is in Appendix B) through which they discussed their feelings, what they miss from school, the topics they liked, the songs they remembered from the lessons, pictures they can relate to, and the superhuman skills they would like to have. Dóra always starts the school year with something similar, too, as a warm up and for the purposes of diagnostic assessment.

What teachers ask from students shows what they value, so looking at feedback that schools ask from students is also informative. Mari shared the assessment of her school, in which students had to rate the following statements about their teachers: [name of the teacher]

Consistent in both teaching and evaluating.

Supports the development of talents and helps integration.

Explains clearly and comprehensibly.

Raises interest in his/her subject, motivates students to work.

His/her lessons are well organized, start and end on time.

Maintains discipline in class.

Prepared, competent, good professional in his/her subject.

Assesses fairly and does not label students.

Respects and treats students as partners, strives for cooperation.

Has a good relationship with students.

Seeks the best solution in case of conflicts.

Admits when he/she is wrong.

Friendly, has a sense of humor.

Patient, understanding.

Participates in extracurricular programs for students.

Helped students' independent learning with diverse activities during remote delivery.

After each semester, teachers receive the averages of their students' answers of each statement and their overall average given in percentages. As mentioned in section 5.1.1, Teachers' Competences for Democratic Culture, in connection with displaying autonomous learning skills, when Mari started teaching, her result was 58%, and in the year of the follow-up interview, she got 94%,

which she is extremely proud of. Her lowest result now is 85% for maintaining discipline in class, which she does not mind at all.

A further consequence of diagnostic assessment through students' feedback is that students can also learn how to express constructive criticism to their teachers. However, it might need some instruction, such as in the case of Fanni, when she paraphrased her student's rude comments, or when Mari managed conflicts between her students and other colleagues (as detailed in section 5.1.2, Teachers' Views on Conflicts and Sensitive Issues).

5.4.4.2 Peer Assessment. As detailed in section 5.1.5, Teachers' Views on Assessment, the interviewed teachers believe that peer assessment is important; however, formulating it in a constructive way is also something that students have to learn. As a result, there are situations or groups in which teachers do not apply any peer assessment, and there are different versions of how they apply it.

Lali asks his students, after working in groups, to assess their teammates' work in online questionnaires. So students know that it is the teacher who sees their peer assessment first. In the questionnaires, they have to mark their groups and roles, and write about the others according to their roles as well (such as how well the researcher or the presentation specialist did their jobs). After that, Lali compares students' feedback with his own notes and gives individual feedback to all the students based on them. He admits that it is a lot of work, but he also thinks that students are grateful for getting such individualized feedback, and also that their opinion is valued, so they take their assessment of their peers seriously as well.

Unlike Lali, who filters through students' assessment of each other, Dóra creates the context for peer assessment, but she lets students manage it in their ways. Coordinating projects, Dóra always assesses the completion of different stages of the projects (at the end of each day, there is an assessment session); however, after she discusses with students what she thinks, there is a dedicated time for students to assess the day for themselves, for which purpose Dóra leaves them alone. She believes that "there are a lot of things that they accept from each other much more easily than from a teacher" (Dóra).

Fanni would like to encourage students' peer assessment, so she incorporated it in her point-collecting system. When students give presentations, there is a shared Google document that everybody can edit, and students can write their assessment there. As they receive points for it, they have to add their names, so it is not anonymous. Fanni's experience is that "students give

each other well worded, correct, and useful feedback”. They usually have a supportive attitude to each other, but they also word criticism. For example, “this part was a little difficult to understand, or there was too much text on the slides, but it’s visible how much you worked with the ppt, well done”. Fanni adds that it also depends on the group because “in one group, they loved watching each other’s [uploaded] presentations and writing feedback to each other while another group was not that active”. She has not experienced hostility, only that some students were not interested and did not choose this way of collecting points.

Edit also uses peer assessment in relation to students’ presentations. All of the students assess all their peers’ presentations based on a rubric that includes “content/input, creativity, fluency, and language, so four times five points”. It results in all of them listening to each other, making notes during the presentation, and then uploading their feedback to a given platform. According to Edit, it is interesting to note that students’ assessment is mostly the same as hers. Cases when it is different are usually when “friends give each other maximum points for input” or “when a presentation is longer or shorter than the necessary, students’ reactions vary” (Edit). She also makes students assess exam recordings, so students assess examinees like them, but they are not their classmates. This is a more difficult task, and students’ assessment differs much more, so Edit finds the discussions about these differences a great learning opportunity.

Ella includes peer assessment for exam preparation, too. They divide the class into small groups who simultaneously role play the exam situation involving students as examiners assessing their peers. According to Ella, students enjoy these role-playing games. “They know that it is useful, that they speak more like this, and have more opportunities to practice. And they enjoy the situation of being the examiner, the teacher, and of course in the next round they switch roles” (Ella). Similar simultaneous small group or pair discussions often happen in Ella’s classes consisting of peer assessment. For instance, “if the task is to summarize an article for one to three points, they listen to each other, and if Berni says that Máté summarized it for three points, I believe it. I’m not going to listen to each and every summary” (Ella). This kind of peer assessment, in which Ella tells the criteria and the point values and students listen to each other and assess each other in pairs, has a lot of benefits. The speaking time of the individual students is increased. As Ella listens to these speaking tasks, she hears how well students in the teacher role ask questions from their peers, which is a crucial aspect of developing students’ communicative competence. All in all, “they leave English classes with a lot of confidence” (Ella). As it can be seen above, peer

assessment is a common feature in the participating teachers' classes, and they believe that it yields several advantages for their students.

5.4.4.3 Self-Assessment. Self-assessment has been mentioned several times in chapter 5.3, Teachers' Goals for Using Alternative Assessment Methods, as the interviewed teachers find it crucial as a part of developing 21st century skills and autonomous lifelong learning. They also agree that it is mainly missing from traditional assessment practices. Edit illustrates it with the following story. When students uploaded their charts including their assessments of their peers' presentations, they crossed out the lines including their own names, so they did not assess themselves only the others. Edit asked them why, and students answered that it had not even occurred to them to fill in their own lines, so according to Edit, this example shows how much students are unused to self-assessment. Edit took this opportunity to discuss the importance of self-assessment with her students and encourage them to assess their own presentations as well next time.

Lali concurs that students should be taught self-assessment, so that "they are aware of their weak and strong points and develop self-reflexion". For this reason, he includes self-reflective questions in his monthly questionnaires such as "What helped me learn? What should I do differently? and so on" (Lali). His experience is that acquiring self-assessment skills takes time, and he can see the progression in students' answers from very vague sentences to pinpointing specific areas of improvement.

As mentioned in connection with diagnostic assessment, the interviewed teachers ask feedback from their students, and just like Lali, others also use this opportunity to incorporate self-assessment into their questionnaires. Fanni, for example, asks her students what their most effective learning strategies are. In addition to assessing the online tools they used, Ilona also asked her students to assess themselves how successful they were during the emergency remote delivery. According to Ilona, students' self-assessment matched with her observations:

I was happy that they were honest and evaluated themselves well. You could see that they knew how much effort they put, or did not put, into something. They could measure the value of things. I think it's really important. (Ilona)

Ilona also experiments with different self-assessing tools such as students signaling with a number how well they do on each task:

Write the appropriate number (1-4) next to each task as you are doing them. Write in your own column only. Change the number as you are improving.

1=I'm a novice. I'm stuck.

2=I'm an apprentice. I may need help in a minute.

3=I'm a practitioner. Good to go.

4=I'm an expert. I'm done and I can help others.

Ilona calls it “learning path” and uses the excel sheet to automatically color code the numbers, so she can have immediate visual feedback on how the whole group is doing (see Appendix H5). So it can also serve diagnostic purposes in addition to students’ self-assessment. Moreover, Ilona adds which activities are core tasks (that are compulsory for students), which ones strengthen students’ different skills (listening, reading, writing, speaking, grammar, vocabulary, and exam preparation appear in the chart), which ones deepen students’ knowledge, and which ones they should do with a computer or smartphone (see Appendix H5).

Dóra is the most prolific interviewee in terms of self-assessment tools. The already mentioned exit ticket or 3-2-1-card can also be used for self-assessment purposes, such as writing “3 things I’ve learned today, 2 questions I have, 1 thing that I don’t understand” (Dóra). In addition to providing information to the teacher, students also have to reflect on their learning process. Similarly to the third column of the KWL-chart, at the end of the topic students can compare their initial ideas (W: what they wanted to learn) to the reality (L: what they have actually learned). According to Dóra, this simple tool can lead to lots of realizations, such as “wow, how many stupid things I wrote in the beginning, and now how much I know about this topic” as she quoted a student of hers.

Dóra also mentioned tools that help in creating a product such as the “shopping list” including the criteria the product should meet (which students can tick off one by one). Similar checklists and rubrics are mentioned by others as well, so that students can self-monitor and assess their work based on them. In connection with projects, Dóra also uses reflective journals that provide students with a structure for carrying out the project in addition to helping questions or sentences to finish, for instance “From this I have learned that... This might be useful because... I will use this when...” (Dóra). Depending on the sentences or questions, they can make students reflect on their learning process, assess a product, or realize the utility of their work.

The last self-assessment tool is the so-called “feedback grid” (Dóra). Its aim is for students to learn from the feedback they receive on their work. It has four parts labeled with four symbols: a plus sign meaning what worked, what was particularly good in the product; a triangle, which is deliberately not a minus sign, meaning what I should change; a question mark meaning what questions I have; and a lightbulb meaning what new ideas I have now, all based on the feedback. (Dóra)

According to Dóra, it develops a growth mindset (Dweck, 2006) in students, so that criticism is something they can learn from and improve themselves based on. She makes students practice to “filter criticism for helpful advice on what they can benefit from” (Dóra). She lists some examples of students formulating for themselves that “I should speak more confidently” or “pay attention to articulation” or “be more structured”, and so on. Generally about self-assessment, Dóra tells her students how important it is to be able to reflect on their own work, so “when you want to ask for a pay raise, you should know the value of your work and not just bullshit your boss”. As the examples above show, self-assessment appears in almost all the interviewed teachers’ practice to fulfill one of their main goals of facilitating students’ autonomous learning.

The last category of further alternative assessment methods is feedback from the teacher to the students, which is detailed in the next section.

5.4.4.4 Feedback From the Teacher to the Students. Alternative assessment methods that have not been categorized into any of the previous sections are further feedback types that teachers provide to students. Giving individual formative feedback appears in all interviews. Fanni says that “I apply a lot of verbal feedback, which during the emergency remote delivery had to become written textual feedback”. Edit also complains about the fact that she had to spend a lot of time in front of the computer as “the individual feedback you tell in class all had to be written down, [...] I also wrote down after each lesson what happened, what students had to do, so I didn’t only post assignments, but lesson plans that students could follow” (Edit). Lali formulates similar ideas: “most of the assessment was built in my assessment system, but obviously I had to give a lot of formative feedback on students’ assignments [which during the emergency remote delivery happened in writing]”. For Mari, it is also self-evident that she gives individual formative feedback as she explains “most of my tasks are not tests but texts that need to be read and to receive detailed feedback, so it’s time consuming, [...] and I also reached out to those who did not submit anything”. As mentioned earlier, during the emergency remote delivery, she chose five assignments to be

assessed each class, which meant not only grading but giving detailed feedback as well. In addition, at the end of the school year, she also gives each of her students long, textual, individual feedback every year. Ilona emphasizes that giving elaborate individual feedback on students' work was so important for her that despite the difficulties of the emergency remote delivery, she also took the time to write the feedback down. As an advantage of the situation, she mentions that as feedback had to be written down, she thought more about it and made more exact points in giving both tasks and feedback. In connection with discussing group work, Ilona also gives feedback to the group as a whole, and she thinks that students can apply it to whatever has been their responsibility. As Júlia puts it "during the emergency remote delivery, I didn't have to travel, so I paid even more attention to giving feedback, and you know, it's a must, so you must give detailed individual feedback to at least students' written work". Ella experienced that when she wrote down the individual feedback through the comment function of Google classroom, her students replied, and it resulted in a constructive discussion. Fanni also gives accounts of such dialogues between teacher and student.

According to Sadler (1989) what makes feedback effective is not only communicating the current level and the goal but also using the information to alter the gap between the two, which is also supported by Wiliam's (2011) meta-analysis of almost 800 studies detailed in section 2.3.2, *Assessment that Supports Students' Development*. There are several examples of the interviewed teachers applying this kind of feedback, which, based on their students' reactions, is also rarely applied traditionally. Fanni finds it crucial to give feedback on how students can develop, and she says that students thank her for the feedback and change the things she points out, so according to her, it seems to be effective. Fanni also asks several questions in anonymous questionnaires, one of which is what the student's most effective learning method is. When a student answered that they had not figured it out, yet, Fanni gave ideas on what they could try. Another example is that next to incorrect answers she writes helping questions and sources where students can find further information on the matter. During the emergency remote delivery, this all happened in writing, so Fanni notes that it is much easier (and less time consuming) for teachers when they can talk about such things with students in person.

Almost all of the interviewed teachers mention similar conversations, in which they discuss how students can develop, which Gamlem and Smith (2013) find the most effective way of assessment from the students' perspective as well. Mari also teaches learning methodology, so she

also helps students with finding how they can learn most effectively. During the emergency remote delivery, Lali and Bea both introduced video recorded formative feedback while correcting students' work. Lali also mentions students thanking him for the feedback as it helped them a lot. Bea's students wrote her things like "Ms, this video is awesome!" or "Thank you so much!" although she did not only say positive things at all. Based on the feedback Bea gave, she also asked students to send back corrected versions of their work, and if it had not happened, she sent reminder messages of waiting for the corrections. Similarly to Dóra, who wrote down individual feedback on students' work, and, for example, when students had to develop exam topics as a preparation for their oral exam, Bea always included what was good, what was missing, and what students had to change, after which students had to rewrite and resend them.

Rubrics have been detailed as tools for students' self-assessment; however, they are mainly applied as the basis for teachers' formative assessment. Júlia uses them to highlight the goal of particular tasks, so she describes what will be assessed and the detailed criteria for them (such as a specific grammar structure, content, design etc.) then refers to them in her feedback. Ilona and Bea regard the easier use of rubrics as an advantage of the emergency remote delivery. Both of them used the built in function of rubrics in Google classroom, with which they could create and reuse rubrics for students' assignments. According to Ilona, it made correcting and giving exact formative feedback to students easier for the teacher and more transparent for the students. Bea thinks that "it makes students much more aware of what they are doing and why they are doing it". She uses rubrics for assessing both oral tests and written assignments, and she experiences the benefits in both cases.

The interviewed teachers use further feedback elements such as ClassDojo points (see link in Appendix B), credit points, stamps, or stickers to achieve further goals. With ClassDojo points, Bea would like to reflect students' effort in class. She really enjoys "the objectivity that it does not depend on who the teacher likes, but it's there black and white how much each student works in class". Lali uses a special credit point system to assess class work, which he compares to the "traditional red point system". In Hungarian schools, besides grading, teachers usually use red and black points (mainly in primary school, which sometimes become plus and minus signs in secondary school) to reward or punish correct or incorrect solutions (in some cases behavior, too), and there are different ways how they are converted into grades (e.g., ten red or black points is worth a grade five or one, or the ratio of red and black points correspond to a grade, etc.). Lali says

that credit points are like red points (without having any kinds of negative points), but with a wider range of usability. Students can still get a grade five for them, but a certain number of points can also be used to buy privileges such as skipping homework or getting a test answer, and so on. He introduced credit points in order to reduce the stress of oral tests. He wanted his students to gain confidence in speaking without being worried about grades, so “it is more of a reward system”. His other goal was to motivate students to work in class. He felt that students had a lot of anxiety using English (e.g., saying something incorrectly), so he wanted to reward their effort. He believes he has achieved his goals as students are not afraid to speak in English anymore, and he does not see any stress or anxiety. According to Lali, rewarding even their small efforts helped in building their confidence. Lali would also like to reward outstanding achievement with stamps and stickers. In addition to credit points, “high achievers can collect stickers, and tests with a 100% correct result are worth kitten stamps”. It might seem childish in secondary education, but Lali recalls events when he forgot the stamps, and students were outraged, so according to him, it does motivate some students.

Almost all the interviewed teachers mention the importance of informal, oral feedback, such as praise during the lessons. They find it so important that even during the emergency remote delivery they continued it not only orally in live online lessons but also in writing. For example, Júlia says: “I wrote at least a ‘great job’ or ‘well done’ or some little motivating statement”. Anna describes her development in how she consciously pays attention to her words. When she started teaching, she used the same expressions as the ones she heard as a student: “don’t do this, stop that, listen to me already, and so on” (Anna). Realizing the futility of her words, she introduced “chat-time”, in which students are given three minutes to chat, at the end of which there is a countdown signaling the end of the “allowed chaos” (Anna). She also emphasizes the importance of agreeing on the rules together with students, so for instance, if students do not stop talking at the end of the countdown, she can refer to the rules they all agreed, try again, and when they keep them, Anna praises them immensely. “This is one example of positive reinforcement, which I believe is crucial to give students when I see that they do something well” (Anna).

In connection with how to best support students’ development, Eszter details her plans for introducing phases or levels that include different rules for point collecting. In the beginning, students might need praise for even starting a task, but later it should become self-evident, and students should be encouraged to choose more and more difficult tasks. She says that

there should be consequences of someone not getting from A to B, and I don't mean minus points or punishment, but just as in life if you don't get to an appointment, you might have to wait a lot for another chance, [...] I would never erase points because in life if you achieve something, nobody can take it away from you either, but somehow I would like to fine tune the consequences of making mistakes. (Eszter)

She believes that different actions have different consequences in life as well, and she would like to make it clearer what the situations are in which students can try as many times as they need to, and introduce challenges with more serious consequences that show students' improvement and develop their autonomy, too, in connection with students taking responsibility for their actions.

As it can be seen, the interviewed teachers use a wide range of further alternative assessment methods. They apply diagnostic assessment not only through placement tests to see students' prior knowledge, but also ask feedback from students continuously through anonymous online forms, exit tickets, or Q&A sessions. The participants' students also assess each other in various forms of peer assessment and themselves using numerous self-assessment tools for students. The interviewed teachers also provide their students with different types of feedback from highlighting students' individual strengths to using detailed rubrics in addition to suggesting how students can improve.

5.4.4.5 Further Alternative Assessment Methods Used in the Action Research. This section details further alternative assessment methods used in the action research in the same order the interviewed teachers' further alternative assessment methods (in chapter 5.4.4) were presented: methods used for diagnostic assessment, peer assessment, self-assessment, and feedback from the teacher to the students. In addition, during the action research there was one instance of registering grades before closing a point-collecting period in advance at mid-term for those students who did not have enough grades for their official midterm grade. For those students who had already collected enough points for a grade five, I registered it, and I asked the others what grade they wanted to achieve during that point-collecting period, and I registered that. All the students collected enough points for the grades given in advance. Students' opinions about this instance and the other assessment methods detailed in this section will be presented in section 5.5.2.3, Reactions to Further Alternative Assessment Methods.

In terms of diagnostic assessment, my primary tool was questionnaires (see Appendix E). My goal with questionnaires administered after each unit was to gather information on students'

opinions, so that I could take them into consideration for making pedagogical decisions such as how to apply different assessment methods, what to change in the next unit, and so on. In connection with mapping students' knowledge, alternatives for pop quizzes (see in section 5.4.2.3, Alternatives for Testing During the Action Research) also served a diagnostic purpose. For example, if I saw that most students could not solve a problem in a mini, I decided to spend more time discussing it instead of going on with the material. Based on the interviews, I experimented with the KWL-chart and exit tickets as well. The KWL-chart included three columns labeled with the three letters: K for "what I already know", W for "what I would like to learn", and L for "what I learned". In the 9thITE group, I assumed that students had already known a lot about the topic (the unit was about basic software), so at the beginning of the unit, I asked students to fill in the first two columns of the chart (what they already knew and what they wanted like to learn) in connection with the topic. Then they had to add the newly learned material to the last column. Unfortunately, the introduction of the emergency remote delivery interrupted this unit resulting in the shift of everyone's focus, so students did not remember much about the beginning of the unit. In the questionnaire responses, only two students could comment on the activity saying that they found it useful for expanding their knowledge. According to my teaching journal, during the following units we gathered what students already knew about a certain topic in our live online classes together (e.g., writing words to the chat or saying them aloud), and I was always surprised how much students contributed. Quoting from my journal: "we collected way more expressions than what was in the book" (p. 30). Exit tickets (see sample in Appendix I7) also had a diagnostic role for me to see what students learned and what questions and opinions they had. For example, on exit tickets, they asked for "more homework", noted that they "should take notes to learn from", and requested the possibility to go through the Kahoot! questions we played in class to practice at home, too. As a result, I introduced taking notes of each other's presentations as a point-collecting opportunity and assigned each Kahoot! we played in class for homework as well. On the other hand, exit tickets also served self-assessment purposes, which will be discussed after peer assessment.

In terms of peer assessment, students had to give feedback to their peers about each other's presentations based on a list of set criteria (see Appendix I8) that we discussed when introducing the opportunity to give presentations. Later, when posters were introduced, we also discussed together how to modify the peer assessment sheet, so that it could be used for giving feedback

about posters as well. In class, students individually filled in an assessment sheet and gave it to the presenter. During the emergency remote delivery, we discussed presentations verbally right after the presentation as part of the online lesson. Each student had to highlight something they liked about the presentation that had not been mentioned before for which they could use items of the assessment criteria (see Appendix I8), and in the next round, they could add constructive criticism in the form about what should be done differently next time; however, in this round it was not compulsory to say something. This way we could also practice how to formulate feedback that was helpful for the presenter. In addition, we also collected useful vocabulary from each presentation and agreed on a list that students wanted to learn, which was later incorporated into their achievement tests.

In terms of self-assessment, during the first unit of the 9thITE class, I asked students to answer questions on paper for themselves, so I did not collect their responses. Then I used exit tickets (see Appendix I7), which I collected. Then I integrated the same questions I asked for the first time into the questionnaires of all groups: “What do you think about your own performance in this period? What went well? Not so well? What would you like to do to make it better?” and I later added the question: “How can I help you?” for diagnostic purposes (see Appendix E). I hoped that by integrating the questions into the questionnaire, students’ self-reflection would improve through regular practice. Students’ responses will be presented and analyzed in section 5.5.2.3, Reactions to Further Alternative Assessment Methods.

The last category of assessment is different types of feedback from the teacher to the students. During the interviews, numerous kinds of feedback emerged, so following the order as they were presented in the previous section, the appearance of such feedback in the action research project is detailed in this paragraph. Providing individual formative feedback verbally appeared both on purpose and in ways that I did not even notice. For example, I was surprised by the observation report of a math teacher trainee writing about me that “she praises a lot” as I was not consciously aware of it. The same observer also noted that students were not hesitant to freely ask several questions during the lesson for which my reactions were something along the lines of “it’s a very good question”. Reflecting on a recorded online lesson, an observer noticed that when a student seemed anxious asking a question, I told him: “it’s okay, I’m also nervous because the lesson is recorded”.

There were similar observations reported from the group of American math teacher trainees reflecting on a more conscious approach on my part:

When a student does not have the correct solution to a problem, rather than just telling them they are wrong or providing them with the right answer, the teacher asks guiding questions to help the students think about the problem they are having in a different way and ultimately come up with the right answer on their own.

Although applying discovery learning is my guiding approach, another math teacher trainee observed the challenges in its realization: “there was a part where she was about to give crucial information to the students, but then held herself back to allow the students to learn on their own. She instead, asked questions to guide the students to the solution”.

My aim is also to provide feedback that does not only communicate the current level of students and the goal but also information on how to alter the gap between them as suggested by Sadler (1989). Some examples of this kind of feedback that I collected during the action research are the following. I also filled in the assessment sheet (see Appendix I8) for each student presentation and poster giving detailed feedback in writing on each product. Similarly to the interviewed teachers, I spent a considerable amount of time on individual or small group discussions with students online during the emergency remote delivery giving feedback to students verbally. In addition, the point-collecting system also provided constant numerical feedback on students’ performance of each task. As the criteria for the different number of points were known to students, they could also infer information based on the number of points received how they can further develop. I also frequently supplemented it with drawing my students’ attention to ways I suggested that they collect points.

5.4.5 Alternative Assessment Methods Used During the Emergency Remote Delivery

During the second term of the 2019/2020 school year, on the Friday evening of March 13, 2020, emergency remote delivery of education was announced from the following Monday (March 16) due to the first wave of the Covid-19 pandemic. As discussed in the literature review, assessment was one of the most problematic elements of emergency remote delivery in Hungary due to the fact that traditional assessment mainly means grading tests, which teachers did not know how to carry out online (Monostori, 2021). As Monostori (2021) summarizes: “very few teachers came to the realization that grading was not the most important element of assessment” (p. 11). Moreover, recommendations of the Educational Authority (Farkas et al., 2021) did not include any

guidelines about grading while it remained the only compulsory element even during the following school year (2020/2021), most of which happened online.

The majority of the interviewed teachers' alternative assessment methods do not depend on the medium, so they could be adapted to remote delivery. Thus, the following section presents the above mentioned problematic areas: the interviewed teachers' grading and testing solutions during the emergency remote delivery. Then their information and communications technology (ICT) use both before and during the emergency remote delivery is discussed. The chapter ends with the alternative assessment methods of the action research during the emergency remote delivery.

5.4.5.1 Grading and Testing During the Emergency Remote Delivery. As detailed in the literature review, Monostori (2021) mentions that if students had enough grades for the second term of the 2019/2020 school year, some teachers closed the assessment of the year and gave the year-end grade based solely on the grades the students had received prior to classes shifting online. Some of the interviewed teachers also used this opportunity and did not worry about grading during the emergency remote delivery at all. However, as the following school year was almost entirely online, everybody had to come up with new ways of grading eventually. Section 5.4.1, Solutions for Grading, presents the variety of grading practices even from before the pandemic. The interviewed teachers' experience is that most of those practices could be maintained during the emergency remote delivery as well. Students collected points that were transformed to a grade at the end of a point-collecting period. However, ways in which students collected points differed greatly before the pandemic, too, so it also resulted in differences during the emergency remote delivery. Some teachers maintained the practice that students received grades exclusively through point collecting and only modified some of the tasks to fit the online setting. Others used point collecting as an option for students to improve their grades, but only if they wanted to, so they did not receive any bad grades, but could collect points with optional tasks for an extra grade five. Finally, the third category was that teachers chose specific activities with which students could earn grades. As a general tendency, the interviewed teachers had different goals in mind and the changes of their assessment practices served those goals as detailed in section 5.3.5, Emergency Remote Delivery Related Assessment Goals. There are some further grading related examples. For instance, Mari and Ella wanted their students to prepare for exams and gave essay tasks graded using the exam criteria. Bea wanted to encourage participation in live online classes, so she gave

“small” grade fives for online class work, which had half the weight of a test (see grade categories with weights, for example, in section 2.1.3, The Pedagogical Programme of the School of the Action Research Project).

The interviewed teachers used live online classes for different purposes as well. For example, in Fanni’s case there were no course books or other materials for her English medium content (EMC) classes, so she held live online lectures and then uploaded her slides for the students. For a passing grade, students had to find and copy-paste information from her slides, and for better and better grades they had to answer more and more difficult questions that required critical thinking and not only the comprehension, but also the application, analysis, and evaluation of the information. Thus Fanni’s assessment reflected the taxonomy of educational objectives by Bloom and Krathwohl (1956). According to Lali, his point-collecting system did not depend on the medium, so it automatically continued through the emergency remote delivery. He held the following two types of online classes. One was for consultation if students had any problems with the material or the tasks they collected the points with. The other was a 25-minute speaking practice in groups of four or five. Students had a schedule, and they had to join the online class only for their group’s speaking time.

Some schools expected teachers to have live online classes according to their in-person timetable, while others put subjects in blocks, and made it possible, for example, to prepare for the upcoming matura exams by writing sample tests at dedicated times, for instance, one subject a day for three-four hours according to the length of the real exam. Some teachers offered extra optional thematic discussions in the afternoon if students did not understand something especially for those who prepared for exams. Mari developed an online lesson structure including a warmup activity together, then students processing the material individually, ending with a summary again as a class, and feedback. Students could volunteer for their individual work to be graded. At the beginning, Mari tried to check everybody’s work, but she soon realized that it was impossible, so she graded five students’ work each lesson. For those who did not volunteer, she organized individual discussions of what to get grades for. Almost all the interviewed teachers gave grades on a voluntary basis, so students had the chance to submit work they wanted to receive grades for (or in Edit’s case: percentages).

As section 5.4.2, Alternatives for Testing presents, all the interviewed teachers had already abandoned traditional testing before the pandemic; however, taking into consideration the special

circumstances of the emergency remote delivery, further adaptations had to be made. They all agreed that asking for lexical knowledge had absolutely no point in such circumstances. Mari focused on applying information students could find online, or elsewhere, and the assessment referred to the quality of the application. Lali encouraged his students to solve tests together in their study groups that he created for the emergency remote delivery. He designed tests to be more difficult, so that one person could not solve them, and also included questions for which students had to use other sources such as finding something in the coursebook or the internet. Although Dóra agreed that there was no use in administering traditional exams in such circumstances, some of her vocational groups had centralized EMC exams, so she took part in creating a question bank from which a program randomly generated a different set of test questions for each student.

In conclusion, the interviewed teachers' grading and testing practices during the emergency remote delivery did not differ much from their alternative assessment methods used before the pandemic (as detailed in sections 5.4.1, Solutions for Grading and 5.4.2, Alternatives for Testing). In addition, their ICT skills also helped them face the adversities of emergency remote delivery.

5.4.5.2 ICT Use Before and During the Emergency Remote Delivery. Information and communications technology (ICT) use played an important part in the interviewed teachers' assessment even before the emergency remote delivery. For this reason, this section introduces digital tools used by the interviewed teachers before the pandemic including what and how they changed during the emergency remote delivery. The links of all the websites and applications mentioned can be found in Appendix B. It is important to note that the interviews did not include any questions about digital tools or ICT use, so the following paragraphs contain only the tools that the interviewed teachers named when talking about alternative assessment.

Most of the interviewed teachers had already used some kind of learning management system (LMS) in order to keep track of their students' points and/or to manage materials. For example, Júlia had organized English lessons using Canvas in computer rooms even before the pandemic, so she and her students were familiar with the LMS, but they still had to extend its usage for vastly different purposes. She digitized a lot of workbook exercises by creating quiz items in Canvas, so that she could individually follow students' work. Students solved these exercises for points that could also be registered within the LMS. Similarly, all of Fanni's groups had already had Google Classroom accounts, so they also used their LMS more widely during the

emergency remote delivery. Beside Canvas or Google Classroom, Microsoft (MS) Teams was introduced as the main LMS in interviewees' schools. Even for teachers who had used ICT before the pandemic, one of the main changes was to hold live online classes. Gamer students in many classes recommended Discord, a platform that they had used for live communication and had the capacity to manage thousands of new users. Zoom also became one of the popular platforms used for live lessons, in addition to Google Meet for those who used Google Classroom, and MS Teams where the whole school chose it as their LMS. The interviewed teachers kept track of their students' points on different platforms, too. Many of them used excel sheets such as Lali and Mari, who used them in a very complex way: based on the teacher's input, the program generated pages for students individually, so the anonymity of students' could be ensured. Others used open leaderboards such as Flippity, ClassDojo, or #school. During the emergency remote delivery, they either continued using the same platforms or integrated the process of collecting points to the new LMS.

As mentioned in section 5.3.3, Teachers' Goals Related to Students' Life Beyond School, the interviewed teachers had aimed to develop their students' digital literacy as one of the 21st century skills even before the pandemic, so they had included tasks with this goal in mind. For instance, Anna's students could record video tests and one of Lali's tasks was video editing that they explicitly mentioned as examples for developing their students' ICT skills. They incorporated digital tools for all the above mentioned alternative assessment categories: testing (section 5.4.2), activities that help process the material (section 5.4.3), and further alternative assessment methods (section 5.4.4) such as diagnostic, peer, and self-assessment, in addition to giving feedback to students. The following paragraphs present the tools the interviewed teachers used in this order.

As discussed in section 5.4.2.1, Alternatives for Oral and Written Tests, the interviewed teachers had already given the chance to students to record their presentations or vlogs instead of traditional oral tests, for which students could use any recording device. This option became more frequent during the emergency remote delivery supplemented by the possibility of live online student presentations. Even in class students had to use visual aids to support the content of their presentations, which was most frequently done using MS PowerPoint, Google Slides, Prezi, Canva, or Mentimeter. The latter made it possible to involve the audience in a presentation. During the emergency remote delivery, it was just as important; moreover, screen sharing provided further opportunities (such as sharing a game play). These either happened in front of the whole class or

in smaller groups when fewer students attended the lesson or using breakout rooms (in Zoom) or separate channels (in Teams).

The option of recording a presentation versus presenting live illustrates the asynchronous versus synchronous nature of online education. This duality characterizes all the task types. For example, all the following alternatives for pop quizzes have live versions and homework options. Teachers used to play Kahoot! in class before the pandemic, which they could also do during online classes by sharing the teacher's screen (instead of projecting the questions); however, the "Kahoot! challenge" option became popular as students individually could go through the questions by themselves. It happened the other way around with Quizlet as earlier it was mainly used for individual learning at home; however, the "Quizlet Live" option made it possible for students to play together in real time. Teachers also used Quizizz, Socrative, Redmenta, Google and MS Forms for alternatives for both pop quizzes and achievement tests. As achievement tests remained the most frequent traditional assessment method, the interviewed teachers noted that their colleagues also used the latter three (Redmenta, Google and MS Forms). Interestingly, Bea added that she had not known Redmenta before and did not get acquainted with it during the pandemic as all her colleagues used it for traditional testing, and she rather applied alternatives to tests. Board games and escape room activities as in-class alternatives to tests could also be adapted to the online setting. The interviewed teachers used Genially for templates, in addition to Flippity or Google Forms, to create escape room types of task.

The interviewed teachers used similar gamified processes in class before the pandemic as well. Iona mentions, for example, the application Actionbound for playing interactive scavenger hunts with the help of students' smartphones. Dóra planned an "investigation" where students had to solve a crime through tasks hidden behind QR codes and different puzzles (see an example in Appendix H3). Mari also created such "escape rooms before [she] had known they were called like that", in which students had to solve a problem somewhere in the classroom and the solution led them to another location with another task, and so on. The common element of these board games, scavenger hunts, investigations, escape rooms, and so on is that the structure of the game enables students to deal with the material autonomously by playing through the game at their own pace without the necessary intervention of the teacher. Going through the material with these kinds of activities during the emergency remote delivery proved to be extremely helpful. In addition to Genially, Flippity, Google and MS Forms, the interviewed teachers used Symbaloo Learning Paths

and Sutori to create them. Only Sutori was a new discovery, they had already used the other ones before.

Either as part of these game structures or separately as individual tasks, the interviewed teachers mentioned using several other sites. Digitizing materials could be done by putting them into one of the templates offered by Learningapps, Live worksheets, Wordwall, or the integrated functions of the LMS (Quizzes and Assignments in Canvas, Google Classroom, or Teams). The interviewed teachers also mentioned assigning videos from Youtube, TED Talks, or materials from BBC Learning English. Their students used different sites for making mind maps, word clouds, infographics, or content related memes (see my students' examples in Appendix I11). The teachers also tried to find online equivalents for games they played in person such as a drawing-guessing game (e.g., Skribbl), creating stories based on pictograms (e.g., Story Dice), sending hidden messages (e.g., Türchen), and other creative tasks (e.g., from Class Tools). They also used platforms on which students could collaborate and create something together such as Google Jamboard, Concept Board (see example in Appendix H4), Padlet, or Slack.

For diagnostic assessment all of the interviewed teachers had asked for feedback from their students using mostly Google Forms, which they either kept during the pandemic, or if the school used MS Teams, they incorporated the feedback into MS Forms as well. Similarly, they either used their usual voting tools (such as Socrative or Mentimeter) or applied the built-in versions of the LMS (e.g., Teams' poll it). Even when they sent materials to students, they tried to make it as varied as possible. For instance, they discovered that it was possible to record a MS PowerPoint presentation with narration. Videos could be made interactive (for example, with the help of EDpuzzle). Exit tickets were possible in Socrative or virtual sticky notes (e.g., Lino). Students could express their opinions through voting (e.g., in Socrative or Tricider). Peer assessment took place in shared Google Docs or Sheets. For self-assessment, the interviewed teachers did not mention any specific digital tools; it was usually incorporated into their students' work in the various platforms. For example, solutions or answer keys were also uploaded and students had to check their work for themselves, or the platforms gave immediate feedback (e.g., Live worksheets, Learningapps, Quizlet, Quizzes, and so on). Teachers' feedback also became an integral part of the LMS. They could comment on their students' digital work in several ways resulting in a lot of communication in the LMS. The links of all the applications and websites listed in this section can be found in Appendix B.

5.4.5.3 Alternative Assessment Methods of the Action Research During the Emergency Remote Delivery. In connection with the action research, this section details first what could remain the same during the emergency remote delivery as it was before shifting to online education. Then modifications of different alternative assessment methods that had to be made due to the new circumstances are elaborated on. At the end, alternative assessment methods that were newly introduced during the action research due to the emergency remote delivery are presented. Students' opinions of these methods in addition to assessment in general are described in section 5.5.3, Reactions to Assessment During the Emergency Remote Delivery.

The emergency remote delivery was introduced during the fifth unit of the action research project. Students received their midterm grade based on the first three units and already had their fourth grade registered. In order to finish the year, students had to receive at least two more grades. Students had already started collecting points during the fifth unit, so I decided that the main point-collecting system would remain during the emergency remote delivery as well. The end of the fifth point-collecting period was postponed, so that students had enough time to adjust and collect enough points in the new setting as well. The last two questionnaires and the focus group interviews inquired about students' experiences of the situation. Beside the main structure of receiving their last two grades as a result of collecting points, some of the activities they could collect points with did not have to be modified either. All the activities that were optional homework tasks earlier could stay in the point-collecting system such as using the website Khan Academy in the 11th grade math classes or creating vlogs in the 9thITE class. In addition to homework tasks, we could also play Kahoot! during the online lessons. Instead of projecting the questions in class, I could share my screen and students could still answer on their own devices from home, after which I could assign each game as homework, too. Students could continue giving presentations; however, my feedback on the presentations together with their peers' assessment happened verbally (as discussed in section 5.4.4.5, Further Alternative Assessment Methods Used in the Action Research).

Slight modifications had to take place, for example, in connection with where to keep track of points. During the action research, we used the website #school (see link in Appendix B); however, students voted for following their points in MS Teams, so that during the emergency remote delivery all the necessary information about the subject could be checked in one platform. For the same reason, I did not continue using different online quiz sites such as Redmenta or

Socrative (see links in Appendix B), but I integrated the tasks as Assignments in MS Teams using MS Forms. However, I was not sure whether filling in these pop quiz alternatives such as minis should happen during the time of the live online lesson or should be assigned for homework. The advantage of the former would be that students had a dedicated time and place for solving them, which might motivate them more; however, the advantage of the latter would be the freedom to deal with them when they fit students' schedules. The other issue I had in connection with these quizzes was how many attempts students should be allowed to have. I speculated that if they could do it only once, then we could discuss them in class, which would be a meaningful usage of the live online lesson; however, if they could solve them as many times as they wanted to, students could practice them until they learned the material similarly to Khan Academy. During the emergency remote delivery, I experimented with all the different versions and asked students' opinions about them in the questionnaires (see their answers in section 5.5.3, Reactions to Assessment During the Emergency Remote Delivery). I also intended to continue using group work online as well. On the one hand, MS Teams did not offer this option at the time. On the other hand, in the 11thME class usually less than half of the group were present in online classes (about four to eight students), so they were not separated into further smaller groups. However, in the 9thITE class, we used Discord (see link in Appendix B), which made it possible to work in pairs and small groups easily.

At the beginning of the emergency remote delivery, the greatest question was how to manage online lessons: when to hold them, which platform to use, and what should happen during the lesson and asynchronously outside of it. I decided to keep all my lessons at the same time as in the original timetable. I made it public in advance what would happen in each class, and students could decide whether they wanted to participate. I did not want to punish anybody for not having the means to attend the lessons; however, I wanted to use every opportunity to support students' learning as much as the circumstances allowed. Students in the 9thITE group asked to use Discord as they had already been familiar with the site, so I also learned its usage, and despite the school's introduction of MS Teams, we continued to hold live lessons on Discord as requested by the students. The 11th grade math classes took place through MS Teams.

In terms of what students could collect points with, I introduced new ways with different purposes in mind. On the one hand, I wanted to reward students for attending the online lessons, so points could be collected by being active in class such as saying a grammatically correct

sentence in the 9thITE class or solving math problems in the 11th grade math classes. The former was similar to throwing Wendy, the plush toy, around in person. The latter was intended to imitate problem-solving math lessons in person. On the other hand, I also provided alternatives through written assignments that students could submit in different ways. As I had only three groups of students during the action research, I could be flexible in the means of how students wanted to submit their work. Most students chose the Assignments function of MS Teams as there could be deadlines and point values added. I could also comment on the point value; however, I could not correct students' work directly within the platform. For this reason, some students preferred working in MS OneNote in Teams as I could directly comment on their work, resulting in a continuous dialogue. As a third version, some students preferred working on paper and sending me photos of their physical notebook pages. Although from the teacher's point of view this was the most difficult to correct, I did not want to force them to use more technology as they were already spending most of their time in front of the computer.

In the 9thITE class, we played Skribbl (see link in Appendix B). In this game students had to draw a word in connection with our topic that the others had to guess and write in, which was similar to the guessing game "activity" in person. Instead of students compiling tests for themselves, I created online escape-room-themed achievement tests for the last two units. However, students could still submit questions for points that might be added to the test. The escape room tests were available for students in Teams as a MS Forms assignment. Each page, students had to solve tasks at the end of which they received a code that led them to the next page. They could also collect secret messages for bonus points. For the 11thME class, I found playful practice exercises in connection with the material that I could assign online for points, and I also created tasks that were related to the real-life pandemic situation. In addition, I experimented with creating different supporting materials such as video and picture explanations of the math content (see Appendix I12). Students' opinions about all of them will be presented in section 5.5.3, Reactions to Assessment During the Emergency Remote Delivery.

From the teacher's point of view, on the one hand, I felt that the fact that the regular live online meetings and the point-collecting systems could be maintained during the emergency remote delivery provided structure and thus the feeling of safety for me and my students as well. On the other hand, the flexibility of the point-collecting system also allowed me to experiment and introduce new ways of assessment taking into consideration students' preferences, too. I was

grateful that I did not have to struggle with assessment during the emergency remote delivery, and I tried to help my colleagues by creating short videos explaining how I carried out alternative assessment online, which I shared in social media. As the follow-up interviews proved, the interviewed teachers had similar experiences.

5.5 Students' Perceptions of Alternative Assessment Methods

In addition to teachers' perspectives, two further research questions focus on students' experiences in connection with alternative assessment: research question 2.2 What are students' perceptions of alternative assessment in EFL and in EMC classes in public secondary education in Hungary? and 2.4 What are students' perceptions of the adaptation of alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary to remote delivery? In order to answer these research questions, the action research project involved 44 students in three groups that I taught during the 2019/2020 school year: a 9th grade beginner IT English (9thITE) group with 17 students as an EFL class, an 11th grade mathematics in English (11thME) group with 16 students and an 11th grade elective advanced mathematics in English (11thAME) group with 11 students as EMC classes (see more information in section 4.2, Participants). As part of the action research project, data was collected through field notes, teaching journals, class observations, student questionnaires after each unit, and focus group interviews with the participating students at the end of the action research project.

Table 10

Number of Students Who Filled in the Different Questionnaires

Group	Entry q.	1st q.	2nd q.	3rd q.	4th q.	5th q.	6th q.
9thITE	17 (100%)	11 (65%)	13 (76%)	11 (65%)	13 (76%)	5 (29%)	8 (47%)
11thME	16 (100%)	16 (100%)	12 (75%)	14 (88%)	10 (63%)	12 (75%)	13 (81%)
11thAME	11 (100%)	-	7 (64%)	-	7 (64%)	-	6 (55%)
Total:	44	27	32	25	30	17	27

Table 10 above shows the number of students who filled in each questionnaire. Out of the 44 participants, one student from the 9thITE group left the school and one student from the 11thAME group left the group during the first semester; in addition, four students did not

participate in the focus group interviews, so 38 students were interviewed. Data on students' perceptions was also gathered from questionnaires before the action research project from 133 students (see more information in section 4.4.1, Student Questionnaires Between 2014 and 2017). The following sections present students' reactions first to point collecting, then to other assessment types such as alternatives for testing, activities that help process the material, and further alternative assessment methods. The chapter ends with students' reactions to assessment during the emergency remote delivery.

5.5.1 Reactions to Point Collecting

In questionnaires (from 2014 and during the action research as well), I asked my students if they wanted to continue using the same point-collecting system in the next period and why. Giving their reasons was an optional open-ended question, as well as adding what they wanted to change about this point-collecting system. I also asked their opinion about the grades they received as a result of point collecting: what grade they had just received, how satisfied they were with the way their work had been assessed by me, why, and how well this grade reflected their knowledge and the energy they had invested in studying. Furthermore, they had to say whether, if they had been graded in the traditional way, the grade would have been better, worse, the same, or they can't decide. The students were invited to add their own explanation as well. As I gathered data about these questions before the action research as well, it is possible to look at a larger sample (n=168) of students' opinions. (See the questionnaires in Appendix E.) After some general information, I present what students like about the point-collecting system, what they would change about it, and their views on the grades they received through point collecting.

Table 11

Number of Students Answering the Question: "In the Next Period, Would You Like to Learn in This Point System?"

Answers	Number of students 2014-2017	Number of students of the action research
Yes	94 (76%)	33 (75%)
No	6 (5%)	1 (2%)
Other	13 (10%)	4 (9%)
No answer	11 (9%)	6 (14%)
All	124 (100%)	44 (100%)

Table 11 presents how many students wanted to continue using the assessment system based on point collecting in the future. I followed individual students' answers through all the questionnaires. In the case of the action research, it means that 33 students answered "yes" in all the one to six questionnaires they filled in (as individual students filled in one, two, three, four, five, or six questionnaires after the entry questionnaire). There was only one student who answered "no" in all the three questionnaires he filled in. In the first two questionnaires, he added that "I don't like the point-collecting system because I'm too lazy, otherwise I like the lessons" and "I think traditional assessment is simpler" (there was no explanation added the third time). The six students who answered "no" before the action research (between 2014 and 2017) wrote the following. One of them added that "we could get more points", three of them asked for getting only grades instead of points, and two did not give any explanations. During the action research, four students chose "other" and gave different textual explanations such as "I don't know yet", "I don't care what point-collecting system we use", "I think it should be decided by the teacher", and "it depends". There were similar answers in the "other" category before the action research as well, in addition to the following two interesting ones: "I would not change it [the point-collecting system], but I think it does not reflect students' knowledge accurately" and "I did not have any problems with it, but I would try out the traditional, too". I did not get any answers from six students during the action research as they did not fill in any of the questionnaires after the entry questionnaire. It is interesting to note that the ratios of different student answers seem to be similar over the years (approximately three quarters of enthusiastic yes, one tenth of other explanations, one tenth of missing answers, and one twentieth of no). It is also worth noting that looking only at the answers of those who filled in the questionnaires during the action research, the 44 students filled in 151 questionnaires in total (on average 3.4 questionnaires per student), from the 151 answers to the question: "In the next period, would you like to learn in this point system?" 144 was "yes", which is 95% of all the answers.

5.5.1.1 What Students Like About the Point-Collecting System. Students' reasons for liking the point-collecting system in the action research project varied. I categorized students' explanations why they would like to continue using the point-collecting system into two groups that I labeled: "just because" and "specific characteristics they like". In the "just because" category there are answers like "just because", "I like it", "because it's better", "it's easier", "I prefer this to the traditional", "because it's good", "ten times better than the traditional grading system", "way

better than getting grades for tests”, “it’s easier than writing tests”, “I’ve got used to it” and “it works”. In the category of “specific characteristics they like”, most students seem to appreciate that they have chances to correct and improve their results. Eleven students wrote something along the lines of “we have much more possibilities to improve”. What they mean by improving can be seen in the context of traditional testing and grading: “if you make a mistake in a test, you have opportunities to correct it”, “you have possibilities to improve your grade”, “I think this assessment system is much better because it’s not like if you get a bad grade that might ruin your whole school report, here you have chances to correct it”, and “because you have a chance to improve, I think all subjects should be graded this way”.

The second most frequently mentioned reason why students would like to continue learning in this point-collecting system during the action research is that their efforts and diligence matter (six mentions). They express it by saying, for example, “even if your test is not the best, but you’re diligent, you can get a good grade” or “I really like that it’s not one test that determines your grade, but a lot of things, so if you do something because you’re diligent, you can get a better grade”. When answering why they wish to continue using the point-collecting system, students also list specific activities (detailed in section 5.5.2, Reactions to Other Assessment Types) in addition to reasons connected to classroom participation (e.g., “I think it’s a good system because your classroom participation is appreciated” or “you have to be active in class”), transparency (e.g., “I’ve never seen such a system before, but I really like how transparent it is with a lot of opportunities”), excitement (e.g., “I like the excitement if I can collect enough points for a grade five”), and long term knowledge instead of cramming (e.g., “because here long term knowledge matters more, and you can’t cram”).

Looking at all the other sources of qualitative data (further open-ended questions in the questionnaires, focus group interviews, class observation reports, and my teaching journal, students describe the point-collecting system in several ways. On the one hand, the point-collecting system itself is frequently mentioned as the thing students like the most about lessons. For example, in the 11thME group answering the question “Thinking back to the past two years / in the past month, what did you like the most in math lessons?” (Appendix E7, question 2), six students answered the point-collecting system. Similarly, it also appears in the 9thITE group’s explanations for why they like the IT English lessons (Appendix E3, question 2) and at the end of the questionnaire: “Here you can detail your answer to any of the questions or anything else that

you have liked or not liked during this period” (Appendix E3, question 11). One of the students from 11thME class adds the following at the very end of the last questionnaire:

This point-collecting system is way more humane than tests. [...] If you work decently through the month, you won't stress about the test, and it's a huge relief. Teachers can also get a much more complete picture of the students. Tests often depend on luck. I'm always stressed about tests, but not about math, and not because it's easy, but because I know that my life doesn't depend on it.

On the other hand, students use several adjectives when talking about the point-collecting system. In addition to “good”, “transparent”, and “humane” that have been quoted above, students mention “logical”, “understandable”, “meaningful”, “easy to follow”, “fair”, “correct”, “realistic”, “feasible”, “colorful”, “fun”, “innovative”, “rewarding” and “motivating” to describe the system.

Transparency and adjectives like “logical”, “understandable”, “meaningful”, “easy to follow” often refer to the fact that students can follow their progression by getting points for each activity they carry out. For example, a student from the 9thITE class says: “it was strange at first because I didn't have anything like this in my previous school, but I really like that it's playful, and we always know what we should do for how many points to achieve which grade”. One student from the 11thME class details the aspect of fairness in the following way:

This [the point-collecting system] is one of my favorite things. You are under less pressure to pursue better and better grades. For example, even if you have a bad day when we write a mini, you don't get into this vicious cycle of feeling worse and worse because you know that you will have the chance to correct it and maintain your average. Of course, it doesn't mean that you don't have to prepare or work during class. It simply means that you don't feel like you were put in a racing stable with people you can't compete with. With this method, the best and the worst students can start with equal chances.

So students feel that the system is fair because it provides equal opportunities to diverse students. Another aspect of fairness is that with the diversity of tasks, students do not feel that they always have to perform everything perfectly as one student from the 11thME class phrases:

I love working in this system because I don't get a grade for each task separately but for the total of lots of them, just like in the matura exam, and it's great because the things that I can do better can compensate for things I'm not so good at.

Similar thoughts appear in the 9thITE class, too: “a great advantage of this system is that it takes out the stress that you continuously have to be on the top in every lesson of every subject”.

Feasibility is connected to fairness in the sense that they all feel that everybody has the possibility to achieve good results. In addition, students feel that it is “easier” and “manageable” to learn like this, such as a student from the 9thITE class who says: “it [the point-collecting system] is good because there are achievable expectations”. It is interesting to contrast with the numerous accounts of students saying how much harder they have to work in this system. Several students say that they prefer this point-collecting system to traditional assessment precisely because they have to work more. One student from the 9thITE class says: “it’s a fact that you have to work harder [...] and be more active in class, but I don’t mind it; actually I really like it”. Another student from the 11thME class tells his story:

The advantage of this point-collecting system is that you have to work a lot in order to get a good grade. Because of this, I fell in love with math. In 10th grade something clicked, and I have liked math ever since. I couldn’t stand it in 9th grade, I don’t know, I had never liked it before, but now I find it fun and interesting. I can’t say any disadvantages.

Another student from the same class further explains it by contrasting it to traditional assessment. “So the advantage is that you must work and that makes you learn because in the other [traditional] system if you’re good at math, you don’t have to do anything just write good tests [...] but here you have to work continuously”. This contrast appears in the 9thITE class as well although with a different outcome: “I don’t feel that I should work more for a good grade, on the contrary, for me it’s easier like this because otherwise I study a lot before tests”. These accounts together with the fact that students often say they have fun during the lessons indicate the presence of the blissful productivity element of gamification.

In addition, students also say that it is motivating for them, which might be explained by the intrinsic motivation caused by the experience of competence as described in Deci and Ryan’s (1985) self-determination theory and illustrated by the following quotation, for example:

What I like about this point-collecting system is the motivation that yes, you can do it, you can be good at math, and you don’t feel left behind the ‘good ones’ who always write perfect tests and you just keep trying to be like them. So it gives me such encouragement

[...] in other subjects I feel like I'm sitting in a constant competition, while here you actually compete with yourself.

As the quotation is from a focus group interview, I know that this student is very hard-working and usually collects points way above what is necessary for a grade five. However, another student who is among those who usually collect the fewest points in the class formulate similar thoughts:

Getting a bad grade is almost as difficult as getting a grade five [in this point-collecting system...]. The problem with me is that I'm not particularly hard-working, so at least it [the point-collecting system] is something that motivates me to do anything because even I have the chance to get a relatively good grade.

In connection with bad grades, another student concludes that this point-collecting system can work for everybody by saying that “whoever gets a bad grade here would get a bad grade in normal grading, too, but even if you don't like it [to be assessed like this], you can still get a good grade with writing a good achievement test and minis”. (Further reactions to alternatives for testing are detailed in section 5.5.2.1.)

Furthermore, a recurring theme is what grades received as a result of collecting points reflect or depend on. On the one hand, students list specific activities, while on the other hand, they express general conclusions. Section 5.5.2, Reactions to Other Assessment Types details several activities, so here I mention those that students relate to the point-collecting system such as “your class participation, homework, and test results all matter” or “it assesses nicely both how much I pay attention in class and how much homework I do”. Another student from the 11thME class explains her attitude:

If I really concentrate in class and understand the material, I can collect points with it multiple times because it appears in the mini and in the achievement test, too, so it motivates me to pay attention. The other thing I like is that there is no compulsory homework, but if I do something at home, it is appreciated even more [...] so it is worth doing it.

Connecting the specific activities with the more general conclusions, one student says: “I am glad that my future does not depend on a few isolated tests, but that my continuous work counts”. As already mentioned above, one of the characteristics of the point-collecting system that students appreciate the most is that their grades reflect the effort they invest in studying. According to a student from the 11thME class:

We can achieve a much greater sense of success because this point-collecting system is really based on diligence, so there are more opportunities for improvement, so it is easier to achieve a good grade than, say, in the other group, where I think I would realistically be [*sic*] a two or three.

It is worth noting the vocabulary usage that he would “be” a certain grade, which also reflects the central role of grades. As this student expressed these thoughts in a focus group interview, I could check that his midterm and year-end grades, received as a result of point collecting, were both four, just as his secondary school leaving matura exam grade at the end of the next year. Students’ 11th grade year-end grades compared to their matura exam grades can be seen in Table 12 in section 5.5.1.3, Connections Between Grades and Point Collecting. According to another student from the same math class: “I think in this [point-collecting] system, it is easier to get a grade five if you care about the tasks, and this can be seen in our grades because those who are lazy don’t get good grades”. In the 11thAME class, a student, who met this alternative assessment for the first time in this elective class, expresses similar thoughts about the point-collecting system:

It is useful if there is a topic that does not go so well because I can still achieve a good result. But I feel that these points reflect my diligence rather than my actual knowledge, and therefore I don’t quite feel that this knowledge would stick as much as I want it to.

This dilemma recurs in all the groups of the action research. In the 11thME class one student says: “it [the point-collecting system] doesn’t specifically assess how much someone knows, but how much they want to do something, and this is both an advantage and a disadvantage”, and in the 9thITE class: “it [the point-collecting system] assesses diligence, so even if you don’t always have the necessary knowledge, with a sufficient amount of practice, you can always get a better grade”. For this reason, some students find collecting points problematic (see next section), but others resolve this dilemma. For example, according to a student:

This point-collecting system is better because it measures how much time and effort we put in studying math, while traditional grading only shows how well we prepare for a given test and it doesn’t matter if the next day after the test I don’t remember half of it.

A lot of students also feel that this point-collecting system “accurately assesses both knowledge and energy put in studying”. In all of the groups of the action research, some students said that it “reflects real knowledge even better”. As I included Likert scale statements reflecting these ideas

in the questionnaires, quantitative data is also presented in section 5.5.1.3, Connections Between Grades and Point Collecting.

The fact that students experience less stress in this alternative assessment system has already appeared in several quotations. As a result, students mention further consequences they attribute to the point-collecting system such as the freedom to make mistakes and the possibility for individual learning paths illustrated by the following quotation from the 11thME class:

I'm really happy that my grade doesn't depend on one occasion [...] here we have lots of possibilities for a good grade, and you can make mistakes, your progress is much more individualized, so if I get a grade five and my classmate gets a grade five, we might have done different things to achieve it.

The same idea also appears in the 9thITE class: "this has been absolutely new for me, I've never experienced anything like this before, and I really like this individualized point collecting". As a further consequence another student from the 11thME class ponders that "tasks seem to be more creative [...] and the system assesses more skills because we work in groups, solve online quizzes, etc. and don't just get grades for tests". Moreover, they report elements that seem to promote lifelong learning. For example, in the 9thITE class a student says that:

it is easier to get good grades, which is an advantage, but I think we also learn a lot in the process, for example when I was preparing my presentation, I learned a lot of background information that I didn't put in my presentation [...] but it was really good to experience that I learn more about something I deal with more.

Another student from the same class said the same in connection with compiling tests for themselves. Section 5.5.2, Reactions to Other Assessment Types further details what students think about these elements such as alternatives for testing. Students in the 11thME class also believe that this alternative assessment facilitates their math learning: "I like the point-collecting system because this way if I can't solve a task immediately, I still have the chance to get a good grade when I later understand it" and "I like that even if math is not my strong suit, as I get points for trying, I feel that it is worth working on the math problems".

These characteristics appeared in students' responses between 2014 and 2017, too (Barbarics, 2017). Thirty-five students mentioned the appreciation of diligence such as "I'm satisfied because it reflects the work I put in studying" or "I feel that it was worth writing so much homework". Twenty-four students found it fair saying, for example, "it's absolutely fair: if I want

to be good, I have the chance to achieve it” or “earlier I felt that I studied in vain if I made a mistake in a test and got a worse grade, but now both my knowledge and diligence is assessed and not only a temporary state of mine”, which is connected to the result that 17 students felt that it reflected their knowledge. Nineteen students wrote about the variety of opportunities, for instance, “I like that we can carry out tasks that are connected to our own interests” or “this totally different approach is fantastic because in many subjects [...] students cannot present their knowledge in their individual ways, only one way” or “this point-collecting system gives more possibilities for students to get a good grade, for example, the mini tests have a good impact by collecting points and revising the material of the previous lesson thus continuously deepening knowledge”, which can be connected to the continuity of assessment mentioned by 14 students: “it’s so good that I can always see how many points I have, where I am in the process” or “I can see my continuous improvement” or “it immediately reflects your knowledge and participation” or “if I make a mistake I know that I can correct it, so I’m not afraid of it”. Finally, students mentioned increased engagement and motivation 15 times and decreased stress levels appeared in 9 answers (Barbarics, 2017).

5.5.1.2 What Students Would Change About the Point-Collecting System. This section presents the results on students’ criticism and suggestions concerning the point-collecting system. In addition to students’ questionnaire answers to “What would you change in this point system?” (see questionnaires in Appendix E), their criticism of the point-collecting system is also detailed based on the collected qualitative data (such as further open-ended questions in the questionnaires, focus group interviews, class observation reports, and the teaching journal).

During the action research, in all of the questionnaires about half of the students (50.8% on average) say that they would not change anything (or even write: “please, don’t change anything” or “I wouldn’t change anything just the number of people applying it - from the teacher's point of view” or “I can’t think of anything, which probably means that there are no big problems that need to be changed”), and others (39.2% on average) mostly continue praising the system (for example by writing “it is great as it is”, “the best”, “I love it”, “it’s fantastic”, or “it’s perfect, but if something comes up, we’ll tell you anyway”); however, one or two suggestions appear in each questionnaire (making up 10% of all the answers). I grouped these suggestions into one of the following categories: more opportunities to collect points, different point boundaries for grades, and dealing with points above the maximum.

In the category of more opportunities to collect points, students either ask for “more opportunities to collect points”, or they provide specific ideas as well. For example, “we could collect more points with in-class activities”, “there could be two grades: one from homework points and another from points collected in class and tests”, or “more difficult tasks for those who are quicker than we don’t check with the class together”. In the 11thAME class, a student who met this alternative assessment for the first time in this elective class would prefer collecting points only with tests: “I’d give points only for tests that would make up a final sum for a final grade”. Three students ask for some urgency solutions if there are only a few points missing for a better grade (e.g., “There could be, say, once a year that if someone misses 1-3 points from a grade 5, it could be collected with some extra tasks”), and one student says in a focus group interview that “it can be really annoying when only a few points are missing for a better grade”.

Probably inspired by the same experiences, others, instead, would change the boundary for a better grade, creating the second category of answers asking for “different point boundaries for grades” such as “less points for a passing grade” or “maybe lower the boundary for a grade five by about five points, but it’s just my personal opinion”. In other open-ended questions and in the focus group interviews, some students confess that they feel that the number of points for a grade five is too far for them, for instance: “sometimes I feel that the amount of work required for a grade five is too much for me” or “everything is perfect just the points for a grade five are a little high, but I guess the reasons are understandable”. Thus for some students it is difficult to reach grade five, while for others it is not challenging enough.

The last category of answers concerns “dealing with points above the maximum”. There are one or two students in each class who collect points way above what is necessary for a grade five. Some of them are satisfied with the situation, for example, by writing “I hope I’ll reach master points this month, too”. While others are not that satisfied. They wish that they “could do something meaningful with the Extra Master Points” or “there could be better things for the master points because I don’t feel it’s worth overachieving for me”.

Further criticism that students express are the following. From the 11thME class, one student writes in the anonymous questionnaire that “I’ve always thought it [the point-collecting system] is better than the normal school assessment, but sometimes there are situations where the normal method would be better”. Unfortunately, the student does not elaborate on what these situations would be. Another student from the same class writes that “to be honest, I’m a little lost

about the points received for independently solved tasks, but I guess it's perfectly fine". Others complain about their classmates (and indirectly my abilities to discipline) such as

the greatest disadvantage of this point-collecting system is that it only works if you take it seriously, [...] and I don't want to hurt anybody, but there are some students who just don't take it seriously, [...] and I hate it when time is wasted on begging them to do the tasks [...] instead of going on, asking questions, or discussing problems.

While most of the students feel that the system is fair by providing equal opportunities to everyone as discussed above, during a focus group interview, one student expressed her disagreement: "maybe it sounds a bit mean, but sometimes I feel like people who don't know the material well enough get good grades only because they're really hard-working". It is worth noting that this student performed highly above the class: going to math competitions, being the only one to take the higher level matura exam in math (with a grade five result), and continuing her tertiary studies in math. It also indicates how traditional assessment awards talent and knowledge, so students socialized in this system might have a difficult time accepting assessment that reflects effort.

The last area of criticism concerns point collecting remaining an extrinsic motivation for students. This criticism only appears in the focus group interview of the 11thME class. One student says that "the only negative thing I can think of is that when I collect enough points for a grade five, I usually lose the motivation to work more". Another student adds that it is also true for the types of tasks by saying: "there is a point beyond which it [the point-collecting system] no longer motivates you: if you collect sufficient amount of points from what you are good at, you will no longer have external motivation to learn what you are not so good at". However, one of their classmates argues with this notion saying that the system is built up in a way that you have to work with everything, or at least he tries to work on all kinds of tasks, while another student also admits trying to avoid some things: "I think it's a disadvantage that I also sometimes select tasks that I can do well and leave out the ones that I can't that much". All in all, it is important for the teacher to consider designing the point-collecting system consciously keeping in mind these student strategies.

In the survey conducted between 2014 and 2017, students expressed similar criticism (Barbarics, 2017). Although all of the students voted for keeping the point-collecting system, 34 students added some comments: 12 of them missed instant grades or traditional tests sometimes and some of them added that tests, especially achievement tests, should have more weight in the

point-collecting system; 10 of them criticized that the system reflects diligence more than knowledge (on the one hand “it’s too easy to get a grade five, you just need to be diligent”, on the other hand “even if you write a good test, you can’t get a good grade, only if you spend a lot of time with tasks at home, too”); and 12 of them asked for more possibilities to collect points (e.g., “more optional tasks such as ‘personal challenges’”, “I would make it harder to get points, but I would give more opportunities”, “more optional tasks that can be done at home” or simply “more homework”) (Barbarics, 2017). Comparing students’ answers from the action research to those of the earlier years, it seems that although I have developed in differentiation and providing individual learning paths, the diversity of students would require even more of these.

5.5.1.3 Connections Between Grades and Point Collecting. Students in Hungary take their matura exams at the end of their secondary studies. The matura exam is a nation-wide centralized and standardized exam in several different subjects and it is graded on the one to five scale as well. At the end of the action research, the mathematics in English class finished 11th grade (in 2020), and they took the matura exam the next year (in 2021 at the end of 12th grade). First, their 11th grade year-end grades that they received at the end of the action research as a result of point collecting exclusively are compared to their matura exam results. Then students’ questionnaire answers in relation to their grades based on point collecting are presented. All the questionnaires (see Appendix E) included the recurring questions of what grade they had just got, how well that grade reflected their knowledge and how well that grade reflected the energy they put in studying, how satisfied they were with the ways the teacher had assessed their work, and in optional open-ended questions they could give their reasons, too. In addition, they had to say whether, if they had been graded in the traditional way, the grade would have been either better, worse, the same, or they can’t decide whether the grade would have been better or worse. There was also an “other:...” option. This section ends with two more Likert scale questions of the student questionnaires asking how much they like the English language or mathematics in general and how much they like the IT English or mathematics in English lessons. Due to the emergency remote delivery, in the last two questionnaires a further Likert scale question was added about how much they like the online IT English or mathematics in English lessons.

Table 12 presents students’ grades as a result of point collecting at the end of the action research in 2020 (11th grade year-end grade) compared to their secondary school leaving matura exam grades in 2021. The average of the group in both cases is exactly 4.375 with three students

getting a grade better and three students getting a grade worse in the matura exam as opposed to their 11th grade year-end grade. Note that Student 4 in Table 12 took the higher level matura exam (both her 11th grade year-end grade and matura exam grade was five).

Table 12

Students' Grades at the End of the Action Research Compared to Their Matura Exam Grades

	11th grade year-end grade	Matura exam grade	Difference
Student 1	4	3	-1
Student 2	4	3	-1
Student 3	4	4	0
Student 4	5	5	0
Student 5	5	5	0
Student 6	5	5	0
Student 7	2	3	+1
Student 8	4	4	0
Student 9	3	4	+1
Student 10	5	5	0
Student 11	5	5	0
Student 12	4	5	+1
Student 13	5	5	0
Student 14	5	5	0
Student 15	5	5	0
Student 16	5	4	-1
Average	4.375	4.375	0

While Table 12 above shows objective grade results, Table 13 below presents students' subjective feeling of what their grade would have been if they had been graded in the traditional way. Mostly students feel that it would be the same (43%) or they cannot decide (30%); however, in a quarter of the cases they feel that their grade would be worse if they were graded traditionally.

These feelings appear in their qualitative answers, too. For instance, the student from the 11thME class whose focus group interview answer is also quoted in section 5.5.1.1, What Students Like About the Point-Collecting System, believes that he “would realistically be around a two or three”. In reality, this student is Student 3 in Table 12 receiving grade fours both as a result of point collecting and at the matura exam. Another student from the same class writes:

I’m always happy with my grade 5 :) In a different system, I would get a worse grade, but I’m glad that my grade isn’t worse just because my mind stops for a while during a test.

If you’re bad, your matura exam will be bad, and that’s punishment enough.

As the quotation is from an anonymous questionnaire, this student’s exam results are unknown; however, most students with grade fives during the action research achieved grade fives at the matura exam as well (only one student did not). There were only three instances (out of 138) when students felt that their grade would have been better if they had been graded traditionally, and they referred to specific situations such as not being able to collect enough points due to missing classes.

Table 13

Number (and percentage) of Students’ Answers to What Their Grade Would Have Been if They Had Been Graded in the Traditional Way

The same	Can’t decide	Worse	Better	Total
60 (43%)	41 (30%)	34 (25%)	3 (2%)	138 (100%)

All students in all the groups filled in the entry questionnaire; however, different numbers of students filled in the different questionnaires after each unit during the action research. As a result, students’ answers to what grades they got results in an average reflecting grades of only those students who filled in the questionnaires instead of the whole group. As I have the actual grades of all the students, comparing the actual averages of the whole group to that of the questionnaire answers provides some background information on who filled in the questionnaires. Table J1 in Appendix J includes all the averages and standard deviations showing that in most cases the averages based on the questionnaires are somewhat higher than the groups’ real averages indicating that more students with better grades tend to fill in the questionnaires; however, there are four cases when it is the other way around (the average of grades reported in the questionnaire

is less than the whole group's average), so it cannot be assumed that it is only students with good grades whose answers appear in the questionnaires.

Table J2 in Appendix J contains the averages and standard deviations of grades reported in the questionnaires compared to how satisfied students are with the assessment. Students had the chance to add their explanations why they feel satisfied or dissatisfied with the assessment. Some of the answers support the correlation by giving the same number for the satisfaction as the grade they got and adding, for example, "Is there anybody who gets a grade 5 and isn't satisfied?" (both the grade and the satisfaction is five), or "I didn't get a five 😞" (both the grade and the satisfaction is four), or "I think my performance is quite fluctuating, if I put myself into it, everything works, but if I don't concentrate or become lazy, I just don't work well" (both the grade and the satisfaction is three), or "I didn't do anything" (both the grade and the satisfaction is one).

However, there are cases when they differ. Some explanations for cases when the satisfaction is higher than the grade are the following. From the 9thITE class, students write: "this is what I deserved" or "the grade reflects my performance", from the 11thME class: "I always start working towards the end of the period to get a good grade, and now I feel like I worked even less, and that's why I didn't get a five" or "I think you were very kind and understanding the whole time, you didn't give up on me even after my disgraceful results, so thank you very much for that". It seems that there is some kind of self-criticism behind these explanations, which might be the reason why a student in the 9thITE class got a grade one and chose number four for how satisfied they were (but did not provide further explanation). In these cases, students seem to be satisfied with the assessment despite their own performance.

Some instances when the satisfaction is lower than the grade are the following. Even though getting a grade five, some students chose four for the satisfaction and explained: "I haven't worked as much as my grade suggests", "I don't think I've deserved the five although I have collected enough points for it", "I am satisfied with the grade itself, but it doesn't reflect my actual knowledge", "I could do much better", or "I have only done the homework and worked during the lessons". These explanations also suggest self-criticism and that the satisfaction (or rather the lack of it) refers to their own performance. A student in the 9thITE class got a grade four and chose number three for how satisfied they were providing the following explanation: "I missed the lesson when we wrote a vocabulary test, and later I didn't have the chance to write it". This clearly indicates dissatisfaction with me not providing him the opportunity. While another student from

the 11thAME group chose one for the satisfaction and added that “actually, I’m not satisfied with myself because I didn’t do anything, but the workouts don’t leave me time for anything, I’m a wreck every day”.

As the examples show, in some cases, students’ satisfaction or dissatisfaction refer to their own performance while in other cases to my assessment. This was one of the reasons why after the second questionnaire I introduced the following separate questions: “What do you think about your own performance in this period? What went well? Not so well? What would you like to do to make it better?”. The analysis of students’ answers to these self-assessment questions are in section 5.5.2.3, Reactions to Further Alternative Assessment Methods.

Students’ explanations why they are satisfied with their grade received as a result of point collecting are very similar to their reasons presented in section 5.5.1.1, What Students Like About the Point-Collecting System. One student from the 9thITE class writes in the first questionnaire: “this whole idea to collect points is very good”. They are satisfied because they find the assessment “fair”, “correct”, “realistic”. However, one student from the 11thME class admits “I rather just hope it’s realistic”. In the 9thITE class one student writes that “unlike earlier, I am finally happy with my grade here”. In the 11thME class one student writes that “this is the grade I would give myself”. They also frequently mention that they are satisfied because they “have possibilities to improve”. The theme of motivation also emerges in the 11thME class: “I work more and more and sometimes I do math even in my free time” or “I’m not much of a science person, but I spend a lot of time with math” and in the 9thITE class too: “finally, a subject in which it is not difficult to get a five. You have to work for the five nevertheless, but at least it doesn’t depend on whether you succeed or not on a 45-minute test, instead active participation in class is important”. In addition, they highlight different aspects, such as they are satisfied with the grade because it reflects “my performance accurately”, “the knowledge I acquired”, “how much effort I put in”, “how much I worked with the material in class and at home”, “my participation in class and knowledge”, or “both my diligence and knowledge”. Quantitative data on how well students think their grades received as a result of point collecting reflects their knowledge and effort is presented in the next paragraphs.

In Appendix J, Table J3 and J4 present all the averages and standard deviations of students’ answers to how much they think their grades reflect their knowledge and effort, respectively. Further averages and standard deviations of students’ questionnaire answers are presented in Table

J5 to the question how much the 9thITE students of the action research like the English language and in Table J6 the IT English lessons, while Table J7 shows how much the 11thME students like math in general and Table J8 shows the math lessons. The 11thME students took their entry questionnaire three years earlier, and their results together with the other groups I taught were published (Barbarics, 2017). These results showed a statistically significant change in all the above mentioned questions when comparing students' entry questionnaire answers to their responses after their first term; however, this change did not last (Barbarics, 2017). So when the point-collecting system was introduced to students, and they compared their grades received in the traditional system earlier to the ones they received as a result of point collecting, during the first term, they felt that their grades based on points reflected both their knowledge and their effort better, and they seemed to like both the subject and the lessons more; however, this difference was not statistically significant after the first term (Barbarics, 2017). Similar tendencies can be detected in the groups of the action research as well.

In the 11thME group, students have known the point-collecting system for three years, and their quantitative answers show rather fluctuating averages (see Tables J3, J4, J7, and J8 in Appendix J). Students' qualitative answers might explain these changes. For example, a student, despite getting a grade five, chose three for how much the grade reflected the knowledge in the 11thME class and explained that "there was a task in the achievement test that I couldn't solve, and another one for which I didn't know the shorter solution". Another student, who was asking for more challenging tasks in another question, added that "for me the tasks are relatively easy, and I collect the points for a grade five easily, so my effort put in studying the subject is almost zero except for working during the lessons" and chose one for how much the grade reflected the effort. As the sample is very small, ten to sixteen students filled in the questionnaires in the 11thME class (see Table 10 at the beginning of chapter 5.5, Students' Perceptions of Alternative Assessment Methods), even one such answer can cause the changes in the average of the group.

In the 9thITE class, there are some statistically significant differences between their entry questionnaire answers reflecting their primary school experiences and their experiences during the action research. For instance, a Mann-Whitney test indicated that students in the 9thITE group think that their midterm grades received as a result of point collecting reflected their English knowledge significantly better than their 8th grade foreign language year-end grade ($U = 35.5$, $p = .00694$); however, the difference was not significant in terms of their effort put in studying the

subject (see averages and standard deviations in Table J4 in Appendix J). In addition, a Mann-Whitney test also indicated that students in the 9thITE group liked IT English lessons significantly better at mid-term than their foreign language lessons in primary school ($U = 40.5$, $p = .01352$); however, the difference was not significant in terms of liking the English language in general (see averages and standard deviations in Table J5 in Appendix J). It is important to note that due to the small sample size, further research is needed to draw more conclusive interpretations, but these findings might inspire future research questions.

5.5.2 Reactions to Other Assessment Types

This section continues following the order of how alternative assessment methods are presented in chapter 5.4, Alternative Assessment Methods and Their Consequences. After students' reactions to point collecting come their reactions to other assessment types: first, their reactions to alternatives for testing (presented in section 5.4.2.3, Alternatives for Testing During the Action Research), then their reactions to activities that help process the material (presented in section 5.4.3.2, Assessment Activities in the Action Research Project), ending with their reactions to further alternative methods (presented in section 5.4.4.5, Further Alternative Assessment Methods Used in the Action Research).

5.5.2.1 Reactions to Alternatives for Testing. Following the order in section 5.4.2.3, Alternatives for Testing During the Action Research, first students' reactions to alternatives for oral tests are presented, then to pop quizzes, ending with achievement tests. For each alternative assessment method, students were asked whether they would like to have it in the future with answer options: yes, no, other... (Table J9 in Appendix J summarizes all the responses), and they could give their reasons, too (see questionnaires in Appendix E). Each questionnaire inquired only about those assessment methods that appeared during that unit, so students could write about them one to three times. In addition to these responses, other relevant data from the questionnaires, focus group interviews, observation reports, and the teaching journal are presented.

Voluntary oral tests were introduced in the 9thITE group during the first term. Eight students (out of 11) answered that they would like to have it in the future as well, mainly because they felt it was a good way to collect points. For example: "it's good that it's not compulsory, but if you feel like it, you can easily collect points with it" or "it can be very useful if you only need 10-20 points for a better grade". Another student added: "I like it because you can find out how much you know". The latter comment suggests that for this student, these voluntary oral tests fulfill

a summative role. Only one student answered no and explained: “I don’t really like oral tests”. Two students chose the other option, one of them writing: “I don’t know” and the other: “only if they’re easy”. Unfortunately, there was no further explanation of what would make these tests easy.

As the beginner 9thITE class gradually spoke more and more, voluntary oral tests became scheduled presentations. First, they happened in class and after the introduction of the emergency remote delivery, online. Out of 26 responses to whether they would like to give presentations in the future, students answered “yes” 19 times, “no” six times, and there was one “other” explanation (Table J9 in Appendix J) that said “it depends on whether the topic is interesting or not”. Those who answered no added that they “don’t like presenting” or “don’t like performing in front of others”, and one student criticized the number of points they could collect with it saying that “it is worth less points than compiling and writing the achievement test, while it [preparing a presentation] takes more time, and you also have to learn the presentation, which is more work”.

All the other students (73%) did not only want to have the opportunity to give presentations in the future, but also praised it on numerous occasions. In their questionnaire answers, students gave the following reasons why they wanted to give presentations. They wrote seven times that it was a great opportunity to collect points, which is a reason that appears in connection with all the activities that students could collect points with. In addition, they highlighted enjoying the process of preparing the presentation, such as “I love making a ppt 🤖”. Five students mentioned that they learned a lot while creating it and one of them added that “you can practice speaking in English”. One student wrote that “although I don’t usually use this option, some people like it, and it’s a very useful way of learning, so I think we should definitely keep it”. The last aspect, which also appeared several times, was that presenting students felt that they were teaching their classmates, for example “I like it because I learn a lot by making it, and because I teach a lot of people by performing it :)”, and the audience also enjoyed it “because learning from a classmate is different”. A class observation by an American teacher trainee (as mentioned in section 4.4.3 Action Research in the School Year of 2019 and 2020), which made it possible to get insight from a native speaker, contained the following about two student presentations:

Both students’ presentations were on social media platforms, the first student’s was on Snapchat, and the second student’s was on Instagram. Both students created PowerPoint slides to use with their presentations. Although each presentation had a significant

number of word choice and grammatical errors, I was impressed by how the information was still mostly intelligible after only having formally studied English for one year.

In fact, as the presentations took place on February 21, 2020, the two presenting students had only been learning English for half a year.

During the emergency remote delivery, presentations continued online (or in the form of vlogs that are detailed in the next paragraph). In the focus group interviews, six students mentioned that one of their favorite things was the online presentations. All of them emphasized the enjoyment of learning from each other:

It wasn't the teacher talking, but other students were talking, and I could learn from them, and then we discussed together what their mistakes were, and I could learn from that, too, and then the questions about what words to take out of it, and so on. I really liked it, and I think it was even better like this than in class.

Others also confirmed that they enjoyed the online presentations more, for example, "I preferred this screen sharing because I don't really like making ppts" or "I liked that I didn't have to go and stand in front of others pointing at a ppt, [...] it's just not my favorite thing to do, so I preferred it online way more".

Students also had the chance to record themselves and send their vlogs to me. Out of 36 responses to whether they would like to have the opportunity to create vlogs in the future, students chose "yes" 22 times, "no" five times, and wrote nine "other" textual answers (Table J9 in Appendix J). There were some very straightforward rejections: "I haven't tried and I don't want to", "it's cringe talking to a camera", or "I'm not interested in vlogging". The nine other textual answers were more neutral such as "I don't know", "maybe, it's not my style, but if somebody enjoys it, it's also a good way to collect points", "as I've never done it, I can't comment on it" or "I don't like memorizing things, but it's a good idea anyway, just making ppts is still my favorite thing because by the time I memorize something I could write it down twice". The latter comment suggests that this student considers presentations for which they have to create a ppt but then improvise the speech more spontaneous, while if it is recorded, they would have to memorize texts. This opinion is in contrast with the experience of other students who created vlogs, who make comments such as "I like that it's more personal than learning a text word for word" or "I think it's not difficult because you just have to look into things a little and then talk about them". Others like it because they find it "imaginative", "diverse", "enjoyable", "interesting, and some people

are more inclined to create this than other things”. In addition, they also think that they “can learn a lot from it”. Despite the joking tone of the following comment: “I haven’t made one yet, but now I want to show off my talent :D” it can be inferred that this student considers vlogs as opportunities to present what they know.

Moving on to written tests, first alternatives for pop quizzes, then alternatives for achievement tests are presented. Mini point-collecting opportunities (i.e., minis as described in section 5.4.2.3, Alternatives for Testing During the Action Research) have been used with the 11thME class since 9th grade, which might be the reason for the 24 “yes” answers out of 25, no “no” answers, and one “other” answer (Table J9 in Appendix J) to the question if they wanted to have minis in the future. The only student choosing the “other” option explained that “sometimes they’re annoying because I don’t feel like writing, and I could just work on the problems instead”. It is interesting to note that some students notice connections between minis and traditional pop quizzes such as “they are good ‘pop quizzes’ and opportunities to collect points” or “they are like pop quizzes, just they help you practice the material”. In addition to collecting points and practicing the material, the main reason why students like them is that they provide formative feedback: “I like them because they show if you understand the material”. In the focus group interview one student says:

minis are really good, I love them, I think it’s one of the best ways to practice the material that you can collect points with, and I can see what I get points for, [...] so I can say “good, I already know this”, and if I get less points for something, I know that I need to work more and practice that.

Others do not seem to mind either if their deficiencies come to light: “I like minis because if there is something I don’t know or can’t apply in ‘practice’, it becomes clear there and as they say: ‘one learns from one’s mistakes’” or “it’s really useful to get feedback on what you don’t know, so it doesn’t turn out in the achievement test at the last minute, and even if you already know it, it’s still a good practice” or “it’s a good way to test what really stuck from the material or if you just memorized it for one lesson”. Students appreciate getting feedback without high stakes: “minis assess our current knowledge, and I like to know how I’m doing, and minis are not a question of life or death but they’re about points with which you can only increase your scores higher and higher”. In addition, “with the help of minis, you can easily prepare for the achievement test”. Furthermore, students’ motivation also emerges: “they motivate me, so I prefer paying

attention continuously throughout the level because if you're hardworking, relatively many points can be collected this way, so you can be calmer at the end of the period" or "I feel more and more when I can solve a problem in class, I will also be able to do it in the mini, which motivates me even more :)". Students' opinions were also asked about minis online, and their reactions are detailed in section 5.5.3, Reactions to Assessment During the Emergency Remote Delivery.

In the 9thITE class, alternatives for traditional pop quizzes were offline vocabulary quizzes and online quizzes such as Kahoot! that took place both in class and online during the emergency remote delivery as well. The latter will be elaborated on in section 5.5.3, Reactions to Assessment During the Emergency Remote Delivery. Students also knew about them in advance and could collect points with them. Out of 13 students nine wanted to have vocabulary quizzes in the future, and two students did not want them and two chose "other" (Table J9 in Appendix J) either because they "did not like them" or wrote "I don't care" one of them adding that "I'm obviously learning from them, but memorizing things is a bit boring". In addition to the usual "it's a good way to collect points", students also mentioned that "they motivate me to learn the expressions" or "vocabulary quizzes are important, or rather learning words is important in order to have a good vocabulary". The formative function of quizzes also appeared: "I can check my knowledge through them". Playing Kahoot! in class received 40 unanimous "yes" votes (Table J9 in Appendix J) including one "other" comment: "I would like to [play Kahoot! in class in the future], but we should get more points even if I am last". Many students wrote something along the lines of "it's fun and exciting and useful for learning at the same time" or "it's entertaining and you can learn a lot from it". Some students highlighted its game aspect: "we can learn through it playfully" or "it's good to compete with the others 😂😁".

Playing Jeopardy! in class also received only one "it doesn't really matter to me" in addition to the unanimous 23 "yes" votes in the 9thITE class (Table J9 in Appendix J). Beside the point-collecting opportunity, students liked it because "it was funny", "comprehensive", "it's easier to learn like this", "I actually learned most of the material then", "it's a great way to summarize the topic", "I could see the whole unit in one", "it refreshed my brain very well before the achievement test", and "I could check how much I know". There was only one less enthusiastic student, who wrote: "it's true that I learned a few things from it, but... I just don't really know...". In the 11thAME group, sample tests served a similar role although in a less playful way. However, based on students' reactions, they still achieved the following results: "it made the preparation [for

the achievement test] so much easier”, “it was good because we knew what we needed to review for the achievement test”, “it helped in narrowing down the curriculum on what to prepare for”, and according to one student: “it was nice to see what the tasks looked like, but it wasn’t much more useful than a simple revision lesson”. Other students would disagree with that by saying: “it is very useful and should be introduced in every subject” and “it helps you assess your own knowledge and realize your shortcomings and provides time to make up for them”. As one of my goals was to practice advanced math in both English and Hungarian, I also asked students’ opinions about the lessons being bilingual. From those students who had already been studying math in English, it was no surprise to receive answers such as “I got used to it”, “I have no problem with it”, or “I’m glad that I’m preparing for the matura exam in both English and Hungarian, and I’m using my professional vocabulary, so I won’t forget it”. It was also reassuring to read students’ feedback who otherwise studied math only in Hungarian: “it’s understandable in English, too”, “nothing special, it’s good”, “it is very useful, as in addition to practicing the given language, it makes things easier if we have to deal with it in a foreign language later on”, or “it’s good because it makes it easier to find literature not only from Hungarian sources”.

During the action research all alternatives for testing were worth points and counted into the point-collecting period. The fact that even achievement tests were not graded separately blurred the distinction between the summative and formative functions of assessment as the following quotation by a student from the 11thME class illustrates: “it was good that you had enough time to solve tests, which helped to really understand the material, so if you hadn’t understood it before, you could through the test”. This implies that students could perceive even achievement tests as opportunities for learning. There were different kinds of alternatives for achievement tests in each group. In the 11thME classes, there were achievement tests with increasingly more difficult tasks (see sample from the 11thAME group in Appendix I3). I asked students’ opinions about the fact that there were optional, challenging “bonus” questions at the end of achievement tests. Even those who did not solve them did not seem to mind as according to one such student:

For those who are looking for a challenge or want to collect more points, it must be good, but I don’t think I’ve solved a single bonus problem so far. (Now I have at least read them in the previous test. I hope this can be considered progress.)

Others considered it an additional chance to collect points or compensate for something that they could not solve from the compulsory section. For instance, “it helps a lot to get more points if I

can find the logic in one of them, while not in some ‘basic’ tasks” or “it’s a good lifeline if I don’t know how to solve one of the [core] problems”. Furthermore, it might even affect how students see themselves as one student writes:

I am usually scared of them because they are bonus problems, but they are not more difficult than the others, so I still have a chance to get many points for the test even if I could not solve one [from the core problems] but the bonus problems went well.

These bonus problems seem to fulfill the role of differentiation as for some students “they are often good for fighting boredom”, others find them “good because they are interesting”, or “they are a challenge and I enjoy them much more [than core problems]”, or “I like them as they are more thought provoking”. There was one student from the 11thAME group who gained immense motivation thanks to these bonus tasks: “after the bonus problems in the last test, I started working on these materials for at least 10-15 hours a week, and I realized that I immensely enjoy them”. Nobody expressed any objections to these alternatives for the achievement tests.

In the 9thITE group students could compile achievement tests for themselves. In response to whether they want to have this opportunity in the future, they answered “yes” 32 times, “no” once, and chose “other” twice (see the numbers in Table J9 in Appendix J next to “students’ own tests”). The student giving the only “no” answer explains: “I don’t want to deal with this, too”. The two “other” answers are more permissive: “it doesn’t really affect me because I don’t have a lot of free time to spend on it, unfortunately 😞, but I don’t want to take away this option from others” and for the other student “it depends on the material”. Their classmates seem to enjoy it for the following reasons. “Obviously because it’s easier, but I still learn what I need to”. “It is much easier for me to get myself to study like this, and I can purposefully learn what I will need, thereby reducing my stress about what will be in the test”. “While I’m putting the test together, I’m already memorizing a lot of things, and if I’m diligent, I don’t have to worry about my grade”. “It’s interesting and really asks what we know”. “I loved putting it together, and then it’s so much better to write my own”. “This is one of the best things about education”. In response to whether they would like to continue using the point-collecting system in the future, one student also adds that they want to because of the opportunity to create tests for themselves. This questionnaire answer might come from the same student who said in the focus group interview that his favorite thing was to “build up achievement tests for ourselves”. Another student also details his experience of creating achievement tests: “I had to search for specific things, and when I searched for

questions in this way, I also found a lot of new information along with the questions, so that's why I really liked it".

Students in the 9thITE group could also create posters in connection with different topics of each unit. In their questionnaire answers to whether they want to create posters in the future, they answered "yes" 15 times, "no" twice, and chose the "other" option once adding up to 18 responses (Table J9 in Appendix J). The two students answering "no" explained their decision in the following way: "I don't like to do crafts, and it is difficult to solve printing at home" and "I'm not that interested in it". The explanation for the "other" is "whoever wants to, feel free to do it, but it's not my favorite because it's too messy, and if you're done, but you think of something good, you can't really edit it anymore". Others appreciated the opportunity to collect points in an additional way and listed some further reasons. For example, "it's exciting", "I have to be creative to make one", "it's good to summarize something through images", "it's enjoyable, playful, and you can learn with it", "it's simpler than a presentation, but it still involves a lot of work", and "I learn a lot by making it, and I can teach others when performing it :)". Teaching others was also mentioned in relation to students giving presentations earlier in this section, and it reappears in relation to further activities that help process the material in the next section.

5.5.2.2 Reactions to Activities That Help Process the Material. Following the structure of section 5.4.3.2, Assessment Activities in the Action Research Project, first students' reactions are presented to further activities that they could collect points with, then to activities that they did not collect points with. Group work falls into both categories as in some cases, students could collect points while working in groups, in other cases, they did not.

In the 9thITE group students could not only play Kahoot! in class, but after the in-class game, they could also practice it at home by individually going through the questions as many times as they needed to. Their reactions to the homework variation is discussed in section 5.5.3, Reactions to Assessment During the Emergency Remote Delivery because students elaborated on their experiences about it more in relation to online classes. Students could also contribute to the game by writing questions. Interestingly this point-collecting option received the most support by 35 "yes" responses out of 35 (Table J9 in Appendix J) to the question whether they wanted to have this opportunity in the future. In addition, students also mentioned appreciating this possibility during the focus group interviews. Sixteen responses in the questionnaires include that "it is a good way to collect points". Eleven responses contain that students can learn while writing the question

and answer options, for instance, “it is the same as compiling your own achievement test, you learn a lot while making it”. Some of them highlight that “it’s useful because we are improving in formulating questions in English”. Even students who did not write any questions support the idea: “unfortunately, I don’t have time for this either, but whoever wants to write, should feel free to do it, I don’t want to ruin it for anybody”. One student writes that “I can also assess my knowledge through this” and another one likes the fact that “the person who wrote the question has an advantage in the game”.

Students in the 11th grade math groups could practice at home with the website Khan Academy. The site did not only appear when answering to whether they wanted to use it in the future, but also in several other responses in all the questionnaires and the focus group interviews as well, which reflects its popularity. Out of 49 responses, students voted for using Khan Academy in the future 40 times, voted against it three times and gave other explanations six times (see Table J9 in Appendix J). Those who did not vote for it provided the following reasons: “I don’t really feel that solving these kinds of tasks is an effective learning method for me”, “it’s a bit fiddly for me, so I prefer to work with tasks given on paper”, “not likable”, and “for me, the inclusion of any kinds of computer technology in learning is a disadvantage as I don’t find solving assignments online useful”. The last comment happened before the introduction of the emergency remote delivery although this student’s opinion did not change as a result of the pandemic either.

Others like Khan Academy for various reasons. They enjoy the aspects of gamification:

I love Khan Academy because I can do the tasks there, get points for them, and develop my cute avatar, and get points for them in math class as well, so it’s a double reward [...] and if you have enough points, you can also develop backgrounds.

Most students appreciate the immediate feedback and the possibility of asking for hints and watching explanations. For example, “it’s good that the man explains it if I get stuck”, “there are difficult tasks, but it’s good that at the end it shows how to solve them and doesn’t just give the correct answer” and “if I mess up, it shows me how it should have been solved, and I can do it over and over again until it’s correct, and that’s how I learn it because it always gives new problems to see if I really understand”. These accounts affirm Sadler’s (1989) view that effective feedback provides information not only about the students’ current level and the goal, but also how to close the gap between the two. In addition, it is also students’ choice when and how much help they

enlist as one student compares it to the mini point-collecting opportunities (alternative pop quizzes): “it is similar to the minis, but you can get help here even before starting the task”.

Another factor that students like is, for example, that it is in English: “unfortunately, or not unfortunately, it’s my favorite way to collect points. It’s good that it explains the tasks in English, and it’s relatively easy to understand”, “it is a very useful summary of what you have learned, plus you’re improving your English vocabulary”, and “it helps to develop your math in English”.

Students also appreciate that they can do it any time, and they want to have more of it: “we could have it more often because that way I was able to practice for the achievement test, which is very important to me, because I can’t do well without practicing” or “I love it, and at the weekend I rather do this than other tasks”; however, it also has the disadvantage of taking their free time, as one student summarizes it: “this way, I can access practice tasks at any time, but not in class, so it is only useful if I’m willing to spend my free time with it” or “it’s a good activity if I have time to fiddle with it at home”. My goal is that students really learn the material while doing the tasks in Khan Academy and not only do them for the points, so in our system they only get points for it if they do the tasks perfectly. Some students understand this connection: “you can earn extra points with these, so it’s a bonus, but it’s quite difficult to do the quizzes 100%, so this way you can also learn from them”, while others are annoyed by it and express further criticism:

I don’t like it [Khan Academy] that much because sometimes I suffer for a long time to get it to 100%, and sometimes I just miss it because of a typo. But the idea itself is very good. I often find the tasks difficult and even after asking for help, I don’t understand them. And it is not always clear how many points are worth what.

Another student also finds the explanations insufficient: “it’s useful, a practice opportunity, I am happy for that, but sometimes we could also discuss its tasks in class because I don’t always understand them at first, sometimes not even with the explanation”. While others say that “I like them because they are really easy”. These answers reflect the diversity in the group and the demand for differentiation. Meeting students’ individual needs can partly be achieved through using such websites, which is also reflected in the answers of the 11th grade elective advanced group.

Students in the 11thME class had been using Khan Academy since 9th grade while most students in the 11thAME group met the website for the first time at the beginning of the action research project. It gained popularity among them as well. Among the reasons, gamification also appeared: “I love getting unreasonable and completely unnecessary points in large quantities to

unlock completely unnecessary backgrounds and profile pictures that practically only I can see :D". At the very end of a questionnaire (see question 14 in Appendix E9) one student added: "I had no problems with the elective advanced mathematics lessons. I especially liked the Khanacademy homework". Most of them also met the point-collecting system for the first time and connected it to this website: "I liked the point-collecting system because of Khan Academy as those tasks could be completed quickly and were like assessments, and in the end, these 'small assessments' made up the big grade" and "I'm satisfied with my grade because I'm not the best in math, but e.g. I spent one of my Saturday mornings with Khan Academy, so my diligence could be added to it [the grade]".

Students in the 9thITE class could collect points with further individual activities such as note taking, solving task sheets, doing listening tasks, writing essays, and so on. Adding their responses to whether they would like to have these activities in the future, 88% of students would like to (see Table J9 in Appendix J). There were two negative responses each for the writing and the listening tasks with the following explanations: "I didn't do the essay homework", "I don't like writing tasks", and "I don't like listening tasks". In the positive responses, it being an easy point-collecting opportunity only appeared five times (12%) in connection with note taking, and all the other explanations referred to the content of the different tasks. For instance, in connection with note taking: "you can remember things better if you also write them down", "it motivates us to pay attention", "I can learn from my notes, and it encourages me to write as much as possible", and connecting it with point collecting: "perhaps the material comes across even better if we take notes, and the points are a good incentive for this". Students expressed the following opinions about writing tasks: "you can learn the most from this", "this improves our grammar a lot", "it's difficult in Hungarian, so especially in English, so that's why you need to practice, and it can be very good if the the topic is good", "it's very enjoyable", and "it's important to be able to write in English". In connection with listening tasks, students wrote: "you can develop a lot from it", "it helps to practice understanding English", "it helps to develop our listening comprehension in English", "it improves our understanding, so it will be easier to speak English in life", "we need to improve our pronunciation", "it makes our ears more attuned to English", and "I like it because it's easy for me".

In addition to individual activities, students frequently worked in pairs or groups of three or four. First, students' reactions to activities they received points for are presented, then their

responses to the ones that did not earn them points are described. In addition to the Jeopardy! game show themed activity described in the previous section, students collected points with treasure hunts and posters made in groups during the lesson. In connection with these activities, only three students mentioned that they liked them because they could collect points with them, all the other explanations referred to the nature of the tasks. There was only one negative answer saying “I don’t like working in groups”. Those who chose the “other” option explained that “it is difficult to work in a team with this class”, “it doesn’t work with everybody because not everyone wants to do the tasks to the same extent”, and “I’d work in groups only if I could be with my friends”. As during the treasure hunt, students had to move in the classroom to different “stations” and solve problems there on a piece of paper, one student did not seem to enjoy the exercise and wrote that “I’d do it only if the papers wander and not us”. Some students seemed to enjoy working with others in general and wrote that “group work is always fun”, “working in groups is always more exciting”, and “it was a lot of fun”. Others appreciated the team building aspect of the tasks: “this way, the class can get used to each other more easily”, “it brings the class together”, “I like that you have to think together”, and some students added that they could also learn: “you can learn a lot from this, and we can even get to know each other”, “it’s easy, exciting, and it teaches”, “it provided me with a lot of useful information, which I was able to use in my achievement test”, “we are learning in a playful way”, and “as it’s a competition, it motivates you more, and you can also learn from it”. As everybody received the number of points their group collected based on the number of correct solutions regardless of the results of the other groups, it seems that students still perceived the activity as a competition.

Students did not seem to miss collecting points from the last category of activities that helped students process the material although they could not collect any points with them. The 9thITE class gave varied reactions to the different activities when asked if they wanted to have that activity in the future (see Table J10 in Appendix J). The following list contains them in the order of popularity. Students liked the most, with 100% support, collecting words, expressions, and sentences in English while throwing Wendy around and creating a mind map together as a group. They found brainstorming with Wendy “fun”, “playful”, and “cute”, adding that “we could also learn through this” and “it helps you practice how to phrase something”. In connection with mind maps they wrote that “it helps you learn”, “it makes the material completely transparent”, “it

helps connect things”, and “it’s a good way to summarize what you’ve learned so far, and it’s easy to learn from it”.

Working in a shared Google document to collect all the material of the unit was also used for summarizing. Seven students (out of 11) answered that they would like to do it in the future, one said “no”, and three chose the “other” option (see Table J10 in Appendix J). The student who said “no” explained that “it gets very confusing and hard to learn from (for me)”. All the three “other” explanations included the condition that their classmates should take the task seriously referring to the fact that as it was edited by the whole group together, somebody always inserted some unrelated pictures or texts disturbing some and inspiring others to do something similar (see my notes in the screenshot of the teaching journal in Appendix D1). Some students even wrote in the questionnaire that “sending memes was fun”. Others enjoyed the collaboration and liked that the material was collected into one document to learn from.

Eight students (out of 13) enjoyed submitting their answers through the website Socrative (see link in Appendix B) in class before the emergency remote delivery, and five students chose “other” and either did not remember the website or claimed to be missing when it was used. Some students who liked it also added that “it’s much easier than submitting things on paper” and “it’s not boring and helps in learning”.

The two least favorite activities, in which less than half of the group would like to have it in the future, were the mistakes maze and the reading comprehension task including teaching others. The penultimate activity was the mistakes maze in which students had to correct their mistakes in pairs by going through a labyrinth (see Appendix I6). The five students (out of 13) did not like it because they found it “too difficult”, which might be caused by the fact that the sentences contained grammatical mistakes that I collected from their work. One student was missing and another wrote in the “other” section that “it was boring and complicated, but that doesn’t make me negative towards it in the future”. According to the six students who liked it, it was “difficult but good”, “interesting”, “challenging”, and “fun, like a puzzle”.

Students’ least favorite activity was teaching each other the material after working with a text individually. Four students liked it (out of 11) mainly because of working together with others. Six students did not like it because “we couldn’t teach each other”, “the others didn’t take it seriously”, “it was too fiddly”, and “it didn’t work out that well, but the idea wouldn’t be bad”. According to students’ feedback on other activities such as their presentations or posters detailed

in section 5.5.2.1, Reactions to Alternatives for Testing, they enjoy teaching each other, so as can be seen in the last comment on this reading comprehension task, teaching each other was not so successful in this activity. Further research is needed to determine the cause, which might be the difficulty of the text or faults in the instruction.

5.5.2.3 Reactions to Further Alternative Assessment Methods. Following the structure of section 5.4.4.5, Further Alternative Assessment Methods Used in the Action Research, students' reactions to peer assessment, self-assessment, and different types of feedback they received from the teacher are presented in this section, in addition to the instance of registering grades before closing a point-collecting period in advance. This instance only happened in the 11th grade math groups before midterm. Although it was not the end of the point-collecting period, some students had already collected enough points for a grade five. According to their questionnaire answers, as a result, some of them did not even realize that they received their grade in advance. Others were content about it, for example, saying "since I got what I wanted and what I achieved, I am happy about it", "I knew that no matter what [...], I would still have the opportunity to reach the level I had chosen", "it was a bit unorthodox, but it was the right solution for the situation". Others were a little more worried about the anticipatory trust: "I was under a little more pressure [to achieve the set grade]" or "I liked the method itself, but it only works for those who can return the favor, for whom it means motivation to reach that certain grade, but others might try to take advantage of this". In spite of the worries, everybody reached the set number of points for the desired grade.

In terms of peer-assessment, students gave feedback to each other's presentations and posters. After introducing the possibility of presentations and posters, there were three units, so students filled in three questionnaires unit after unit. In response to the question whether they would like to give feedback to each other in the future, the ratio of their answers in the three questionnaires changed in the following way. There were "other" options chosen only the first time saying "I think it's unnecessary, but if it makes someone happy, so be it" and "it would be useful if everyone took it seriously". The ratio of "yes" votes went from 38% to 80% ending with 87.5%, and the ratio of "no" votes from 46% to 20%, ending with 12.5% in the three questionnaires, respectively (the sum of the number of their answers can be seen in Table J10 in Appendix J). These changes over time suggest that unit by unit students realized the benefits of giving feedback to each other. Their open-ended questions also reflect this shift. First, the majority of answers were dissenting such as "I don't like writing", "nobody reads them", "we don't accept

each other's recommendations anyway", and "I think it's unnecessary as everyone marks something else, so you don't really get a clear result". For the second and the third time, there was only one negative answer on each occasion saying that "the others just randomly choose something, so it's not realistic" and "it doesn't really work".

On the other hand, positive views increased. There were some students who wrote for the first time that "we can help each other a lot with this", "the presenter can learn from this", "this way I can get feedback about my performance", and "it's good feedback [...] because we can see 'ourselves' from several points of view". Similar opinions appeared later supplemented by more details such as "it's good to know what other people think about my work", "presenters have the chance to learn from their mistakes", and "presenters can more easily recognize their mistakes and also the strong points of the presentation". In the last questionnaire students wrote: "it's very instructive", "this is how we can learn from our mistakes", "it always helps you in what and how to be better", and "it's really good for the presenter, so you know how to improve". Moreover, these ideas appeared in the focus group interviews, too, as mentioned in connection with presentations in section 5.5.2.1, Reactions to Alternatives for Testing. One student highlighted that his favorite parts were the presentations when "everyone shared their screen, talked about it, and then in the end, we discussed what was good about it and how it should be done better, which helped a lot".

Students' reactions to self-assessment show a great variety as well. As mentioned in section 5.4.4.5, Further Alternative Assessment Methods Used in the Action Research, in the 9thITE group I first asked students to answer questions ("What do you think about your own performance in this period? What went well? Not so well? What would you like to do to make it better?") on paper only for themselves. Then in the following questionnaire I asked their opinion about it. In response to whether they would like to do it in the future, out of the 13 answers only four chose "yes", seven chose "no", and there were two "other" choices (see Table J10 in Appendix J). Students voting for "yes" wrote that "it's useful to me", "I've summarized my tasks", "it helps me judge my performance", and "just because". All the others found it "unnecessary", "not helping", or wrote that they did not enjoy assessing themselves. One of them added that "I'm doing self-assessment anyway" and two students did not see the point in praising themselves: "I would rate myself as good" and "if I say I'm great, it sounds pretentious, even if in exceptional cases I am". After this first attempt, I introduced exit tickets (see sample in Appendix I7) with a slightly higher success

(see numbers in Table J10 in AppendixJ). Out of 11 responses, five students voted for and five against having them in the future, and one student chose “other”. The explanation for the “other” option said “it must be useful for the teacher, I don’t care if we have it”. This student seems to realize the diagnostic function of the exit ticket; however, not the self-assessment aspect of it. Similarly to the previous self-assessment attempt, students voting against found it “unnecessary” and did not see its point: “I didn’t like it, I don’t know why we need it”, “I’m lazy, so I didn’t really want to fill it out”, and “it didn’t provide me with extra knowledge, what I knew I knew without writing it down”. One of the positive answers reflects one of its intended purposes: “I can see how much I’ve improved”, and the other also refers more to its diagnostic role: “I like that the teacher asks for our opinion about our education”.

After these attempts, I introduced the self-assessment questions (“What do you think about your own performance in this period? What went well? Not so well? What would you like to do to make it better?”) in the questionnaires of all groups, so students would fill them in after each unit. One of the disadvantages of students not using consistent pseudonyms through the questionnaires was that I could not follow individual student’s anonymous progress, for example, in their self-assessment responses. However, I could look at general tendencies. In connection with self-assessment questions in the 9thITE group, students answered 749 words in 50 responses, which is 15 words per response on average, while in the 11thME group, students answered 1643 words in 61 responses, which is 27 words per response on average. So more 11th graders than 9th graders filled in the questionnaires, and they also gave longer answers.

In terms of content, students in the 11th grade groups pointed out specific materials or math problems that went well or not so well, which served a diagnostic purpose, too, so I could see how students felt about certain topics. In connection with their own performance, there were accounts of being proud of themselves, such as “math went pretty well this period, I got master points for the first time, so ‘something clicked’ regarding math” or “I’ve got used to it, I’ve realized how it works, now it’s easier for me to adapt, I don’t get upset if I can’t do something because I know that I will by the time of the test”. These accounts show a lot of self-reflection, just as in this other example: “I did well compared to the class, but not compared to myself” or

When I sat down to do math, I always solved what I set out to solve. Sometimes I had to use help, but I never gave up on a problem because it didn’t work at first. Maybe I should just get myself to deal with the problems more. I was always able to do them when there

was a deadline, a specific task, but it is difficult for me to start dealing with them on my own. I think it would help if we got more tasks uploaded as assignments.

Similarly to this student, others also phrase exact things they would like to do: “I should practice more, I could take better notes (with more transparent task descriptions), and I shouldn’t give up on solving a problem too soon”, “I should request for help more and work in groups”, “what I liked, went well, so I should deal a little more with topics that I don’t necessarily enjoy learning”, or “keep a deadline diary”.

In the 9thITE group, there was an increase in students’ self-reflection. In the questionnaires of the first term most answers were very general such as “I’m satisfied”, “everything went well”, “nothing [should be done differently]”, “level 2 [went well]”, and so on. There was only one specific activity mentioned: “I learned the most during the Kahoot! games”. In addition, what students think they should do were also rather general: “prepare more for the lessons”, “learn more”, “spend more time with it”, or “pay more attention during lessons”. During the second term, more and more exact tasks and areas that went well, or not so well, appeared such as “compiling the achievement test”, “the tasks sheets in class”, “my presentation”, “the poster”, “the sudoku” (see in Appendix I10), “definitions”, “making the vlog”, “giving feedback to each other”, “listening tasks” for what went well, and “listening tasks”, “the essay”, “writing tasks”, “grammar”, “fluently speaking in English” for what did not go so well. In addition what students would like to do also had more details in the questionnaires of the second term: “learning the words and abbreviations”, “for my next presentation I need to do some more research”, “I shouldn’t forget citing my sources on the poster”, and “next year, I’d like to do a language exam, so I should start preparing for it”; however, a lot of “everything”, “nothing”, and similarly general responses also remained in the second term as well.

There were no direct questions in connection with any kinds of feedback from the teacher, so this paragraph presents instances of students reflecting on feedback from me in their questionnaire or focus group interview answers. In the focus group interview, a student from the 9thITE class said in connection with the point-collecting system that “what I like is that there are no minus points, only positive feedback, while in other subjects if you get a grade one, you can never make up for it”. Another student also complained about difficulties in other subjects and told me that “there was nothing bad about this class, in fact, you even wrote to me separately asking if I wanted to make up for some missing points and stuff like that”. In the quotations above in

connection with what they would like to do differently, my formative feedback on their work also appeared (e.g., grammar mistakes in their writing tasks, missing sources from posters, more background information in the presentation, and so on). In the 11thME class, one student addressed me in connection with why they were satisfied with their grade saying that “whenever I had a mistake, you always told me and helped me correct it”. In the 9thITE class, students could not really answer the question how I could help them as they wrote “I don’t know” or “you can’t”. While 11th graders could articulate some specifics, in addition to including my approach to them in some cases: “so that if I have an idiotic question, you will answer it, just like you always do”, “I think that you are always open and helpful”, “I am 100% satisfied with this course, so just let everything stay exactly the same”, “PLEASE, don’t change”, “I’m not sure because you explain things clearly and logically in class, but afterwards I forget it without realizing it, and when I need this knowledge again, I realize that I don’t have it anymore”, and “I think the lessons provide enough help, and if I still don’t understand something, I would ask you after the lesson if you have some time and energy to help me understand what I still don’t understand”. As the interviewed teachers mentioned, the role of feedback immensely increased during the emergency remote delivery, so in addition to discussing students’ reactions to it, the following section contains students’ opinions about all their assessment related experiences of online education during the action research project.

5.5.3 Reactions to Assessment During the Emergency Remote Delivery

The structure of this section is identical to the structure of section 5.4.5.3, Alternative Assessment Methods of the Action Research During the Emergency Remote Delivery. First, students’ opinions on what they perceived to remain the same even after the introduction of emergency remote delivery are presented. Then their reactions to the different modifications that we had to make are contrasted with their experiences in person. At the end, students’ questionnaire and focus group interview answers are detailed about assessment methods introduced as a result of the emergency remote delivery.

In connection with collecting points, students agreed that “the advantage of this assessment system is that it works both online and offline, so we haven’t had to change much” as a student from the 11thME class summarized it. A student from the 9thITE class wrote in a questionnaire explaining why they were completely satisfied with the IT English lessons online that “we are not graded, but we have to collect points with tasks that don’t take much time, but we still learn a lot

from them". In the interviews almost all students said something along the lines of "this class is just as good online as it was before, not much has really changed". Some of them added further reasons: "IT English was just as good as when we went to school because the lessons stayed playful", "these classes remained just as playful [as before] and I also liked the achievement tests" (more about the tests later in this section), "we didn't have torture ourselves but could still learn the material, [...] and the online lessons were fun, so it was easy to pay attention", "I'm so glad that we kept the point-collecting system [...] this was a much better solution, these lessons were good compared to others", "these classes were exceptionally enjoyable", and many of them just said "I didn't have any problems with the IT English lessons online, they just stayed the same". Similarly 11thME students said: "I can't think of any disadvantages, math lessons remained the same" and added similar explanations: "math lessons haven't changed much, and it's a big advantage because I think other subjects changed a lot online, they became harder, but math hasn't become harder than in school", "more or less everything remained the same, which is good because other classes deteriorated, it became more difficult to study them, but not math", "during other lessons I knew we were not going to do anything, so I listened to teachers talking while I was working out, but not in math because here we had to participate", "math lessons were the same because we could ask questions any time, and we discussed the problems just like in school", "it was not the teacher talking, but we had to work, and that's way math lessons were way better than other classes" and "I think we even had more opportunities to collect points online than in the school". Students of the 11thAME group also added their reasons: "personally I really like these classes because I wasn't under the constant stress of having to do a lot of homework", "I liked that we didn't get too much homework, just enough to practice, but to have time to write other homework as well (towards the end of the year, there was even some free time left)", "I love the atmosphere of the online lessons as it is the same as in school", "online classes are just as calm and cozy as in school", and "I liked that we could progress with the material online, too".

In addition, 11thME students also mentioned the use of Khan Academy, Kahoot!, and bonus tasks that they could continue collecting points with during the emergency remote delivery as well saying that "Khan Academy is much better and more useful than anything else in any other class so far", "Khan Academy remained the coolest thing in the world", "it's good that there was Kahoot! and Khan Academy every month, it would be nice to keep them and introduce more similar things", "I liked the Kahoot!, so there could be more gamification because sometimes the

lesson consists of only problem solving”, and “this is my best class, especially since there are unlimited bonus tasks for each week, and no one bothers me even if I don’t do a single task and even if I only work on them for 72 hours”. Kahoot! was frequently mentioned by the 9thITE students as well in connection with the emergency remote delivery: “I liked that we used Kahoot! the same way as before” and

I think most students enjoy a playful lesson more than a normal one, for example, it was really good when we played Kahoot! [...] I think most classes could be organized in a playful way so that they include group work, maybe a little competition, and some extra motivation such as a small grade five, and Kahoot! has the playful element, the competition, and the motivation to collect points, so I think these can be enjoyed while we are also learning from it.

Another student from the 9thITE class wrote at the end of a questionnaire where they could add anything they wanted to that “I liked how we were graded and Kahoot! the most”. I asked them whether they wanted to have Kahoot! for homework and out of 16 responses, 14 chose “yes” and two “no” (see Table J9 in Appendix J). The two negative answers were “because I was lazy and didn’t do it” and “I prefer them in class because I learn a lot from them and it’s also funny when we do them together and excitedly wait for the right answer :)”. The others wrote the same reasons why they like it online and for homework as the ones in class detailed in section 5.5.2.1, Reactions to Alternatives for Testing.

However, some students also found collecting points more difficult during the emergency remote delivery, for instance, from the 9thITE class: “for me it’s more difficult to perform from home as I always get distracted” and “I can pay attention easier in person than through a screen”, and from the 11th grade math classes: “it [point collecting] is a little harder for me online”, “it’s more difficult to get myself to work on the assignments because during normal teaching there is a fixed time when you can deal with them, but at home your attention is quickly diverted”, “I have more motivation when sitting in class than at home”, “I feel more left alone with my questions”, “it’s more difficult to talk about a problem without being able to show it and see mine and the teacher’s work on the same paper”, and “I hate the whole online education”. Students listed several problems they had during the emergency remote delivery that were not connected to the action research project. As discussing them would be beyond the scope of the present dissertation, the following paragraphs detail students’ reactions only to further alternative assessment methods.

Although as a general tendency, students felt that alternative assessment and the online lessons of the action research remained similar to their experiences at school, there were smaller and greater changes that they also reflected on. As mentioned in section 5.5.2.1, Reactions to Alternatives for Testing, some students even preferred giving presentations online to in person and for this reason said that “IT English stayed the same; moreover, in some sense it even got better”. They also enjoyed the discussions following the presentations more than filling in the paper-based peer assessment sheets. In connection with group work, half of the students wanted to continue having pair and group work online, while half of them did not (see Table J10 in Appendix J). Those who were against it wrote that “it was boring”, “you can’t really work together without seeing the others”, “group work online is not good, it’s much better in person”, and “I don’t like online education”. While those who voted for group work online as well wrote: “it was good”, “that was the best”, “it brings the class more together”, and “I would like to, but it is a little more difficult to solve tasks this way through the Internet”.

In connection with the mini point-collecting opportunities in the 11thME class, out of 38 responses, 31 students wanted to have them online as well, five did not, and two chose the “other” option (see Table J9 in Appendix J). All the “other” and the negative opinions expressed the same idea that they like minis but only if they can write them in person. Two of them further explained that Khan Academy and other online quizzes should be online, but not minis. Those who voted for it added similar reasons as described in section 5.5.2.1, Reactions to Alternatives for Testing such as “we should definitely have them as they show the gaps in our knowledge”. In addition, one student wanted to see the correct answers immediately after filling them in, while others preferred if they could do them infinitely many times to practice until a given deadline. Their opinions on how to discuss the solutions varied. Some students wanted to go through all the questions during the next lesson, others would be happy with the solutions explained in text or video format and only talk about things they did not understand.

I experimented with different methods that help processing math content in general and asked students’ opinions about them. I recorded whole 45-minute lessons or only those parts of lessons in which I explained a problem or answered a question. I created narrated PowerPoint presentations to show how to solve specific tasks. I uploaded screenshots of the “whiteboard” of the online lessons, which were in fact image files saved in the program Paint as that seemed to be the easiest solution (see Appendix I12).

Eighteen responses out of 25 asked for having video explanations in the future, four chose not to, and three added “other” explanations, while for picture explanations 13 out of 13 students wanted to have them (see Table J10 in Appendix J). Those who were against video explanations wrote the following: “I’d prefer to discuss them during the lesson because then you can also ask questions”, “I don’t need them because I’ll ask my questions in class if I don’t understand something”, “things are more understandable ‘offline’”, “it takes too much time to watch them”, and “I wouldn’t be motivated to pay attention in class”. Others appreciated the freedom to watch the recordings any time, look for missing information, and they found them helpful for learning. As a result, some students were more fond of the emergency remote delivery, such as this student from the 11thME class:

My brain works much better in the afternoons and evenings, so I can solve tasks then with investing less energy and time. No need to discipline anyone. My own room, pajamas, and being my own boss are more comfortable (of course, this also comes with responsibility). If I don’t understand something, I can watch the videos as many times as I want. However, I still have the option to ask my questions during the lessons if something is not clear.

In connection with pictures, everybody agreed that they were useful. Some students preferred videos, but wrote that “although pictures don’t provide as much help as videos, they can be interpreted much faster, so if someone just needs a little help, a picture is perfect”. Others preferred pictures to videos: “it’s faster to get information from pictures than from videos, and for most things it’s enough to take a picture, which is less complicated”, “it helped me a lot that all the data was summarized in a picture”, “I get an easy-to-understand and immediate answer if I’m stuck”, “it’s easier to understand a problem with visual aid”, and “this is extra help, if these are available, you don’t have to ask that many questions”. One student was not happy with the quality of the pictures (see samples in Appendix I12), so he suggested the following in the focus group interview:

They are not very nice in Paint, so if we keep these picture explanations next year, it could be a task for us to construct them nicely, and then the whole group could use our work, but if they remain in Paint, it’s no big deal, just an idea.

I was also interested in how many videos and pictures they used later for studying (see questions in Appendix E8), and Table 14 shows the number of students’ responses.

Table 14

Number of Students' Answers to How Many Video and Picture Explanations They Watched

No. of videos	No. of answers	No. of pictures	No. of answers
0	4 (16%)	0	0 (0%)
1-2	8 (32%)	1-2	2 (15%)
3-4	9 (36%)	3-5	4 (31%)
5-6	4 (16%)	6-10	4 (31%)
7-8	0 (0%)	10-20	3 (23%)
Total	25 (100%)		13 (100%)

As mentioned in section 5.4.5.3, Alternative Assessment Methods of the Action Research During the Emergency Remote Delivery, I held all the lessons according to the original timetable as live online lessons and the students seemed to highly appreciate this. They highlighted in all the groups that the “regularity”, “consistency”, and the fact that they “could always count on these lessons was very helpful”. However, one student would have liked fewer lessons: “I think the number of live lessons is a bit much, especially since you work a lot with us beside them, too. Video calls are still a bit strange, but I’m glad that this skill of mine is also improving”. On the other hand, some students even enjoyed online lessons more, for example, for the following reasons: “there is no noise, no discipline problems, no waiting for the teacher to answer my question, which is why I feel that online lessons are better”. For others it depended on what the lessons were about: “I liked those lessons when we discussed problems, but not so much when we had to solve them right then. I prefer to deal with tasks when I feel like it”. Some students were satisfied with the balance: “We covered everything. The emphasis was on understanding and learning as much as possible from the tasks. I liked that we focused on whether we had any questions about what we were working on”.

Most students missed connecting with each other in person, so this element also had to be incorporated into the lessons on which students also reflected: “the lessons were to the point, but it was great to chat a little at the beginning”, “the lessons were good, but mostly because we had some human interaction”, “I only missed the free discussions before class in the school”. Answering to what was memorable to them about the online lessons, students from the 11thAME

group said the following: “The fact that you used your camera in every lesson, so I knew that there was someone behind the sound, and waving goodbye at the end of the lessons was also memorable because none of the other teachers did it”, “I really enjoyed all the online lessons, and for me they actually went better than in the school. I think partly because I didn’t have other classes before”, and “lessons were easy to follow, and they were very useful. I was relieved that you could explain everything just as well as in school.”

Closely related to assessment, the issue of cheating was also brought up by 11th grade math students. This problem was not mentioned by the 9thITE students, which might be because they did not have any assignments that would resemble traditional testing during the emergency remote delivery. The alternatives for their achievement test will be discussed later in this section. One student from the 11thME class described the situation in the following way:

In most subjects, we practically had to teach ourselves because there were no live lessons, or they were completely incomprehensible, and then we got a test based on a material that hadn’t been properly taught, so we either just guessed or cheated. But not in math, here we had all our lessons regularly according to our timetable, and we were actually working during the lessons, which was really good.

As a disadvantage of the emergency remote delivery, another student admitted not studying for tests as he would in school and learning less in general as a result: “I wasn’t preparing for tests because you sit at home, open the book... In math, I couldn’t apply this ‘method’, but in other subjects I had to when I saw how difficult tests were, so I didn’t learn that much”. Others also elaborated on their experiences with our math tests during the emergency remote delivery, and one of them linked the absence of cheating to the problem-solving nature of the tests in a questionnaire answer:

Considering the fact that tests were online, I don’t think it was possible to cheat because you had to think a lot. I think writing these tests was just like our tests under normal circumstances. I liked that we had some multiple-choice questions, too, I thought they would be easy, but they weren’t... :D

While another student believed that it was because assignments had low stakes as a result of the point-collecting system: “in math, I think there’s no point in cheating, especially online, because there are no bad consequences if I don’t do something well as the worst case scenario is that I have to work a bit more”. They also enjoyed the playful online practice exercises saying: “they made

me want to do math again”, “the more interesting a task is, the more you want to solve it”, “everyone loves playing games, and now that I’m home alone during the day, it’s the easiest way to immerse myself in math”, and adding a personal message to the questionnaire answer: “I liked the Pandemic tasks, and the game with the graphs, too. Hang in there, Ms!”

Students enjoyed playful activities in the 9thITE class as well. I asked their opinion about the drawing game Skribbl with which we practiced the vocabulary of the unit during a live online lesson. From the eight responses, seven students enjoyed it, and one student did not (see Table J10 in Appendix J), adding that “I like drawing but not on the computer”. Others praised that “it’s a great and easy way to learn words”, “it was really funny”, “it’s a great game”, and “DEFINITELY play it in the future, I loved it”. Students could not collect any points with this activity; however, they could by saying correct sentences applying a certain grammar structure or vocabulary from the unit. It seemed to achieve my goal of increasing students’ active participation during the online lessons as they wrote in their questionnaire answers the following: “I liked that in IT English you had to be there to collect points [...] that was really good, it motivated me to participate” and “getting points for being active in class motivates students who are otherwise not motivated to study to pay attention in class”. Another student also wrote that he liked it because “what I improvise reflects my real knowledge”.

In the 9thITE class, instead of students compiling achievement tests for themselves, I designed escape-room-themed achievement tests. Students could submit test questions, some of which I incorporated into the escape room test. Out of 13 students, 13 wanted to have such tests in the future and 11 wanted to write test questions (see Table J9 in Appendix J). The two students who did not wish to write test questions in the future added that “I suffered a lot with them, it’s fiddly :D” and “I didn’t write any questions as I consider it unnecessary”. Those who voted for the possibility of writing test questions in the future listed the usual “good point-collecting opportunity” and “you can learn from it” reasons. In addition, students found it fun: “quizzes are fun in themselves, and this can increase the ‘excitement’ even more” and an “an interesting task while I’m studying the material” also noting that “formulating questions is different than just answering them, and you learn it very well in the meantime”. Escape-room-themed achievement tests were immensely popular and appeared not only in students’ questionnaire answers but also during their focus group interviews: “I liked these achievement tests that were built up like you were going on a path, a virtual excursion” and “I really liked our achievement tests because they

were playful and not so traditional”. In their questionnaire answers they also wrote that “they were funny”, “way more enjoyable and still educational”, “adding codes made the test more playful”, “this online test was a lot of fun, and sometimes there were very hard puzzles, it was still worth doing it”, and “I liked it, but we should get rid of word grid tasks. They don’t require English knowledge [...] and are really annoying. Last time, my family and I searched for the last word for an hour and a half”. The mentioned word grid task can be seen in Appendix I13.

In conclusion, students participating in the action research project seemed to be satisfied with most of the alternative assessment methods. In connection with being graded as a result of point collecting, participants listed several advantages. Students appreciated that they had several chances to correct and improve their results, that their efforts and diligence mattered, that they got continuous feedback and could follow their progress. They also found the point-collecting system more transparent, logical, easy to follow, and fair. Fairness appeared in two ways. On the one hand, students felt that the point-collecting system provided equal opportunities to diverse students as everybody had the possibility to achieve good results. On the other hand, due to the diversity of opportunities, individual students did not feel that they always had to perform everything perfectly, which made them experiment more, take risks, and make mistakes. Students were also more motivated by the alternative assessment system. They enjoyed the fact that they had to work more, which might be explained by experiencing less stress and more success. Students appreciated that their assessment was more individualized and facilitated their learning process better. They felt that the grade they received as a result of point collecting reflected both their diligence and knowledge. Only about 10% of students formulated criticism. One such criticism was that these grades reflected students’ diligence more than their knowledge, which some students found unfair. The other main criticism was that for some students collecting points remained an extrinsic motivation. Students also suggested modifications for the assessment system such as asking for more opportunities to collect points. In connection with the emergency remote delivery, students participating in the action research agreed that “the advantage of this assessment system is that it works both online and offline, so we haven’t had to change much”.

6. Conclusion

This last chapter of the dissertation summarizes the findings by going through all the research questions and includes descriptions of the limitations of the study including further research possibilities. The main limitation of the research is that it is examining immensely complex issues through small sample studies. However, as the main aim of the research is to gain insights into the use of alternative assessment methods in English as a foreign language (EFL) and in English medium content (EMC) classes in public secondary education in Hungary, it is an exploratory study and does not intend to result in generalizable data; however, a deeper understanding of the stakeholders' views of assessment will hopefully be transferable to other professional contexts. The Conclusion ends with pedagogical and policy implications and suggestions on the basis of the research.

6.1 The Findings Relating to the Research Questions

The first research question (1.1 What do teachers mean by assessment and alternative assessment in particular in EFL and in EMC classes in public secondary education in Hungary?), is related to definitions of assessment. Participants of the teacher interviews mention several aspects of assessment from the literature as reviewed in section 2.2, Categorizations and Definitions of Assessment Types. The interviewed teachers define assessment as information, evaluation, measurement, or feedback highlighting its different purposes. For example, assessment provides information for students about their work and also for teachers about theirs, measurement that creates comparisons among students, classes, schools, and so on, which provides information for other stakeholders such as parents and schools, too. The interviewed teachers share examples for both norm- and criterion referenced assessment (Gipps, 1994) as well as assessment *of* learning, assessment *for* learning, and assessment *as* learning (Earl, 2006). All diagnostic, formative, and summative purposes of assessment (Rea-Dickins, 2000) appear in the interviews. There seems to be a common understanding of what traditional assessment means in public secondary education in Hungary: the one to five compulsory grading defined by the law that traditionally happens through written or oral tests. Participants define alternative assessment as anything that is different from that with a great variety of their different practices determined by their pedagogical goals.

The twelve interviewed teachers claim to use numerous alternative assessment methods relating to the second research question (1.2 What alternative assessment methods do teachers claim to use in EFL and in EMC classes in public secondary education in Hungary?). Their main

alternative for grading is using different point-collecting systems. Participants also apply alternatives for both oral and written testing. Instead of the traditional oral tests in which students are randomly chosen and asked by the teacher in front of the class, students can volunteer and know the date and the exact criteria in advance. The interviewed teachers prefer live or recorded student presentations to oral tests, which some of them call vlogs. Traditionally unannounced and graded pop quizzes are also turned into opportunities for formative feedback without high stakes for students.

The interviewed teachers introduce alternatives for achievement tests as well. Students can write tests in pairs or groups, compile tests for themselves, or the tasks of the test are in order of increasing difficulty making differentiation possible by, for example, having optional challenging tasks at the end of the test. In addition to alternatives for achievement tests, interviewees use further methods for summative purposes such as essays, translations, extended reports, reflective journals, project journals, further project work, creative individual or group products, posters, and portfolios. An additional category is achievement tests presented in a gamified framework such as a board game, for which students can create task cards, questions, or even design the whole game; a treasure hunt; a mystery game; a quiz show, for example, Jeopardy! or Who wants to be a millionaire?; or an escape room activity that can happen both in class or online.

The teachers also apply assessment activities that help process the material. They provide countless opportunities for students to collect points for their grade in addition to tasks for their own sake such as different kinds of collaborative group work. Diagnostic assessment frequently appears including feedback that participating teachers ask from students through anonymous online forms, exit tickets, or Q&A sessions. The interviewed teachers include various forms of peer assessment and students' self-assessment, too. They also provide diverse forms of feedback to their students: from highlighting students' individual strengths to using detailed rubrics in addition to individual formative feedback that informs students not only about their current level and the goal but also how they can alter the gap between the two (Sadler, 1989).

Answering the third research question (1.3 What are teachers' views of using alternative assessment in EFL and in EMC classes in public secondary education in Hungary?), several themes emerged. The interviewed teachers display competences for democratic culture (Council of Europe, 2016), and their attitudes, skills, knowledge, and values affect their views as their student-centered approaches influence their teaching and assessment practices. For example, they choose

their assessment methods based on the pedagogical goals they would like to achieve. Their views on learning also affect how they teach and assess. They believe in creating a safe environment for learning by minimizing stress and making the learning process enjoyable. In connection with their subjects, their view of English as a lingua franca results in focusing more on developing communication. As a general tendency, the interviewed teachers believe in giving a lot of freedom of choice to students within a set framework, such as selecting from varied tasks and materials involving students' interests. The interviewed teachers use alternative assessment as a result of their views mentioned above, in addition to the further views more directly linked to assessment. They believe that through alternative assessment, they have wider possibilities in facilitating the learning process. They express their views on what assessment should be like: fair, transparent, individualized and comparing students to themselves, supportive, and motivating, in addition to focusing on students' needs and developing assessment methods to facilitate their improvement.

The fourth research question (1.4 What are teachers' motivations and purposes for using alternative assessment in EFL and in EMC classes in public secondary education in Hungary?) asks about teachers' motivations and purposes. The interviewed teachers' motivations and purposes for using alternative assessment come from two main sources: their criticism of traditional assessment and the further pedagogical goals they would like to achieve through assessment. Participants find the compulsory grading in Hungary inadequate or even impedimental for reaching different pedagogical goals, hence their need for alternative assessment. They find traditional grading practices unfair and believe that grades in themselves cannot provide adequate feedback either to students on their work or to teachers on theirs. As a result, they would like to give grades to students that reflect students' effort and energy invested in studying, their development compared to themselves, in addition to their performance and knowledge of a certain section of the material as compared to standards described in the curriculum. They would also like their students to understand what is behind a certain grade and find it fair.

Furthermore, participants criticize the exclusively summative nature of grade giving in contrast to the need for supporting students' development. They are also concerned that traditional assessment affects students' learning strategies. As they are required to memorize lexical knowledge, they apply rote learning of materials, which participating teachers find difficult to change. Moreover, the interviewed teachers have diagnostic purposes, such as obtaining

information about what students already know and what they still need to practice in a transparent way, so that the information is not only accessible for the teacher but also for the students.

In addition to summative and diagnostic goals, most goals described by the participating teachers are in connection with formative assessment. They would like to reflect on the learning process, and not just the outcome. Their aim is to facilitate students' learning through assessment, such as providing help on how to improve and achieve set goals. It leads to one of their main purposes: enabling students to reflect on their own work and become self-monitoring learners, thus developing students' autonomy.

The interviewed teachers also expressed purposes related to learning and subjects in school in general, students' lives outside of school, the teachers' own perspectives, and goals that emerged during emergency remote delivery, too. Teachers aim to develop intrinsic motivation in their students by fulfilling their students' needs for autonomy, competence, and relatedness (Deci & Ryan, 1985) through, for instance, differentiation providing individualized possibilities for students.

In connection with educational goals, they would like to prepare their students for exams, such as the matura exams or language exams, and their further studies of tertiary education, or work, and lifelong learning in general. Consequently, the participating teachers wished to develop complex skills and competences. Based on the list of Trilling and Fadel (2009), the following 21st century skills appear as developmental goals in the interviews: critical thinking, problem solving, communication, collaboration, creativity, innovation, digital literacy, initiative and self-direction, social and cross-cultural interaction, flexibility and adaptability, leadership and responsibility, and productivity and accountability skills. All the interviewed teachers were very concerned about how their students could cope with the pandemic, which fundamentally influenced their goals by placing their students' support in the center.

As the teachers' introduction of alternative assessment highly depends on their pedagogical goals, when answering the fifth research question (1.5 What are teachers' experiences with using alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary?), their experiences are also linked to their evaluation of being able to reach their purposes. The interviewed teachers have found that students receiving grades through their various point-collecting systems can achieve the goal of giving the compulsory grades that also reflect students' effort and energy invested in studying and their development compared to themselves,

in addition to their performance and knowledge of a certain section of the material as compared to standards described in the curriculum. In the interviewed teachers' experience, their students find their point-collecting systems fair, and they carry out discussions so that students understand the system. Participants believe their alternative assessment methods develop learner autonomy, which seemed to be proven during the emergency remote delivery. They also experience that their alternative assessment methods inspire other colleagues and can result in some changes in the pedagogical programme of some of the participating schools as well. In connection with alternatives for tests, the interviewed teachers experience that their students are less stressed and enjoy the freedom and flexibility they have in different ways of presenting what they know.

The teachers also experience more freedom and flexibility from their own point of view as they perceive more chances for differentiation and involvement of a wider range of methods to apply and skills to be assessed. Through providing continuous feedback on students' work, the participants believe that they can facilitate students' development more effectively. Due to the variety of assessment activities that help process the material, students are more motivated by experiencing autonomy through their freedom of choice. Students also seem to enjoy the learning process more. In addition, they have the chance to experiment and see which learning methods work for them and learn accordingly. This seems necessary as the interviewed teachers experience a great diversity in students' answers. According to the interviewed teachers, some of their students show signs of intrinsic motivation through experiencing the above mentioned autonomy, competence, and relatedness by, for example, performing better after studying and preparing together in study groups. However, not all students seem to develop intrinsic motivation; according to the teachers' experiences, point collecting can also stay an extrinsic motivation for some students.

Applying diagnostic assessment, the teachers can acquire information for their pedagogical decisions in addition to experiencing further benefits. As students also have to provide feedback to their teachers and their peers, they have the opportunity to learn and practice how to give constructive feedback and formulate their criticism in a way that others and themselves can learn from it. Students also seem to enjoy learning from each other and improve in several skills through such situations, for instance, when organizing mock exams.

Regarding the teachers' point of view, the last research question reflects on the emergency remote delivery situation (1.6 How can alternative assessment methods in EFL and in EMC classes

in public secondary education in Hungary be adapted to remote delivery?). One of the main findings of the study is that most of the alternative assessment methods do not depend on the medium, and for this reason could easily be adapted to remote delivery. In fact, the closer to traditional grading practices assessment was, the more difficult it was to adapt online. The interviewed teachers also applied various solutions for grading and testing. They mostly kept their point-collecting systems and adapted the ways in which points could be collected by students. However, there were also examples of grading single assignments chosen by students. The participants reported different ways of handling live online classes as well, such as giving lectures for EMC classes where students had no coursebooks, providing consultation opportunities if students had problems with the material or the tasks they collected the points with, using live calls for speaking practice or exam preparation, and also varying interactive tasks together as a class and individual tasks to process the material. As all the interviewed teachers had already abandoned traditional testing before the pandemic, they had to make only smaller adaptations taking into consideration the special circumstances of the emergency remote delivery. For instance, they did not test students' lexical knowledge; instead, students had to apply, analyze, and evaluate (reflecting the higher order thinking skills of Bloom et al., 2001) the information they accessed on the internet or they were encouraged to solve tests together.

Looking at the students' point of view, the findings answering the first research question describe students' experiences of school assessment in general (2.1 What are students' experiences in connection with assessment in EFL and in EMC classes in public secondary education in Hungary?). In questionnaires prior to the introduction of any alternative assessment practices, based on 150 students' 415 expressions that they associate with school assessment, more than two-thirds of the students (67%) mention grades. The second most frequent association is tests (written or oral) including 12% of the words. These associations, in addition to further questionnaire responses and discussions with observers, all reflect the prevalence of grading and the most frequently used forms of assessment being oral and written tests in traditional public education in Hungary. Moreover, students write more than twice as many negative associations as positive ones and detail their criticism against traditional assessment in EFL and EMC classes. The participating students mostly criticize the practice of traditional grading. They do not find it fair because, according to them, graded tests do not reflect their diligence, class participation, homework, the fair amount of work they invest in studying, their real knowledge, their continuous work, or their

various skills. Students feel that tests require them to remember lexical knowledge that they have to cram and only show their performance at a given moment. They suggest instead textual feedback, corrections with explanations, personal discussions with the students and their parents, and more oral tasks, projects, and cooperative tasks. They also express the need for flexibility, individualization, and lots of different opportunities for students to choose from. Students often ask for individualized assessment, support for their learning processes, and the possibility to express themselves in varied ways that suit them as part of their assessment.

The findings show that alternative assessment practices can fulfill these student needs. After the introduction of alternative assessment methods, students answered the second research question (2.2 What are students' perceptions of alternative assessment in EFL and in EMC classes in public secondary education in Hungary?) in the following way. Students seem to enjoy point-collecting systems for several reasons. Most students feel that their efforts and diligence matter and appreciate the chances to correct and improve their results. They find the point-collecting systems transparent, logical, understandable, meaningful, and easy to follow, which might refer to the fact that students can follow their progress by getting points for each activity they carry out. Students also report point-collecting systems being fair for providing equal opportunities to diverse students. Another aspect of fairness is that with the diversity of tasks, students do not feel that they always have to perform everything perfectly.

In addition, students feel that the point-collecting system is easier and manageable, which might be a connection between feasibility and fairness in the sense that they all feel that everybody has the possibility to achieve good results. Several students say that they prefer point-collecting systems to traditional assessment because they have to work more, they are more motivated, and experience less stress, which reflects the blissful productivity element of gamification (Rab, 2013). Students appreciate that assessment is more individualized and facilitates their learning. The grades they receive as a result of point collecting depend on their class participation, homework, and test results all together.

However, students do not agree on the ratio of how much these grades contain their diligence and knowledge of the given material. Some students are satisfied, while others express concerns that these grades reflect more their diligence than their knowledge. Further disagreement concerns the fact that for some students it is difficult to reach grade five through collecting points, while for others it might not be challenging enough. The last piece of criticism that students

experience is in line with teachers' views that some students, after collecting the necessary points to reach a certain grade, lose their motivation; thus, collecting points for grades seems to be extrinsically motivating for some of them, but not intrinsically.

The last two research questions focus on students' experiences of assessment during the emergency remote delivery in general and in connection with alternative assessment. For research question 2.3 (What are students' experiences in connection with assessment during emergency remote delivery in public secondary education in Hungary?), students' main experience in general was that most of their teachers were replicating online the same practices they carried out in person. This general principle was true in many aspects: those teachers who held frontal classes face-to-face also gave lectures online and tried to apply the same traditional testing and grading procedures as well. Students experienced various problems with such lectures and the following tests. According to them, as tests mostly required remembering lexical knowledge, they often used their notes or the internet, which they categorized as cheating. As a response, some teachers asked for the use of cameras, set short time limits, or gave a higher number of tasks. The overwhelming amount of tasks was a recurring experience that students reported in addition to teachers using different platforms and ways of assessment that students found difficult to follow. As the study reflects only on the first wave of the pandemic, further research is needed to determine students' and teachers' experiences of the following school year that also happened mainly in remote delivery.

The students' perceptions of alternative assessment during the emergency remote delivery is presented in response to the last research question (2.4 What are students' perceptions of the adaptation of alternative assessment methods in EFL and in EMC classes in public secondary education in Hungary to remote delivery?). One of the main findings of the study is that according to students, "the advantage of this assessment system is that it works both online and offline, so we haven't had to change much". Most students expressed that the reason why they liked these alternative assessment methods is that they could be maintained during the emergency remote delivery. Students participating in the action research felt that their learning process was supported through assessment. They did not feel the need for cheating in tests for different reasons, for example, their alternatives for tests did not have high stakes and required problem solving skills instead of memorization. Some students even preferred online solutions, such as the freedom to solve tasks according to their own time management, giving presentations through screen sharing

and having live discussions afterwards, and collecting points in more varied ways than in person. Students also enjoyed live online lessons in which they were actively involved. Their reactions to online group work and further assessment activities online varied. As the study focused only on the emergency remote delivery period, further research is needed to explore assessments carried out online for longer periods of time. Although all stakeholders faced various adversities during the emergency remote delivery, the findings of the study show that alternative assessment methods might have mitigated some of the problems.

6.2 Pedagogical and Policy Implications and Suggestions

The final section of the dissertation restates the research niche, presented in Chapter 3, and formulates pedagogical and policy implications and suggestions based on the findings. Hungary did not opt to take part in the OECD's country reviews (2013), so the OECD report (2013) could only describe the existence of central frameworks for internal summative assessment on all levels of education, the existence of standardized central examinations at the end of secondary education, and the lack of central frameworks or guidelines for formative assessment on all levels of education. As only few studies have been published in connection with assessment in Hungary focusing on ways that aim to fulfill other roles than the compulsory grade giving (Hubai & Lázár, 2018), research was needed to determine what kinds of assessment, if any, takes place in Hungarian public secondary education apart from what is legally required. The study presents thirteen teachers' alternative assessment methods in depth; however, further research is needed to determine how widespread the use of such methods is.

Based on my participants' experience, they are the exceptions to the rule when they experiment with alternative assessment methods. As Mari expresses this in an interview: "it is difficult to always swim against the current". The reason why they might feel this way could be attributed to the duality in Hungarian legislation. On the one hand, international trends in assessment are reflected in the texts of policy documents (Hubai & Lázár, 2018): providing learning support, developing students, and carrying out ongoing evaluation and analysis of educational processes (Act CXC, 2011, Section 64), or applying differentiation as a basic principle, taking into consideration the development of talented students, and incorporating different types of assessment (National Core Curriculum, 2012) are all described as vital. On the other hand, output requirements do not reflect these ideals. Although the effectiveness of the one to five grading has been questioned for more than a century (Kemény, 1912), it remains the only

compulsory element of assessment. Moreover, the requirements of the curriculum (National Core Curriculum, 2012) and the standardized central examinations (Government Decree 100/1997) also contradict these ideals. As another participant, Anna alerts:

this system doesn't care about the student, it is not the student that is in the focus, but the material, the lexical knowledge... and as long as the output requirements only focus on these, there is no use for all the skills and competence development in the curriculum, [...] as nobody cares about them.

According to the OECD (2013), policy priorities should take a holistic approach, align assessment with educational goals, focus on improving classroom practices and build on teacher professionalism, avoid distortions such as teaching to the test, and put students at the center. Although “the use of innovative assessment approaches remains quite limited within the national assessment frameworks of OECD countries” (OECD, 2013, p. 149), I believe there is a need for developing such assessment frameworks and adjusting output requirements accordingly. The study presents how the interviewed teachers align their assessment methods with their educational goals and improve their classroom practices, so policy makers could rely on such teacher professionalism. The findings of the present research strongly suggest that it is not enough to include ideals in the texts of official policy documents if practice can hardly follow it. I believe that the holistic approach would mean, on the one hand, providing autonomy to teachers and schools to make professional decisions taking into consideration their local contexts. On the other hand, support should also be given for the development and implementation of such innovative assessment frameworks.

The Covid-19 pandemic also shed light on the need for implementing nationwide innovative assessment frameworks. As described in the literature review, in section 2.5.3, Emergency Remote Delivery in Hungary, most teachers “identified assessment and grading as the most problematic issue [... namely] how the grades would be given [..., and] in the 2019/2020 school year, very few teachers came to the realization that grading [should not be] the most important element of assessment” (Monostori, 2021, p. 11). Further research is needed to map the views of teachers in the whole country in connection with assessment. However, as output requirements did not change during the remote delivery periods either, even if teachers would not consider grading as the most important element of assessment, it remained the only mandatory one.

Teachers, schools, and families did not receive much help during the emergency remote delivery period. According to Láncoš and Christián (2021), schools could not implement the recommendations published by the Educational Authority, on the one hand because they “came with considerable delay [...] since the schools were bound to transition to digital teaching mid-March, the recommendations issued mid-April were no longer considered relevant by the respondent schools” (p. 90), and on the other hand, “the recommendations were too generally framed and were difficult to adapt to the diverse situations in which the different (public, private or church-run) schools and the affected students and families found themselves when the closures were announced” (p. 90). It is also true for the second version of the recommendations (Oktatási Hivatal, 2020, August 25). Then the 277-page-long volume, entitled *Collection of Digital Pedagogical Methodological Recommendations* (Farkas et al., 2021), was published by the Educational Authority during the school year of 2020/2021, which was already the second semester of remote delivery.

The content of the *Collection of Digital Pedagogical Methodological Recommendations* (Farkas et al., 2021) is in line with the findings of this study. This is the reason for its detailed discussion in section 2.5.4, *Recommendations of the Educational Authority Focusing on Assessment*. It repeatedly emphasizes that the current situation only accelerated the necessary paradigm shift caused by the advancements of the 21st century and urges that “traditional” practices need to be changed. The following is referred to as traditional: the source of information is the teacher, the focus is on transmitting pieces of information from the teacher to the students, the students are passive recipients of this information transmission, and assessment is mainly evaluating acquired knowledge in written or oral tests (Farkas et al., 2021). Instead, Farkas et al. (2021) recommend that teachers renew their approaches and practices by, “for example, providing interactivity, personalization, support for learning instead of or in addition to teaching, opportunities for alternative learning and knowledge acquisition, acceptance of knowledge gained elsewhere, supportive assessment approaches, and encouraging community learning” (p. 87). The exact assessment methods detailed in the recommendations (Farkas et al., 2021) can be tools in this renewal; however, the authors of the recommendations neglect the issues of the compulsory grading or the lexical knowledge based curriculum and other factors that hinder such renewal. All in all, publications of the Educational Authority support the necessary paradigm shift in education;

however, if they stay on the level of theoretical suggestions without exact policy changes reflecting these ideals, they do not and will not result in the renewal of practices on a large scale.

On a small scale, however, both international (Vogt & Tzagari, 2014) and Hungarian (Radnóti, 2006) studies show that teachers express the need for further training in assessment that would help them meet the challenges of the 21st century such as assessing complex skills and competences. The findings show that it is possible to carry out such assessment while complying with existing rules and regulations. Further research is needed to determine the effectiveness of these alternative assessment methods in different contexts (e.g., having more, or having fewer external regulations) and the potentials that lie in the professional development of teachers' assessment literacy. However, according to the interviews analyzed in the present study, teachers base their assessment methods on their pedagogical goals influenced by their views and experiment with different practices until they reach their goals. Their student-centeredness shines through their constant involvement of their students in the decision making processes reflecting a democratic approach in practice. They model the behavior they require from their students through lifelong learning.

All the interviewed teachers and myself as the researcher in this study are dedicated to practice and disseminate alternative assessment practices. Professional conferences, workshops, online and in-person communities such as the Pestalozzi Fridays (Lázár, 2015) provide such development opportunities on a small scale. However, my dream would be to teach in an education system where policy makers do not only promote democratic goals in theory, but also in practice by including them in the content of the curriculum, the output requirements, the assessment frameworks, and the professional development of teachers, providing both the indispensable autonomy to teaching staff and the necessary support for innovation and teachers' professional development.

References

- Act CXC on National Public Education (2011). Ministry of Human Capacities.
http://www.ilo.org/dyn/natlex/docs/ELECTRONIC/106832/131356/F-1702001629/act_national_education.pdf
- Act CCIV on National Higher Education (2011). Ministry of Human Capacities.
https://www.mab.hu/wp-content/uploads/Nftv_angol_2Sept2016_EMMI-forditas.pdf
- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment. *Language Teaching*, 34(4), 213–236. <https://doi.org/10.1017/S0261444800014464>
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques* (2nd ed.). Jossey-Bass Publishers.
- Bachman, L. F. (1989). Assessment and evaluation. *Annual Review of Applied Linguistics*, 10, 210–226. <https://doi.org/10.1017/S0267190500001318>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Baehr, M. (2005). Distinctions between assessment and evaluation. In S. W. Beyerlein, C. Holmes, & D. K. Apple (Eds.), *Faculty guidebook* (pp. 441–444). Pacific Crest.
- Barbarics, M. (2014). *Possible uses of the assessment system of gamification in ELT*. [Unpublished master's thesis]. Eötvös Loránd University.
- Barbarics, M. (2015). Iskolai értékelés gamification alapokon [School assessment based on gamification]. *Oktatás-Informatika*, 7, 43–61.
<https://www.eltereader.hu/kiadvanyok/digitalis-nemzedek-konferencia-2015/>
- Barbarics, M. (2016). Gamifikáció a magyar oktatásban [Gamification in Hungarian education]. In E. Gabos (Ed.), *A media hatása a gyermekekre és fiatalokra VIII* (pp. 63–69). Nemzetközi Gyermekmentő Szolgálat Magyar Egyesület.
- Barbarics, M. (2017). Alternatív, személyre szabott értékelés gamification alapokon [Alternative, individualised assessment based on gamification]. In T. Polonyi & K. Abari (Eds.), *Digitális tanulás és tanítás* (pp. 135–157). Debreceni Egyetemi Kiadó.
<https://www.researchgate.net/publication/320740707>
- Barbarics, M. (2018). Gamifikáció a demokratikus iskolai értékelésért [Gamification for democratic assessment in schools]. In E. Gabos (Ed.), *A media hatása a gyermekekre és fiatalokra IX* (pp. 22–28). Nemzetközi Gyermekmentő Szolgálat Magyar Egyesület.

- Barbarics, M. (2019a). Secondary school teachers' lifelong learning of assessment: Autonomy in developing alternative assessment methods. *Journal of Adult Learning, Knowledge and Innovation*, 3(2), 61–72. <https://doi.org/10.1556/2059.03.2019.06>
- Barbarics, M. (2019b). Innovatív értékelési módok az iskolában [Innovative ways of school assessment]. In T. Polonyi, K. Abari, & F. Szabó (Eds.), *Innováció az oktatásban* (pp. 105–130). Oriold és Társai Kiadó.
- Barbarics, M., Rózsahegyiné, V. É., & Wintsche, G. (2019). *A játékok fejlesztő hatása* [The developmental effect of games]. Eötvös Loránd Tudományegyetem. http://mindenkiiskolaja.elte.hu/wp-content/uploads/2019/09/A-j%C3%A1t%C3%A9kok-fejleszt%C5%91-hat%C3%A1sa_INTERA.pdf
- Barrett, M., Byram, M., Lázár, I., Mompoin-Gaillard, P., & Philippou, S. (2014). *Developing intercultural competence through education*. Council of Europe. <https://www.coe.int/t/dg4/education/pestalozzi/Source/Documentation/Pestalozzi3.pdf>
- Báthory, Z. (1997). *Tanulók, iskolák - különbségek* [Students, schools - differences]. OKKER Kiadó.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box – assessment for learning in the classroom. *Phi Delta Kappan*, 86(1), 9–21. <https://doi.org/10.1177/003172170408600105>
- Bloom, B. S. (1968). Learning for mastery. *Instruction and Curriculum*, 1, 1–12.
- Bloom, B. S. (1969). Some theoretical issues relating to educational evaluation. In R. W. Tyler (Ed.), *Educational evaluation: New roles, new means* (pp. 26–50). University of Chicago Press.
- Bloom, B. S., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., & Wittrock, M. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals* (Vol. 1). Longman.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and practices*. Pearson Education.

- Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653–675. <https://doi.org/10.2307/3587999>
- Burns, A. (1999). *Collaborative action research for English language teachers*. Cambridge University Press.
- Burns, A. (2015). Action research. In B. Paltridge & A. Phakiti (Eds.), *Research methods in applied linguistics* (pp.187–204). Bloomsbury Publishing.
https://www.researchgate.net/publication/282199978_Action_research
- Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78(3), 210–216. <https://doi.org/10.1037/0022-0663.78.3.210>
- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.
- Chicago Manual. (2017, April 3). *Chicago Style for the Singular They*. CMOS Shop Talk.
<https://cmosshoptalk.com/2017/04/03/chicago-style-for-the-singular-they/>
- Cohen, A. D. (2001). Second language assessment. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language* (pp. 515–534). Heinle & Heinle/Thomson Learning.
- Cohen, A. D. (2022). *Assessment*. Andrew D. Cohen's Website.
<https://sites.google.com/a/umn.edu/andrewdcohen/publications/assessment>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge.
- Council of Europe. (2016). *Competences for democratic culture: Living together as equals in culturally diverse democratic societies*. Council of Europe Publishing.
<https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=09000016806ccc07>
- Council of Europe. (2018). *Reference framework of competences for democratic culture: Descriptors of competences for democratic culture* (Vol. 2). Council of Europe Publishing. <http://rm.coe.int/prems-008418-gbr-2508-reference-framework-of-competences-vol-2-8573-co/16807bc66d>
- Council of the European Union. (2020, June 26). *Council Conclusions on countering the COVID-19 crisis in education and training*. Official Journal of the European Union.
[https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020XG0626\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020XG0626(01))

- Coyle, D., Hood, P., & Marsh, D. (2010). Assessment issues in CLIL. In *CLIL: Content and Language Integrated Learning* (pp. 112–132). Cambridge University Press.
<https://doi.org/10.1017/9781009024549>
- Csapó, B. (1978). A mastery learning elmélete és gyakorlata [Theory and practice of mastery learning]. *Magyar Pedagógia*, 1, 60–73.
- Csapó, B. (1992). Educational testing in Hungary. *Educational Measurement: Issues and Practice*, 11(2), 5–8. <https://doi.org/10.1111/j.1745-3992.1992.tb00233.x>
- Deci, E. L., & Ryan, R. M. (1985). *Self-determination and intrinsic motivation in human behavior*. Plenum Press.
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. E. (2011). From game design elements to gamefulness: Defining ‘gamification’. *Proceedings of the 15th International Academic Mindtrek Conference*, 11, 9-15. <https://doi.org/10.1145/2181037.2181040>
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford University Press.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.
- Earl, L. M. (2006). *Rethinking classroom assessment with purpose in mind: Assessment for learning, assessment as learning, assessment of learning*. Government of Manitoba.
- Easton, L. (2017, March 24). *Making a case for a singular ‘they’*. AP Definitive Source.
<https://blog.ap.org/products-and-services/making-a-case-for-a-singular-they>
- Ernst-Slavin, G., Moore, M., & Maloney, C. (2002). Changing lives: Teaching English and literature to ESL students. *Journal of Adolescent & Adult Literacy*, 46(2), 116–131.
- European Centre for the Development of Vocational Training. (2020, July 7). *Hungary: national vocational education and training (VET) responses to Covid-19*. Cedefop.
<https://www.cedefop.europa.eu/en/news/hungary-national-vocational-education-and-training-vet-responses-covid-19>
- European Commission. (2022, June 16). *Glossary*. Eurydice.
<https://eurydice.eacea.ec.europa.eu/national-education-systems/hungary/glossary>
- Farkas, A., Földeáki, A., Főző, A. L., Frész, A. J., Genáhl, K. J., Horváth, Á., Jánossy, Zs., Kapcsáné, N. J., Krajcovicz, Á., Neumann, V., Pintér, G., Sió, L., Szabados, T., Szalay, S. Zs., Szilágyi, Á., Timár, B., & Tóth, T. (2021). *Digitális pedagógiai módszertani ajánlások gyűjteménye* [Collection of digital pedagogical methodological recommendations]. Oktatási Hivatal.

https://www.oktatas.hu/pub_bin/dload/kozoktatas/tavoktatas/modszertani_gyujtemeny_ki_eg.pdf

- Flick, U. (2009). *An introduction to qualitative research* (4th ed.). Sage.
- Fromann, R. (2017). *Játékoslét – A gamifikáció világa* [The world of gamification]. Typotex.
- Gaal, P., Szerencses, V., & Velkey, Z. (2020, November 10). *HSRM Countries: Hungary - Policy responses*. European Observatory on Health Systems and Policies.
<https://eurohealthobservatory.who.int/monitors/hsrm/hsrm-countries/hsrm/hungary/preventing-transmission/physical-distancing/>
- Gál-Berey, Cs. (2022). *Szakmai program* [Vocational programme]. Budapesti Műszaki Szakképzési Centrum.
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150–169.
<https://doi.org/10.1080/0969594X.2012.749212>
- Genesee, F., & Upshur, J. (1996). *Classroom-based evaluation in second language education*. Cambridge University Press.
- Gimenez, J. (1996). Process assessment in ESP: Input, throughput and output. *English for Specific Purposes*, 15, 233–241. [https://doi.org/10.1016/0889-4906\(96\)00007-5](https://doi.org/10.1016/0889-4906(96)00007-5)
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. The Falmer Press.
- Gipps, C., & Stobart, G. (2003). Alternative assessment. In T. Kellaghan & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 549–576). Kluwer Academic Publishers. https://doi.org/10.1007/978-94-010-0309-4_33
- Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldane.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519–521. <https://doi.org/10.1037/h0049294>
- Golnhofer, E. (2003). A pedagógiai értékelés [The pedagogical assessment]. In I. Falus (Ed.), *Didaktika* (pp. 234–248). Nemzeti Tankönyvkiadó.
- Government Decree 100/1997 (VI. 13.) on the Issuance of the Procedural Rules of the Upper Secondary School Leaving Examination. <https://njt.hu/jogszabaly/1997-100-20-22>

- Government Decree 326/2013 (VIII. 30.) on the Teacher Promotion System and on the Implementation of the Act XXXIII of 1992 on the Legal Status of Public Servants in Public Education Institutions. <https://njt.hu/jogszabaly/2013-326-20-22>
- Guba, E. G. (1981). Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communication and Technology*, 29(2), 75–91. <https://doi.org/10.1007/BF02766777>
- Hamayan, E. V. (1995). Approaches to alternative assessment. *Annual Review of Applied Linguistics*, 15, 212–226. <https://doi.org/10.1017/S0267190500002695>
- Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70(2), 151–179. <https://doi.org/10.2307/1170660>
- Holliday, A. (2007). *Doing and writing qualitative research*. Sage.
- Hubai, K. (2016). *Alternative assessment in the classroom* [Unpublished dissertation]. Eötvös Loránd University.
- Hubai, K., & Lázár, I. (2018). Assessment of learning in the Hungarian education system with a special focus on language teachers' views and practices. *Working Papers in Language Pedagogy*, 12, 83–93.
- Kagan, S., & Kagan, M. (1994). *Cooperative Learning* (2nd ed.). Kagan Cooperative Learning.
- Kamberelis, G., & Dimitriadis, G. (2005). Focus groups. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 887–907). Sage.
- Kemény, G. (1912). A tanulók osztályozása [Grading students]. *A Gyermek*, 9-10, 434–435.
- Kemény, G. (1934). *Iskolai értékelés és kiválasztás* [Assessment and selection at school]. Merkantil Nyomda.
- Kemmis, S., & McTaggart, R. (Eds.). (1982). *The action research planner*. Deakin University Press.
- Kocadere, S. A., & Çağlar, S. (2015). The design and implementation of a gamified assessment. *Journal of E-Learning and Knowledge Society*, 11(3), 85–99. <https://doi.org/10.20368/1971-8829/1070>
- Kohn, A. (2011). The case against grades. *Effective Grading Practices*, 69(3), 28–33.
- Kohonen, V. (1997). Authentic assessment as an integration of language learning, teaching, evaluation and the teacher's professional growth. In A. Huhta, V. Kohonen, L. Kurki-

- Suonio, & S. Luoma (Eds.), *Proceedings of LTRC '96: Current developments and alternatives in language assessment* (pp. 7–22). University of Jyväskylä.
- Kovács, J. (2015). Célnyelven tanult tantárgyi tartalom értékelése a korai CLIL-ben [The evaluation of content knowledge in early CLIL]. *Neveléstudomány*, 2, 17–29.
<http://nevelestudomany.elte.hu/index.php/archivum/20152-2/?lang=en>
- Láncos, P. L., Christián, L. (2021). Domestic soft law regulation during the COVID-19 lockdown in Hungary: A novel regulatory approach to a unique global challenge. *European Journal of Risk Regulation*, 12(1), 77–92. <https://doi.org/10.1017/err.2020.115>
- Lapan, S. D., Quartaroli, M. T., & Riemer, F. J. (Eds.). (2012). *Qualitative research*. Jossey-Bass.
- Larsen-Freeman, D., & Long, M. H. (2014). *An introduction to second language acquisition research*. Routledge.
- Lázár, I. (2015). Tegyük mi is azért, hogy tanulni és tanítani jó legyen! Pestalozzi Péntek - egy nemzetközi hálózat hazai műhelyei [Let's act for a more enjoyable teaching-learning experience. Pestalozzi Fridays - The local workshops of an international organization]. *Autonomy and Responsibility Journal of Educational Sciences*, 1(2), 52–63.
<https://journals.lib.pte.hu/index.php/arjes/issue/view/551/221>
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day. *Educational Leadership*, 63(3), 18–24.
- Massler, U. (2011). Assessment in CLIL learning. In S. Ioannou-Georgiou & P. Pavlou, (Eds.), *Guidelines for CLIL implementation in primary and pre-primary education* (pp. 114–136). PROCLIL.
- Maxwell, J. A. (1992). Understanding the validity in qualitative research. *Harvard Educational Review*, 62(3), 279–300. <https://doi.org/10.17763/haer.62.3.8323320856251826>
- Maykut, P., & Morehouse, R. (1994). *Beginning qualitative research: A philosophic and practical guide*. The Falmer Press.
- Mihai, F. M. (2010). *Assessing English language learners in the content areas: A research- into-practice guide for educators*. University of Michigan Press.
- Ministerial Decree 20/2012 (VIII. 31.) on the Operation of Public Education Institutions and on the Use of Names of Public Education Institutions. Ministry of Human Capacities.
<https://net.jogtar.hu/jogszabaly?docid=a1200020.emm>

- Molnár, Gy., Námesztovszki, Zs., Glušac, D., Karuović, D., & Major, L. (2020). Solutions, experiences in online education in Hungary and Serbia related to the situation caused by Covid-19. *11th IEEE International Conference on Cognitive Infocommunications*. 601–608. <https://doi.org/10.1109/CogInfoCom50765.2020.9237844>
- Monostori, J. (2021). *The school year 2020-2021 in Hungary during the pandemic: Country report*. European Commission, Joint Research Centre, Publications Office. <https://doi.org/10.2760/062696>
- Nádori, G., & Prievara, T. (2011). *Kis-nagy IKT könyv* [Small-big ICT book]. TanárBlog. <http://tanarblog.hu/letoelthet-tananyagok/1979-kis-nagy-ikt-koenyv>
- Nádori, G., & Prievara, T. (2012, March 23). *PIL Akadémia tananyagok* [Partners in Learning Academy materials]. TanárBlog. <http://tanarblog.hu/letoelthet-tananyagok/2914-pil-akademia-tananyagok>
- National Core Curriculum (2012). Government Decree 110/2012. (VI. 4.) on the issuing, introduction and implementation of the National Core Curriculum. Ministry of Human Capacities. <https://njt.hu/jogszabaly/2012-110-20-22>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- OECD. (2013). *Synergies for better learning: An international perspective on evaluation and assessment*. OECD Reviews of Evaluation and Assessment in Education. <https://doi.org/10.1787/22230955>
- OECD. (2015). *Education policy outlook: Hungary*. OECD. <https://www.oecd.org/education/Hungary-Profile.pdf>
- Oktatási Hivatal. (2020, August 25). *A tanulási folyamat nyomon követése és az értékelés* [Monitoring the learning process and assessment]. https://www.oktatas.hu/koznevelas/ajanlas_szemelyes_talalkozas_nelkuli_oktatas_neveles_modszereire/tanulasi_folyamat_nyomon_kovetese
- Oktatási Hivatal. (2021, June 29). *Ajánlás a személyes találkozás nélküli oktatás-nevelés lehetséges módszereire és eszközeire* [Recommendations for possible methods and tools of education for remote delivery].

https://www.oktatas.hu/koznevelas/ajanlas_szemelyes_talalkozas_nelkuli_oktatas_nevelas_modszereire

- Olechowski, R. (2003). Alternatív teljesítményértékelés – az iskola humanizálása [Alternative performance assessment – humanizing the school]. In I. Bábosik & R. Olechowski (Eds.), *Tanítás – tanulás – értékelés* (pp. 215–235). Peter Lang.
- Orbán Viktor a Kossuth Rádió „Jó reggelt, Magyarország” című műsorában [Viktor Orbán in the “Good morning, Hungary” broadcast by Radio Kossuth]. (2020, March 13). Magyarország Kormánya. Retrieved June 7, 2022, from <https://2015-2019.kormany.hu/hu/a-miniszterelnok/beszedek-publikaciok-interjuk/orban-viktor-a-kossuth-radio-jo-reggelt-magyarorszag-cimu-musoraban-20200313>
- Orbán Viktor rendkívüli bejelentése a koronavírussal kapcsolatos kormányzati intézkedésekről [Viktor Orbán's emergency announcement about government measures related to the coronavirus]. (2020, March 13). Koronavirus.gov.hu. Retrieved June 7, 2022, from <https://koronavirus.gov.hu/cikkek/orban-viktor-rendkivuli-bejelentese-koronavirussal-kapcsolatos-kormanyzati-intezkedesekrol>
- Pappamihiel, N. E., & Mihai, F. (2006). Assessing English language learners' content knowledge in middle school classrooms. *Middle School Journal*, 38(1), 34–43. <https://doi.org/10.1080/00940771.2006.11461563>
- Pink, D. (2009). *Drive: The surprising truth about what motivates us*. Riverhead Hardcover.
- Polonyi, T., & Abari, K. (Eds.). (2017). *Digitális tanulás és tanítás* [Digital learning and teaching]. Debreceni Egyetemi Kiadó. <https://www.researchgate.net/publication/320740707>
- Prievara, T. (2015). *A 21. századi tanár* [The 21st century teacher]. Neteducatio.
- Prievara, T., & Nádori, G. (2018). *A 21. századi iskola* [The 21st century school]. Enabler.
- Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103, 683–700. <https://doi.org/10.1037/a0023911>
- Rab, Á. (2013). A gamifikáció lehetőségei a nem üzleti célú felhasználások területén, különös tekintettel a közép-és felsőoktatásra [Possibilities for applying gamification outside of a business context, especially in secondary and tertiary education]. *Oktatás-Informatika*, 3.

- <http://www.oktatas-informatika.hu/2013/03/rab-arpad-a-gamifikacio-lehetosegei-a-nem-uzleti-celu-felhasznalasok-teruleten-kulonos-tekintettel-a-kozep-es-felsooktatásra/>
- Radnóti, K. (2006). Milyen oktatási és értékelési módszereket alkalmaznak a pedagógusok a mai magyar iskolában? [What teaching and assessment methods do teachers use in Hungarian schools today?]. In Z. Kerber (Ed.), *Hidak a tantárgyak között* (pp. 131–167). Országos Közoktatási Intézet.
- Radó, P. (2020). Social selection in education: the wider context of the segregation of Roma pupils in Hungary. *CPS Working Paper Series*, 4, 1–24.
<https://cps.ceu.edu/publications/working-paper/social-selection-education-wider-context-segregation-roma-pupils-hungary>
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28(1), 4–13.
<https://doi.org/10.1002/bs.3830280103>
- Rea-Dickins, P. (2000). Classroom assessment. In T. Hedge (Ed.), *Teaching and learning in the language classroom* (pp. 375–401). Oxford University Press.
- Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education*, 13(2), 191–209. <https://doi.org/10.1080/0305498870130207>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144. <https://doi.org/10.1007/BF00117714>
- Santiago, P., McGregor, I., Nusche, D., Ravela, P., & Toledo, D. (2012). *OECD Reviews of Evaluation and Assessment in Education*. OECD. www.oecd.org/edu/evaluationpolicy
- Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE — Life Sciences Education*, 13(2), 159–166. <https://doi.org/10.1187/cbe.cbe-14-03-0054>
- Scriven, M. (1966). *The methodology of evaluation*. Purdue University.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. <https://doi.org/10.3102/0013189X029007004>
- Short, D. J. (1993). Assessing integrated language and content instruction. *TESOL Quarterly*, 27(4), 627–56. <https://doi.org/10.2307/3587399>
- Smith, R., & Rebolledo, P. (2018). *A handbook for exploratory action research*. British Council.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage.

- Surbhi, S. (2017, October 21). *Difference between assessment and evaluation*. Key Differences. <https://keydifferences.com/difference-between-assessment-and-evaluation.html>
- Szederkényi, É., Gassany, S. P., & Sevim, B. C. (2021). Pathways to work-based learning in the COVID-19 era. In K. T. Fodorné (Ed.), *Tanulás - tudás - innováció* [Learning - knowledge - innovation] (pp. 405-415). MELLearn.
- Tobin, T. J., Mandernach, B. J., & Taylor, A. H. (2015). *Evaluating online teaching: Implementing best practices*. Jossey-Bass.
- Tracy, S. J. (2010). Qualitative quality: Eight “big-tent” criteria for excellent qualitative research. *Qualitative Inquiry*, 16(10), 837–851. <https://doi.org/10.1177/1077800410383121>
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times*. Jossey-Bass.
- Tsagari, D. (2004). Is there life beyond language testing? Introduction to alternative language assessment. *CRILE Working Papers*, 58, 1–23. <https://pdfs.semanticscholar.org/19ad/ddb4879992814f8ebbc323a8d6f2dd491a4f.pdf>
- Tyler, R. W. (1970). *Basic principles of curriculum and instruction* (30th ed.). University of Chicago Press.
- Ur, P. (1991). *A course in language teaching*. Cambridge University Press.
- van Roy, R., & Zaman, B. (2017). Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory. In M. Ma & A. Oikonomou (Eds.), *Serious games and edutainment applications* (Vol. 2). (pp. 485–509). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-51645-5_22
- Vidákovich, T. (1990). *Diagnosztikai pedagógiai értékelés* [Diagnostic pedagogical assessment]. Akadémiai Kiadó.
- Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402. <https://doi.org/10.1080/15434303.2014.960046>
- Vuorikari, R., Velicu, A., Chaudron, S., Cachia, R., & Di Gioia, R. (2020). *How families handled emergency remote schooling during the Covid-19 lockdown in spring 2020*. Publications Office of the European Union. <https://doi.org/10.2760/31977>

- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49–65. <https://doi.org/10.1080/0969594042000208994>
- Yeh, S. S. (2010). Understanding and addressing the achievement gap through individualized instruction and formative assessment. *Assessment in Education: Principles, Policy & Practice*, 17(2), 169–182. <https://doi.org/10.1080/09695941003694466>

Appendix A

Hungarian-English glossary of related expressions as used in the dissertation

bizonyítvány: school report
 e-KRÉTA: national compulsory e-gradebook system (in short: e-gradebook)
 emelt szintű oktatás (középiskolában): advanced education
 érdemjegy, jegy, osztályzat: grade
 érettségi bizonyítvány: secondary-school leaving certificate
 érettségi vizsga: secondary-school leaving examination, matura exam
 értékelni, minősíteni: to assess
 értékelés, minősítés: assessment
 év vége: end of the year
 év végi: year-end (adj.)
 dolgozat: test
 [...] fakt: elective advanced [subject]
 fejlesztő értékelés: formative assessment
 feleltetés: oral test
 félév: term, mid-term (in secondary), semester (in tertiary education)
 félévi: midterm (adj.)
 gimnázium: secondary general grammar school
 helyi tanterv: local curriculum
 kéttannyelvű: bilingual (adj.)
 kísérettségi: minimatura exam
 mérés: measurement
 nyelvi előkészítő: preparatory language year
 osztályfőnök: form teacher
 osztályozó értekezlet: assessment meeting
 pedagógiai program: pedagogical programme
 pótvizsga: retake exam
 röpdolgozat: pop quiz
 szakközép iskola, szakgimnázium: vocational secondary school

számonkérni: to evaluate, to test

számonkérés: evaluation, testing

tantermen kívüli digitális munkarend: emergency remote delivery

távoktatás: distance education

témazáró dolgozat: achievement test

teszt: quiz

Appendix B

Links of websites and applications mentioned in the dissertation

Actionbound: <https://en.actionbound.com/>

BBC Learning English: <https://www.bbc.co.uk/learningenglish/>

Bubbl.us: <https://bubbl.us/>

Canva: <https://www.canva.com/>

Canvas: <http://canvaslms.com>

Checkli: <https://www.checkli.com/>

Classcraft: <https://www.classcraft.com/hu/>

ClassDojo: <https://www.classdojo.com/>

Class Tools: <https://classtools.net/>

Conceptboard: <https://conceptboard.com/>

Diagrams.net: <https://www.diagrams.net/>

Discord: <https://discord.com/>

EdPuzzle: <https://edpuzzle.com/>

Edubase: <https://www.edubase.net/>

FlashBack: <https://www.flashbackrecorder.com/>

Flippity: <https://flippity.net/>

Formative: <https://goformative.com/>

G Suite for Education: <https://events.withgoogle.com/g-suite-for-education-hu/>

Genially: <https://genial.ly/>

Google Classroom: <https://classroom.google.com/>

Google Docs: <https://www.google.com/docs/about/>

Google Forms: <https://www.google.hu/intl/hu/forms/about/>

Google Jamboard: <https://jamboard.google.com>

Google Meet: <https://meet.google.com/>

Google Sheets: <https://www.google.com/sheets/about/>

Google Slides: <https://www.google.com/slides/about/>

eDIA: <http://www.edia.hu/>

Kahoot!: <https://kahoot.com/schools-u/>

Khan Academy: <https://www.khanacademy.org/>
LearningApps: <https://learningapps.org/>
Lino: <https://en.linoit.com/>
Live worksheets: <https://www.liveworksheets.com/>
Mentimeter: <https://www.mentimeter.com/>
Mindmeister: <https://www.mindmeister.com/>
MindMup: <https://www.mindmup.com/>
Mindomo: <https://www.mindomo.com/hu/>
MS Meet: <https://www.microsoft.com/en-us/microsoft-teams/group-chat-software>
MS Forms: <https://www.microsoft.com/hu-hu/microsoft-365/online-surveys-polls-quizzes?rtc=1>
MS OneNote: <https://www.microsoft.com/hu-hu/microsoft-365/onenote/digital-note-taking-app?ms.url=onenotecom&rtc=1>
MS PowerPoint: <https://www.microsoft.com/en-us/microsoft-365/powerpoint>
MS Teams: <https://teams.microsoft.com/>
National Competency Assessment:
<https://www.oktatas.hu/kozneveles/meresek/kompetenciameres/feladatsorok>
National Public Education Platform: <https://www.nkp.hu/>
OneNote Class Notebook: <https://www.onenote.com/classnotebook?omkt=hu-HU>
Padlet: <https://hu.padlet.com/>
Prezi: <https://prezi.com/>
Redmenta: <https://redmenta.com/>
Screencast-O-Matic: <https://screencast-o-matic.com/>
Skribbl: <https://skribbl.io/>
Slack: <https://slack.com/>
Socrative: <https://www.socrative.com/>
Story Dice: <https://davebirss.com/storydice-creative-story-ideas/>
Sutori: <https://www.sutori.com/>
Symbaloo Learning Paths: <https://learningpaths.symbaloo.com/>
TED Talks: <https://www.ted.com/>
Tricider: <https://www.tricider.com/>
Türchen: <https://tuerchen.com/en/>

Quizizz: <https://quizizz.com/>

Quizlet: <https://quizlet.com/>

Wakelet: <https://wakelet.com/>

Webex: <https://www.webex.com/>

Windows webcam: <https://support.microsoft.com/hu-hu/help/17444/windows-camera-app-webcams-help>

Wordwall: <https://wordwall.net/hu>

YouTube: <https://www.youtube.com/>

Zoom: <https://zoom.us/>

#school: <https://hashtag.school/>

Appendix C

Interview guides

Appendix C1

The piloted, main interview guide

Dear Colleague,

Thank you very much for participating in this interview contributing to my PhD research. I am Márta Barbarics. I study at ELTE, Faculty of Education in the Language Pedagogy Programme. My research is about assessment in secondary schools. I would like to get acquainted with the views of teachers on assessment, so there are no right or wrong answers, as I am interested in your personal experience and opinion. The data will be used for research purposes only. You remain anonymous. If you are interested, I am happy to share the results with you. If you agree to the interview being recorded, we can start. Participation is voluntary, so you can refuse to answer any time. Thank you very much!

First of all, let me ask you some background data.

- How old are you?
- Where did you graduate (which university, which programme)?
- Which subject(s) do you teach?
- Which grades (age groups)?
- Where (which school)?
- How long have you been teaching there?
- Did you teach anywhere else before?
- Have you ever lived, worked abroad?

We will be talking about assessment in more detail. People mean different things by assessment. What does assessment mean to you?

- What would you call traditional assessment?
- Compared to this, what would you call alternative assessment?
- What ways of assessment do you use?

From these ones which ones would you call alternative assessment? (From here, we go through the different ways one by one with the following questions.)

- Can you describe how it happens? (What do you assess? What do you give feedback on?)
- Why did you introduce this form of assessment?
- What would you like to achieve by using it?
- What was the reaction of your students when you introduced it?
- How have you been using it? (How long have you been using it? Have you ever modified something about it? If yes, what and why?)
- What kind of advantages and disadvantages of it do you see?
- How effective do you think this way of assessment is? (What makes it effective?)
- Has it brought about any changes in students' behavior? (attitude, motivation, engagement results, and so on)

Is there anything we haven't talked about and you think it is connected to this topic and you would like to share it?

Thank you very much for the interview!

Appendix C2

The follow-up interview guide

Dear Colleague,

Thank you very much for participating in this interview contributing to my PhD research. In our first interview, I wanted to get acquainted with the views of teachers on assessment. Now I am interested in the effects of the emergency remote delivery, so there are no right or wrong answers, as I am interested in your personal experience and opinion. The data will be used for research purposes only. You remain anonymous. If you are interested, I am happy to share the results with you. If you agree to the interview being recorded, we can start. Participation is voluntary, so you can refuse to answer any time. Thank you very much!

First, let me ask you some background data.

- Where do you teach? (Has it changed since our last interview?)
- What kinds of classes did you teach during the emergency remote delivery?
- How many lessons did you have with them?
- What were the advantages of remote delivery?
- What were the disadvantages of remote delivery?

Now let's move on to assessment.

- Was assessment easier or more difficult for you during the emergency remote delivery? Why?
- Could you continue using... (here I insert the exact assessment methods mentioned in the first interview, and go through all the questions below for each)
 - Did you have to modify it? How?
 - What was the reaction of your students?
 - What kind of advantages and disadvantages of it do you see?
 - How effective do you think it was? (What makes it effective?)
 - Has it brought about any changes in students' behavior? (attitude, motivation, engagement results, and so on)
- Have you applied any other assessment methods during the emergency remote delivery?
 - Why did you introduce it?
 - Can you describe how it happens?
 - (if any of these is missing, I ask about it: What do you assess? What do you give feedback on? Why did you introduce this form of assessment? What would you like to achieve by using it? What was the reaction of your students when you introduced it? How have you been using it? What kind of advantages and disadvantages of it do you see? How effective do you think this way of assessment is? Has it brought about any changes in students' behavior?)
- Which of these assessment methods do you intend to use in the future as well? Why?

(I add further individual questions that I would like to follow-up based on the first interview.)

Is there anything we haven't talked about and you think it is connected to this topic and you would like to share it?

Thank you very much for the interview!

Appendix C3

The interview guide for those who quitted public education

Dear Colleague,

Thank you very much for participating in this interview contributing to my PhD research about assessment in public secondary education. I am interested in what has happened since our previous interview, so there are no right or wrong answers. The data will be used for research purposes only. You remain anonymous. If you are interested, I am happy to share the results with you. If you agree to the interview being recorded, we can start. Participation is voluntary, so you can refuse to answer any time. Thank you very much!

First, let me ask you some background data.

- Since our previous interview in [insert the data for each person], where and for how long have you been teaching?
- When did you leave public education? Why? (possible follow-up question: how was it connected to assessment?)
- Looking back at your experiences using alternative assessment, what was the reaction of the leadership of the school and your colleagues?
- What do you think worked well?
- If you were teaching again, what would you change?
- Would you like to teach in public education again?
- What conditions would be ideal for you to go back?

(I add further individual questions that I would like to follow-up based on the first interview.)
Is there anything we haven't talked about and you think it is connected to this topic and you would like to share it?

Thank you very much for the interview!

Appendix D: Samples from the instruments of the action research

Appendix D1

Sample from the teaching journal

Action Research Napló ☆ 📅 ☁

File Edit View Insert Format Tools Extensions Help

1 2 3 4 5 6

←

9D - 1. óra

Szeptember 9. hétfő

11A - 2-3. óra

9D - 2. óra

Szeptember 11. szerda

11A - 4. óra

Szeptember 13. péntek

11A - 5. óra

9D - 3. óra

Szeptember 16. hétfő

11A - 6-7. óra

9D - 4. óra

Szeptember 18. szerda

11A - 8. óra

Szeptember 20. péntek

11A - 9. óra

9D - 5. óra

- Pontrendszer (extra mester pontok is!!!)

Szeptember 16. hétfő

11A - 6-7. óra

- 8:55-10:35 (2-3. óra) - A205 - nagyterem
- Nincs hiányzó
- Kisérettségit írtak 6-an (2-3-masok: ~~Övölgyesi, Póti, Gyári, B. Miklós, M. M. és T. M.~~ próbálta meg 4-esről 5-ösre javítást)
- OKTV-re készültek 3-man (~~M. J. S., S. M., L. M.~~)
- Többiek a tavalyi feladatokból csináltak még (konkrétan felírtam a táblára a feladatok számát, többé-kevésbé dolgoztak, mindenki megcsinált 1-2 feladatot)

TIPP: jövő órától össze kell állítani az alap feladatokat + extrákat differenciálva

9D - 4. óra

- 11:40-12:25 (5. óra) - A222 - gépterem
- Nincs hiányzó
- Hashtag school regisztráció
- Google doksiban dolgozni: eleinte egész jól ment, aztán elkezdtek megint ökörködni, úgyhogy elég rossz állapotba került végül a dokumentum
- Kahoot kérdéseket küldeni (50 pont / helyes kérdés + 4 válaszlehetőség) - max 10 kérdés / ember
- Pénteken felelés (10 pont / helyes mondat)

TIPP: ne a teljes osztály dolgozzon egy doksiba és arra is kapjanak pontot, megbecsüljék Kahoot kérdéseknél most még nem vonok le pontot a nyelvtan miatt, javítom, de legközelebb már igen!

KILÉPŐCETLI: 3-2-1 exit ticket: 3 facts I learned, 2 questions I have, 1 opinion I have

Appendix D2

Sample from a proforma class observation instrument

A pedagógus neve: Barbarics Márta

Az óralátogatás helye: A206-os terem

Műveltségi terület: Matematika

Tantárgy: Matematika angol nyelven

Az óra témája: Vegyes feladatok gyakorlása és átbeszélése

Az osztály: 11A-1cs-mat

Az óralátogató neve: ~~.....~~

Dátum: 2019. november 11.

Idő	Az óra menete	Megjegyzések
1. tanóra	<p>A tanár kijelöli a megoldandó feladatokat.</p> <p>A diákok önállóan, illetve párban (vagy hármassával) szépen dolgoznak. A tanár közben monitoroz, ha probléma van, segít és rávezeti a diákokat a megoldásra.</p>	<p>Az a benyomásom, hogy a tanulók már hozzászoktak az órán alkalmazott értékelés rendszeréhez. Gondolom ezt abból, hogy többen már automatikusan jelentkeztek, hogy a tanár ellenőrizze le a megoldásaikat – tudták, hogy a tanár feljegyzi ezeket és pontot ad értük.</p>

Appendix D3

Sample from posts in MS Teams after the introduction of emergency remote delivery

Microsoft Teams **9D IT English with Ms Barba...**

Általános Bejegyzések Fájlok Points + Csoport Értekezlet

Barbarics Márta 2020. 03. 17. 20:36
Hi guys, you can find the same document with the tasks here as well (under the files of Level 5 - basic software).
Válasz

2020. március 20.

Barbarics Márta 2020. 03. 19. 17:38
Dear IT English group, I'll have a teachers' meeting tomorrow, so we still won't have a live lesson tomorrow 😞 Instead I'd like to ask you to write down what I wanted to talk about: How was this week for you? What have you been doing? What is your daily routine now? What learning methods have worked for you? What do you suggest for these following weeks? How would you like to collect points? (I'll put these questions into your personal Homework page under Class Notebook.) See you on Monday 😊
Kevesebb megjelenítése

Az összes összecsukása

[Redacted] 2020. 03. 20. 12:06
This is my worst week since the school started. We got a lot of homework from every subjects. Every teachers want to teach us because they don't want to fall behind with the educational material so they give us a lot of homework. I sit down to the computer at 8 o'clock, and I stand up at 16 o'clock, so it's a little bit too much for me.
Another problem's the Microsoft Teams. It's too necessary for me. I can't find the homeworks, I can't see what the teachers write here, I don't get any notifications, any emails. It's not a user friendly platform. I think the Google Classroom is really better than this. :/
Sorry, I don't want to be a cryin' machine, but this thing, whats others call distance learning is a sh*t... :(
Kevesebb megjelenítése

Barbarics Márta 2020. 03. 20. 12:12 Szerkesztve
Dear [Redacted], thanks for writing. I know it's really difficult now because nobody really knows how to do it well, but be patient, we will figure it out, and it will be better! Help your teachers, for example, give polite feedback about the situation. I'm sorry we couldn't have a live meeting today, but we will talk about all these things on Monday. Feel free to ask me if you have any questions! Have a very nice weekend ❤️ 😊

[Redacted] 2020. 03. 20. 12:22
Sorry for the feedback, I do not want to insult anyone, I just write down what I feel now.
I hope this'll be better and enjoyable.
Have a nice weekend! <33

Appendix E

Student questionnaires

Appendix E1

Entry questionnaire for the 9thITE group compared to the pre-PhD entry questionnaires

Same as pre-PhD

Different from pre-PhD questionnaires

Your answers will be used (anonymously) exclusively in my doctoral dissertation. There are no good or bad answers as I'm interested in your honest opinion. Filling in this questionnaire takes about 10-15 minutes. Thank you very much for your help!

Ms Barbarics

Pseudonym:

Date:

1. (3.) Do you play computer games? If yes, which ones? And in what language?
2. In your opinion, how important is it to learn English? (1= not important, 5 = very important)
3. (1.) How much do you like the English language? (1 = I hate it, 5 = I love it)
4. How have you learnt English so far? (Check all that applies.)
 - Primary school lessons
 - Private teacher
 - By myself (e.g. from movies, games, etc.)
 - Other:...
 - I haven't learned English, yet.
5. (4.) What are the first three words or expressions that come to mind when you hear "school assessment"?

From here, questions about foreign language refer to English if you had English lessons in 8th grade, but if you didn't, answer in connection with the foreign language you learned.

6. In primary school (8th grade) how much did you like foreign language lessons? (1 = I hated them, 5 = They were one of my favourites.)
7. What was your final grade in foreign language (at the end of 8th grade)? VS. (2.) What grades do you usually get in it?

Appendix E2

A sample questionnaire from the 9thITE group

Kérdőív

A kérdőívben leírt véleményedet (név nélkül) kizárólag a doktori kutatásomhoz fogom felhasználni. Nincsenek jó vagy rossz válaszok, az őszinte véleményedre vagyok kíváncsi. A kérdőív kitöltése kb. 10-15 percet vesz igénybe. Nagyon köszönöm a segítségedet!

Barbarics Tanárnő

Álnév: *Zoli* Dátum: *2019.09.13.*

Kérlek, hogy olyan álnevet válassz, amire még egy hónap múlva is emlékezni fogsz, ha nem bízol a memóriádban, válassz nagyon különleges álnevet és / vagy írd is fel valahova magadnak :)

1. Szoktál számítógépes játékokkal játszani, ha igen, mikkel? És milyen nyelven?
Igen szoktam: World of Warcraft, Subnautica, The crew 2, Plants vs. Zombies battle for Neighborville (angol nyelven)

2. Szerinted mennyire fontos ma angolul tudni? Karikázd be az alábbi skálán, ahol 1 = egyáltalán nem fontos, 5 = nagyon fontos!

1 2 3 4 **5**

3. Mennyire szereted az angol nyelvet? Karikázd be az alábbi skálán, ahol 1 = utálok, 5 = nagyon szeretem!

1 2 3 **4** 5

4. Hogyan tanultál eddig angolt? (Jelöld be mindegyiket, amelyik igaz rád!)

- általános iskolai órán
- külön tanárnál
- önállóan (pl. filmekből, számítógépes játékokból stb.)
- egyéb:
- nem tanultam eddig egyáltalán

5. Mi az az első 3 szó vagy kifejezés, ami eszedbe jut arról, hogy „iskolai értékelés”?

betűszámot; leadandó; dolgozat

Innentől, ha tanultál 8.-ban angolul, akkor az idegennyelvvvel kapcsolatos kérdések az angolra vonatkoznak, ha más idegennyelvet tanultál, akkor arra a nyelvre.

6. Általános iskolában (8. osztályban) mennyire szeretted az idegennyelv-órákat? Karikázd be az alábbi skálán, ahol 1 = utáltam, 5 = egyik kedvenc tantárgyam volt!

1 2 3 4 **5**

7. Mi volt a 8. osztályos év végi jegyed idegennyelvből?

5

8. Mennyire tükrözte a 8. osztályos év végi jegyed az idegennyelv-tudásod? Karikázd be az alábbi skálán, ahol 1 = egyáltalán nem tükrözte, 5 = teljes mértékben tükrözte!

1 2 3 ④ 5

9. Mennyire tükrözte a 8. osztályos év végi jegyed az idegennyelv tanulásába fektett energiád? Karikázd be az alábbi skálán, ahol 1 = egyáltalán nem tükrözte, 5 = teljes mértékben tükrözte!

1 2 3 4 ⑤

10. Mennyire voltál elégedett azzal, ahogy 8. osztályban a tanárok értékelték a munkádat és eredményeidet az alábbi tantárgyakból? (Az értékelés alatt az osztályzást és minden egyéb írásbeli vagy szóbeli visszajelzést is értünk.) Karikázd be az alábbi skálákon, ahol 1 = egyáltalán nem voltam vele elégedett, 5 = teljesen elégedett voltam vele és indokold is a választod!

idegennyelvből: 1 2 3 4 ⑤

Miért? Sokat tanultam az Angolra és volt, hogy sebecs ment. De végül 5-ös lett.

matematikából: 1 2 3 4 ⑤

Miért? Mindig értettem a matekat és jelles is voltam belőle.

magyarból: 1 2 3 4 ⑤

Miért? A magyarral nagyon ügy vagyok mint az Angollal. Sokat dolgoztam és végül 5-ös lett.

történelemből: 1 2 3 4 ⑤

Miért? A történelem logikus és értem és 5-ös voltam belőle.

Válassz még egy tantárgyat, aminek az értékelésével nagyon elégedetlen vagy nagyon elégedett voltál és írd le, miért!

Választott tantárgy: Informatika 2 3 4 ⑤

Miért? A tanárom sok versenyre vitt és

szereztem a tantárgyat 5-ös lettem belőle is (mint mindenki -ről)

11. Ha tehetnéd, mit változtatnál az iskolai értékelésen?

A felesleges versstanulást, és a kötelező olvasmányokat törölném.

12. Ide írd, ha valamelyik kérdésre részletesebben szeretnél válaszolni, vagy írd egy konkrét példát, történetet az értékeléssel kapcsolatban!

Appendix E3

The first questionnaire after assessment based on gamification is introduced in the 9thITE group compared to the pre-PhD questionnaires reflecting on assessment based on gamification

Same as pre-PhD

Different from pre-PhD questionnaires

This is the next questionnaire for my doctoral research. Just like the one a month ago, this is also anonymous, but the difference is that I won't only use your answers for my dissertation, but based on what you write, we can improve our lessons and assessment together. For this reason, please, write your honest opinion, so that we can continue our work together more efficiently. Thank you very much!

Ms Barbarics

Pseudonym:

Date:

1. How much do you like the English language? (1 = I hate it, 5 = I love it)

2. Based on your experiences so far, how much do you like the IT English lessons? (1 = I hate it, 5 = it's one of my favourite subjects)

Why?

3. In the next period, would you like to... (yes - no - other, then why?)

- a. have the opportunity to compile the test for yourself?
- b. have a Jeopardy! quiz show (summary game)? (picture)
- c. play Kahoot! during the lesson?
- d. practice with Kahoot! at home?
- e. have the opportunity to write Kahoot! questions?
- f. have exit tickets? (picture)
- g. have a treasure hunt (4-station group game)?
- h. have voluntary oral tests?
- i. collect sentences in English while throwing Wendy around?
- j. work in Google Documents?

4. In the next period, would you like to learn in this point system?

- Yes
- No
- Other:

Why?

5. What would you change in this point system?

6. What grade have you just got from IT English?

7. How satisfied are you with the ways I assessed your work in IT English in this period? (1 = “I am not satisfied at all” to 5 = “I am absolutely satisfied.”)

Why?

8. If you had been graded in the traditional way, your grade would have been...

- better
- worse
- the same
- you can't decide whether your grade would have been better or worse
- Other:...

9. How well does this grade reflect your IT English knowledge of this period? (1 = “It doesn't reflect it at all” to 5 = “It reflects it perfectly.”)

10. How well does this grade reflect the energy you put in studying IT English during this period? (1 = “It doesn't reflect it at all” to 5 = “It reflects it perfectly.”)

11. Here you can detail your answer to any of the questions or anything else that you have liked or not liked during this period.

Appendix E4

A sample questionnaire from the 9thITE group

Kedves Csoportom!

Ez a doktori kutatásomhoz felhasznált kérdőív második része. Az egy hónappal ezelőttihez hasonlóan ez is anonim. A különbség annyi, hogy az információkat nem csak a kutatásomhoz fogom felhasználni, hanem azok alapján, amiket írtok, közösen alakíthatjuk az órákat, az értékelést. Ezért kérek, őszintén írd le a véleményed az elmúlt időszakról, hogy még eredményesebben folytathassuk a közös munkát! Köszönöm szépen! Barbarics Tanárnő

Álnév: *Sajtóvadás* Dátum: *2019.10.14*

- Mennyire szereted az angol nyelvet? (Jelöld be az alábbi skálán, ahol 1 = utálom, 5 = nagyon szeretem.)
1 2 3 4 5
- Az eddigi élményeid alapján mennyire szereted az IT angol órát? (Jelöld be az alábbi skálán, ahol 1 = utálom, 5 = nagyon szeretem.)
1 2 3 4 5
- Miért? (ide írd, ha szeretnéd bővebben kifejtetni miért szereted vagy sem az órát)
A tanárnő érdekesen tartja az órákat.
- Szeretnéd-e, hogy a következő időszakban is legyen lehetőség saját magadnak összeállítani dolgozatot?
 a. Igen
b. Nem
c. Attól függ:
- Miért?
Sokkal könnyebb rá tanulni.
- Szeretnéd-e, hogy a következő időszakban is legyen összefoglaló játék?
 a. Igen
b. Nem
c. Attól függ:
- Miért?
Nagyon ~~szépen~~ ~~és~~ nagyon sok dolgot ~~is~~ ~~is~~ tanultam meg.
- Szeretnéd-e, hogy a következő időszakban is legyen Kahoot! az órán?
 a. Igen
b. Nem
c. Attól függ:
- Miért?
Sokat tanulok azon az órán és sok naptár lehet benne

10. Szeretnéd-e, hogy a következő időszakban is legyen Kahoot! otthoni gyakorlásra?

- a. Igen
- b. Nem
- c. Attól függ:

11. Miért?

Dezerteg dolgot lehet tanulni így.

12. Szeretnéd-e, hogy a következő időszakban is legyen lehetőség Kahoot! kérdéseket írni?

- a. Igen
- b. Nem
- c. Attól függ:

13. Miért?

Szövegben tanulni és az érvényesíteni az órán.

14. Szeretnéd-e, hogy a következő időszakban is legyen kilépőcetli?

- a. Igen
- b. Nem
- c. Attól függ:

15. Miért?

(Itt) Jelszók, hogy a tanári felhívni a véleményünk a tanítással kapcsolatban.

16. Szeretnéd-e, hogy a következő időszakban is legyen négy állomásos csapat játék?

- a. Igen
- b. Nem
- c. Attól függ:

17. Miért?

Szó dolgot lehet így tanulni

18. Szeretnéd-e, hogy a következő időszakban is legyen lehetőség önkéntes szóbeli feleletre?

- a. Igen
- b. Nem
- c. Attól függ: *Ami, hogy könnyűbb lesznek e?*

19. Miért?

Hogy ke lehet sok szót jöjjön

20. Szeretnéd-e, hogy a következő időszakban is legyen angol mondatok gyűjtése Wendy dobálása közben?

- a. Igen
- b. Nem
- c. Attól függ:

21. Miért? *Mert Wendy cubi*

22. Szeretnéd-e, hogy a következő időszakban is legyen Google dokumentumban dolgozás?

- a. Igen
- b. Nem
- c. Attól függ:

23. Miért?

Nagyon összehasonlít és nehéz róla tanulni (nehéz)

24. Szeretnéd-e a következő anyagrészt is ilyen pontozásos rendszerben tanulni?

- a. Igen
- b. Nem
- c. Attól függ:

25. Mit változtatnál ezen a pontozásos rendszeren? *Semmit*

26. Hányast kaptál most IT angolból?

1 2 3 4 5

27. Mennyire vagy elégedett azzal, ahogy az elmúlt időszakban értékeltem a munkád IT angolból? (Jelöld be az alábbi skálán, ahol 1 = egyáltalán nem vagyok vele elégedett, 5 = teljesen elégedett vagyok vele)

1 2 3 4 5

28. Miért? (ide írd, ha szeretnéd bővebben kifejtetni miért vagy elégedett vagy elégedetlen ezzel az értékeléssel)

Sok esélyt ad a tanárnak a javításra, vagy a pont-szerésre

29. Ha hagyományos osztályzással kaptál volna most jegyet, akkor...

- a. jobb jegyed lenne.
- b. rosszabb jegyed lenne.
- c. ugyanilyen jegyed lenne.
- d. nem tudod eldönteni, hogy jobb vagy rosszabb jegyed lenne.
- e. egyéb:

30. Mennyire tükrözi a kapott IT angol jegy az elmúlt időszakban elsajátított IT angol tudásod? (Jelöld be az alábbi skálán, ahol 1 = egyáltalán nem tükrözi, 5 = teljes mértékben tükrözi)

1 2 3 4 5

31. Mennyire tükrözi a kapott IT angol jegy az elmúlt időszakban az IT angol tanulásába fektetett energiád? (Jelöld be az alábbi skálán, ahol 1 = egyáltalán nem tükrözi, 5 = teljes mértékben tükrözi)

1 2 3 4 5

32. Ide írd, ha valamelyik kérdésre részletesebben szeretnél válaszolni, illetve bármi mást a fentiekén kívül, ami tetszett vagy nem tetszett az elmúlt időszakban!

31. Sokkal kevesebbet tanultam, de mégis jó jegyet kaptam

Appendix E5

Questionnaires after each grade in the 9thITE group compared to each other

Same in all questionnaires

Differences between questionnaires

This is the next questionnaire for my doctoral research. Just like the previous one, this is also anonymous, and I won't only use your answers for my dissertation, but based on what you write, we can improve our lessons and assessment together. For this reason, please, write your honest opinion, so that we can continue our work together more efficiently. Thank you very much!

Ms Barbarics

Pseudonym:

1. How much do you like the English language? (1 = I hate it, 5 = I love it)

2. Based on your experiences so far, how much do you like the IT English lessons? (1 = I hate it, 5 = it's one of my favourite subjects)

Why?

3. In the next period, would you like to... (yes - no - other, then why?)

- have the opportunity to compile the test for yourself? (1, 2, 3)
- have a Jeopardy! quiz show (summary game)? (1, 2)
- play Kahoot! during the lesson? (1, 2, 3, 5: during the online lesson or for homework?)
- practice with Kahoot! at home? (1, see 5 above)
- have the opportunity to write Kahoot! questions? (1, 2, 3)
- have exit tickets? (1 -> self-assessment later)
- have a treasure hunt (4-station group game)? (1, 2 -> group work later)
- have voluntary oral tests? (1 -> vlog later)
- collect sentences in English while throwing Wendy around? (1)
- work in shared Google Documents? (1 -> see next)
- have a Google Document containing all the material of the level? (2, 3 -> then all)
- have mistakes maze? (2)
- have listening tasks? (2, 6: in class or for homework?)
- use Socrative to give your answers in class? (2)
- create a mind map together? (2)
- work in groups (creating a poster together with roles)? (3 -> 6: group work in general)
- teach each other (after working with a text individually)? (3)
- have the opportunity to create vlogs? (3, 4, 5, 6)
- have the opportunity to give presentations? (4, 5, 6)
- give feedback to each other about your presentations? (4, 6)

- get points for making notes about each other's presentations? (4)
- have the opportunity to create posters? (4, 5)
- give feedback to each other about your posters? (5)
- have a vocabulary quiz? (4 -> integrated into the final test in 5, 6)
- have self-assessment? (4 -> integrated into the questionnaire)
- have the opportunity to write test questions? (5, 6)
- have a similar final test? (5, 6) - narrative / escape room style
- collect points by being active in class? (6)
- have pair/group work in person? (6)
- have pair/group work in Discord if we have to be online? (6)
- have essay writing tasks? (6)
- play skribbl (drawing a word in connection to our topic that the others should guess) (6)

4. In the next period, would you like to learn in this point system? - Yes / No / Other:...
Why?

5. What would you change in this point system?

6. What grade have you just got from IT English?

7. How satisfied are you with the ways I assessed your work in IT English in this period? (1 = "I am not satisfied at all" to 5 = "I am absolutely satisfied.")
Why?

8. If you had been graded in the traditional way, your grade would have been...
...better / ...worse / ...the same / you can't decide whether your grade would have been better or worse / Other:...

9. How well does this grade reflect your IT English knowledge of this period? (1 = "It doesn't reflect it at all" to 5 = "It reflects it perfectly.")

10. How well does this grade reflect the energy you put in studying IT English during this period? (1 = "It doesn't reflect it at all" to 5 = "It reflects it perfectly.")

11. (introducing it after Level 2, making them separate questions after Level 3 :)

- a. What do you think about your own performance in this period? What went well?
- b. Not so well?
- c. What would you like to do to make it better?

12. Here you can detail your answer to any of the questions or anything else that you have liked or not liked during this period.

Appendix E6

Additional questions in the questionnaires in the 9thITE group

Midterm extra questions (together with the questions after the 3rd grade):

- What was your half yearly grade from IT English?
- How satisfied are you with the ways I assessed your work in IT English during this half year? (1 = “I am not satisfied at all” to 5 = “I am absolutely satisfied.”) Why?
- How well does this grade reflect your IT English knowledge of this period? (1 = “It doesn’t reflect it at all” to 5 = “It reflects it perfectly.”)
- How well does this grade reflect the energy you put in studying IT English during this period? (1 = “It doesn’t reflect it at all” to 5 = “It reflects it perfectly.”)
- What should we change in the next half year?

After the 4th grade extra questions:

- How would you collect points in the next period (learning about basic software)?
 - Creating a final test for myself (1000 points)
 - Writing a short test put together by the teacher (500 points)
 - Giving an individual presentation (about 8-10 minutes, 1000 points)
 - Giving a presentation in pairs (about 4-5 minutes/person, 500 points)
 - Giving a presentation in a group of 3 (2-3 minutes/person, 2-300 points)
 - Recording a vlog about the whole material (100-500 points)
 - Presenting an individual poster summarizing the whole material (1000 points)
 - Creating a poster about a subtopic (depending on the material: 100-500 points)
 - Other:

After the 5th grade (and the beginning of emergency remote delivery) extra questions:

- Based on your experiences so far, how much do you like the IT English lessons online? (1 = I hate it, 5 = it’s one of my favourite subjects) Why?
- We started this level in person by making notes about operating systems into two columns: “What I already know / New”. What do you think about this task?
- What do you think about the current online IT English lessons?
- Where would you like to follow your points? hashtag.school // Teams // Other
- How can I help you? (After what would you like to do to make it better?)

After the last grade and the end of the year extra questions:

- How did you collect points during Level 6 (I added “Why?” after each type of activity.)
 - Online class participation
 - Presentation
 - Vlog
 - Listening tasks in connection with gaming
 - Essay writing about gaming
 - Creating test questions
 - Level 6 test
 - Other:
- In what other ways would you like to collect points?
- What did you think about the online IT English lessons?
- What was your final grade from IT English? What do you think about this grade?

Appendix E7

Questionnaires after each grade in the 11thME group compared to each other

Same in all questionnaires

Differences between questionnaires

This is the next questionnaire for my doctoral research. Just like the previous one, this is also anonymous, and I won't only use your answers for my dissertation, but based on what you write, we can improve our lessons and assessment together. For this reason, please, write your honest opinion, so that we can continue our work together more efficiently. Thank you very much!

Ms Barbarics

Pseudonym:

1. How much do you like mathematics (in general)? (1 = I hate it, 5 = I love it) Why? (only from the second questionnaire)
2. Thinking back to the past two years / in the past month, what did you like the most in math lessons? And why?
3. Thinking back to the past two years / in the past month, what did you like the least in math lessons? And why?
4. How much do you like the math lessons? (1 = I hate it, 5 = it's one of my favourite subjects) Why? (only this in the first questionnaire)
5. What do you think about...
 - class work (that you were working on different tasks: preparing for the minimatura, solving problems from last year, working on competition tasks etc.)
 - the changes in the points system (5* - Super Master from 60 points, 5 - Master from 46 points, 4 - Good from 33 points, 3 - So-so from 27 points, 2 - Pass from 21 points)? (1, 2, 3, 4, 5, 6)
 - getting all the problems of the level at once? PIC (2, 3)
 - the minis? (3, 4, 5)
 - Khan Academy? (3, 4, 5)
 - the amount of time spent solving problems? (3, 4)
 - discussing problems together in class? (3, 4)
 - the test? (5, 6)
 - the online lessons? (5, 6)
6. In the next period, would you like to learn in this point system?
 - Yes

- No
- Other:

7. Why?

8. What would you change in this point system?

9. What grade have you just got from math?

10. How satisfied are you with the ways I assessed your work in math in this period? (1 = "I am not satisfied at all" to 5 = "I am absolutely satisfied.")

11. Why?

12. If you had been graded in the traditional way, your grade would have been...

- better
- worse
- the same
- you can't decide whether your grade would have been better or worse
- Other:...

13. How well does this grade reflect your math knowledge of this period? (1 = "It doesn't reflect it at all" to 5 = "It reflects it perfectly.")

14. How well does this grade reflect the energy you put in studying math during this period? (1 = "It doesn't reflect it at all" to 5 = "It reflects it perfectly.")

(introducing the following questions (15-18) after the 3rd grade:)

15. What do you think about your own performance in this period? What went well?

16. Not so well?

17. What would you like to do to make it better?

18. What should we change (in connection with assessment, or the lessons, or the tasks, and so on)?

19. Here you can detail your answer to any of the questions or anything else that you have liked or not liked during this period.

Appendix E8

Additional questions in the questionnaires in the 11thME group

Midterm extra questions (together with the questions after the 3rd grade):

- What was your half yearly grade from math?
- How satisfied are you with the ways I assessed your work in math during this half year? (1 = “I am not satisfied at all” to 5 = “I am absolutely satisfied.”)
- Why?
- How well does this grade reflect your math knowledge of this period? (1 = “It doesn’t reflect it at all” to 5 = “It reflects it perfectly.”)
- How well does this grade reflect the energy you put in studying math during this period? (1 = “It doesn’t reflect it at all” to 5 = “It reflects it perfectly.”)
- What should we change in the next half year? (in connection with assessment, or the lessons, or the tasks, and so on)

After the 4th grade extra questions:

- What grade have you just got from math?
- Have you got this grade in advance? Yes - No
- How do you feel about getting a grade in advance?

After the 5th grade (and the beginning of emergency remote delivery) extra questions:

- How much do you like the math lessons online? (1 = I hate them, 5 = I love them)
- Why?
- In the next period, how would you like to submit your work to me? (You can mark more than one option and add your own ideas, too.)
 - through continuous communication in One Note (Class Notebook)
 - through Assignments with deadlines
 - Other:
- In the next period would you like to... (yes - no - other:) Why?
 - have minis? If yes, in what form? (discuss it right afterwards, or you can fill it in any times, and so on)
 - have Khan Academy problems?
 - have video explanations?
 - have playful practice exercises?
- In the last period, how many of my video explanations have you watched?
 - 0
 - 1-2
 - 3-4
 - 5-6
 - 7-8
- How can I help you? (After “what would you like to do to make it better?”)

After the last grade and the end of the year extra questions:

- How did you collect points in this last period? (Mark everything that you received points for.)
 - uploaded answers in One Note (Class Notebook)

- submitted answers in Assignments
- answers sent in other ways (e.g. Facebook messenger, Teams chat)
- online mini
- Khan Academy
- participation in competitions
- Level 6 test
- Other:
- Next year would you like to... (yes - no - other:) Why?
 - have minis? (If we are in school, on paper, otherwise online)
 - have online quizzes even if we are in school?
 - have Khan Academy problems?
 - have video explanations if we stay online?
 - have picture explanations if we stay online?
- In the last period (in May, in connection with Level 6), how many of my video explanations have you watched?
 - 0
 - 1-2
 - 3-4
 - 5-6
- In the last period (in May, in connection with Level 6), how many of my picture explanations / help have you looked at?
 - 0
 - 1-2
 - 3-5
 - 6-10
 - 10-20
 - more than 20
- Next year in what other ways would you like to collect points?
- What did you think about the online math lessons?
- What was your final grade from math?

Appendix E9

Questionnaires in the 11thAME group

This is a questionnaire for my doctoral research. It is anonymous, and I will use your answers for research purposes, but based on what you write, we can also improve our lessons and assessment together. For this reason, please, write your honest opinion, so that we can continue our work together more efficiently. Thank you very much!

Ms Barbarics

1. Who is your math teacher? (teaching the non-elective math classes)

2. Do you play computer games? If yes, which ones? And in what language? (Only in the first questionnaire)

3. What are the first three words or expressions that come to mind when you hear “school assessment”? (Only in the first questionnaire)
4. What grade have you just got?
5. How satisfied are you with this assessment? (1 = “I am not satisfied at all” to 5 = “I am absolutely satisfied.”)
6. Why?
7. How well does this grade reflect your math knowledge of this period? (1 = “It doesn’t reflect it at all” to 5 = “It reflects it perfectly.”)
8. How well does this grade reflect the energy you put in studying math during this period? (1 = “It doesn’t reflect it at all” to 5 = “It reflects it perfectly.”)
9. In what ways do you think assessment has been traditional and in what ways new during this period? (Only in the first questionnaire)
10. What do you think about...
- the sample test a week before the real test? (1)
 - working for a grade registered in advance? (1)
 - the class being bilingual? (1)
 - the online lessons? (2, 3)
 - point-collecting system of this period (from 150 points 5*, from 120 points 5, from 90 points 4, from 60 points 3, from 30 points 2)? (2, 3)
11. In the next period, would you like to learn in this point system?
- Yes
 - No
 - Other:
12. If yes, where shall we keep track of your points? (only in the second questionnaire)
- hashtag.school (everybody sees everybody’s points)
 - excel sheet in Teams (everybody sees everybody’s points)
 - excel sheet in Teams with pseudonyms
 - under Points in OneNote (everybody sees only their own)
 - Other:
13. What would you change in this point system?
14. Here you can detail your answer to any of the questions or anything else that you have liked or not liked during this period.

Extra questions in the second questionnaire:

- What do you think about your own performance in this period? What went well?
- Not so well?
- What would you like to do to make it better?

- What should we change (in connection with assessment, or the lessons, or the tasks, and so on)?
- How can I help you?

Extra questions in the third (last) questionnaire:

- What did you like the most?
- What did you like the least?
- What was the most memorable good experience in connection with the online lessons?
- What was the most memorable bad experience in connection with the online lessons?
- Would you like to choose this course next year, too?
 - Yes
 - No
 - I don't know, yet
 - Other:
- What should we change next year? (in connection with assessment, or the lessons, or the tasks, and so on)

Appendix E10

Information about the groups and the questionnaires

Questionnaire	Entry	1 st	2 nd	3 rd	4 th	5 th	6 th
9thITE	2019	2019	2019	2020	2020	2020	2020
Date	Sept. 13	Oct. 14	Dec. 9	Jan. 20	March 2	Apr. 19	June 7
No. of students	n=17	n=11	n=14	n=11	n=13	n=5	n=8
11thME	2016	2019 Oct.	2019 Dec.	2020	2020	2020	2020
Date	Sept. 9	16	10	Jan. 22	March 11	Apr. 18	June 6
No. of students	n=17	n=16	n=12	n=14	n=10	n=12	n=13
11thAME							
Date	2020	2020	2020	2020	-	-	-
No. of students	Feb. 5	Feb. 5	Apr. 29	June 7			
	n=7	n=7	n=7	n=6			

Appendix F

The consent form signed by the parents

Appendix F1

The consent form signed by the parents in Hungarian

Tisztelt Szülők!

Barbarics Márta vagyok a Budapesti Műszaki Szakképzési Centrum Petrik Lajos Két Tanítási Nyelvű Vegyipari, Környezetvédelmi és Informatikai Szakgimnáziumában angol nyelv és matematika tanár. Emellett az Eötvös Loránd Tudományegyetem Neveléstudományi Doktori Iskolájában a Nyelvpedagógiai Programban folytatok doktori kutatásokat a középiskolai értékelés témakörében. Ebben a tanévben a 9.D osztály általam tanított szakmai angol csoportját is szeretném a doktori kutatásomba bevonni.

Kérem, járuljanak hozzá, hogy gyermekük a csoport tagjaként részt vegyen a kutatásban, ami a tanév során néhány kérdőív kitöltését, esetleg egy-egy interjút jelent majd. A kérdőíveket és interjúkat mind név nélkül, be nem azonosítható módon, kizárólag kutatási célra fogom felhasználni. Bármilyen további kérdés esetén, kérem, forduljanak hozzám bizalommal az iskola által meghirdetett fogadóórákon személyesen vagy a barbarics.marti@petrik.hu e-mail címen írásban!

Köszönettel: Budapest, 2019. szeptember 2.

NYILATKOZAT

Alulírott

(a szülő vagy gondviselő teljes neve nyomtatott betűkkel)

engedélyezem, hogy gyermekem

(a gyermek teljes neve nyomtatott betűkkel)

részt vegyen a fent említett kutatásban.

..... Dátum:

szülő (gondviselő) aláírása

Appendix F2

The English translation of the consent form signed by the parents

Dear Parents,

My name is Márta Barbarics. I teach English and mathematics in Budapesti Műszaki Szakképzési Centrum Petrik Lajos Két Tanítási Nyelvű Vegyipari, Környezetvédelmi és Informatikai Szakgimnáziuma. I am also a PhD candidate in the Language Pedagogy Programme of Eötvös Loránd University. My research is about assessment in secondary schools. This year I would like to include the 9th grade IT English group I teach into my research.

Please, give your consent that your child as part of the group could participate in my research. It would involve filling in questionnaires and taking part in interviews. Both questionnaires and interviews will be administered anonymously and used exclusively for research purposes. Should you have any further questions, please, feel free to contact me in person during my office hours advertised by the school or in writing to the following e-mail address: barbarics.marti@petrik.hu

Thank you!

..... Budapest, September 2, 2019.

CONSENT FORM

I

(name of parent in all capitals)

hereby consent to the participation of my child

(name of child in all capitals)

in the above described research.

.....

Date:

parent's signature

Appendix G

Description of the point-collecting system for the 9th grade IT English group

Dear Students,

You can collect points in four or five weeks depending on the length of the topic. At the end of this period you get a grade based on your points in the following way:

0-699 points = grade 1 (fail)

700-899 points = grade 2 (pass)

900-1099 points = grade 3 (so-so)

1100-1599 points = grade 4 (good)

and from 1600 points = grade 5 (Master)

(from 2000 points = 5* “Super Master” :)

You can collect points in the following ways:

- the final test is always worth at least 1000 points
- point-collecting minis in the beginning of lessons for about 200-300 points
- homework
- classwork
- online tasks
- above 2000 points you can collect extra Super Master points
- you can also get grades for the extra Super Master points, or you can buy privileges:
 - 200 points = a yes-no question in a test
 - 400 points = an extra day in a deadline
 - 800 points = 5 minutes chosen class activity
 - 1600 points = grade 5 OR a game lesson for you
 - 5000 points = in class activity for the whole group (game class, pizza ordering etc.)
 - 10000 points = out of class activity for the whole group (going for ice cream during the lesson, excursion in the afternoon etc.)

You can always check your points at www.hashtag.school

Love,
Ms Barbarics

Appendix H

Samples of alternative assessment methods mentioned in the teacher interviews

Figure H

Screenshot of some quotations labeled as “examples of alternative assessment methods” from the transcript of the pilot interviews coded in the atlas.ti program

Show quotations coded with példák	
ID	Name
2:15	pontrendszer
2:16	pluszt gyűjtenek
2:17	szóbeli visszajelzéseket az órán
2:18	videófelelet
2:19	páros felelet
2:20	csoportos dolgozat
2:21	csoportmunka
2:116	klasszikus témazáró
2:117	a kisdolgozat, amit ő állít össze magának
2:118	az a röpi, amit én rakok össze a tankönyv kérdéseiből
2:119	kiselőadás
2:120	videofelelet
2:121	csoportmunka, ami egy féléves projekt
2:122	az extra, ami kvázi bármi kreatív
2:123	az órai munka, az a füzet vezetésére
2:124	a lízing
2:129	a beadandót
3:9	mixed, kevert, összevegyített értékelést használok, ami félig jegyekre.
4:4	sok szöveges értékelésben van részük
2:184	pozitív visszajelzést
2:185	dumaidó
2:187	visszaszámlálás
2:189	az én értékelésem, hogy ők értékelnek engem
2:191	redmentás otthoni dolgozatot
2:192	egymás értékelése
2:193	önértékelés
3:15	játékosított jegy, vagy pont-jegy, vagy XP-jegy
3:18	keresztrejtvényt, szókeresőt, szókérdőt, activity játékot stb
3:19	kreatív típusú feladatok
3:20	pl. rajzolj egy képet
3:21	rajzolhatnak képregényt
3:22	illusztrálják a párbeszédet a könyvben
3:23	videovágás

- 3:42 kredit pontok
- 3:45 voltak ilyen kollektív elemei is
- 3:47 rangsor, igen, ahol mindenki látta
- 3:57 élménypedagógiai vonal, ha úgy vesszük
- 3:61 projektek
- 3:62 önértékelés
- 3:63 értékelik egymást
- 3:65 csoportbeli szerepükért is kapnak egy jegyet, hogy azt milyen jól való...
- 3:67 ez mind online kérdőívekkel zajlik, amit nagyon nagy macera karbantart...
- 3:73 visszajelzést kérek tőlük minden lecke végén, a hónap végén, ilyen ref...
- 3:80 nem szoktam nyelvtan dolgozatot íratni jegyre
- 3:82 szimuláljuk a dolgozatírást
- 3:84 nem ellenőriztük a munkafüzetet együtt
- 3:85 mi tudja őket motiválni házfeladat írásra és végül kiderült, hogy az...
- 3:90 az összes diák 5-ösről indul
- 3:91 nyelvvizsga
- 3:92 mirét ne lehetne mindenkinek más célja
- 3:93 egyéni célkitűzés
- 3:94 3 felsőfokú nyelvvizsgázóim kiültek az ebédlőbe
- 3:95 kiscsoportos megbeszélések minden héten
- 3:107 mikrogamifikációs elemek
- 3:108 cicás pecsétet kapnak, meg matricá
- 3:109 tehetség gondozás része, a high-achiever tanulókat pluszban jutalmazod
- 4:29 órai munkát
- 4:30 számonkérések
- 4:31 beadandók
- 4:37 írnak egy mini dolgozatot
- 4:41 törzsanyag, ez az, ami kötelező
- 4:43 tudásrendszerező feladatok
- 4:44 készítsen gondolattérképet
- 4:45 hasonlítsa össze egy táblázatban két különböző korszakot
- 4:46 készítsen el egy tételt, tehát ami ilyen összefoglaló
- 4:47 Mindig vannak szövegalkotási feladatok, ezek általában esszék
- 4:48 mindig vannak úgynevezett kreatív feladatok
- 4:100 Írj két irodalmi szereplő között egy sms váltást
- 4:101 készíts illusztrációt
- 4:102 képzeld el, hogy ennek a történetnek mi lehet a folytatása
- 4:103 csinálj meg egy learningapps-t
- 4:147 Szoktam kérni azt, hogy lőjék be, hogy szerintük a dolgozat hanyas lesz...
- 4:148 egymást javítják, értékelik
- 4:149 saját dolgozatot javítani
- 5:11 gamifikáltan
- 5:15 szöveges értékelés
- 5:16 szóbeli konzultációk

- 5:124 kódfejtés, mystery games, akár story szinten, akár szavakkal, bárhog...
- 5:125 vocab fejlesztés az ilyeneken keresztül ment
- 5:126 hangtan is, amit ilyen piktogramos szimbólumokból tanultak meg
- 5:127 ne kelljen az IPA kódokat használni
- 5:128 két nap felkészülés és egy nap teljesítés
- 5:129 első nap mondjuk bevezeted a témát
- 5:130 következő nap olyan, mint egy főpróba a nagy előadás előtt
- 5:131 a harmadik napon pedig teljesítettek
- 5:132 ez minden héten ismétlődött
- 5:133 sose mentünk rá a teljesítés napra úgy, hogy előtte ne próbálta volna...
- 5:134 A teljesítésben egy külön pontcsomagot lehetett megszerezni
- 5:135 a hó végi összpontokba minden egyes óra beleszámított
- 5:136 Megkérdeztem őket az elején, hogy hogy szeretnék
- 5:149 utolsó órán volt egy ilyen Q&A session
- 5:176 szóbeli tevékenységek
- 5:177 monológot kellett mondani
- 5:178 dialogizálni kellett
- 5:179 előadás is volt, amit csak nekem tartottak
- 5:180 ezek közül választani lehetett és akkor ők dönthették el
- 5:181 írásbeli, általános teszt
- 5:182 Lehetett fogalmazást írni
- 5:183 fordításokat
- 5:184 bármi, ami írásbeli, műfajilag szerintem az összes ilyen dolog megjele...
- 5:185 poszter készítés
- 5:186 feladatlap, amin fokozatosan nehezdedtek a feladatok

Appendix H1

Examples of Lali's cards that student can buy with their points



Appendix H2

Example of Ilona's point-collecting system

The future (7A,C/2016-7) Számonkérés - Értékelés

Feladat (Task)	Határidő (Deadline)	Pontszám (Points)	Pontozási útmutató (Rubric)
4 5-min tests	10.25, 11.08, 14, 21, 28	5*2=10	1 point 50-70% 2 points 80%+

Unit 2 test	12.06	10	1 point for each 10%
Transport mindmap (writing)	11.29	4	<p>0 point - Több szó, kifejezés hiányzik, egyetlen plusz kifejezést sem adott hozzá, több szó nem illik a témához, több kifejezés egyértelműen rossz helyen szerepel, a szavak olvashatatlanok vagy többnyire hibásak.</p> <p>1 point - Sok hibával írsz, a kértnél kevesebb szót adsz az ábrához, a tananyag kifejezései közül 1-2 hiányzik, 1-2 kifejezés egyértelműen rossz helyre került vagy nem tartozik a témához.</p> <p>2 points - Több hibával, de megfelelő mennyiségű kifejezés, szerepelnek az újonnan tanult szavakat.</p> <p>3 points - Előfordul 1-2 hiba, de megfelelő mennyiségű kifejezés, a tanult szókinccset jól használod.</p> <p>4 points - Szinte hibátlan gyűjtemény, minden kért kifejezés szerepel. Tetszetős megjelenítés. Határidőre beadva.</p>
Role play (speaking)	11.22	4	<p>0 point - A tankönyv segítségével sem tudod eljátszani a szerepet érthetően és hosszabb gondolkodás nélkül.</p> <p>1 point - A tankönyv segítségével hosszabb gondolkodási idő nélkül el tudod játszani a szereped, kiejtésbeli pontatlanságok előfordulnak.</p> <p>2 points - Kisebb szünetekkel és segítséggel el tudod játszani a szereped. Pontatlanságok és kiejtésbeli problémák előfordulnak.</p> <p>3 points - Kisebb szünetekkel el tudod játszani a szereped. Pontatlanságok és kiejtésbeli problémák előfordulnak.</p> <p>4 points - Folyamatosan és kifejezően játszod a szereped kevés hibával.</p>
Interactive notebook	12.06	4	<p>0 point - Rendezetlen, olvashatatlan, hiányos jegyzet.</p> <p>1 point - Olvasható, de hiányos jegyzet sok hibával.</p> <p>2 points - Rendezett, áttekinthető jegyzet, de előfordul hiányzó tartalom, nem naprakész.</p> <p>3 points - Rendezett, áttekinthető, hiánytalan, naprakész jegyzet.</p> <p>4 points - Rendezett, szinte hibátlan és teljes jegyzet. Jól áttekinthető, a tartalomjegyzék naprakész. Saját ötleteket, plusz információt tartalmaz. Szépen díszített.</p>
Altogether		32	Maximum pontszámot csak a határidőre elvégzett feladatra lehet kapni.
Extras	12.02		<p>CB 21/4 OR <i>Detective of the year</i> olvasmány szavainak gondolattérképe. OR Translation: CB 21/4 or WB 15/4 - max. 4 points</p> <p>The London Underground Extra (handout) - 4 points</p>

Óraterv

Hét/ Dátum	Óra- szám	Feladatok	Készségfejlesztés	Nyelvhelyesség	Eszközök	Házi feladat	Megjegyzés
43/9 10.24	32	2A Journey into Space	Reading comprehension: recognition of vocabulary, picture-word matching, application to common knowledge Vocabulary: mindmaps	Vocabulary: space, orbs, tools, people, sci-fi, the Earth, the Moon, the Sun, star, planet, astronaut, galaxy, rocket, satellite, spaceship, space station, exploration. Dangerous, cheaper, probably	book board	Wb 14/1,2,3	
10.25	33	2A Grammar	Problem solving: deduction.	Will for future.	book board	Wb 15/4	5-min test
10.26 10.27	34	2A writing & speaking	Reading comprehension: comprehension questions	question word order	book board	Wb 5/6	
10.27 10.28	35	2B Detective of the year	Spelling; vocabulary-collocation; reading comprehension: true/false/not mentioned	solve, case, award, straightaway, jealous, president	book, CD		
45/10 11.07	36	2B Grammar & speaking	Speaking: respond to a problem or idea	will for immediate decisions All right. OK. Don't worry. Yes, please. Good idea. touch, saucepan, go bowling, present, wash up, answer, send, take sy in the car, decision, decide	book, cards (Wb 17/3) internet, CD	Wb 16-17/2	plickers
11.08	37	The Universe	Listening for specific words and collocations; recognising clines; gap-fill	Milky Way, a galaxy, the Sun, a star, the Earth, a planet, Laniakea a supercluster, The Great Attractor, the centre of a supercluster, Virgo, a galaxy cluster, Perseus Pisces	internet, projector, ppt, cards, handout, loudspeaker		5-min test: 2B vocab, will future +/- /? https://drive.google.com/drive/folders/0B0kCQnB8LSbdVWFYeUtxZmRqZEK
11.09 11.10	38	2C Your future	Reading comprehension-questionnaire; Vocabulary	another, prediction, village, outdoors, hospital, none of these, appear on TV, aliens visit the Earth, disease, war, happen, intelligent	book	WB 18/2	
11.10 11.11	39	2C Places	Vocabulary	prepositions: in, on	book	Wb 19/4	

Hét/ Dátu m	Óra- szám	Feladatok	Készségfejlesztés	Nyelvhelyesség	Eszközök	Házi feladat	Megjegyzés
46/11 11.14	40	2C Predictions	Listening for specific information & speaking: express opinion - I think, I hope, I don't think, I agree, I don't agree, ...probably..., perhaps...	study at home with computer, send e-mail, people need houses and food, robots need electricity, wild tiger, see animals in the zoo, go shopping, order food on the internet, shops deliver food to your house, work in a factory, play sports, take holidays	book	Wb 19/5	5-min test: 2C/questionnaire vocab, prepositions of place
11.15	41	Fortune telling	Writing and speaking.	Questions.	handout	WB 19/6	pickers: what has happened in the story so far?
11.16 11.17	42	2D Martin's party	Listening for gist. Vocabulary: useful expressions. WB 20/1	I'm really looking forward to it. You're kidding! What's the problem? He/She can't make it. Are you sure? Are you having a good time? Would you like a drink? Here goes.	book, internet, tablet, loudspeaker	learn the story+WB 20/1 by next Tuesday 20/1	
11.17 11.18	43	2D Offering help	Speaking: situations	<i>will</i> for offers	book	WB 21/3	
47/12 11.21	44	Let me help you Nasreddin and the grave	Speaking: situations. Writing: lists Reading: ordering paragraphs	<i>will</i> for offers, organise a party, tell a story	handout	WB 21/4	5-min test: 2C/prediction vocab
11.22	45	2D Martin's party	Speaking - Role play.				marked
11.23 11.24	46	2D Intentions	Speaking & writing.	<i>going to</i> for intentions	book	WB 21/5	
11.24 11.25	47	2 Culture: transport	Reading: T/F/DS, matching Speaking: discussion	vocabulary: vehicle, airport, cart, Channel Tunnel, horse, train, toll, terminal, take off, station, railway, motorway, land, driver	book	mindmap for travelling + space travel + 5 words	
48/13 11.28	48	2 Culture: The London Underground	Reading: gapfill, wordsearch, T/F	vocabulary: ticket, pigeon, lift, platform, tube train, mice, map, track	handout	add words to mindmap	5-min test: going to future for intentions

Hét/ Dátu m	Óra- szám	Feladatok	Készségfejlesztés	Nyelvhelyesség	Eszközök	Házi feladat	Megjegyzés
11.29	49	2 English across curr.: Science	The solar system - reading. Speaking: 20 questions	numbers, comparatives, superlatives	book internet, projector, tablet		plickers quiz
11.30 12.01	50	2 English across curr.: Science	Temperature - math	numbers, sums, comparatives, superlatives	handout		
12.01 12.02	51	Revision					
49/14 12.05	52	Revision					
12.06	53	Unit 2 test					
12.07 12.08	54	Revision units 1&2					
12.08 12.09	55	Revision test 1					Separate mark

Megvalósítás / Reflexiók

Appendix H3

An example of Dóra's rebus puzzles

a+  -rl+  -ft

s+  -x+  -le+  v=r

 -c+e+  -lo+  -ck+p+  t=m

Appendix I

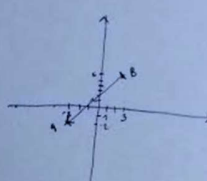
Samples of alternative assessment methods used during the action research

Appendix II

Samples of mini point-collecting opportunities in the 11thME group

Mini

- Points $A(-3; -2)$ and $B(3; 6)$ are given,
 - find the coordinates of vector \vec{BA} ,
 - find the length of vector \vec{BA} .
- The following vectors are given $\vec{a}(8; -4)$ and $\vec{b}(2; 4)$.
 - Calculate their scalar product.
 - Find the angle between them.
- Find the angle between vectors \vec{AB} and \vec{AC} if $A(-2; -1)$, $B(1; 3)$, and $C(4; -9)$.
- Vectors \vec{a} and \vec{b} are perpendicular, find the value of x if $\vec{a}(1; 12-x)$ and $\vec{b}(7; 9-2x)$.



$\vec{BA} = (-6; -8)$
 length: $(-6)^2 + (-8)^2 = e^2$
 $100 = e^2$
 $10 = e$

1. $s = 2x \cdot b_x + 2y \cdot b_y = 8 \cdot 2 + (-4) \cdot 4 = 0$
 $\angle = 90^\circ$

3. $\vec{AB} = (3; 4)$ $\vec{AC} = (6; -8)$
 $s = 2x \cdot b_x + 2y \cdot b_y = \vec{AB} \cdot \vec{AC} + \vec{AB}_y \cdot \vec{AC}_y = 3 \cdot 6 + 4 \cdot (-8) = -14$
 $s = |\vec{AB}| \cdot |\vec{AC}| \cdot \cos \alpha$
 $\cos \alpha = \frac{s}{|\vec{AB}| \cdot |\vec{AC}|} = \frac{-14}{5 \cdot 10} = -\frac{14}{50} = -\frac{7}{25}$
 $\alpha \approx 109,26^\circ$

$\cos \alpha = 0 \Rightarrow \alpha = 90^\circ$

$\vec{a} = (1; 12-x)$ $\vec{b} = (7; 9-2x)$
 $|\vec{a}|^2 = 1^2 + (12-x)^2$ $|\vec{b}|^2 = 7^2 + (9-2x)^2$
 $|\vec{a}| = \sqrt{1 + (12-x)^2}$ $|\vec{b}| = \sqrt{7^2 + (9-2x)^2}$

$|\vec{a}| \cdot |\vec{b}| \cdot \cos \alpha = a_x \cdot b_x + a_y \cdot b_y$
 $\sqrt{1+(12-x)^2} \cdot \sqrt{7^2+(9-2x)^2} \cdot 0 = 1 \cdot 7 + (12-x) \cdot (9-2x)$
 $0 = 7 + (108 - 9x + 108 - 22x)$ *This was the good*
 $0 = 223 - 31x + 2x^2$
 $0 = (223 - 31x + 2x^2) \cdot (108 - 22x - 9x + 2x^2 + 108 - 9x + 2x^2)$
 $0 = (223 - 31x + 2x^2) \cdot (223 - 66x + 4x^2)$
 $x_{1/2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
 $x_{1/2} = \frac{31 \pm \sqrt{(-31)^2 - 4 \cdot 2 \cdot 223}}{2 \cdot 2}$
 $x_{1/2} = \frac{31 \pm \sqrt{785}}{4}$

1. $a_1 = -3$ $a_n = 2 \cdot (a_{n-1} - a_{n-2})$
 $a_2 = -2$
 $a_3 = 2$
 $a_4 = 8$
 $a_5 = 12$
 $d = 26$

2. $a_1 = -18$ $a_2 = 6$ $a_3 = 30$ $a_4 = 54$ $a_5 = 78$
 $S_n = \frac{a_1 - a_n}{2} \cdot n$
 $S_5 = \frac{-18 + 78}{2} \cdot 5 = 150$

3. $a_1 = 2$ $a_2 = 6$ $a_3 = 18$ $a_4 = 54$ $a_5 = 162$
 $a_n = a_1 \cdot q^{n-1}$ $q = 3$
 $S_5 = \frac{2 + 162}{2} \cdot 5 = 205$
 or $q = -3$ $S_5 = 262$

Appendix I2

Sample of a vocabulary quiz in the 9thITE group

Level 4 - Vocabulary Test

1. Translate the following words from English to Hungarian.
 - a. Newsgroup
 - b. Consent
 - c. Agreed boundaries
 - d. STIs
 - e. Illegal activities
2. Translate the following words from Hungarian to English.
 - a. Tűzfal
 - b. Nem tervezett terhesség
 - c. Felhasználói felület
 - d. Színvak
 - e. Böngésző előzmények
3. Explain the meaning of the following words (in English).
 - a. Modem
 - b. Hotspot
 - c. FTP
 - d. Assertive communication
 - e. Privacy settings
 - f. Layers of the internet
 - g. IP address

For bonus points, recall information from the presentations.

Appendix I3

Sample of a test with increasingly more difficult tasks in the 11thAME group

Advanced Level Test 1

1. There are 20 red, 16 blue, 12 white, and 8 green balls in a box. At least how many should you randomly pick (without replacement) so that you surely have among them
 - a. green
 - b. red or blue
 - c. red and blue
 - d. two different colors
 - e. at least 4 blue
 - f. at least 3 from all the colors
 - g. at least 3 from one color?
2. List the elements of sets A, B, and C if you know the following about them:

$A \setminus B = \{2, 3, 9, 10, 15\}$	$B \setminus C = \{6, 7, 12, 13\}$
$C \setminus A = \{1, 4, 8, 11, 14\}$	$A \cap B \cap C = \{5\}$
$A \cup B = \{2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15\}$	$ C = 9$
$A \cup C = \{1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 14, 15\}$	
3. Prove the following divisibility for all the natural numbers: $4 \mid 7^n + 3^{n+1}$
4. Find x (the base of the number system) if
 - a. $1034_5 = 100_x$
 - b. $354_{10} = 542_x$
5. Find the missing digits (x and y), so that the following divisibility rules are true:
 - a. $2 \mid 324x2_5$
 - b. $3 \mid 310x2_4$
 - c. $10 \mid 3x25y_6$

+1. Prove that among five integers, there is always three whose sum is divisible by 3.

+2. Prove the following for all natural numbers:

$$1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + 3 \cdot 4 \cdot 5 \dots + n \cdot (n + 1) \cdot (n + 2) = \frac{n(n + 1)(n + 2)(n + 3)}{4}$$

+3. Give four sets for which all the following are true:

- any two has common elements
- the intersection of any three is empty set
- the cardinality of all of them equals
- they have the least possible number of elements

+4. Prove the following statement: in the base-7 number system if the sum of the digits of a number is 18, it is divisible by 6.

+5. Find the missing digits (x and y), so that the following divisibility rule is true:

$$56 \mid 31x52y_8$$

+6. Are there any numbers greater than 1 that consist of only 1s (as digits) in both the decimal and the binary number system?

Appendix I4

Sample of an achievement test compiled by a student in the 9thITE group

Test #3

1.Task: What do the abbreviations mean? (/5)

HDD: _____

SSD: _____

CD: _____

DVD: _____

Blu-ray: _____

2. Task: Match the pictures with their names. (/4)



DVD

CD

Blu-ray

Floppy disk

3.Task: Write down the storage capacity of the items in sentences. (/5)

CD: _____

DVD: _____

Floppy: _____

Blu-ray: _____

HDD: _____

4.Task: Write 3 important things about flash memory. (/3)

5.Task: Circle the flash memories. (/5)

SSD HDD CD Pendrive Blu-ray

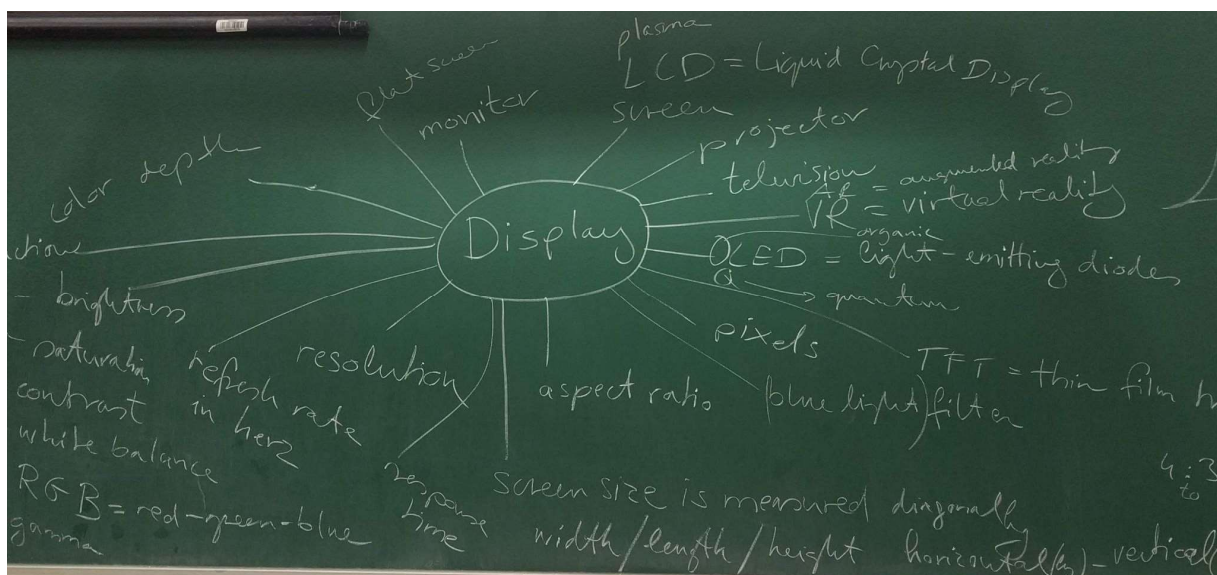
6.Task: Write 3 important things about magnetic storage. (/3)

7.Task: Write 2-2 sentences for comparison between flash memory and magnetic storage. (/4)

8.Task: Describe the benefits of Blu-ray. (/3)

Appendix I5

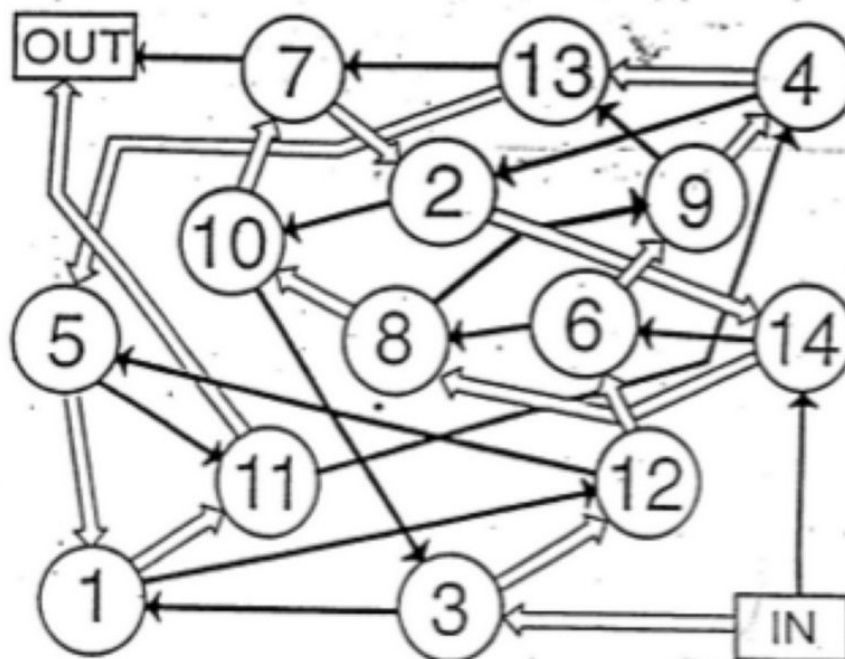
Sample of a mind map created together with the 9thITE group



Appendix I6

Sample of a Mistakes Maze activity from the 9thITE group

Mistakes Maze



Correct or incorrect?

You should start at IN and finish at OUT. If you think a sentence is correct, follow the white arrow. If you think it is wrong, follow the black arrow. You must visit each number once. Good luck!

IN: "Buses can carry data between to things."

- 1 - The computer hardware represents the physical parts of the computer.
 - 2 - The keyboard have function keys.
 - 3 - Key combinations can execute commands.
 - 4 - An image scanner transforms things to a digital form.
 - 5 - The central processing unit does math calculations.
 - 6 - The CPU the brain of the computer.
 - 7 - The computer case is made of plastic or metal.
 - 8 - The PSU have its own cooling fan.
 - 9 - Laptops use a touchpad instead of a mouse.
 - 10 - The PC is mean personal computer.
 - 11 - Some keyboards have Hungarian letters.
 - 12 - The headphones is output device.
 - 13 - Optical disc drives is a peripheral.
 - 14 - The motherboard was connect all of the hardware.
- OUT? :)

Appendix I7

Sample of an Exit Ticket from the 9thITE group

3-2-1 Exit Ticket

Three facts I learned:

- 1.
- 2.
- 3.

Two questions I have:

- 1.
- 2.

One opinion I have:

Appendix I8

Peer assessment sheet for students' presentations

.....'s presentation evaluated by.....

Content:

Was the presentation interesting?	--	-	?	+	++
Was the content clear, understandable?	--	-	?	+	++
Did the presentation have an intro., main part, and summary?	--	-	?	+	++
Was the grammar correct?	--	-	?	+	++
Was the vocabulary correct?	--	-	?	+	++

Presentation style:

Was the presenter engaging, enthusiastic?	--	-	?	+	++
Was the presenter's tone and rhythm good?	--	-	?	+	++
Was the pronunciation understandable?	--	-	?	+	++
Was the presentation easy to follow?	--	-	?	+	++
Did the presenter keep eye contact with the audience?	--	-	?	+	++

Visuals:

Did the visuals look good?	--	-	?	+	++
Did the visuals support the message of the presentation?	--	-	?	+	++
Were the slides logically built up?	--	-	?	+	++
Were the texts and pictures good (length, size, colors etc.)?	--	-	?	+	++

All in all impression:

-- - ? + ++

Why?

Appendix I9

Samples of students' posters from the 9thITE group

Kurzgesagt is a German word what means short or in a nutshell and a small team who make short animation from science.

They make video from human's nature and invention.

Also make videos from big things like the beginning end the ending of the universe or small thing like an ant or atom's

KURZGESAGT — IN A NUTSHELL —

They make the most complicated science also understandable for kid's, adult's and scientists

The videos have good music and understandable storyline

Safe Internet Usage

Be Careful What You Download

A top goal of cybercriminals is to trick you into downloading malware—programs or apps that carry malware or try to steal information. This malware can be disguised as an app: anything from a popular game to something that checks traffic or the weather. As PCWorld advises, don't download apps that look suspicious or come from a site you don't trust.

Choose Strong Passwords

Passwords are one of the biggest weak spots in the whole Internet security structure, but there's currently no way around them. And the problem with passwords is that people tend to choose easy ones to remember (such as "password" and "123456"), which are also easy for cyber thieves to guess. Select strong passwords that are harder for cybercriminals to demystify.

Make Sure Your Internet Connection is Secure

When you go online in a public place, for example by using a public Wi-Fi connection, PCMag notes you have no direct control over its security. Your vulnerable endpoint is your local Internet connection. Make sure your device is secure, and when in doubt, wait for a better time (i.e., until you're able to connect to a secure Wi-Fi network) before providing information such as your bank account number.

Keep Your Privacy Settings On

Marketers love to know all about you, and so do hackers. Both can learn a lot from your browsing and social media usage. But you can take charge of your information. As noted by Lifehacker, both web browsers and mobile operating systems have settings available to protect your privacy online. Major websites like Facebook also have privacy-enhancing settings available.

Sources:
<https://www.getsafeonline.org/protecting-your-computer/safe-internet-use/>
<https://usa.kaspersky.com/resource-center/preemptive-safety/top-10-internet-safety-rules-and-what-not-to-do-online>

Appendix I10

Sample from a sudoku task and its solution by a student from the 9thITE group










Source: p. 69 from Esteras, S. R. (2008). *Infotech: English for computer users* (4th ed.).







Cambridge University Press.















Instructions

This Word Sudoku is a variation on the normal Sudoku. Instead of using the numbers 1 to 9, we are using words and icons. There are nine WP functions and their equivalent icons, so we are playing with nine pairs. In order to complete the grid, you can use each function or the equivalent icon only once in each row, each column, and in each of the 3x3 boxes. The icons can only be used in the coloured boxes.

Word processing functions and icons

-  Align Left
-  Print Preview
-  Insert Table
-  Drawing
-  Bullets
-  Insert Hyperlink
-  Columns
-  Undo
-  Open

								
Align Left			Insert Table					
		Undo			Print Preview			
Print Preview								
								Columns
Undo	Insert Hyperlink	Open						Drawing
						Open	Bullets	
			Columns					Insert Hyperlink
				Insert Table	Insert Hyperlink	Drawing		

Insert hyperlink	Drawing	Columns	Bullets	Undo	Open		Align left	Print Preview
Align Left	Open	Print preview	Insert Table	Insert Hyperlink	Columns			Bullets
Bullets	Insert table	Undo	Drawing	Align left	Columns	Print Preview		
Columns	Print Preview	Bullets	Open		Drawing	Insert Hyperlink	Undo	Align left
Drawing	Align left	Insert table				Bullets	Open	Columns
Undo	Insert Hyperlink	Open		Columns	Insert table	Print preview	Insert table	Drawing
Print preview		Insert Hyperlink	Undo	Drawing	Align left	Open	Bullets	Undo
	Bullets	Drawing	Columns	Open	Bullets	Align left	Print preview	Insert Hyperlink
		Align left	Print preview	Insert Table	Insert Hyperlink	Drawing	Columns	Insert Table

Appendix I11

Samples from mathematics related memes sent by students from the 11thME group

$$f(x) = \text{🍕} \quad g(x) = \text{🍍}$$

$$f(g(x)) = \text{🍕🍍}$$

$$g(f(x)) = \text{🍍🍕}$$

$$\begin{cases} x + x + x = 60 \\ x + y + y = 30 \\ y - 2z = 3 \end{cases}$$

$$z + x + y = ?$$



Engenheiro Sincero

$$\text{🌺} + \text{🌺} + \text{🌺} = 60$$

$$\text{🌺} + \text{🌺} + \text{🌺} = 30$$

$$\text{🌺} - \text{🌺} = 3$$

$$\text{🌺} + \text{🌺} + \text{🌺} = ?$$



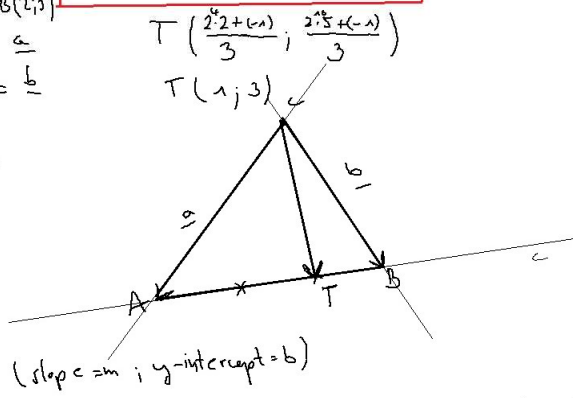
Appendix I12

Samples from picture explanations for the 11thME group

$$241) a) \begin{cases} A(a_1, a_2) \\ B(b_1, b_2) \\ C(c_1, c_2) \end{cases} \left\{ \begin{array}{l} M_{AB} \left(\frac{a_1+b_1}{2}; \frac{a_2+b_2}{2} \right) \\ S \left(\frac{x \cdot \frac{a_1+b_1}{2} + c_1}{3}; \frac{x \cdot \frac{a_2+b_2}{2} + c_2}{3} \right) \end{array} \right.$$

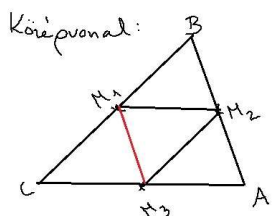
only for the centroid: $S \left(\frac{a_1+b_1+c_1}{3}; \frac{a_2+b_2+c_2}{3} \right)$

b) $\vec{CA} = (-9; -3) = \frac{a}{3}$
 $\vec{CB} = (-6; 3) = \frac{b}{3}$
 $\vec{CT} = (-7; 1) = \frac{c}{3}$

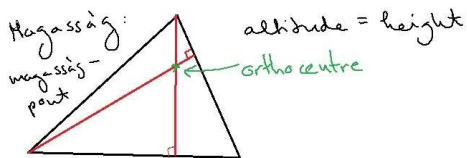
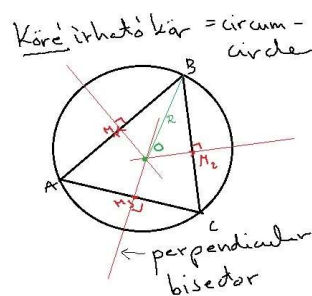


hw (c) equations of the lines of the sides
 $y = m \cdot x + b$ (slope = m ; y-intercept = b)

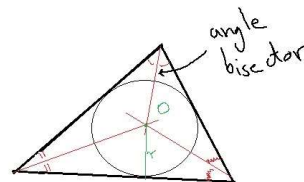
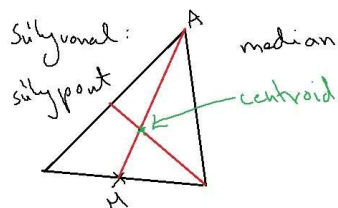
$n_1 x + n_2 y = n_1 p_1 + n_2 p_2$ (normal vector = (n_1, n_2) ; point on the line (p_1, p_2))



midlines (mid segments)
 $M_1 M_3 \parallel AB$
 parallel
 $2 \cdot M_1 M_3 = AB$



Béírtató kör = inscribed-circle




Appendix I13


Sample from a word grid task of the 9thITE group's escape-room-themed achievement test
 Link: <https://learningapps.org/watch?v=ptmdo7hv320>


Task

Find all the game genres and write down your message and code at the end.

OK





Account settings: Márta Barbarics 

Browse Apps
Create App
Create collection
My Stuff

Game genres

2020-05-29

F	I	R	S	T	-	P	E	R	S	O	N		S	H	O	O	T	E	R	V	M
X	Z	K	E	R	O	L	E	-	P	L	A	Y	I	N	G		G	A	M	E	T
X	D	R	X	I	M	R	X	T	L	Q	Y	T	M	V	F	S	C	N	J	C	Q
B	Q	J	F	T	U	G	E	F	H	L	Y	V	U	Q	M	J	F	F	K	N	A
K	E	K	B	I	O	Y	B	K	P	U	Z	Z	L	E	Q	K	T	L	P	W	F
A	X	X	I	M	O	V	T	L	X	U	V	W	A	D	V	E	N	T	U	R	E
L	L	V	G	I	X	T	F	C	L	F	A	C	T	I	O	N	Y	H	F	O	K
Q	H	X	W	Y	Z	T	Q	I	S	R	A	C	I	N	G	Q	I	G	M	L	F
Y	E	A	H	H	F	X	I	N	J	W	S	P	O	R	T	S	N	M	C	M	X
I	T	A	O	H	K	U	F	I	G	H	T	I	N	G	L	E	B	R	N	X	Z
C	W	P	N	A	T	S	V	C	V	E	R	J	D	D	R	I	I	T	A	K	M
L	J	G	Q	E	G	T	C	U	E	L	A	F	H	I	A	W	M	P	P	B	A
J	P	U	N	D	A	J	X	M	T	B	T	R	T	K	N	U	T	I	S	F	R
L	M	J	N	K	K	Y	I	H	A	R	E	W	C	V	K	W	M	J	F	Z	L
L	R	A	D	B	C	N	M	F	F	L	G	U	M	A	S	X	H	T	X	K	J
U	I	B	D	X	X	P	A	O	X	L	Y	L	W	W	B	F	N	L	W	H	I

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

Appendix J

Data tables of students' questionnaire answers

Table J1

Data of students' actual grades and the grades they reported in the questionnaires

actual grade	average	SD	questionnaire	average	SD
9thITE			9thITE		
no.1 (n=16)	4.75	0.45	no.1 (n=11)	4.8	0.42
no.2 (n=16)	4.25	0.86	no.2 (n=14)	4.17	0.94
no.3 (n=16)	4.4	0.91	no.3 (n=11)	4.64	0.67
midterm (n=16)	4.5	0.63	midterm (n=11)	4.82	0.4
no.4 (n=16)	4.25	1.34	no.4 (n=13)	4.08	1.44
no.5 (n=16)	4.06	1.34	no.5 (n=5)	4.8	0.45
no.6 (n=16)	4.25	1.18	no.6 (n=8)	4.88	0.35
year-end (n=16)	4.38	0.96	year-end (n=8)	4.88	0.35
11thME			11thME		
no.1 (n=16)	4.31	0.87	no.1 (n=16)	4.31	0.87
no.2 (n=16)	4.5	1.1	no.2 (n=12)	4.83	0.39
no.3 (n=16)	4.19	0.83	no.3 (n=14)	4.5	0.52
midterm (n=16)	4.38	0.72	midterm (n=14)	4.43	0.76
no.4 (n=16)	4.38	0.89	no.4 (n=10)	4.5	0.71
no.5 (n=16)	4.06	1.12	no.5 (n=12)	4.5	0.67
no.6 (n=16)	3.81	1.52	no.6 (n=13)	3.61	1.61
year-end (n=16)	4.38	0.89	year-end (n=13)	4.31	0.95
11thAME			11thAME		
no.1 (n=10)	4.8	0.42	no.1 (n=7)	4.86	0.38
no.2 (n=10)	4.2	1.14	no.2 (n=7)	4.86	0.38
no.3 (n=10)	3.8	1.55	no.3 (n=6)	4.17	1.6
year-end (n=10)	4.5	0.71	year-end (n=6)	4.5	0.84

Table J2

Data of students' grades reported in the questionnaires compared to how satisfied they are with these grades (from 1 = "not satisfied at all" to 5 = "absolutely satisfied")

grades	average	SD	satisfaction	average	SD
9thITE			9thITE		
no.1 (n=11)	4.8	0.42	no.1 (n=11)	4.73	0.47
no.2 (n=14)	4.17	0.94	no.2 (n=14)	4.38	0.77
no.3 (n=11)	4.64	0.67	no.3 (n=11)	4.73	0.47
midterm (n=11)	4.82	0.4	midterm (n=11)	4.82	0.4
no.4 (n=13)	4.08	1.44	no.4 (n=13)	4.31	1.18
no.5 (n=5)	4.8	0.45	no.5 (n=5)	5	0
no.6 (n=8)	4.88	0.35	no.6 (n=8)	5	0
11thME			11thME		
no.1 (n=16)	4.31	0.87	no.1 (n=16)	4.81	0.4
no.2 (n=12)	4.83	0.39	no.2 (n=12)	4.92	0.29
no.3 (n=14)	4.5	0.52	no.3 (n=14)	4.57	0.65
midterm (n=14)	4.43	0.76	midterm (n=14)	4.79	0.43
no.4 (n=10)	4.5	0.71	no.4 (n=10)	4.9	0.32
no.5 (n=12)	4.5	0.67	no.5 (n=12)	4.92	0.29
no.6 (n=13)	3.61	1.61	no.6 (n=13)	4.62	1.12
11thAME			11thAME		
no.1 (n=7)	4.86	0.38	no.1 (n=7)	5	0
no.2 (n=7)	4.86	0.38	no.2 (n=7)	4.57	0.53
no.3 (n=6)	4.5	0.84	no.3 (n=6)	4.33	1.63

Table J3

Data of averages and standard deviations of students' questionnaire answers to how much they think their grade reflects their knowledge (1 = it doesn't reflect it at all, 5 = reflects it perfectly)

9th grade	Entry (n=17)	1 st (n=11)	2 nd (n=14)	3 rd (n=11)	4 th (n=13)	5 th (n=5)	6 th (n=8)
average	3.44	4.64	4.00	4.27	3.85	4.60	4.25
SD	1.06	0.50	0.91	0.65	1.41	0.55	1.04

11th grade	1 st (n=16)	2 nd (n=12)	3 rd (n=14)	4 th (n=10)	5 th (n=12)	6 th (n=13)
average	4.00	4.08	4.14	4.20	4.17	3.85
SD	0.82	1.00	0.86	0.92	0.83	1.07

Table J4

Data of averages and standard deviations of students' questionnaire answers to how much they think their grade reflects their effort (1 = it doesn't reflect it at all, 5 = it reflects it perfectly)

9th grade	Entry (n=17)	1 st (n=11)	2 nd (n=14)	3 rd (n=11)	4 th (n=13)	5 th (n=5)	6 th (n=8)
average	3.29	3.73	3.77	3.82	3.38	4.00	3.63
SD	1.31	1.10	1.09	1.08	1.04	0.71	1.41

11th grade	1 st (n=16)	2 nd (n=12)	3 rd (n=14)	4 th (n=10)	5 th (n=12)	6 th (n=13)
average	4.00	3.42	3.79	4.00	4.42	4.08
SD	0.89	1.44	1.05	1.25	1.00	1.38

Table J5

Data of students' answers to how much they like the English language (1 = I hate it, 5 = I love it)

	Entry (n=17)	1 st (n=11)	2 nd (n=14)	3 rd (n=11)	4 th (n=13)	5 th (n=5)	6 th (n=8)
average	3.8	4.36	4.15	4.27	4.46	4.80	4.88
SD	0.51	0.48	0.55	0.47	0.52	0.45	0.35

Table J6

Data of students' questionnaire answers to how much they like the IT English lessons* (1 = I hate it, 5 = it's one of my favorite subjects) *in the entry questionnaire: foreign language lessons

	Entry (n=17)	1 st (n=11)	2 nd (n=14)	3 rd (n=11)	4 th (n=13)	5 th (n=5)	6 th (n=8)
average	3.47	4.82	4.31	4.45	4.31	in school 4.4 online 3.20	in school 4.63 online 4.50
SD	0.98	0.39	0.48	0.52	0.63	in school 0.89 online 1.64	in school 0.52 online 0.76

Table J7

Data of students' questionnaire answers to how much they like math in general (1 = I hate it, 5 = I love it)

	1 st (n=16)	2 nd (n=12)	3 rd (n=14)	4 th (n=10)	5 th (n=12)	6 th (n=13)
average	4.00	4.08	4.14	3.80	3.92	3.92
SD	0.97	0.67	0.86	0.92	0.79	0.95

Table J8

Data of students' questionnaire answers to how much they like the math lessons (1 = I hate it, 5 = it's one of my favorite subjects)

	1 st (n=16)	2 nd (n=12)	3 rd (n=14)	4 th (n=10)	5 th (n=12)	6 th (n=13)
average	no data*	4.33	4.14	3.90	in school 4.17 online 3.5	in school 4.46 online 3.77
SD	no data*	0.65	0.77	0.99	in school 0.72 online 0.90	in school 0.88 online 1.09

* the first questionnaire did not contain the question

Table J9

Data of the number of students' answers to whether they would like to do the following activities (that they can collect points with) in the future

	Yes	No	Other	Sum
voluntary oral test	8 (73%)	1 (9%)	2 (18%)	11 (100%)
presentation	19 (73%)	6 (23%)	1 (4%)	26 (100%)
vlog	22 (61%)	5 (14%)	9 (25%)	36 (100%)
mini (in class)	24 (96%)	0 (0%)	1 (4%)	25 (100%)
vocabulary quiz	9 (69%)	2 (15%)	2 (15%)	13 (100%)
Kahoot! (in class)	39 (98%)	0 (0%)	1 (2%)	40 (100%)
Jeopardy!	23 (96%)	0 (0%)	1 (4%)	24 (100%)
students' own tests	32 (91%)	1 (3%)	2 (6%)	35 (100%)
poster	15 (83%)	2 (11%)	1 (6%)	18 (100%)
writing Kahoot! questions	35 (100%)	0 (0%)	0 (0%)	35 (100%)
Khan Academy	40 (82%)	3 (6%)	6 (12%)	49 (100%)
note taking	12 (100%)	0 (0%)	0 (0%)	12 (100%)
listening	18 (85%)	2 (10%)	1 (5%)	21 (100%)
writing	6 (75%)	2 (25%)	0 (0%)	8 (100%)
treasure hunt	20 (87%)	1 (4%)	2 (9%)	23 (100%)
Kahoot! (homework)	14 (88%)	2 (12%)	0 (0%)	16 (100%)
mini (online)	31 (82%)	5 (13%)	2 (5%)	38 (100%)
escape room test	13 (100%)	0 (0%)	0 (0%)	13 (100%)
writing test questions	11 (85%)	2 (15%)	0 (0%)	13 (100%)

Table J10

Data of the number of students' answers to whether they would like to do the following activities (that they cannot collect points with) in the future

	Yes	No	Other	Sum
group work (in class)	5 (62%)	1 (13%)	2 (25%)	8 (100%)
group work (online)	4 (50%)	4 (50%)	0 (0%)	8 (100%)
Wendy	11 (100%)	0 (0%)	0 (0%)	11 (100%)
group mind map	13 (100%)	0 (0%)	0 (0%)	13 (100%)
editing a shared doc	7 (64%)	1 (9%)	3 (27%)	11 (100%)
mistakes maze	6 (46%)	5 (39%)	2 (15%)	13 (100%)
teaching each other	4 (36%)	6 (55%)	1 (9%)	11 (100%)
Socratic	8 (62%)	0 (0%)	5 (38%)	13 (100%)
peer assessment	16 (64%)	8 (32%)	1 (4%)	25 (100%)
self-assessment	4 (31%)	7 (54%)	2 (15%)	13 (100%)
exit ticket	5 (46%)	5 (46%)	1 (8%)	11 (100%)
video explanation	18 (72%)	4 (16%)	3 (12%)	25 (100%)
picture explanation	13 (100%)	0 (0%)	0 (0%)	13 (100%)
Skribbl	7 (88%)	1 (12%)	0 (0%)	8 (100%)