



Cirlan, Elena

ETHICAL CLASSROOM ASSESSMENT: DIFFERENCES FOR PRE-SERVICE AND IN-SERVICE TEACHERS IN FINLAND

Master's Thesis in Education  
FACULTY OF EDUCATION  
Learning, Education and Technology

2017

University of Oulu Faculty of Education

Elena Cirlan: Ethical classroom assessment: differences for pre-service and in-service teachers in Finland

Master's Thesis in Education, 63 pages, 7 appendices

May 2017

---

Abstract

The assessment is a demanding and important activity which has a significant impact on students' learning and motivation (White, 2009; Brookhart, 1993, 1994), curriculum development, and the teaching process as a whole (Harlen, 2007; Lyon, 2013). Since the teachers play a central role in deciding upon assessment techniques and grades, it is their responsibility, in the first place, to assure that the assessment is versatile, feasible, fair, and ethical. Consequently, pre-service and in-service teachers' assessment training and knowledge concerning the assessment ethical codes and guidelines are of a major importance. The previous studies conducted by Green et al. (2007) and Liu et al. (2016) found that there is no professional agreement between pre-service and in-service teachers and in many cases, teacher's ethical perceptions differed from the ethical principles specified in literature.

The aim of the current study is to examine Finnish pre-service and in-service teachers' perceptions about the ethicality of the classroom assessment practices. Moreover, it was important to detect whether all teachers follow the same guidelines when it comes to the ethicality of the classroom assessment. The participants of the study are pre-service teachers of the Primary Teacher Education Master Degree Programs of the University of Oulu and in-service teachers of the University of Oulu Teacher Training Schools, Linnanmaa, and Koskela.

The results of this study show that on a less than a third of the scenarios, teachers within each group disagreed. And three out of 35 items displayed significant statistical disagreement between the groups of teachers. Based on these results, can be concluded that the instructions given to Finnish pre-service and in-service teachers regarding the classroom assessment practices, and their ethical judgement is in the most cases in line with the theories and guidelines. However, the findings also suggest that further instructions related to score pollution, the assessment of collaborative work, the use of the versatile assessment, and the awareness of the instructions regarding assessment conveyed in the Curricula, should be more carefully addressed in pre-service and in-service teacher education.

Keywords: classroom-assessment, ethics, in-service teachers, pre-service teachers, Finland

## CONTENTS

<b>1. INTRODUCTION</b> .....	<b>1</b>
<b>2. THEORETICAL FRAMEWORK</b> .....	<b>3</b>
<b>3. LITERATURE REVIEW</b> .....	<b>6</b>
3.1 Ethics and classroom assessment .....	8
<b>4. RESEARCH CONTEXT - FINNISH EDUCATION</b> .....	<b>12</b>
4.1 The Finnish National Core Curriculum, Local Curriculum, and Assessment .....	12
4.2 The principles of the Finnish educational system .....	14
4.3 Pre-service teacher education in Finland .....	16
4.4 In-service teacher education in Finland .....	17
<b>5. AIM AND RESEARCH QUESTION</b> .....	<b>19</b>
<b>6. METHODOLOGY</b> .....	<b>20</b>
6.1 Participants .....	20
6.2 Instrument.....	20
6.3 Data analysis and the Research Method.....	22
<b>7. RESULTS</b> .....	<b>24</b>
7.1 The degree of agreement within each group of teachers.....	24
7.1.1 Bias / Fairness.....	24
7.1.2 Communication about Grading.....	26
7.1.3 Confidentiality.....	27
7.1.4 Grading Practices.....	28
7.1.5 Multiple Assessment Opportunities.....	31
7.1.7 Summary.....	33
7.2 Analysis of the significance of difference between the teachers' groups .....	34
<b>8. DISCUSSION AND CONCLUSION</b> .....	<b>35</b>
8.1 Practical Implications.....	39
8.2 Directions for future research.....	40
<b>9. EVALUATION OF THE STUDY</b> .....	<b>42</b>
9.1 Research Validity and Reliability .....	42
9.2 Ethical Issues.....	43
9.3 Study limitations .....	43
<b>10. REFERENCES</b> .....	<b>45</b>
<b>11. APPENDIX</b> .....	<b>57</b>
Table 3. Scenarios in Bias/Fairness.....	57
Table 4. Scenarios in Communication about Grading.....	58

Table 5. Scenarios in Confidentiality.....	59
Table 6. Scenarios in Grading Practices.....	60
Table 7. Scenarios in Multiple Assessment Opportunities.....	62
Table 8. Scenarios in Test Administration.....	63

**TABLES & FIGURES**

Table 1. Total Number of Questions and Scenarios.....	22
Table 2. Areas of Agreement and Disagreement .....	23
Figure 1. Answers for item 8.....	24
Figure 2. Answers for Item 15 .....	25
Figure 3. Answers for Item 26 .....	26
Figure 4. Answers for item 11.....	26
Figure 5. Answers for item 17.....	27
Figure 6. Answers for item 10.....	28
Figure 7. Answers for item 12.....	28
Figure 8. Answers for item 38.....	29
Figure 9. Answers for item 34.....	29
Figure 10. Answers for item 24.....	30
Figure 11. Answers for item 29.....	30
Figure 12. Answers for item 37.....	31
Figure 13. Answers for item 30.....	31
Figure 14. Answers for item 33.....	32
Figure 15. Answers for item 5.....	33

## 1. INTRODUCTION

The assessment activity occupies one third of the professional time of teachers (Popham, 2004; Crooks, 1988) and has a great impact on students' learning, motivation (White, 2009; Brookhart, 1993, 1994), curriculum development, and the overall teaching process (Harlen, 2007; Lyon, 2013). Pope (2006) conveyed that morality and ethics should be highly considered in the process of classroom assessment. Interestingly, there still is a gap in the literature that investigates how ethics supports the provision of fair and qualitative assessment.

The issues related to the lack of ethical guidelines for the assessment practices resulted in multiple incidents published in the United States' press. The importance of this issue, motivated researchers to study the core problem of ethics in academic assessment. Thus, the focus of the recent studies was to examine the teachers' perceptions regarding various ethical issues in classroom assessment (Green, Johnson, Kim, & Pope, 2007). Green et al. (2007) highlighted that grades and other assessment techniques should show the level of student mastery of the set of instruction objectives. On the other hand, modification of grades for student effort, behavior, and late work put in danger the ethicality of assessment, polluting the grades. Moreover, the findings of this study showed that the pre-service and in-service teachers agreed with each other on less than half of the scenarios related to ethical issues in assessment used in the study. This suggests that there is no professional agreement in the field of assessment. Consequently, there is a need of further investigation in order to understand the existing gap between these two groups of teachers.

Johnson, Green, Kim, and Pope (2008) examined principals and principal candidates' understanding of the ethicality of the assessment practices applied in the United States. The results of this study revealed that there is no consensus between them as well. Therefore, Johnson et al. (2008) recommended a more profound professional development regarding ethical assessment. Tierney (2013) found that teachers "emphasized different aspects of fairness with the most prominent involving student's opportunities to learn and demonstrate learning, transparency, the classroom environment, critical reflection, and the tension between equal and equitable treatment in assessment (p. 55)." A recent study by Liu, Johnson, and Fan (2016) represents a comparison of Chinese and U.S. pre-service teacher's perceptions regarding the ethicality of classroom assessment. The researchers reported that there is no

global agreement regarding assessment and ethics. Moreover, teacher's ethical perceptions, in many cases, differed from the ethical principles specified in earlier literature.

The present study aims to investigate the ethicality of classroom assessment in Finland by comparing the perceptions of pre-service and in-service teachers. In this study, I have formulated the following research question: *what is the degree of agreement between the groups of pre-service and in-service teachers regarding the ethicality of classroom assessment?* In order to answer this question, I have developed and implemented a novel 35-scenario survey for pre-service and in-service teachers in elementary schools of Oulu, Finland. I built this study based on similar studies described above.

The Finnish educational system is considered one of the best in the world (the results of multiple international comparisons proved it, OECD, 2003, 2010, 2014). Consequently, it is important to find out whether the results obtained in this study are similar to or different from those of the previous studies conducted in other countries. Moreover, to the best of my knowledge no similar studies have been conducted in Finland to investigate the issue of ethicality of classroom assessment practices. In this sense, this study initiates a new research realm in the Finnish educational context concerning the ethicality of classroom assessment.

## 2. THEORETICAL FRAMEWORK

From a theoretical standpoint, ethics as a term has its roots in ancient Greek and refers to human conduct. Plato described ethics as a field of philosophy that refers to what “we ought to do” and its concern is to define what is right and wrong when it comes to human behavior. Singer (1993) suggested that ethics should not be perceived as a set of interdictions or as an “ideal system” that cannot be applied in practice. Once an ethical judgement cannot be applied in practice it means it lacks comprehensiveness in theory. Interestingly, Aristotle (284-322 BC) argued that by following our own choice, standards, and rationality we reach the meaning of moral behavior. Thus, only when the conditions of free decision and choice-making are fulfilled, humans can think and behave ethically (Brandt & Rose, 2004). Another aspect that defines a behavior as ethical or unethical is its accordance or contradiction to one’s obligations. This explains why there may be a persistent gap between the ethical principles and behavior. Behavior requires specific judgement in a specific context and situation; thus a general agreement for each case would be hardly achieved (Hostetler, 1997; Strike & Soltis, 1998; Kant, 1790).

Furthermore, according to Herbart (1917), there are two scientific branches of knowledge that the science of pedagogy relies on: ethics and psychology. The first identifies the goals, whereas the second examines the means and the challenges. Taking into account these aspects and the core mission of education – *to guide the development of an individual to reach the best results not only in a certain field but also as a sympathetic and respectful human being* – we may conclude that education is an ethical activity. Mahony (2009) and Singer (1993, 1994) pointed out that such ethical behavior as honesty and integrity is pertinent for both teachers and students.

Many studies show that classroom assessment is a field where there is no general professional agreement, each teacher/professor being the one who decides upon the ethicality of assessment. Thus, teachers’ grading schemas vary from one case to another. For instance, they can include different aspects of assessment as neatness or class participation that are not considered directly connected to the mastery of learning goals. Guidelines for ethical assessment are provided in the United States by such sources as The Joint Committee on Standards for Educational Evaluation (JCSEE, 2003) that “require that student evaluations be ethical, fair, useful, feasible, and accurate” (JCSEE, p. 3), the Code of Ethical Conduct and Statement of Commitment of the National Association for the Education of Young Children

(2005). In Finland, such guidelines are provided by the National Core Curriculum that asks for “fair and ethical assessment” (p. 49). In addition, there are the Local and School Curriculums that present more specific guidelines regarding the education as a whole and regarding assessment. The meaning of these guidelines is to serve as supporting tools for teachers when it comes to solving the dilemmas related to assessment.

McMillan (2000) proposed 11 assessment principles that represent the knowledge teachers and school administrators must possess. He also asked for versatile, valid, fair and ethical, efficient, and feasible assessment. Following such principles, teachers and administrators could build a basis for sound assessment practices. On the other hand, by not following these principles, teachers put in danger the ethicality of assessment and, hence, the students’ motivation and educational development.

It is important that teachers have a broad understanding regarding the ethics in classroom assessment and their appropriateness for certain situations. According to Gipps (1994), the existing guidelines of ethical assessment are “general and almost superficial” (p. 144). In addition, Green et al. (2007) suggested that the element of applicable ethical guidelines to classroom assessment has not been addressed in depth and that such guidelines may not be applicable in specific situations. Consequently, it is significant that more research in the field of assessment ethicality be carried out in order to come up with reasoned guidelines and tools that would support teachers’ assessment activities.

Besides the fact that previous literature offered limited attention to teacher’s ability to apply ethical guidelines in particular contexts (Green et al., 2007), teacher’s knowledge concerning ethical codes and guidelines may be doubtful because they either lack formal assessment training or their training is dated (Impara, Plake & Fager, 1993; Plake & Impara, 1997; Stiggings, 1999). Thus, the results of the present research may give an insight regarding the Finnish teachers’ assessment literacy and the knowledge they have about the existing guidelines related to the classroom assessment field.

A few studies that attempted to study in more depth ethical issues in academic assessment are those of Airasian (2005), Taylor and Nolen (2005). These researchers examined the professional standards, ethical principles in education, the ethicality of the standardized test preparation activities, and the assessment literature. They found two main directions or two general guiding principles that could potentially explain the ethical dilemmas depicted in these documents. The first principle is “Do no Harm”, which is perceived as a general ethical



principle, applicable in people's daily activities. In regards to this principle, there is a need of broad understanding of what harm is and how to choose from different harms when it comes to the teacher-student relationship. For instance, in Campbell's (2003) study, a teacher had to change students' grades in order to better the school's performance statistics. In this situation, the teacher had to choose either to harm the student by altering the grade or to "harm" the school's image by not raising the students' grades.

The "Do no Harm" principle derives from the "Golden Rule" - "Assess as Ye Would Be Assessed" (Payne, 2003) and is in line with the fundamental premise indicated in JCSEE (2003). In the Student Evaluation Standards, the ethical guidelines specify that one must protect the rights of those whose work is evaluated (p. 7). In this sense, by considering the student as the main "beneficiary" in the assessment process instead of the teacher and/or the school's image (or the parents' satisfaction) may prioritize the student's rights above those of the other stakeholders of the educational process.

The second general principle for classroom assessment is to "Avoid Score Pollution" (Popham, 1991; Haladyna, Nolen, & Haas, 1991). This principle was first proposed for the ethical standardized test preparation. However, Green et al. (2007) considered that this principle can also be applied in other classroom assessment practices - for example, the grading practices. Brookhart (2004) and Smith, Smith, and DeLisi (2001) argued that the assessment that includes grades should reflect only the level of mastery of the set objectives. Thus, whether a teacher considers students' effort, behavior, or late work, when assigning grades, denotes the possibility that these grades do not accurately reflect the level of mastery of the academic goals. Moreover, if teachers do not use a blind grading system, they might unconsciously pollute the scores by giving higher grades to students that they prefer and vice versa.

### **3. LITERATURE REVIEW**

In educational research, teacher ethics have been considered since the mid 1980's (Colnerud, 1997). Several researchers suggested that teachers' ethical expertise should be addressed and expanded. Thus, pre-service and in-service teachers must receive the proper instructions when it comes to assessment ethicality. Moreover, they should be aware of the existing guidelines for assessment and be knowledgeable when it comes to their application in practice.

Importantly, each profession implies a set of rules regarding behavior which are considered as the ethical rules of the profession (Thorndike, Cunningham, Thorndike & Hagen, 1991). The professionalization process stimulated the development of a new body of research in the field of ethics of teaching (Colnerud, 2006). In this sense, the research in the field of ethics of teaching emphasized moral values that are perceived as a personal philosophy and led to the development of professional standards for teachers (for example, the standards issued by the General Teaching Council for Northern Ireland where the teacher's ethical code (GTCNI (2007) p. 5) is formulated).

Since the professional standards were formulated, the attempts to compare the ethics of the teacher profession with those of other professions are always prevalent. However, the teacher profession has unique ethical conditions that must be considered. In this regard, Fenstermacher (1986) highlighted three distinctions. The first one is that in contrast to doctors (medical practitioners), teachers do not keep a knowledge gap between themselves and their pupils. The second difference derives from the teacher's duty of understanding and helping the pupil. Consequently, the maintenance of a social distance between the teacher and the pupil is not possible. The last distinction is that teachers cannot fulfill the responsibility of teaching alone, pupils' cooperation and feedback being a must for a sound educational process.

Tom (1980), Strike (1990), and Clark (1990, 1991) argued that the power and responsibilities towards the pupils and the risk of exerting the power in a manipulative way by teachers should be addressed in more depth. For instance, Buzzelli and Johnson (2002) and Ryan (1997) conveyed that because of the authority involved in the process of judging a student's work, grading is perceived as a morally demanding aspect of teaching. In addition, Terhart (1987, 1994) observed that teachers must be able to use such skills as self-control and self-monitoring in the classroom activities where external control is lacking. The aptitude of applying this set of skills (self-control and self-monitoring) assures that the manipulative use

of authority is diminished. By doing so, the teacher takes the role of a facilitator and not the role of the only source of knowledge in the classroom.

Since pupils through the end of high school (in most countries) are obliged to go to school, their relationship with the teacher(s) is not voluntary. Thus, the educational system should make sure that pupils are treated fairly and correctly in order to ground the obligation of children to attend school (Terhart, 1987; Soder, 1990; Bergem, 1993). The education system can ensure that by applying a rigorous selection of teacher candidates - by giving a sound teacher education - and by continually educating the in-service teachers. Interestingly, Sockett (1990) concluded that teachers must be trusted by the society to be able to carry out their duties within the educational system. The trust built between the society and teachers is a result of teachers' professionalism and maturity exhibited in all the educational processes. Another interesting conclusion was that, in most of the cases, the society and pupils show respect for a teacher when (s)he is a knowledgeable expert in the field and possesses strong pedagogical skills.

The responsibility of making the mission of education achievable lies mainly on teachers. Moreover, the mission regarding educating moral norms is molded by the teachers' approach when interacting with pupils. Hansen (2001b) pointed out that "any action a teacher undertakes in the classroom is capable of expressing moral meaning that, in turn, can influence students" (p. 286). Furthermore, teachers are usually seen as moral role models. Popham (1991) specified that educators' ethical obligations derive from their agreement of serving *in loco parentis*. "Educators... take on an ethical responsibility to serve as models of ethical behavior for children" (p. 13). In this sense, teachers consider that it is morally important to show respect towards themselves and their pupils (Campbell, 2000, 2003). By behaving and acting respectfully, the teacher shows to the students how to be respectful towards themselves as well as towards the others. Consequently, a strong bond between the teacher and students is built and the mutual trust is strengthened. This idea explains why the trust for teachers is manifested not only by society and parents, but also by pupils.

Alongside with respect, there are other factors that lead to teacher-student mutual trust, such as: consistency, unambiguousness, and the seriousness of dealing with one's work. Campbell (2000, 2003) specified that teachers should protect the pupil's dignity contributing to the building of mutual trust by excluding embarrassment, humiliation, and/or mockery of a pupil or anyone else in public. Furthermore, Colnerud (2006) suggested that many teachers have a

negative effect on pupils. The problem that Jackson, Boostrom, and Hansen (1993) identified is that there are teachers who do not realize the moral impact they exert and the importance of their actions on pupils. Usually, pupils analyze carefully teachers' reactions, facial expressions, and gestures. As a result, because of the lack of the self-control and self-monitoring skills, many teachers have an unwanted impact on pupils without perceiving it.

Beside the interactions and relationships that teachers have with their students - imposed by the profession itself - there are also the interactions that occur between the teachers, their colleagues, and parents. According to Colnerud (1997), due to these interactions, there are five norms that may create ethical dilemmas that teachers face in case of conflicts: interpersonal, professional, institutional, social conformity, and self-protection. Moreover, the author divided the interpersonal norm into five sub-norms: protection from harm, respect for integrity, respect for autonomy, justice, and veracity. The researcher also suggested that the moral conflicts that occur at the interpersonal level are considered to be the most abstract. However, all five norms conflict with each other and it is impossible to separate them. As a result, one may conclude that being a morally good teacher is difficult due to the multitude of interactions and personal factors that (s)he faces.

### **3.1 Ethics and classroom assessment**

Teachers play a central role in deciding upon assessment techniques and grades. Therefore, various researchers argue that it is crucial for teachers to be assessment literate (Impara, Plake, & Fager, 1993). In the context of education, assessment literacy relates to deep understanding of principles of fair assessment (Popham, 2004; Stiggins, 2002) by using various practices, theories, and philosophies that enhance teaching and learning (DeLuca & Bellara, 2013). Therefore, the use of multiple assessment tools is not only an indicator of fairness and assessment validity, but also an indicator of teacher's assessment literacy and broad knowledge.

Classroom assessment comprises the summative and formative assessment practices applied by teachers for measuring students' learning. According to Pope (2006), classroom assessment involves both morality and ethical issues. Ethics in the classroom guarantee the "quality and fairness of assessment". According to Green et al. (2007), the assessment practices that interrelate with ethics are: communication about grading, multiple assessment opportunities, bias/fairness, confidentiality, grading practices, test administration, and

standardized test preparation. These practices co-exist in a close interdependence. For instance, the lack of confidentiality influences the fairness of the assessment and biasness impacts grading practices. In regards to the grading practices, Tierney (2015) suggested that grade alteration (the decrease or increase of the grade) is a widespread phenomenon that persists in educational systems, but it becomes evident only in cases when it is not fair for students. Nevertheless, teachers' grading practices may be influenced by such factors as: subject area (Duncan & Noonan, 2007; McMillan, 2001; Resh, 2009), level taught (Guskey, 2009; Randall & Engelhard, 2009), teachers' gender, subject proficiency, and perception of the subject matter (Biberman-Shalev, Sabbagh, Resh, & Kramarski, 2011). Other studies show that even such factors as student's gender and socioeconomic background may also have an impact on teacher's assessment decisions (Elwood, 2006; Klapp Lekholm, 2011; Organisation for Economic Co-operation and Development, 2012; Peterson & Kennedy, 2006). Moreover, the study conducted by Tierney (2015) conveyed other three reasons that may lead to grade alteration: compassion; access to opportunity; and desire of teachers to give life lessons. As a result, teachers' judgement and, hence, their assessment could be influenced consciously or unconsciously by many other factors than their own beliefs. In this context, the teachers' ability to monitor the ethicality of their own judgement and decisions as well as the ability of following the ethical guidelines for assessment is important.

Multiple studies also show that teachers consider the non-achievement factors as effort when assigning grades (Cox, 2011; Duncan & Noonan, 2007; Hunter, Mayenga, & Gambell, 2006; Sun & Cheng, 2014). Interestingly, Brookhart (1993) conveyed that effort usually is not considered equally for every student. Thus, the higher-achieving students' grades would not be lowered by the lack of effort, but the lower-achieving students' grades would be inflated because of effort. In this sense, there is the perception that higher-achieving students do not need to put much effort in order to complete the tasks, whereas lower-achieving ones may need more effort and time. Thus, effort is considered as a decisive factor in some marginal cases - when it makes a difference whether the student gets the satisfactory mark in order to pass the grade.

The multiple assessment methods principle, discussed by Green et al. (2007) and McMillan (2000), has an interesting implementation in Finland. In accordance with this principle, the Finnish National Core Curriculum asks for a versatile assessment and that the educational system develop pupils' self-assessment and peer-assessment skills. Teachers may use such practices in addition to the assessment techniques that (s)he may apply personally (e.g., tests,

inquires, essays). As a result, by applying different evaluation methods, teachers receive more compelling evidence regarding the student's achievement and, in this way, make the overall assessment more reliable and fair.

The National Core Curriculum also requires that pupils be informed about the assessment objectives and procedures in advance. This ensures that the trust, which "is a central tenet of ethical professional practice for all members of the teaching profession" (OCT, 2013 p. 3), is built and maintained. Mutual trust is further enhanced by the aspect of confidentiality of the assessment. Consequently, by treating each student's work with consideration and by addressing each student learning particularities in a private manner, the teacher contributes greatly to the strengthening of the teacher-student bond.

It is generally suggested that ethical classroom assessment activities be in line with the professional guidelines for teachers. Green et al. (2007) recommended that, in order to provide fair and ethical assessment, teachers should take classroom assessment courses that explore in more depth the ethical issues and various factors that influence ethical behavior. Importantly, pre-service teachers should be instructed how to deal with the ethical assessment issues that might occur in a classroom setting. Nevertheless, research shows that beginning teachers have little knowledge about ethical assessment, which may be affecting their relationship with the pupils (DeLuca & Bellara, 2013). In addition, Airasian (1984) expressed that there are two areas that teachers focus on during classroom assessment: academic achievement and social behavior. The social behavior being of a greater importance in elementary school. Moreover, there are such factors as pupil's temperament and personality that teachers also assess when it comes to social behavior. Beside the effort factor, the behavior also influences greatly teachers' assessment decisions. In this sense, it is decisive for pre-service teachers to also be involved for a reasonable period of time in practice activities while being students. Such practice gives them the opportunity to learn to be more critical towards their own judgements and to develop their reflective abilities in accordance the ethical issues discussed above as well as the social aspect of behavior.

Finally, Stiggins and Conklin (1992) conveyed that the nature of assessment may also be strongly influenced by the roles of students set by the teacher, the teacher's expectations, and the desired teacher-student interaction. Based on these observations, I can conclude that the nature of classroom assessment may differ from one class to another depending on teacher's beliefs and teacher-student interactions. Furthermore, Cizek, Fitzgerald, Shawn and Rachor

(1995) summarized that “many teachers seemed to have their own individualistic values and beliefs about teaching” (p. 160). Thus, it is apparent that the ethical issues related to classroom assessment also differ from one teacher to another.

## **4. RESEARCH CONTEXT - FINNISH EDUCATION**

Since the present study was carried out in Finland, it is important to understand in more detail the context of this research. In this chapter I discuss the guidelines regarding assessment presented in such educational policies as the National and Local Curriculum, the main principles of the Finnish educational system, and the qualifications of the pre-service and in-service teachers. All these factors, the educational policies, the main principles of education, the education of pre-service and in-service teachers, influence greatly teachers' ethical and professional understanding and, therefore, the findings of this study.

### **4.1 The Finnish National Core Curriculum, Local Curriculum, and Assessment**

The normative part of the basic education in Finland consists of the Basic Education Act and Decree, Government Decrees, the National Core Curriculum, the Local Curriculum, and the Annual Plans of Individual Schools. As mentioned in the National Core Curriculum for Basic Education, the main purpose of assessment is "to guide and encourage learning and to develop pupil's capacity for self-assessment" (p. 49). In order for the assessment to achieve its goals, it must meet certain key characteristics. The assessment must support and promote pupils' motivation and eagerness to learn as well as their enthusiasm to participate actively in the learning process. The assessment must serve as a source of information for teachers' instruction planning, it must be versatile, fair, and ethical.

The general principles of assessment, stipulated in the Core Curriculum specify that there are three main areas that the assessment must be focused on: pupil's learning, working aptitudes, and behavior. The assessment of learning focuses on the pupil's progress that is evaluated in comparison to earlier achievements and the set objectives. The achievements of a specific pupil are not comparable to the achievements of other pupils. These aspects ensure the conformity of the measurement of the academic achievement with the set objectives. Moreover, they ensure the confidentiality aspect and the individual approach of the assessment.

The development of pupils' working skills is one of the central goals of primary education. The working skills include the skills to plan, regulate, and evaluate their own work as well as to act responsibly and to collaborate constructively when interacting. The assessment of the working skills is part of the grade and it is based on each subject objectives and goals related to the development of the working abilities. According to the guideline offered by the



Curriculum, the effort can be considered as a working skill and, thus, as part of the grade. However, because it is not clearly mentioned that effort is or is not considered as a working skill, this guideline leaves room for different interpretations and ambiguousness.

The assessment of behavior does not influence the grade or the verbal assessment given for a subject, it is rather treated as a separate entity. In this regard, teachers must make sure, from the start, that the objectives and the assessment criteria for behavior is known and understood by pupils. This aspect of the assessment offers a clear ethical support for teachers. Thus, the issue of considering behavior as part of the grade assigned for a subject should be minimal.

The interaction between teachers and students is an important factor that contributes to a sound assessment. Therefore, the feedback given to pupils is of a major importance. The key features of a constructive feedback are: it must be focused on the pupil's successes and progress; it must be in accordance with the pupil's age and capabilities; and it should not take into account the pupil's personality or temperament. In the Local Curriculum of the city of Oulu it is suggested that the feedback given by teachers must convey information not only regarding the successes and the progress of the pupils, but also the failures. The experiences of failure must be treated in a respectful manner and must promote learning. By following these instructions, teachers are able to build a strong student-teacher bond as well as trust, which are important factors for learning and development.

Moreover, since the school and the teachers' feedback have a considerable effect on students' self-efficacy beliefs, it is suggested that a wide range of assessment techniques should be applied. Thus, teachers must collect information regarding students' progress in different areas of learning and in different learning situations by considering students' various ways of learning and working. That is why it is imperative that teachers are assessment literate and able to apply multiple assessment techniques.

During the studies, pupils need to be guided to observe their own and their peers working abilities and the factors that influence the outcomes, thereby developing their self- and peer-assessment skills. In addition, teachers must make sure that the goals and objectives of the assessment are understood by the pupils and the discussions about pupils' progress and achievements of the goals support the development of self-assessment skills. In Oulu, assessment is developed from the aspect of the learning process. The goals and the assessment criteria as well as the continuing assessment with multiple methods are highlighted. From the beginning, students are informed about the objectives of the learning process and how they

can self-affect their own learning process by setting their own goals. Accordingly, setting common goals for both teacher and students and making the criteria clear and common for all is part of the assessment culture.

Strong cooperation between home and the school is also considered part of a good assessment. Students and caretakers can take part in the learning and assessment practices by participating in analytical discussions, thus building mutual trust. It is of major importance to ensure a sound communication between teachers, students, and the caretakers. Moreover, a deeper understanding of all the particularities of the evaluation are perceived by all three participants (teachers, pupils, and caretakers) by practicing joint assessment analysis and open conversations.

As specified in the Core Curriculum, there are two types of assessment in the basic education in Finland: assessment during the studies and final assessment. The assessment during the studies refers to the feedback and assessment applied till the final assessment. The aim of the assessment during the studies is to guide and support learning as well as to develop self-assessment and peer-assessment skills. The assessment notes, reports, and discussions about the progress and development of the pupil represent an important part of the assessment during the studies. The assessment during the studies is formative and summative. The final assessment is the assessment carried out at the end of the school year and it represents an overall assessment of the progress and performance of the pupil. It also includes information regarding the decision to promote the pupil to the next grade or to retain him/her in the same grade. The school year report can contain both a verbal assessment and a numerical grade.

## **4.2 The principles of the Finnish educational system**

There are several distinctive principles that the Finnish educational system is based on. The first principle is the one of equity. It has been one of the main principles of the Finnish education policy since the late 1960's. This principle is at the basis of all the study cycles: from early childhood to higher education and adult education. The main goal of the Finnish education policy is to assure equal opportunities for all citizens to receive education, regardless of such factors as age, gender, origin, mother tongue, or financial situation. Therefore, education in Finland is considered as one of the fundamental rights of all citizens (Kumpulainen & Lankinen, 2012; Niemi, 2014; OECD, 2006).

The second principle regards the National curriculum system, the local freedom and the aspect of responsibility (Niemi, 2015). The educational system in Finland is decentralized, thus the local educational authorities are in charge for the provision and quality of education. The national curriculum system provides values and objectives for each educational level while the local educational authorities and schools have the autonomy to organize education and implement the core curriculum (Halinen & Holappa, 2013). As a result, teachers have the freedom to choose the way they teach, the learning materials they use, and the assessment methods they apply. Finally, together with their principals, teachers bear the responsibility for the local school-based curriculum. The freedom of teachers and the responsibility that comes with it also bears an ethical connotation. By choosing the learning materials, teaching methods, and the assessment tools, teachers have to plan carefully their work and consider all possible risks.

The third principle of the Finnish educational system is related to the support of special needs. This translates into the pillars of basic education system: inclusive ideology and the strategies of support for different learners. Learners must be supported from the early educational stages so that they overcome the learning difficulties as they progress. In this sense, the teachers' responsibility is to observe and recognize the needs of a pupil for special support (Niemi, 2015). Those pupils that need more aid than the standard aid offered to all students are provided with intensified support that is organized in accordance with an individual learning plan (Vainikainen, 2014). Usually these pupils are provided with support in their own classes and only in unique cases are transferred to special classes or schools. Thus, special needs teachers, social workers, nurses, and school psychologists collaborate with regular teachers to make study-plans for special needs pupils.

The fourth principle is related to national evaluation system. The Finnish education policy is different than the policy applied in other countries because the Finnish education does not include national standardized tests (Niemi, 2014). Sahlberg (2015) pointed out that, in teachers' view, the standardized tests would lead to "teaching to test." So, the lack of standardized tests gives teachers the possibility to support more carefully the student learning by using different teaching and assessment methods that are in line with students' needs. It also leads to the absence of the ethical issues related to it, that persist in other countries (e.g. USA)

### **4.3 Pre-service teacher education in Finland**

Finland represents an example of a country where the education provided to the pre-service teachers is successful. This is supported by the fact that Finland has been among the top-performing countries in terms of educational achievement in numerous international comparisons (OECD, 2003; 2010; 2014). Such results are based on the requirement that Finnish teachers must complete both the Bachelor's (three years) and Master's (two years) studies.

The main distinctive feature of the Finnish teacher education from many other countries is that students conduct research in addition to their regular academic studies (Jyrhämä & Maaranen, 2012; Niemi & Jakku-Sihvonen, 2006). All the pre-service teachers take part in research seminars or projects, learn about various research methods, and write Bachelor's and Master's theses. In this way, the prospective teachers learn about the most innovative research in the field of teaching and learning and, thus, gain scientific literacy. They are trained to be analytical and critical thinkers, as well as to develop systematically their teaching and learning environments (Niemi & Nevgi, 2014). Due to this educational system, the Finnish pre-service teachers are very well equipped with a strong set of skills when graduating.

The Finnish universities provide different education programs for various categories of teachers (e.g., elementary teachers, subject related teachers for lower and upper secondary schools, special needs teachers, school study counselors, and vocational and adult teachers). The pre-service teachers can choose the study modules and design their own study-plans according to their career path in the educational system. Nevertheless, every pre-service teacher must complete the basic criteria of the academic studies and pedagogical studies with teaching practice. In addition, teachers have the possibility to continue their studies after graduation in order to advance their qualifications for specific subjects, pedagogy skills, or multidisciplinary learning. Some teachers also have the possibility to widen their knowledge by continuing with doctoral studies. Typically, the professors in teacher-training universities follow this learning path. This factor plays an important role since the pre-service teachers are instructed and guided during the practice periods by skilled and knowledgeable professors. An intrinsic aspect of the teacher-education programs is that theory and practice are interrelated. Jyrhämä (2006) concluded that there are three stages of the teaching practice: orientation, intermediate practicum, and the advanced practicum. It is the fundamental role of

professors in the teacher-training schools and the local partner schools to supervise and support the student's critical thinking and reflection during these practice periods.

#### **4.4 In-service teacher education in Finland**

The autonomy and freedom to make decisions are the main factors that in-service teachers are entitled to according to the Finnish educational system. At the same time, they are expected to put the maximum effort in enhancing students' learning outcomes and overall well-being. Importantly, it is required from teachers to be ethically committed to the profession and prepare students for lifelong learning. In this sense, the values of the teaching profession development highlighted by The Finnish National Board of Education (FNBE) (Rajakaltio, 2014, p. 5) are: life-long learning; knowledge and research based orientation; effectiveness; the future needs and competences in education. It is also important to mention that the teacher profession is highly valued and respected by the Finnish society. This is a result of the rigorous prospective teacher selection and the sound qualification and training the teachers go through. As a result, the society trusts, appreciates, and respects teachers for their valuable work.

According to Tali's review (OECD, 2013), there are few educational programs for in-service teachers in Finland, in comparison with other countries. This is the result of the school-based development projects that are not usually considered as traditional in-service training. The local providers (municipalities or cities) are responsible for the access to and quality of the educational trainings for in-service. Officially, there are three compulsory in-service training days for each teacher each year. However, in addition to these, the schools may organize internally different seminars and trainings.

The majority of universities have education centers for in-service teachers' training. These education centers provide projects with a more extended development process instead of short courses. The purpose of these projects is to give the opportunity to teachers to reflect on their own work, to learn new skills, and to share ideas with colleagues. Moreover, research is a fundamental part of the teacher education in Finland; therefore, the trainings maintain the continuation of the research-based and research-informed vector in teacher's careers. These activities help teachers to learn about the most recent research related to the subjects they teach and to the pedagogical aspects. Lastly, the overall goal of the in-service training is to

positively affect the motivation and learning of students as well as the professional growth and well-being of teachers.

## 5. AIM AND RESEARCH QUESTION

The aim of this study was to compare the pre-service and in-service teachers' perceptions about the ethicality of the classroom assessment practices. Also, it set to examine whether such factors as work experience (characteristic for in-service teachers) and the recent familiarization with the up to date research regarding teaching skills (characteristic for pre-service teachers) make a difference in the way teachers perceive the ethicality of certain classroom assessment practices. Taking into account these differences between pre-service and in-service teachers, it was important to detect whether all the teachers follow the same guidelines when it comes to the ethicality of the classroom assessment. The pre-service teachers were involved in this study because they represent the next generation of teachers who will have a great impact on education. Thus, I consider important to be aware of whether there are substantial differences related to the ethical principles and guidelines followed by teachers in both groups.

In this study, I explored the following research question: *what is the degree of agreement between the pre-service and in-service teachers regarding the ethicality of the classroom assessment?* The strong agreement between the groups of teachers would suggest that they follow the similar ethical principles and guidelines while performing the classroom assessment. On the other hand, a weak agreement between the groups of teachers would indicate that they do not follow the same guidance or similar ethical principles while performing the classroom assessment. This would denote the need for further research and dialog.

## **6. METHODOLOGY**

The present study is a quantitative research based on a 40-question survey. In this chapter of the thesis, I discuss the description of the participant pool that consists of pre-service and in-service teachers. I continue this discussion by presenting the instrument used in this study and its question categories, followed by a detailed report of the procedures that the questionnaire was exposed to (translation and adaptation). Finally, I convey the detailed description of the data analysis procedures and the research method of the study.

### **6.1 Participants**

In the present study, I employed a convenient sampling procedural strategy. The participant pool for this study consisted of 45 pre-service and 34 in-service teachers. The pre-service teachers were predominantly women (73%) with an average age of 26. In addition, the pre-service teachers were 1<sup>st</sup> (21%) and 2<sup>nd</sup> (79%) year students of Primary Teacher Education Master Degree Programs of the University of Oulu. Importantly, the students were engaged in courses of Ethics and/or Ethics and Education as well as courses related to the Evaluation processes and procedures during their undergraduate and graduate studies.

The in-service group included primary school teachers of Linnanmaa, the University of Oulu Teacher Training School, and Koskela, the University of Oulu Teacher Training School. The average age of this sub-sample was 48 years, with 74% being women who had an average teaching experience of 23 years. The grade level taught by these teachers ranged from the first to the sixth grade. Importantly, participation in the study was voluntary and all the participants consented to release their information.

### **6.2 Instrument**

The scenarios, that represent the instrument of this research were designed and developed by Green et al. (2007) as well as Johnson et al. (2008) and later adapted by Liu et al. (2016). For the current study, I selected scenarios from both the original and the adapted versions. I contacted the leading author of the original studies and obtained official permission to readapt and translate the scenarios to Finnish. The translation was done and then revised by a group of teachers and professors who are Finnish native speakers. Moreover, I deleted or modified those scenarios that did not reflect the Finnish primary school context. For example, the



original scenarios, among other, comprised the category *Standardized Test Preparation*; however, because there are no standardized high-stakes tests in the Finnish educational system, but rather only one that is taken by those that aim to enter higher education (Hendrickson, 2011), this category was deleted.

In addition, because the original scenarios were developed in the United States and reflect the assessing characteristics and processes that are normally applied there, the scenarios related to state level or district level test administration were either deleted or revised and changed to classroom level testing to better fit the Finnish context. The final version of the survey consisted of 40 questions (*see Table 1*). Taking into consideration the original topics that the scenarios refer to (per Green et al., 2007, and Johnson et al., 2008), I divided them into seven categories from which I included only six in the current study. The six categories considered in this study were: 1) Biasness / Fairness, 2) Communication about Grading, 3) Confidentiality, 4) Grading Practices, 5) Multiple Assessment Opportunities, and 6) Test Administration. For the detailed list of scenarios within each category see Tables 3 – 8 in the Appendix.

The first part of the survey consisted of four demographic questions - gender, age, study program, and year of graduation - for pre-service teachers as well as four demographic questions - gender, age, teaching experience, and grade level taught - for in-service teachers. The 35 scenarios represented the second part of the survey. The scenarios illustrate various classroom assessment methods which the respondents answered as being either “*ethical*” or “*unethical*” in accordance to their understanding of the ethicality of the procedure described in the scenario. At the end of the survey, there was one optional question which asked for the participant’s opinion regarding the study or a particular scenario they found difficult to answer to.

The survey was carried out during the autumn semester of 2016 and was delivered in two ways: online and on paper. The online version was distributed to the in-service teachers through *Webropol*, an online survey tool that is used by the researchers of the University of Oulu. In order to participate, the respondents had to click on a personal link sent to them by e-mail. The paper version of the survey was handed out to pre-service teachers during a class and then collected after they filled it in.

Table 1

*Total number of questions and scenarios*

Category	Number of Questions / Scenarios
Demographic questions	4
Bias/Fairness	5
Communication about Grading	5
Confidentiality	4
Grading Practices	12
Multiple Assessment Opportunities	6
Test Administration	3
Optional Question	1
Total	40

### **6.3 Data analysis and the Research Method**

As explained above, this study is a continuation of the work of Green et al. (2007), Johnson et al. (2008), and Liu et al. (2016). The research method of the study is quantitative. Consequently, the data was analyzed by using mathematical and statistical methods. Since there were two groups of respondents - pre-service and in-service teachers - I first analyzed what was the degree of agreement within each group and then between the groups. In order to analyze the data of this study, I used the IBM SPSS Statistics 24 statistical software.

Following the same procedure as in the previous studies, I employed the frequency test with the rule that if at least 80% of the teachers answered the respective scenario as either ethical or unethical then this indicated strong agreement among participants for the respective scenario. In addition to this, I added one more area of agreement that included the range between 70% and 80% (*see Table 2*). In the previous studies, if 50% of teachers labeled a practice as ethical and 50% of teachers labeled it as unethical, the researchers considered this as the highest level of disagreement between the two groups. Finally, those items with percentages ranging from 30% to 70%, in the study of Liu et al. (2016), and from 50% to 70%, in the study of Green et al. (2007), were considered in the disagreement area. In the present study, I have modified the range for the disagreement area slightly. I considered that the scenarios with percentages ranging from 55% to 70% indicated disagreement and those in the range from 45% to 55% indicated strong disagreement.

Table 2

*Areas of Agreement and Disagreement*

	Strong Agreement Area	Agreement Area	Disagreement Area	Strong Disagreement Area
Frequency of responses either “ethical” or “unethical”	>=80%	70% - 80%	55%- 70%	45%-55%

The reason for changing the ranges of percentages was to provide more clarity about the level of disagreement. A split of 60-40 provides a weaker level of disagreement than a split of 55-45. As a result, I decided to treat the frequencies in the small vicinity of 50% as strong disagreement, whereas those that extended beyond 55% till 70% as a weaker disagreement. For instance, if 49% of pre-service teachers considered that a teacher always checks the name of the student whose essay test he is grading as ethical and 51% consider the procedure as unethical then this scenario fell in the strong disagreement area.

In order to examine the differences between pre-service and in-service teachers’ perceptions about the ethicality of the procedures described in the scenarios, I applied the Fisher’s Exact Test. I chose this approach because the current study used a small sample, which would not allow a high level of accuracy for the chi-squared test (Field & Field, 2009), which was used in earlier studies. Based on the above considerations, I expressed the null hypothesis in the following way: *there is no difference between pre-service and in-service teachers’ perceptions about the ethicality of the assessment methods.*

In each of the six categories, within every group of teachers, I firstly analyzed the scenarios that represented a substantial difference in perceptions about the ethicality of the assessment practices. I also discuss the scenarios that show strong agreement among the teachers of the same group. Finally, based on the information received from the analysis within each group of teachers, it was possible to investigate and compare the patterns between groups.

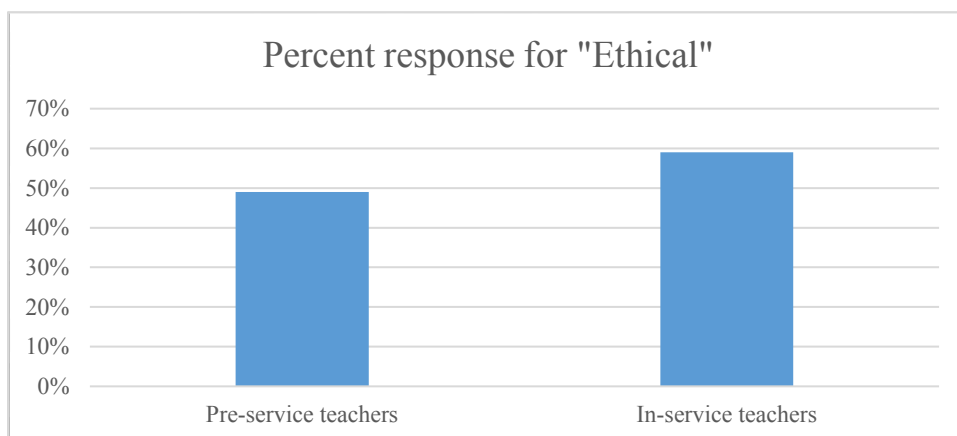
## 7. RESULTS

I report the main findings of the research in this chapter. Firstly, I present the results within each group of teachers for each category. The main findings shown by both groups of participants are displayed in the charts. Secondly, I provide a concise summary of the results. Lastly, I discuss the scenarios that indicate a significant statistical difference between the two groups of teachers.

### 7.1 The degree of agreement within each group of teachers

#### 7.1.1 Bias / Fairness

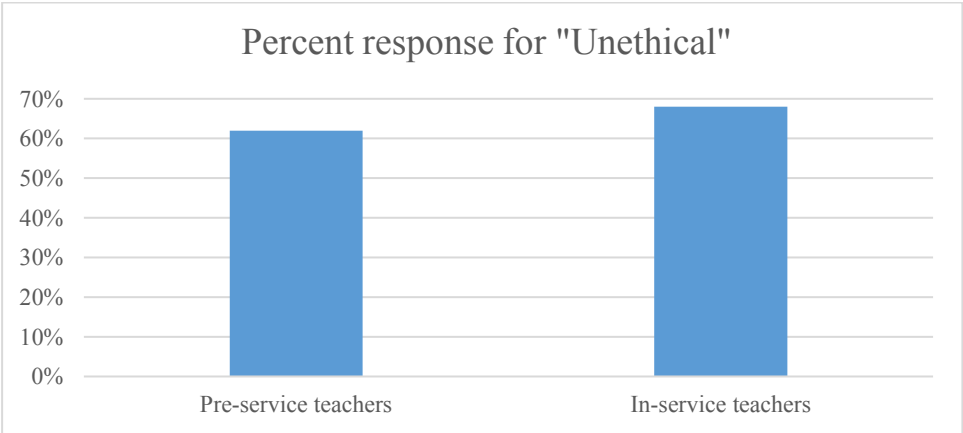
The “Bias / Fairness” category consists of 5 scenarios. The perceptions about the ethicality of the procedures described within both groups of teachers were divided for items 8, 15 and 26. These items fell either in the “strong disagreement” or “disagreement” areas. In item eight, 49 percent of pre-service teachers and 59 percent in-service teachers considered whether a teacher always checking the name of the student whose essay test he/she grades as ethical (*see Figure 1*). Based on these results, the pre-service teachers fall within the strong disagreement area, whereas the in-service teachers fall in the weaker disagreement area.



*Figure 1.* Answers for item 8: “A teacher always checks the name of the student whose essay test he is grading.”

In item 15, 62 percent of pre-service teachers and 68 percent of in-service teachers rated that when an elementary teacher addresses only students’ strengths when writing narrative report cards to enhance student’s self-esteem as unethical (*see Figure 2*). According to these results, both the pre-service and in-service teachers fall in the same area of weaker disagreement. In item 26, 53 percent of pre-service teachers and 44 percent of in-service teachers rated whether a teacher who knows that a student had a bad week because of problems at home bumps the student’s participation grade up a few points to compensate for his bad score on a quiz as ethical (*Figure 3*). The results of this item show a similar pattern as those in item eight, in which the pre-service teachers’ responses fall in the strong disagreement area, whereas the in-service teachers fall in the weaker disagreement area.

The analysis of the other 2 items within this category reveals that both groups of teachers had a strong agreement in their perception of ethicality (*see Table 3 in the Appendix*). Both groups considered that assigning very few card grades with a maximum grade because of the belief that students’ work is rarely perfect - item 4 - is unethical (the frequencies were 84 and 82 percent for the pre-service and in-service teachers, respectively). Similarly, 80 percent of pre-service teachers and 88 percent of in-service teachers labeled as ethical the situation when a teacher changes one student’s course grade from a 9 to a 10 because tests and papers showed that the student had mastered the course objectives even though he had not completed some of his homework assignments.



*Figure 2. Answers for Item 15: “To enhance self-esteem, an elementary teacher addresses only students’ strengths when writing narrative report cards.”*

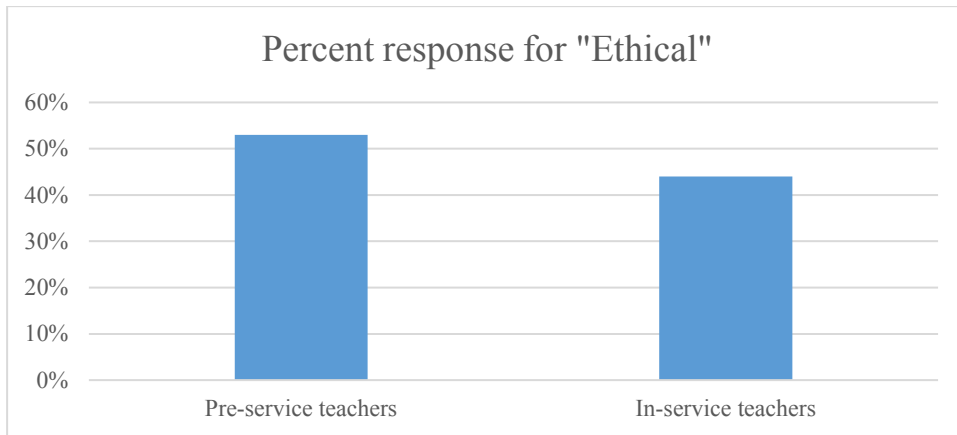


Figure 3. Answers for Item 26: “A teacher who knows a student had a bad week because of problems at home bumps the student’s participation grade up a few points to compensate for his bad score on a quiz.”

### 7.1.2 Communication about Grading

In the content category “*Communication about Grading*”, the perceptions of both groups of teachers were divided just for item 11. Interestingly, in the case of pre-service teachers the perceptions were strongly divided, whereas in the case of the in-service teachers the ethicality perceptions on this item were less divided. Thus, the practice of a teacher who uses always for the final test a few surprise items about topics that were not on the study guide was considered ethical by 49 percent of pre-service teachers and by 65 percent of in-service teachers (see Figure 4).

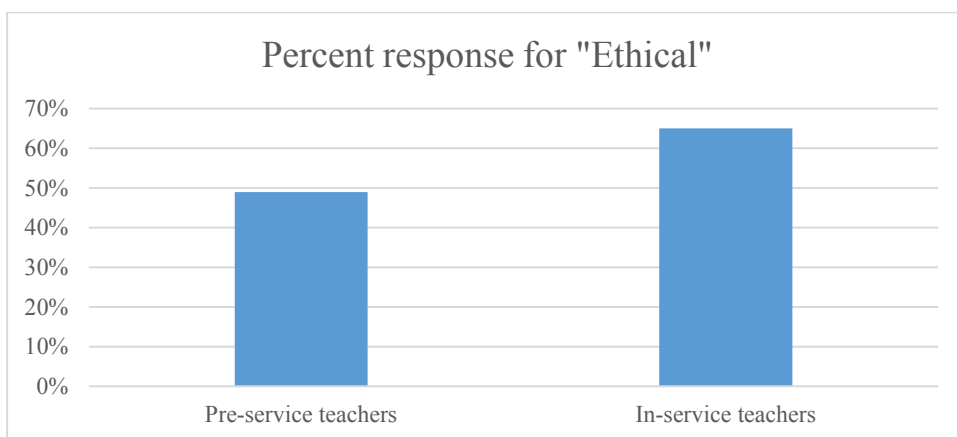


Figure 4. Answers for item 11: “For the final test, a teacher always uses a few surprise items about topics that were not on the study guide.”

In the case of the other four items of this category (items 9, 16, 22, and 27), the degree of agreement regarding the ethicality of the procedures described ranged between 96 percent and 100 percent for the pre-service teachers group and between 91 percent and 97 percent for the in-service teachers group (see Table 4 in the Appendix). According to these results, there seems to be small difference in the ethicality perceptions for each scenario between the two groups.

### 7.1.3 Confidentiality

The “Confidentiality” category consists of 4 Items. Item 17 fell in the area of disagreement for both groups of teachers. In this item, 65 percent of pre-service teachers and 38 percent of in-service teachers considered the practice ethical (see Figure 5).

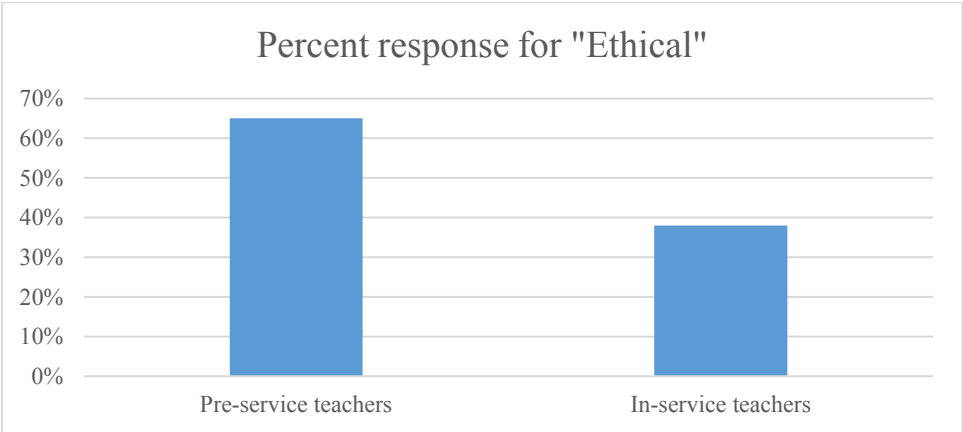
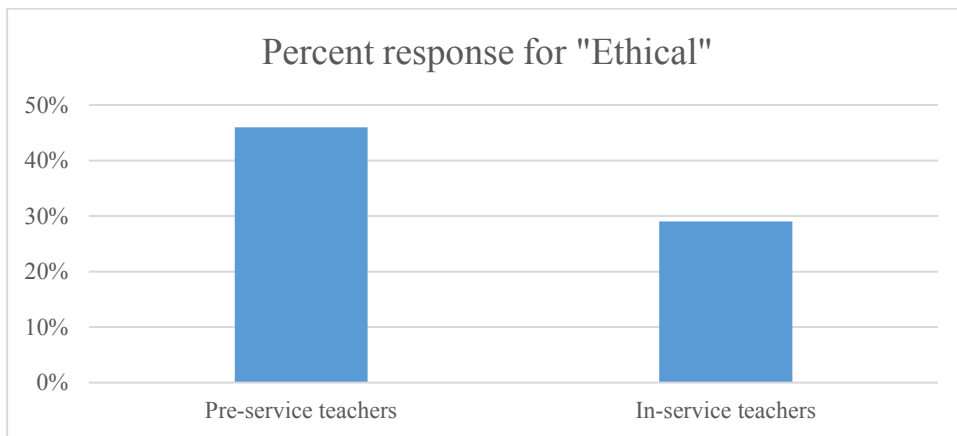


Figure 5. Answers for item 17: “At a high school, mathematics teachers do not grade all tests or homework. Instead, teachers let students grade each other’s math homework and then share the results in groups. Teachers believe teamwork will help students learning.”

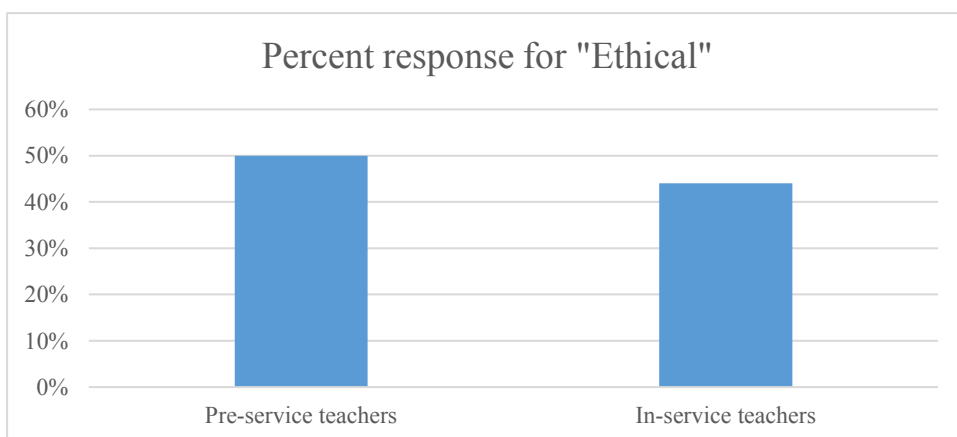
In the case of the other three items, the degree of agreement ranged from 84 percent to 100 percent for the pre-service teachers group and from 94 percent to 100 percent for the in-service teachers group (see Table 5 in the Appendix). The results of this category suggest that there is an overall level of agreement on the ethicality of the respective practices, except for item 17 in which the perception of ethicality appears to change drastically between the two groups.

#### 7.1.4 Grading Practices

As shown in *Table 6 in the Appendix*, the pre-service group of teachers agreed on nine out of twelve items, whereas the in-service group of teachers agreed on ten items. The pre-service teachers strongly disagreed in the case of items 10, 12, and 38, whereas the in-service teachers had a weaker disagreement in the case of items 12 and 38. In item 10, 46 percent of pre-service teachers and only 29 percent of in-service teachers found the practice ethical (*see Figure 6*). In item 12, 50 percent of pre-service and 44 percent of in-service teachers found the practice ethical (*see Figure 7*). Finally, in item 38, 46 percent of pre-service teachers and 38 percent of in-service teachers found the practice ethical (*see Figure 8*).



*Figure 6.* Answers for item 10: “A teacher uses student peer ratings as 40% of the grade on an oral report.”



*Figure 7.* Answers for item 12: “A teacher weights homework heavily in determining grades.”



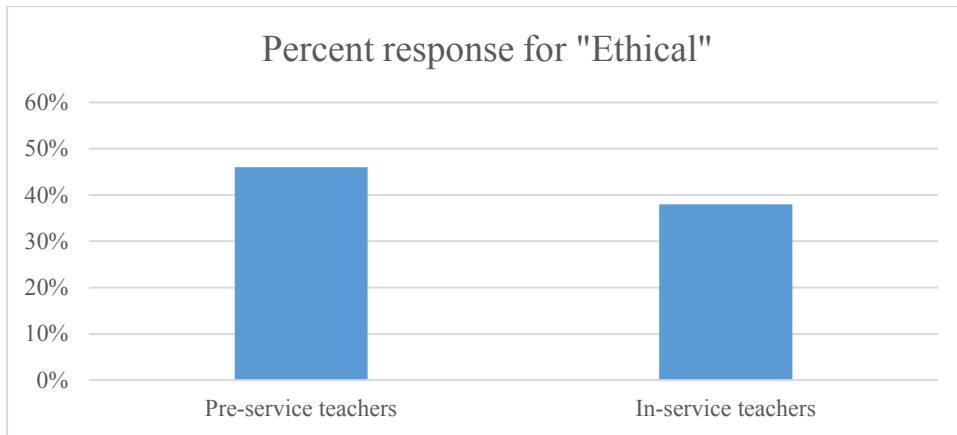


Figure 8. Answers for item 38: “A teacher lowers grades for disruptive behavior.”

Additional variation appears in items 24 and 34. The pre-service teachers strongly agreed upon the ethicality of the practice in these two items (91 and 96 percent, respectively), however the in-service teachers answered in a proportion of 76 and 79 percent, respectively (see Figures 9 and 10). The responses for item 29, although they fall in the strong agreement area, appear to be at the limit: 98 percent of the pre-service teachers perceived this scenario as ethical, whereas only 82 percent of the in-service teachers agreed upon this practice’s ethicality (see Figure 11).

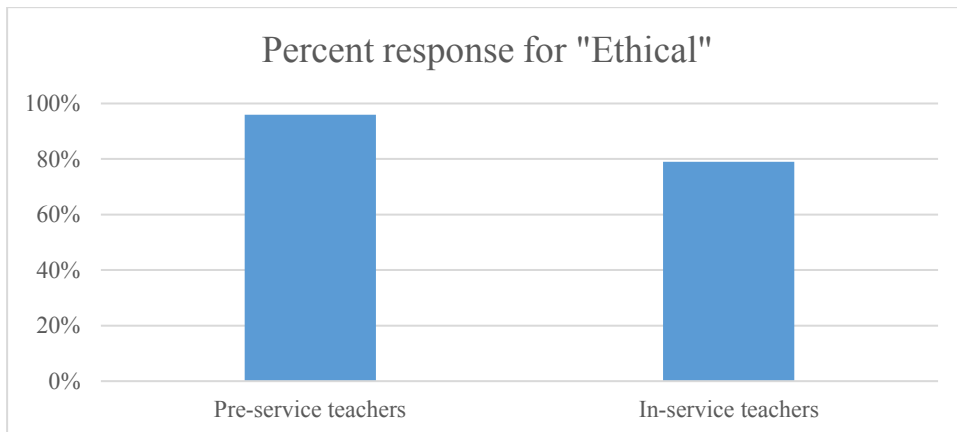


Figure 9. Answers for item 34: “A teacher considers student effort when determining grades.”

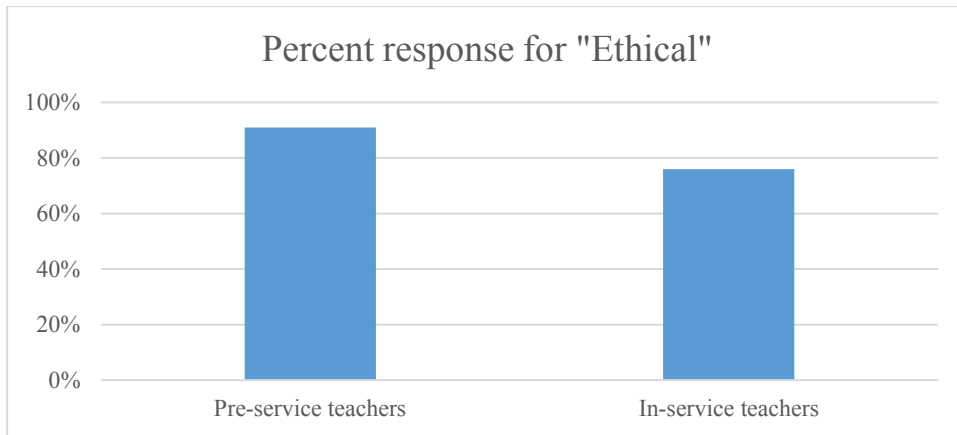


Figure 10. Answers for item 24: “When a child is caught cheating on a test, rather than assigning a zero, the teacher gives the student an alternate version of the same test.”

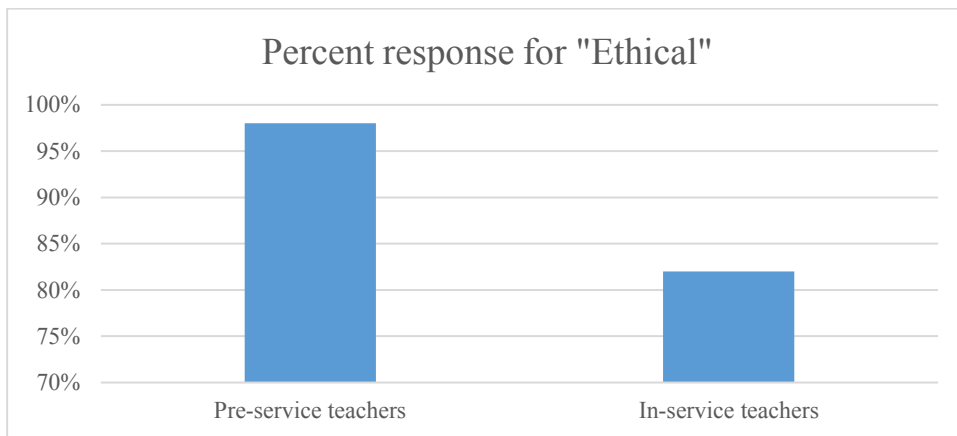


Figure 11. Answers for item 29: “A teacher considers students’ growth in assigning grades.”

A similar observation was in the case of item 32: 74 percent of the in-service teachers considered this practice unethical, whereas 85 percent of the in-service teachers viewed it as unethical. The final area of variation appears to be in item 37. In this item, 78 percent of pre-service and 94 percent of the in-service teachers considered the situation as unethical (*see Figure 12*). The responses for the other three items (6, 18, and 35) fell in the area of strong agreement for both groups of teachers.

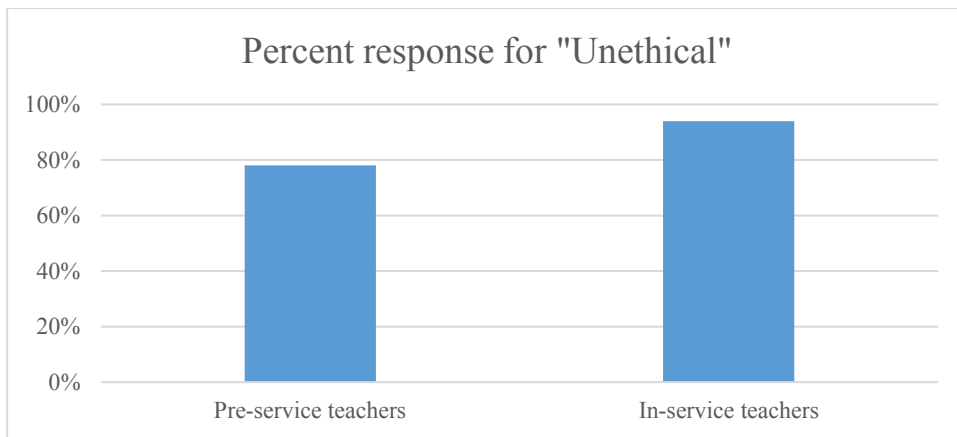


Figure 12. Answers for item 37: “A teacher lowers grades for late work by one number grade for each day.”

### 7.1.5 Multiple Assessment Opportunities

This category has six scenarios. In both groups, the responses of the participants fell within the agreement area for five of the scenarios and in the disagreement area in the case of one scenario (although it was not the same scenario across the two groups). Interestingly, items 30 and 33 appear to switch the level of agreement across the two groups. In item 30, the responses fell in the agreement area for the pre-service teachers (71 percent found the situation as unethical), whereas for the in-service teachers they fell in the disagreement area (68 percent found the situation as unethical). In the case of item 33, however, 59 percent of pre-service teachers and 71 percent of the in-service teachers found the scenario unethical (*see Figures 13 and 14*).

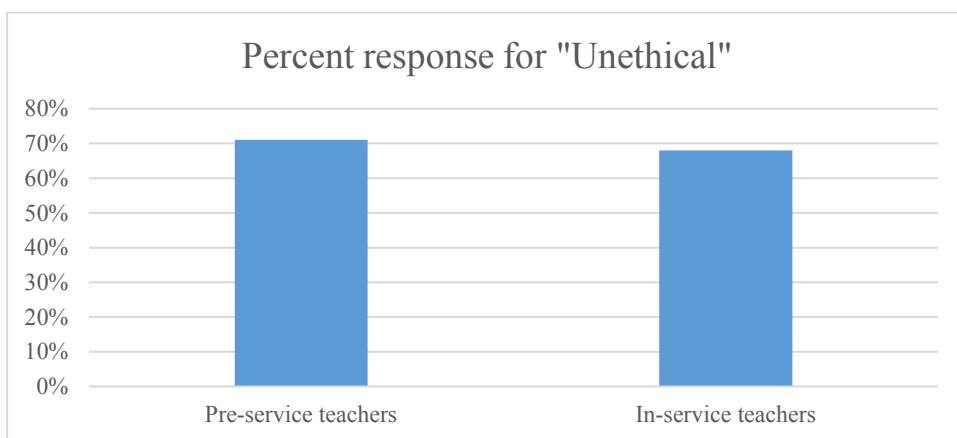


Figure 13. Answers for item 30: “A second-grade teacher uses observations as the sole method to assess what students have learned.”

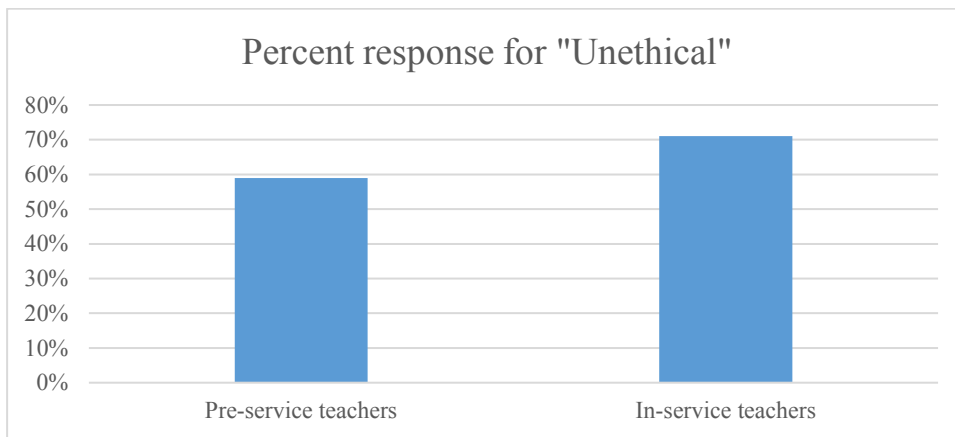


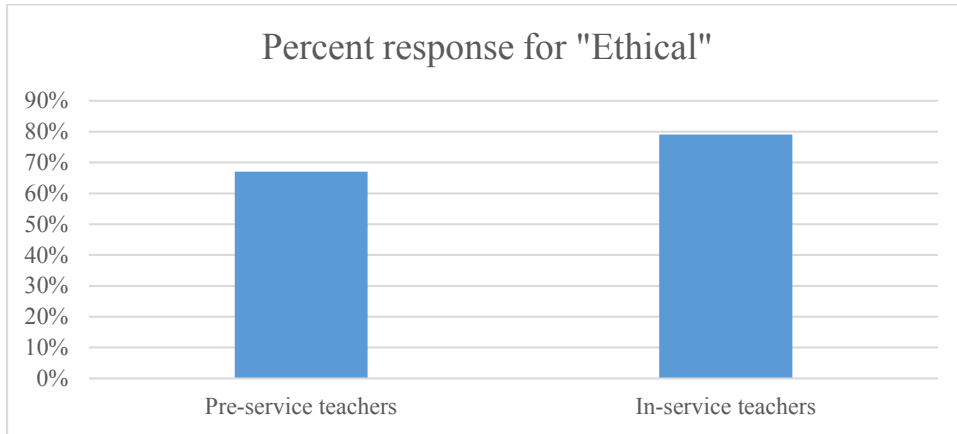
Figure 14. Answers for item 33: “While assessing group work, a teacher gives the same mark to each group participant.”

Most of the pre-service teachers (91 percent) found the procedure depicted in item 13 as ethical, whereas only 74 percent of the participants from the in-service group found it ethical. In both groups, over 90 percent of the participants considered the use of observational checklists, anecdotal notes, and interviews when assessing students as ethical (item 7). Also, both groups agreed (84 percent for pre-service teachers and 91 percent for the in-service teachers) that the practice of a high school social studies teacher to base students’ final semester grade on 2 multiple-choice tests is unethical. Finally, both groups of teachers completely agreed (100 percent) on the ethicality of the use of many types of assessment: multiple-choice tests, essays, projects, portfolios (item 19).

#### 7.1.6 Test Administration

The category of “Test Administration” consists of 3 items (*see Table 8 in the Appendix*). Within the group of pre-service teachers, 2 items (14 and 20) showed strong agreement in the direction of the response and one of the items (item 5) fell in the category of disagreement with 67 percent believing the scenario as being ethical. In the in-service teachers group, the situation is almost identical with the exception of item 5. The degree of agreement upon the ethicality of the situation ranged between 88 percent and 97 percent in the case of items 14 and 20. However, in the scenario from item 5, 79 percent of the in-service teachers found the situation as ethical (compared to 67 percent in the pre-service group). As a result, the

response for item 5 in the case of in-service teachers fell in the agreement area, whereas that of pre-service teachers fell in the disagreement area (see Figure 15).



*Figure 15. Answers for item 5: “While administrating a test, a teacher notices that a student is copying answers from another student. The teacher stops at the student’s desk, quietly reminds her that cheating is not appropriate, and then the teacher continues walking around the room. At the end of the day the teacher sent a note home about the cheating.”*

#### 7.1.7 Summary

Within the group of pre-service teachers, 21 items (out of 35) fell in the strong agreement area, four items in the agreement area, four items in the disagreement area, and six items in the strong disagreement area. In the case of the in-service teachers’ group, the responses for 21 items fell in the strong agreement area, for six items fell in the agreement area, and for eight items in the disagreement area (there were not items presenting strong disagreement in the responses for the in-service group of teachers). From the number of items that conveyed divided opinion (either strong or not), 7 were common for both groups of teacher’s. The category that showed the most frequent occurrence of division in the perceptions of the pre-service teachers group is the “Bias / Fairness” category (three of the five scenarios had responses falling either in the strong or weaker disagreement areas). The rest of the categories showed high levels of agreement in the group of pre-service teachers: “Communication about Grading” (four of five scenarios), “Confidentiality” (three of four scenarios), “Grading Practices” (nine of twelve scenarios), “Multiple Assessment Opportunities” (five of six scenarios), and “Test Administration” (two of three scenarios). In the case of in-service teachers, the category with the highest frequency of disagreement was also “Bias / Fairness” (three out of five). Similarly, the remaining categories showed high levels of agreement:

“Communication about Grading” (four of five scenarios), “Confidentiality” (three of four scenarios), “Grading Practices” (ten of twelve scenarios), “Multiple Assessment Opportunities” (five of six scenarios), “Test Administration” (all scenarios).

## **7.2 Analysis of the significance of difference between the teachers’ groups**

There were three scenarios that showed significant statistical differences between the two groups of teachers. First, item 17 from the category “Confidentiality” had a statistically significant difference for the level of responses in the two groups (65 percent of pre-service teachers considering the practice as ethical and 62 percent of in-service teachers considering the practice as unethical) at the five percent level ( $p$ -value of the Fisher’s Exact Test was .023). Second, item 29 from the category “Grading Practices” registered a statistically significant difference for the level of responses in the two groups (98 percent of pre-service teachers labeled the practice ethical, whereas 82 percent of the in-service teachers labeled it ethical) with a  $p$ -value of .038 based on the Fisher’s Exact Test. Third, item 34 from the same category, “Grading Practices”, also registered a statistically significant difference in responses between the two groups. In the case of the pre-service teachers 96 percent labeled the scenario as ethical, whereas 79 percent of in-service teachers labeled it unethical). The Fisher’s Exact Test revealed that the difference was significant at the five percent level ( $p$ -value = .034).

## **8. DISCUSSION AND CONCLUSION**

The present study aimed to investigate pre-service and in-service teachers' perceptions about the ethicality of classroom assessment practices and to examine whether there are substantial differences between the two groups of teachers. First, I found that on a less than a third of the scenarios teachers disagreed within both groups. Moreover, the analysis reveals that only in three out of 35 items there was significant statistical disagreement between the groups. Consequently, the answer to the research question regarding the degree of agreement between the two groups of teachers, can be answered positively. There appears to be a strong degree of agreement between the two groups of in-service and pre-service teachers in Oulu, Finland. This finding indicates that they follow similar ethical principles and guidelines while carrying out the classroom assessment. Moreover, the results show that such differences as work experience and the recent familiarization with the up-to-date research regarding teaching skills did not make a difference in the way teachers perceive the ethicality of the proposed classroom assessment practices.

A significant conclusion is that the results obtained in the present study are different from those obtained by Green et.al (2007), in which teachers in the United States were divided on the ethics depicted in 25 percent of the scenarios. The results are also different from those obtained by Johnson et al. (2008) where educational leaders disagreed on 33 percent of the items, and from the results of Liu et al. (2016), where 36.1 percent of Chinese and 33.3 percent of American pre-service teachers were divided in their opinion.

It is notable that the assessment guidelines stipulated by the Finnish National Curriculum, the Local Curriculum, and the School Curriculum are not vague and general as Gipps (1994) concluded. Moreover, it appears that they are applicable for specific situations, thus supporting teachers' decisions. The efficacy of the guidelines is a result of the clearly defined structure of the three main areas that the assessment must be focused on in Finnish schools: pupil's learning, working aptitudes, and behavior. The pupil's learning and working aptitudes are considered as part of the grade while the behavior is assessed as a separate entity. Each subject has the specific objectives for each of these dimensions and the behavior has separate objectives that are made known and explained to pupils. Moreover, in Finland, in order to motivate pupils to be actively involved in the learning process as well as in the assessment procedures the development of pupils' capacity for self-assessment and peer-assessment is one of the main aims of the assessment process. Importantly, the feedback given to students is

fundamental because it comprises information about the pupil's successes; thus, the Curriculum also conveys clearly the key features of the constructive feedback.

The National curriculum system is one of the distinctive principles that the Finnish educational system is based on (Niemi, 2015). While it provides values and objectives for each educational level, the local educational authorities and schools have the autonomy to organize the education and the implementation of the core curriculum (Halinen & Holappa, 2013). The decentralized model of the educational system that is applied in Finland gives local freedom, putting the local educational authorities in charge for the provision and quality of education. These distinctive characteristics of the Finnish educational system and policy making have an important impact on the eloquence of the educational strategies, guidelines, and principles.

As pointed out by researchers (Impara, Plake & Fager, 1993; Plake & Impara, 1997; Stiggings, 1999), teachers' knowledge regarding the assessment ethical codes and guidelines may be uncertain because they either lack formal assessment training or their training is dated. In Finland, the pre-service teachers' education is well planned and structured. It comprises Bachelors and Masters studies and the obligation to conduct research and practice (Jyrhämä & Maaranen, 2012; Niemi & Jakku-Sihvonen, 2006). The educational programmes provided to the in-service teachers, however, are fewer than in other countries (OECD, 2013); officially, in-service teachers have only three in-service training days each year. In addition, there are school-based development projects that have as a purpose the development of teachers' skills, the support of reflection on their work and on new research in the field of education, and sharing of new ideas. All these factors described above (the National Core Curriculum guidelines, the decentralized model of the educational system, the pre-service and in-service teachers' education) influence greatly teachers' pedagogical skills, their assessment literacy, and their perceptions about the ethicality of the classroom assessment procedures hence the results of this research.

Even though the findings of this study present a strong agreement between teachers, there are a few scenarios that do show a split of opinion regarding the ethicality of the assessment practices. Thus, it is important to clarify, in accordance with the theoretical guidelines, what should be the correct judgement that needs to be applied in those cases.

There are three scenarios that indicate disagreement within the groups of teachers in the category "Bias / Fairness". Two of these scenarios (items 8 and 26) refer to "Avoid Score



Pollution” guiding principal for classroom assessment that was recommended by Popham (1991) and Haladyna et al. (1991). In case teachers do not apply the “blind grading system”, as described in item 8, and they prefer certain students, unconsciously, they can temper the students’ grades (Green et al., 2007). According to this guideline, the correct response choice for item 8 should be “unethical”.

In addition, the findings of Tierney (2015) convey that there are three reasons why teachers alter students’ grades: the need for compassion, the wish to provide opportunity for students, and the aim to teach a life lesson. Thus, in item 26, the need for compassion reason is confirmed by the teachers’ answers and the correct answer for this item should be “unethical”. Furthermore, the Finnish National Core Curriculum points out that the feedback pupils receive must be focused on the pupil’s successes and progress, which means that it can be concluded that the correct answer for item 15 should be “ethical”.

In the category “Communication about grading,” item 11 displays disagreement within the groups of teachers. According to existing literature, teachers must be trusted by the society (Sockett, 1990) and by the pupils because they constitute their role models (Popham, 1991). Among the other factors that lead to trust are consistency and unambiguousness. These factors must be taken into account when considering the practice of using surprise items about topics that were not on the study guide for the summative test as ethical or unethical.

Similarly, in the category group “Confidentiality,” there was one item (item 17) that showed discrepancy among teachers’ perceptions within each group as well as between groups. In addition to the fact that the National Core Curriculum stipulates that education develops pupil’s aptitudes for self- and peer-assessment by reflecting on their own and on their peers’ learning, it also conveys that the assessment applied should be versatile and should avoid the comparison of pupils’ achievements. In case of students grading each other’s math homework, firstly, other assessment methods, beside the peer-assessment method, should also be applied. Secondly, by grading their peers’ homework, unconsciously, students can pollute the grades based on their preferences and by comparing the work they assess with that of the other peers. At last, by applying such an assessment method, the principle of confidentiality is also violated. Thus, these guidelines must be considered seriously when labeling the assessment practice described in item 17 as “ethical”.

In the fourth scenario category “Grading Practices,” there are three items (items 10, 12, and 38) that present differences in the teachers’ perceptions, within each group and two items

(items 29 and 34) that present significant statistical difference between the groups. For both groups, two of the three items (items 12 and 38) are similar but the third item (item 10) shows significant difference just in the group of pre-service teachers. Item 10 refers to the same guideline discussed in the previous scenario category, “Avoid score pollution”. Teachers must think carefully how heavily they weight peer-assessment for the grade formation because of the possible bias that may be introduced by peer-assessment.

As specified in the Core Curriculum, and by McMillan (2000), the assessment must be versatile (implying various assessment techniques). Consequently, by considering homework as the most dominant factor in determining grades (item 12), this would not be in line with the theoretical guidelines. Moreover, it would be an indicator that the teacher is not assessment literate and does not understand the fair assessment principles; hence, (s)he may not be using various assessment practices, philosophies, and theories (DeLuca & Bellara, 2013).

In the Finnish National Core Curriculum it is specified that “Behavior is assessed as a separate entity in a report, and it does not affect the grade or verbal assessment given for a subject.” Pertaining to this guideline, the results for the item 38 - “A teacher lowers grades for disruptive behavior”, show that 46.5 percent of pre-service teachers and 38 percent of in-service teachers labeled it as “ethical”. In terms of items 29 and 34, however, the theoretical guideline at the base for an ethical or unethical choice should be “Avoid score pollution”. It is specified that the assessment that includes grades should consider only the level up to which students learned and achieved the goals and objectives of instruction (Brookhart, 2004; Smith, Smith, & DeLisi, 2001). As a result, if the grades are modified because of behavior issues, late work, or student effort, they do not reflect the level of achievement of the instruction goals. In this sense, item 6 refers to this specific guideline. Even though the difference of the answers given for this item is not considerable, it is worth mentioning because 14 percent of both the pre-service and in-service teacher groups considered as “ethical” the practice of assigning a zero mark for late work. Similarly, item 32 does not present a significant difference within and between the groups; nevertheless, 26 percent of the pre-service and 15 percent of the in-service teachers considered as “ethical” to add extra points to each student’s mid-term score to make sure that most students still have a chance to pass at the end of the semester. This result confirms again the findings that Tierney (2015) presented, where one of the reasons teachers alter the grades is the wish to provide opportunity for students.

In the category “Multiple assessment opportunities,” items 30 and 33 fell in the area of disagreement for either the pre-service or in-service teachers. Both items refer to the theoretical guidelines regarding the versatile assessment and the score pollution. Thus, the procedure of using just one method of assessment or giving the same mark to all group participants should not be considered “ethical”. In this sense, Woolfolk (2003) specified that in order to avoid score pollution the grade assigned for collaborative work should involve an individual component in addition to the common grade for the whole group.

Finally, there was only one item (item 5) that presented differences of opinion in the category group “Test Administration.” This was only observed in the case of pre-service teachers. Considering that cooperation between home and school is an important factor of a sound assessment culture, as mentioned in the Finnish Core Curriculum, and that teachers should not embarrass or publicly humiliate any pupil (Campbell, 2000, 2003), the correct choice for item 5 should be “ethical”.

## **8.1 Practical Implications**

In my opinion, the findings of the present study show that the instructions given to Finnish pre-service teachers regarding the classroom assessment practices, and their ethical judgement, are in line with the theories and guidelines. Similarly, the results reveal that in-service teachers’ education and competence is according to the standards stipulated in all official documents. This leads to a constant updating regarding the latest research in educational sciences and the needed competences. On the other hand, the findings of this study emphasize that further instructions related to score pollution, the assessment of collaborative work, the use of the versatile assessment, and the awareness of the instructions regarding assessment given in the Curricula, should be more carefully addressed in pre-service and in-service teacher education.

In addition, the results of the present study denote that the clarity of the Finnish assessment guidelines and the education process that the Finnish pre-service and in-service teachers receive is well-planned and effective. Considering the differences between the results of the present study and the results showed by the previous studies, I can conclude that the educational policy makers of the other countries may want to follow the example of the Finnish education to minimize the issues related to the ethicality of the classroom assessment. Furthermore, the results of the previous studies, as well as the present one, showed that

teachers considered highly the students' effort when assigning grades. However, researchers in the field of assessment pointed out that grades should reflect exclusively the achievement level but not the effort and other factors as growth, late work, or conduct (Brookhart & Nitko, 2008; Oosterhof, 2009). In contrast with these guidelines regarding pupils' growth, the Finnish National Core Curriculum stipulates that growth or progress is part of the assessment of learning, thus, it is considered an achievement factor. Pupils' effort is considered a working skill which is also assessed as part of the grade in Finland. However, since it is not clearly specified in the National Core Curriculum, I believe that this element should be addressed more clearly in order to avoid different interpretations.

The Core Curriculum specifies that late work is an indicator of lack of planning and regulation skills, which are considered essential working skills. Therefore, it is considered as part of the grade in elementary school. However, this guideline, given in the Core Curriculum, is not in line with the guideline of Smith and DeLisi (2001), which conveys that the assessment that includes grades should reflect only the level of mastery of the set objectives. Thus, by considering students' effort, behavior, or late work, when assigning grades, this indicates that the grades do not accurately reflect the level of mastery of the academic goals. In regards to pupils' conduct/ behavior, it is assessed as a separate entity, according to the guidelines given in the Curriculum, and it does not influence the grade given for a subject. In conclusion, the research in the field of classroom assessment should investigate further whether it is ethical or unethical / fair or unfair / effective or ineffective to apply the guidelines conveyed by the Finnish National Curriculum or those conveyed by some other researchers.

## **8.2 Directions for future research**

The current research is based on scenarios designed and developed in the U.S. educational context. Consequently, the future research that will address the ethical issue of the classroom assessment in Finland, should elaborate scenarios typical for the Finnish educational context. In this sense, future work should take a closer look at the cultural differences of classroom assessment in various countries of the world and how these cultural differences impact particular educational policies. Moreover, the current research had as participants Finnish pre-service and in-service teachers of elementary school education. It may be relevant to investigate in future research the ethicality perspectives of teachers from other educational levels and potentially compare them to those of elementary teachers.

In addition, as suggested by Liu et al. (2016), future research should include another scenario category that tackles the issue of ethics of caring. This, the importance of the ethics of carrying, is also suggested by Rogers and Webb (1991) as well as Beets (2011). The dimensions of care and justice and their interactions which results in ethical dilemmas is further described by Colnerud (2006). Additionally, I would recommend using a mixed method of investigation by including interviews as an additional instrument beyond the questionnaire. This may allow researchers to gather more information about the teachers' perceptions regarding classroom assessment.

Moreover, this study addressed the in-service teachers from teacher training schools. Many of these teachers are Ph.D. graduates. As a result, this factor may have biased the findings. For future studies, it would be interesting to address also teachers from other schools and with different levels of education.

Finally, in previous studies, as well as in the present study, the 80 percent threshold was considered as the lower limit for the strong agreement area. For future studies, that should also include a bigger sample, I recommend investigating what should the appropriate threshold between strong and weak agreement as well as strong and weak disagreement be. This could potentially be achieved based on statistical significance testing of the different thresholds, which may provide a more scientific argument for the specific agreement / disagreement thresholds.

## **9. EVALUATION OF THE STUDY**

Educational research studies also address issues pertaining to the study's validity and reliability. This is required in order to establish the degree of relevance for the obtained results. In this section I discuss these two concepts and provide further insights that pertain to the current study.

### **9.1 Research Validity and Reliability**

Content validity looks at whether the instrument comprises all the content that it is intended to, taking into account the variables. On the other hand, reliability looks at the consistency of the instrument. In this regard, the instrument of this study was developed previously and used in similar studies by Green et al. (2007) then adapted and used in another study by Liu et al. (2016). For the present study, the questionnaire was re-adapted to the Finnish educational context and translated into Finnish by a group of experts, all Finnish native speakers. As for the internal consistency, the questionnaire used in previous studies was grouped in 7 categories based on the topics of the scenarios, whereas the present study kept six of the original categories and eliminates the "Standardized test preparation" category due to the lack of standardized testing in Finland.

The total number of participants for this study was 79 (45 pre-service and 34 in-service teachers). This sample size is optimal and reliable, the survey can be replicated in other regions of the country / world, and the results can be generalized to the overall issue of assessment ethicality. The pre-service teachers were students of Primary Teacher Education Master Degree Programs of the University of Oulu. The in-service teachers were primary school teachers of the University of Oulu Teacher Training Schools. This factor might have influenced the findings since many teachers from teacher training schools are PhD graduates.

The evaluation of the results is justified and appropriate. The frequency test was analyzed by following the example of previous studies. Nevertheless, in contrast to the previous studies, in the present research the Fisher's Exact Test was applied to investigate the significance of the differences between pre-service and in-service teachers' perceptions (instead of the chi-square test, which was used in earlier studies). This decision was taken because of the small sample, which makes it difficult to use the chi-squared test accurately.

## **9.2 Ethical Issues**

The research that involves human subjects must obey the ethical guidelines that implicate guaranteed anonymity and confidentiality, the maintained dignity of the participants, and the informed consent (Couchman & Dawson, 1990; Morse, 1991; Polit & Hungler, 1991; Christakis, 1992; Punch, 1994). Brown (1993) pointed out that all human beings have the right to privacy. Hence, research participants may not want the details about themselves, or their opinions and attitudes, to be made publicly available. Consequently, the data for this study was collected anonymously and the participation for the study was voluntary. Both groups of participants were firstly briefed about the purpose of the study, the voluntary way of participation, and the structure of the survey.

Another right that every individual that participates in research has is the informed consent (Couchman & Dawson, 1990). Thus, the paper version of the questionnaire, that was handed to the pre-service teachers contained the consent of using their responses for research purposes, which asked for their signature at the bottom. Moreover, the data collection from the in-service teachers was firstly approved by the school principal who received the request for research and then was approved after a thorough assessment of the questionnaire (by the school principals). Before initiating the data collection procedure, I discussed the particulars of the current study with the school principals as well as the teachers in order to obtain their consent. All the participants consented to release their information for research purposes. The questionnaire was sent by e-mail as a link generated by Webropol (an online survey tool).

## **9.3 Study limitations**

This study comprised one optional question (the 40<sup>th</sup> question) which inquired about the participant's opinion regarding the study or a particular scenario they found difficult to answer to. By analyzing these answers, I found that many respondents needed more details regarding the practice described and other descriptive factors to make the right decision regarding the ethicality of the scenario. Additionally, respondents admitted that in many cases it was difficult to select "ethical" or "unethical" and suggested that a scale should give the possibility to answer more accurately.

Another limitation is that this study relies mostly on U.S literature when considering the theoretical background and the ethical guidelines because there are little materials regarding the Finnish classroom assessment practices, polices, guidelines available in English.

Moreover, 35 scenarios cannot address all the ethical dilemmas that teachers struggle with when applying classroom assessment, thus more scenarios should be development for the future studies (potentially in stricter accordance to the Finnish policies in the educational system).

A final limitation of the study may be the low number of participants. Although the overall sample size, 79 participants, may be enough to run statistical hypothesis tests, the sub-samples of the two groups of teachers may be too small for within group analysis. A larger sample size may also allow for different types of hypothesis testing that may be easier to compare with previous studies on this topic.



## 10. REFERENCES

- Airasian, P. (2005). *Assessment in the classroom: A concise approach (2nd ed.)*. Boston, MA: McGraw-Hill Company.
- Airasian, P. W. (1984). Classroom assessment and educational improvement. Paper presented at the conference, *Classroom Assessment: A Key to Educational Excellence*, Northwest Regional Educational Laboratory, Portland. OR.
- Crisp, R. (1982) *Aristotle, Nicomachean ethics*. St. Anne's College, Oxford, England: Oxford University Press.
- Beets, P. A. D. (2011). Strengthening morality and ethics in educational assessment through Ubuntu in South Africa. *Educational Philosophy and Theory*, 44, 68–83.
- Bergem, T. (1993). Examining aspects of professional morality. *Journal of Moral Education*, 22, 297–312.
- Biberman-Shalev, L., Sabbagh, C., Resh, N., & Kramarski, B. (2011). Grading styles and disciplinary expertise: The mediating role of the teacher's perception of the subject matter. *Teaching and Teacher Education: An International Journal of Research and Studies*, 27(5), 831–840.  
<http://dx.doi.org/10.1016/j.tate.2011.01.007>
- Brandt, D., & Rose, C. (2004). Global networking and universal ethics. *AI & Society*, 18 (4), 334–343.
- Brown, K. H. (1993). Descriptive and normative ethics: class, context and confidentiality for mothers with HIV. *Social Science & Medicine*, 36(3), 195–202.
- Brookhart, S. (2004). *Grading*. Upper Saddle River, N.J.: Pearson.

- Brookhart, S. M. (1993). Teachers' Grading Practices: Meaning and Values. *Journal of Educational Measurement*, 30(2), 123–142.
- Brookhart, S. M. (1994). Teachers' Grading: Practice and Theory. *Applied Measurement in Education*, 7(4), 279–301.
- Brookhart, S. M., & Nitko, A. J. (2008). *Assessment and grading in classrooms*. Upper Saddle River, N.J.: Pearson.
- Buzzelli, C. A., & Johnston, B. (2002). *The moral dimensions of teaching: Language, power, and culture in classroom interaction*. New York: RoutledgeFalmer.
- Campbell, E. (2000). Moral and ethical exchanges in classrooms, paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 24–28 April.
- Campbell, E. (2003). *The ethical teacher*. Maidenhead, England.: Open University Press.
- Christakis, N. A. (1992). Ethics are local: engaging cross-cultural variation in the ethics for clinical research. *Social science & medicine*, 35(9), 1079–1091.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational assessment*, 3(2), 159–179.
- Clark, C. M. (1990). The teacher and the taught: moral transactions in the classroom, in: J. Goodlad, R. Soder & K. Sirotnik (Eds) *The moral dimensions of teaching*. San Francisco, CA: Jossey-Bass.
- Clark, C. M. (1991). Educating the good teacher. Keynote address presented at The Fifth Conference of the International Study Association on Teacher Thinking, Guildford, 23–27 September.

- Colnerud, G. (1997). Ethical conflicts in teaching, *Teaching and Teacher Education*, 13, 627–635.
- Colnerud, G. (2006). Teacher Ethics as a Research Problem: Syntheses Achieved and New Issues. *Teachers and Teaching: Theory and Practice*, 12(3), 365–385.
- Couchman, W., & Dawson, J. (1990). *Nursing and health-care research: the use and applications of research for nurses and other health care professionals*. London, England: Scutari Press.
- Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, 40(1), 67–87.
- Crooks, T. J. (1988). The impact of classroom evaluation on students. *Review of Educational Research*, 58, 438–481.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, 64, 356–372.
- Duncan, R. C., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research*, 53(1), 1–21.
- Elwood, J. (2006). Gender issues in testing and assessment. In C. Skelton, B. Francis, & L. Smulyan (Eds.), *The SAGE handbook of gender and education* (pp. 262–278). Thousand Oaks, CA: SAGE Publications.  
<http://dx.doi.org/10.4135/9781848607996.n20>
- Fenstermacher, G. D. (1986). Philosophy of research on teaching: three aspects, in: M. C. Wittrock (Ed.) *Handbook of research on teaching*. New York, N.Y.: Macmillan.
- Field, A., & Field, A. P. (2013). *Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll* (4th ed.). London, England: Sage.

- General Teaching Council for Northern Ireland Teaching: The reflective profession GTCNI, Belfast (2007).
- Gipps, C. V. (1994). *Beyond testing: Towards a theory of educational assessment*. Washington, DC: The Falmer Press.
- Green, S., Johnson, R., Kim, D., & Pope, N. (2007). Ethics in classroom assessment practices: issues and attitudes. *Teaching and Teacher Education*, 23(7), 999–1011.
- Guskey, T. R. (2009). *Bound by tradition: Teachers' views of crucial grading and reporting issues*. Paper presented at the American Educational Research Association, San Francisco.
- Haladyna, T. M., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20, 2–7.
- Halinen, I., & Holappa, M.-S. (2013). Curricular balance based on dialogue, cooperation and trust – The case of Finland. In W. Kuiper & J. Berkvens (Eds.) *Balancing Curriculum Regulation and Freedom across Europe. CIDREE Yearbook* (pp. 39–62). Enschede: SLO Netherlands Institute for Curriculum Development.
- Hansen, D. (2001b). Teaching as a moral activity, in: V. Richardson (Ed.) *Handbook of research on teaching*. Washington, DC.: American Educational Research Association.
- Harlen, W. (2007). Formative classroom assessment in science and education. In J. McMillan (Ed.), *Formative classroom assessment* (pp. 116–135). New York, N.Y.: Colombia University Teachers College Press.
- Hendrickson, K. A. (2012). Assessment in Finland: A Scholarly Reflection on One Country's Use of Formative, Summative, and Evaluative Practices. *Mid-Western Educational Researcher*, 25(1/2), 33–43.

- Herbart, J. F. (1917). *Konturer till föreläsningar i pedagogik* [Outlines of lessons in pedagogy]. Stockholm: Albert Bonniers Förlag. Original work published 1835.
- Hostetler, K. D. (1997). *Ethical judgment in teaching*. Boston, MA.: Allyn and Bacon.
- Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing, 11*(1), 42–65.
- Impara, J. C., Plake, B. S., & Fager, J. J. (1993). Educational administrators' and teachers' knowledge of classroom assessment. *Journal of School Leadership, 3*(5), 510–521.
- Jackson, P. W., Boostrom, R. & Hansen, D. (1993). *The moral life of schools*. San Francisco, CA: Jossey-Bass.
- JCSEE (2003). *The student evaluation standards*. Thousand Oaks, CA: Arlen Gullickson, Chair, Corwin.
- Johnson, R., Green, S., Kim, D., & Pope, N. (2008). Educational leaders' perceptions about ethical assessment practices. *The American Journal of Evaluation, 29*(4), 520–553.
- Jyrhämä, R., & Maaranen, K. (2012). Research-orientation in a teacher's work. In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *Miracle of education: The principles and practices of teaching and learning in Finnish schools* (pp. 97–110). Rotterdam, NL.: Sense Publishers.
- Jyrhämä, R. (2006). The function of practical studies in teacher education. In R. Jakkusihvonen & H. Niemi (Eds.), *Research-Based Teacher Education in Finland – Reflections by Finnish Teacher Educators* (pp. 51–70). Turku, Finland: Finnish Educational Research Association.

- Kant, I. (1790). *Critique of judgement* (J.H. Bernard Trans). New York, N.Y.: Hafner Publishing Company, Inc. 1951.
- Lekholm, A. K. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian journal of educational research*, 55(6), 587–608.
- Kumpulainen, K., & Lankinen. T. (2012). Striving for educational equity and excellence: Evaluation and assessment in Finnish basic education. In H. Niemi, A. Toom, & A. Kallioniemi (Eds.), *Miracle of education: The principles and practices of teaching and learning in Finnish schools* (pp. 69-81). Rotterdam: Sense Publishers.
- Liu, J., Johnson, R., & Fan, X. (2016). A comparative study of Chinese and United States pre-service teachers' perceptions about ethical issues in classroom assessment. *Studies in Educational Evaluation*, 48, 57–66.
- Lyon, E. G. (2013). Learning to assess science in linguistically diverse classrooms: Tracking growth in secondary science preservice teachers' assessment expertise. *Science Education*, 97, 442–467. doi:10.1002/sce.21059
- Mahony, P. (2009). Should 'ought' be taught? *Teaching and Teacher Education*, 25, 983–989.
- McMillan, J. H. (2000). Fundamental assessment principles for teachers and school administrators. *Practical Assessment, Research & Evaluation*, 7(8), 1–9.
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, 20(1), 20–32.  
[http:// dx.doi.org/10.1111/j.1745-3992.2001.tb00055.x](http://dx.doi.org/10.1111/j.1745-3992.2001.tb00055.x)
- Morse, J. M. (1991). Strategies for Sampling In: Morse J. M, editors. *Qualitative Nursing Research: A Contemporary Dialogue*.

- National Association for the Education of Young Children (2005). Code of ethical conduct and statement of commitment. Retrieved December 27, 2005, from <http://www.naeyc.org/about/positions/pdf/PSETH05.PDF>.
- Niemi, H. (2015). Teacher professional development in Finland: Towards a more holistic approach. *Psychology, Society and Education*, 7(3), 279–294.
- Niemi, H. (2014). Purposeful policy and practice for equity and quality - a Finnish case. In S. K. Lee, W. O. Lee., & E. L. Low (Eds.), *Education Policy Innovations: Levelling Up and Sustaining Educational Achievement* (pp. 103–121). Singapore: Springer Singapore.
- Niemi, H., & Nevgi, A. (2014). Research studies and active learning promoting professional competences in Finnish teacher education, *Teaching and Teacher Education*, 43, 131–142.
- Niemi, H. & Jakku-Sihvonen, R. (2006). Research-based teacher education. In R. Jakku Sihvonen, & H. Niemi (Eds.), *Research-based teacher education in Finland. Reflections by Finnish teacher educators* (pp. 31–50). Turku, Finland: Finnish Educational Research Association.
- Niemi, H., & Nevgi, A. (2014). Research studies and active learning promoting professional competences in Finnish teacher education, *Teaching and Teacher Education*, 43, 131–142.
- OECD (2003). Organization for Economic Co-operation and Development. First results from PISA 2003. Executive Summary. Retrieved Feb. 5, 2015 from <http://www.oecd.org/education/preschoolandschool/programmeforinternationalstudentassessmentpisa/34002454.pdf>.
- OECD (2006). Equity in Education. Thematic Review. Finland Country Note. Retrieved Feb. 5, 2015 from [http://www.oecd.org/document/3/0,2340,en\\_2649\\_34531\\_36296195\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/3/0,2340,en_2649_34531_36296195_1_1_1_1,00.html).

- OECD (2010). PISA 2009 Results: What students know and can do: Student performance in reading, mathematics and science (Volume I). Paris, France: OECD.
- OECD (2014). PISA 2012 Results: (Volume V). Paris: OECD. Retrieved Feb. 5, 2015 from <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-v.htm>.
- OECD (2013), A Teachers' Guide to TALIS 2013: Teaching and Learning International Survey, available at <http://dx.doi.org/10.1787/9789264216075-en>.
- Ontario College of Teachers [OCT]. (2013). Exploring ethical professional relationships: A self-reflective resource. Toronto, ON. Retrieved from <http://www.oct.ca/public/professional-standards>.
- Oosterhof, A. (2009). *Developing and using classroom assessments*, 4th ed. Upper Saddle River, N.J.: Merrill.
- Opetushallitus. (2016). *National core curriculum for basic education 2014*. Helsinki, Finland: Finnish National Board of Education.
- Organisation for Economic Co-operation and Development. (2012). *Grade expectations: How marks and education policies shape students' ambitions, PISA*. OECD Publishing. <http://dx.doi.org/10.1787/9789264187528-en>
- Oulun kaupungin sivistys- ja kulttuuripalvelut (2016)*. Perusopetuksen opetussuunnitelman perusteet ja Oulun kaupungin paikalliset linjaukset, Vuosiluokat 1.6. Haettu 6.3.2017 sivulta <https://www.ouka.fi/opetussuunnitelma>
- Payne, D. A. (2003). *Applied educational assessment* (2nd ed.). Belmont, CA: Wadsworth Publishing.
- Peterson, S. S., & Kennedy, K. (2006). Sixth-grade teachers' written comments on student writing: Genre and gender influences. *Written Communication*, 23(1), 36–62.



- Plake, B., & Impara, J. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment*. San Diego, CA.: Academic Press.
- Polit, D. F., & Hungler, B. P. (1991). *Nursing Research Principles and Methods*, 4th ed. JB Lippincott.
- Pope, N.S. (2006). Do no harm to whom? An examination of ethics and assessment, *South Atlantic Philosophy of Education Society Yearbook*, 25–31.
- Popham, W. J. (1991). Appropriateness of teachers' test- preparation practices. *Educational Measurement: Issues and Practice*, 10(4), 12–15.
- Popham, W. J. (2004). Why assessment illiteracy is professional suicide. *Educational leadership*, 62(1), 82–83.
- Punch, M. (1994). Politics and ethics in qualitative research, in, Denzin, N. & Lincoln, Y.(eds) *Handbook of Qualitative Research*.
- Rajakaltio, H. (2014). Towards renewing school. The action model of the school development - Integrating in-service-training and the development process. (In Finnish), Reports and 2014:9. Reports and reviews 2014:9. *The Finnish National Board of Education*, Helsinki, Finland.
- Randall, J., & Engelhard, G. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement*, 46(1), 1–18.  
[http:// dx.doi.org/10.1111/j.1745-3984.2009.01066.x](http://dx.doi.org/10.1111/j.1745-3984.2009.01066.x)
- Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal*, 12(3), 315–325.  
<http://dx.doi.org/10.1007/s11218-008-9073-z>
- Rogers, D., & Webb, J. (1991). The ethic of caring in teacher education. *Journal of Teacher Education*, 42(3), 173–181.

- Ryan, A. G. (1997). Professional obligation: A dimension of how teachers evaluate their students. *Journal of Curriculum and Supervision*, 12(2), 118–134.
- Sahlberg, P. (2015). *Finnish lessons 2.0: What can the world learn from educational change in Finland?* (Second edition.). New York, N.Y.: Teachers College Press.
- Singer, P. (1993). *About ethics: in practical ethics*, 2nd ed. Cambridge, England: Cambridge University Press.
- Singer, P. (1994). Introduction: in ethics. Oxford, U.K.: Oxford University Press.
- Smith, J. K., Smith, L. F., & DeLisi, R. (2001). *Natural classroom assessment: Designing seamless instruction and assessment*. Thousand Oaks, CA.: Corwin Press.
- Sockett, H. (1990). Accountability, trust and ethical codes of practice. In Goodlad, J., Soder, R., & Sirotnik, K. (Eds.) *The moral dimensions of teaching*. San Francisco, CA.: Jossey-Bass Inc. Publishers.
- Soder, R. (1990). The rhetoric of teacher professionalization. In Goodlad, J., Soder R., & Sirotnik, K. (Eds.) *The moral dimensions of teaching*. San Francisco, CA.: Jossey-Bass Inc. Publishers.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83, 758–765.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, N.Y.: State University of New York Press.
- Strike, K. A. (1990). Teaching Ethics to Teachers: What the Curriculum Should Be About. *Teaching and Teacher Education*, 6(1), 47–53.
- Strike, K. A., & Soltis, J. F. (1998). *The ethics of teaching* (3rd ed.). New York, N.Y.: Teachers College Press.

- Sun, Y., & Cheng, L. (2014). Teachers' grading practices: Meaning and values assigned. *Assessment in Education: Principles, Policy & Practice*, 21(3), 326–343.
- Taylor, K., & Nolen, S. (2005). *Classroom assessment: Supporting teaching and learning in real classrooms*. Upper Saddle River, N.J.: Pearson Education, Inc.
- Terhart, E. (1987). Vermutungen über das Lehrerethos für Pädagogik. *Zeitschrift für Pädagogik*, 33, 787–804.
- Terhart, E. (1994). The ethics of school teachers: Between administrative control, professional autonomy and public interest. Paper presented at the fourth international conference on ethics in the public service, Stockholm, Sweden. June, 1994.
- Thorndike, R., Cunningham, G., Thorndike, R., & Hagen, E. (1991). *Measurement and evaluation in psychology and education*, 5<sup>th</sup> ed. New York, N.Y.: McMillan.
- Tierney, R. D. (2013). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation*, 43, 55–69.
- Tierney, R. D. (2015). Altered grades: A grey zone in the ethics of classroom assessment. *Assessment Matters*, 8, 5–30.
- Tom, A. R. (1980). Teaching as a Moral Craft: A Metaphor for Teaching and Teacher Education. *Curriculum Inquiry*, 10(3), 317–323.
- Vainikainen, M-P. (2014). Finnish primary school pupils' performance in learning to learn assessments: A longitudinal perspective on educational equity. Faculty of Behavioral Sciences. *Department of Teacher of Education Research Report 360*. University of Helsinki.
- White, E. (2009). Are your assessments literate? Some fundamental questions regarding effective classroom-based assessment. *OnCUE Journal*, 3(1), 3–25.

Woolfolk, A. (2003). *Educational psychology: Developing learners. 4th ed.* Upper Saddle River, N.J.: Merrill Prentice-Hall.

## 11. APPENDIX

Table 3. Scenarios in Bias/Fairness

Item	Scenario	Pre-service teachers (% Ethical, N= 45)	In-service teachers (% Ethical, N= 34)	Fisher's Exact Test ( <i>p</i> -value)
4	Two teachers teach different sections of the same course. One teacher assigns very few report card grades with a maximum grade because of her belief that students' work is rarely perfect.	16	18	1.000
8	A teacher always checks the name of the student whose essay test he is grading.	49	59	.495
15	To enhance self-esteem, an elementary teacher addresses only students' strengths when writing narrative report cards.	38	32	.643
21	As a teacher finalizes grades, she changes one student's course grade from a 9 to a 10 because tests and papers showed the student had mastered the course objectives even though he had not completed some of his homework assignments.	80	88	.375

26	A teacher who knows a student had a bad week because of problems at home bumps the student's participation grade up a few points to compensate for his bad score on a quiz.	53	44	.498
----	---	----	----	------

Table 4. Scenarios in Communication about Grading

Item	Scenario	Pre-service teachers (% Ethical, N= 45)	In-service teachers (% Ethical, N= 34)	Fisher's Exact Test ( <i>p</i> -value)
9	A teacher tells students which materials are important to learn in preparing for a class test.	96	97	1.000
11	For the final test, a teacher always uses a few surprise items about topics that were not on the study guide.	49	65	.178
16	At the beginning of the course, a teacher shares with students the rubrics for each task. The teacher leads students in a discussion about the rubrics, makes changes to the rubrics according to students' feedback, and gives students the final versions to guide their completion of the course tasks.	100	91	.076
22	A teacher states how she will grade a task when she assigns it.	100	97	.430

27	A middle school principal directs teachers to give students a written policy that explains how grades are calculated in their classes.	96	94	1.000
----	--	----	----	-------

Table 5. Scenarios in Confidentiality

Item	Scenario	Pre-service teachers (% Ethical, N= 45)	In-service teachers (% Ethical, N= 34)	Fisher's Exact Test ( <i>p</i> -value)
17	At a high school, mathematics teachers do not grade all tests or homework. Instead, teachers let students grade each other's math homework and then share the results in groups. Teachers believe teamwork will help students learning.	65	38	.023
23	An elementary teacher discloses to the parents of a student their child's score on a test.	84	97	.128
28	To motivate students to perform better, a science teacher always announces that he is distributing scored tests to students in order of points earned, from the top score to the bottom score.	0	6	.182

31	To calm the fears of distraught parents, a teacher compares their child's achievement scores with the results of the student's cousin who is also in the class.	0	0	-
----	---	---	---	---

Table 6. Scenarios in Grading Practices

Item	Scenario	Pre-service teachers (% Ethical, N= 45)	In-service teachers (% Ethical, N= 34)	Fisher's Exact Test ( <i>p</i> -value)
18	A middle school history teacher offers extra credit opportunities to all of his classes except the advanced class.	7	3	.630
24	When a child is caught cheating on a test, rather than assigning a zero, the teacher gives the student an alternate version of the same test.	91	76	.114
6	A teacher assigns a zero for late work.	16	15	1.000
10	A teacher uses student peer ratings as 40% of the grade on an oral report.	46	29	.161
12	A teacher weights homework heavily in determining grades.	50	44	.650
29	A teacher considers students' growth in assigning grades.	98	82	.038



32	A physics teacher is concerned that most students did not perform well on the mid-term test. Based on the results, it has become mathematically impossible for about 70% of students to earn a passing grade. Thus, the teacher adds extra points to each student's mid-term score to make sure most students still have a chance to pass at the end of the semester.	26	15	.267
34	A teacher considers student effort when determining grades.	96	79	.034
35	A physical education teacher gives a student a zero as a homework grade for not returning a field trip form requiring a parent's signature.	5	9	.648
36	To encourage lively discussion in English III, a teacher counts class participation as 30% of the final grade.	76	88	.246
37	A teacher lowers grades for late work by one number grade for each day.	22	6	.060
38	A teacher lowers grades for disruptive behavior.	46	38	.496

Table 7. Scenarios in Multiple Assessment Opportunities

Item	Scenario	Pre-service teachers (% Ethical, N= 45)	In-service teachers (% Ethical, N= 34)	Fisher's Exact Test ( <i>p</i> -value)
13	For a group project, a teacher bases each student's grade on the group's product a mural of a historical event – and a heavily weighted essay completed independently by each student.	91	74	.063
7	An early elementary teacher uses observational checklists, anecdotal notes, and interviews (student conferences) in assessing students.	91	97	.384
19	A teacher assesses student knowledge by using many types of assessments: multiple-choice tests, essays, projects, portfolios.	100	100	-
25	A high school social studies teacher bases students' final semester grade on 2 multiple-choice tests.	16	9	.500
30	A second-grade teacher uses observations as the sole method to assess what students have learned.	29	32	.807
33	While assessing group work, a teacher gives the same mark to each group participant.	41	29	.346

Table 8. Scenarios in Test Administration

Item	Scenario	Pre-service teachers (% Ethical, N= 45)	In-service teachers (% Ethical, N= 34)	Fisher's Exact Test ( <i>p</i> -value)
5	While administrating a test, a teacher notices that a student is copying answers from another student. The teacher stops at the student's desk, quietly reminds her that cheating is not appropriate, and then the teacher continues walking around the room. At the end of the day the teacher sent a note home about the cheating.	67	79	.311
14	While administering a test, a teacher notices that most students missed the same question. The teacher reminds all students to check their answers of that question one more time.	96	97	1.000
20	While administering a test, a high school teacher notices that a student has skipped a problem and is recording all of her answers out of the sequence on the answer form. The teacher shows the student where to record the answer she is working on and instructs the child to put the answer to each question with the same number on the answer sheet.	88	88	1.000