



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

교육학석사학위논문

Korean English Teachers' Beliefs about
Performance Assessment and their Practices

한국 영어교사들의 수행평가에 대한 신념과 실행

2021년 8월

서울대학교 대학원

외국어교육과 영어전공

민 경 은

Korean English Teachers' Beliefs about Performance Assessment and their Practices

by
Kyungeun Min

A Thesis Submitted to
the Department of Foreign Language Education
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts in Education

At the
Graduate School of Seoul National University

August 2021

Korean English Teachers' Beliefs about Performance Assessment and their Practices

한국 영어교사들의 수행평가에 대한 신념과 실행

지도교수 소 영 순

이 논문을 교육학 석사 학위논문으로 제출함

2021년 6월

서울대학교 대학원

외국어교육과 영어전공

민 경 은

민경은의 석사학위논문을 인준함

2021년 7월

위 원 장

부위원장

위 원

Korean English Teachers' Beliefs about
Performance Assessment and their Practices

APPROVED BY THESIS COMMITTEE:

KITAEK KIM, COMMITTEE CHAIR

HYUNWOO KIM

YOUNGSOON SO

ABSTRACT

Korean English Teachers' Beliefs about Performance Assessment and their Practices

Kyungeun Min

English Major, Dept. of Foreign Language Education

The Graduate School of Seoul National University

Since assessment plays a central role in teaching and learning, performance assessment has been implemented in Korea to assess and encourage high-level, complex thinking skills since 1999. However, Korean society's meritocratic system made it difficult for teachers to implement performance assessment in the desired way due to the loose link to the college entrance exam and doubts on reliability. In order to enhance successful implementation of meaningful performance assessment, scrutinizing teachers' beliefs about performance assessment, which are proven to affect assessment practices, and detecting their relationship with practices are essential.

The purpose of this study is to investigate teachers' beliefs about performance assessments, the relationship between beliefs and practices, and the contextual factors which influence the degree of alignment between beliefs and

practices. Exploring teachers' beliefs and practices regarding performance assessment will shed light on more satisfactory implementation of valid performance assessment.

For this study, 109 Korean English teachers in secondary schools voluntarily participated in an online questionnaire of 52 items. The questionnaire consisted of three subsections: Section A about purposes of performance assessment, Section B regarding methods and techniques, and Section C concerning teachers' feedback, grading, and reporting of grades. Based on the result of cluster analysis of the questionnaire, two teachers from each cluster were randomly chosen for the subsequent semi-structured interview which lasted about twenty minutes. Since there were three clusters identified, a total of six teachers participated in an individual interview. They discussed contextual factors at the micro level, i.e. the teacher or classroom level, and at the meso level, i.e. the school level.

The results of the study suggest that most of the Korean English teachers held a positive belief about performance assessment as Assessment for Learning. They generally believed that performance assessments should improve teaching and learning while employing various methods and providing constructive feedback. However, some noticeable conflicting patterns of beliefs and practices existed for different aspects of performance assessment, and major contextual

factors were investigated that affect the degree of alignment between teachers' beliefs about performance assessment and their actual practices. The identified factors were teacher training and coping style against students' feedback at the micro level, and colleagues and complaints from students and/or parents about grades at the meso level.

Despite several limitations of the sample size and methodological issues, this study contributes to an understanding of the relationship between beliefs and practices regarding performance assessment and various contextual factors that influence the degree of alignment between beliefs and practices. Practical suggestions are also identified for teachers in Korean secondary schools to implement valid performance assessments more successfully.

Key Words: teachers' belief, belief and practice, performance assessment, contextual factors, Korean English teachers

Student Number: 2019-21002

TABLE OF CONTENTS

ABSTRACT	i
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
CHAPTER 1. INTRODUCTION	1
1.1 The Motivation and Purpose of the Study.....	1
1.2 Research Questions	4
1.3 Organization of the Thesis	5
CHAPTER 2. LITERATURE REVIEW	6
2.1 Performance Assessment	6
2.1.1 Assessment for Learning	7
2.1.2 English Performance Assessment	9
2.1.3 English Performance Assessment in Korea	10
2.2 Teachers' Beliefs about Assessment.....	13
2.2.1 Teachers' Beliefs and Assessment Practices	13
2.2.2 Alignment between Beliefs and Assessment Practices	15
2.2.3 English Teachers' Beliefs in Korea.....	19
2.2.4 Contextual Factors, Beliefs, and Practices	21
2.3 Limitations of Previous Research	26

CHAPTER 3. METHODOLOGY	28
3.1 Participants.....	28
3.1.1 Respondents for Questionnaire	29
3.1.2 Participants for Semi-Structured Interview	30
3.2 Instruments.....	30
3.2.1 Questionnaire	31
3.2.2 Semi-Structured Interview Questions	33
3.3 Procedures.....	34
3.3.1 Data Collection	35
3.3.2 Data Analysis	36
3.3.2.1 Quantitative Analysis.....	36
3.3.2.2 Qualitative Analysis.....	37
CHAPTER 4. RESULTS	38
4.1 Teachers' Beliefs and Assessment Practices	38
4.1.1 Instructional Purpose of Performance Assessment.....	40
4.1.2 Methods and Techniques	43
4.1.3 Feedback, Grading, and Reporting of Grades	47
4.2 Patterns of Beliefs and Practices	49
4.3 Contextual Factors	63
4.3.1 Contextual Factors at the Micro Level.....	65
4.3.1.1 Training.....	65
4.3.1.2 Coping Style against Students' Feedback.....	69
4.3.2 Contextual Factors at the Meso Level	74

4.3.2.1 Colleagues	75
4.3.2.2 Complaints about Grade from Students/Parents	78
CHAPTER 5. DISCUSSION	85
5.1 Beliefs and their Difference from Practices	85
5.2 Contextual Factors	92
CHAPTER 6. CONCLUSION.....	103
6.1 Major Findings and Implications	103
6.2 Limitations and Suggestions for Future Research	106
REFERENCES.....	108
APPENDICES.....	117
ABSTRACT IN KOREAN.....	125

LIST OF TABLES

Table 3.1 Profiles of Teacher Participants in Questionnaire	29
Table 4.1 Mean Comparison of Three Subsections.....	39
Table 4.2 Instructional Purposes	41
Table 4.3 Methods and Techniques.....	45
Table 4.4 Feedback, Grading, and Reporting of Grades	48
Table 4.5 Demographic Information of Clusters	52
Table 4.6 Descriptive Statistics of Clusters	53
Table 4.7 Agreement on Subsequent Interview	63
Table 4.8 Profiles of Teacher Participants in Interview	64

LIST OF FIGURES

Figure 2.1 Teacher cognition, schooling, professional education, and classroom practice	22
Figure 2.2 Conceptual framework connecting teacher’s assessment practices with contextual factors at the micro, meso, and macro level.....	24
Figure 4.1 Dendrogram of cluster analysis	51
Figure 4.2 Comparing beliefs of clusters	56
Figure 4.3 Comparing practices of clusters.....	57
Figure 4.4 Comparing differences of clusters.....	59
Figure 4.5 Comparing beliefs and practices of ‘Typical Teachers’	61
Figure 4.6 Comparing beliefs and practices of ‘Teachers in Dilemma’	62
Figure 4.7 Comparing beliefs and practices of ‘Progressive Teachers’.....	62

CHAPTER 1.

INTRODUCTION

This study aims to investigate beliefs about performance assessment of Korean English teachers and their actual practices in classrooms. This chapter presents the motivation and the organization of this thesis. Section 1.1 describes the motivation and purpose of the study, and Section 1.2 offers the research questions. The last Section 1.3 outlines the organization of the thesis.

1.1 The Motivation and Purpose of the Study

Assessment plays an essential and imperative role in teaching and learning (Cheng, Rogers, & Hu, 2004) in that it determines how the progress goes on, whether students are successful or whether a curriculum works well or not (Serpil & Derin, 2017). Being assessment a cornerstone for better learning (Elwood & Klenowski, 2002), teachers dedicate a considerable part of their preparation time to devising instruments, recording, marking, and combining results in formal and informal reports in their everyday teaching (Cheng et al., 2004). In fact, Rogers (1991) estimated that teachers dedicate between a quarter and a third of their instructional time to assessment-related

works. Despite this importance and essentialness of assessment, there is some inefficiency when it comes to real classroom practice (Serpil & Derin, 2017).

This is especially true in high-stakes assessment contexts, such as the one in Korea. Teachers in Korea are often caught in a situation where they conduct assessments which are not pedagogically sound (Bang & Chun, 2011). The typical teaching-to-the-test practices are pervasive in many classes, and these unsound activities sometimes bring score gains which do not mirror actual improvement in abilities, but rather an immediate result of coaching of test-taking strategies (Gebriel, 2017). To correct this crack between learning and assessment, performance assessment was introduced in Korea in 1999 as part of the 7th Education Curriculum reform. It gained more weight under the 2015 Revised National English Curriculum (RNEC) which underscores communicative competency, information processing competency, self-management competency, and community competency (S. H. Kim, 2017). This was anticipated to wipe out the prevailing skeptics of not properly assessing learners' English language skills in public English education and to facilitate learning through assessment. However, the reality fell short of expectations. Even many years after the introduction of performance assessments, it is reported that students only experience assessment, without learning (Lee & Sung, 2017).

In order to enhance the chance of successful implementation of

meaningful performance assessment, examining teachers' beliefs about performance assessment is paramount. Understanding the assessment beliefs of English teachers is pivotal since it can help promote the quality of assessment practices as well (Elshawa, Abdullah, & Rashid, 2017). It is proven that teachers' beliefs about the purpose of assessment affect how assessment is realized in classroom settings (Brown & Remesal, 2017), and surprisingly, beliefs influence practices more than teaching experience or schools' socioeconomic context do (Griffiths, Gore, & Ladwig, 2006). Indeed, the way teachers conceive of assessment can have a strong impact on how they make decisions and how they act in different situations (Pishghadam, Adamson, Sadafian, & Kan, 2014). Furthermore, providing opportunities to critically examine beliefs which shape practice contributes to the professional development of teachers themselves (James & Pedder, 2006).

In addition to teachers' beliefs, their assessment practices should also be scrutinized to examine what comprises effective assessment which leads to learning. Research over many decades has revealed that teachers' beliefs and their actual assessment practices don't always show a high degree of alignment (Guadu & Boersma, 2018; James & Pedder, 2006; Muñoz, Palacio, & Escobar, 2012; Winterbottom, Taber, et al., 2008). James and Pedder (2006) even reported that it is a routine for teachers to work within gaps between beliefs and practices as part of the typical experience in classrooms. Therefore,

identifying and diagnosing relationships between beliefs and practices are crucial to promote professional development of teachers regarding assessment (Winterbottom, Brindley, et al., 2008).

Another important issue to consider is the role of context. Contextual factors have been attested to occupy an influential role in the extent of alignment between beliefs and assessment practices (Beach, 1994; Tabachnick & Zeichner, 1986). Hence, various contextual factors have to be addressed to fully investigate teachers' beliefs and their actual assessment practices in classrooms. Discovering the underlying determinants that affect the translation of teachers' beliefs into practices is priceless. A lack of empirical studies on the relationship between beliefs and assessment practices of Korean English teachers and on contextual factors affecting the degree of alignment between beliefs and practices stimulates the present study.

1.2 Research Questions

With the aforementioned problems present, this research investigates teachers' beliefs about performance assessments, the relationship between beliefs and practices, and the contextual factors which influence the degree of alignment between beliefs and practices. It is guided by the following research questions:

1. What are the English teachers' beliefs about the performance assessment?
2. How consistent are the teachers' practices with their beliefs about performance assessment?
3. What are the detectable patterns of the relationship between teachers' beliefs and practices?
4. What are the contextual factors that affect the degree of alignment between beliefs and practices?

1.3 Organization of the Thesis

This thesis is organized as follows. Following the present chapter of introduction, Chapter 2 reviews the literature on performance assessment and teachers' beliefs about assessment offering a theoretical framework. Chapter 3 describes the methods and designs of data collection. It includes participants, instruments, and procedures of data collection and data analysis. Chapter 4 reports descriptive statistics of teachers' beliefs, relationships between beliefs and practices, and patterns of beliefs and practices. Results from the analysis of contextual factors are added. Chapter 5 discusses significant issues from the analysis along with a detailed explanation. Finally, Chapter 6 summarizes the main findings, presents pedagogical implications, and addresses the limitation of the study with suggestions for future research.

CHAPTER 2.

LITERATURE REVIEW

This chapter explains the theoretical framework for the present study by examining previous literature related to the topic. Section 2.1 provides a literature review of performance assessment with its base as Assessment for Learning and elaborates on English performance assessment in the Korean context. Section 2.2 defines teachers' beliefs and practices of assessment and reviews studies examining relationships between beliefs and practices in assessments along with the influence of contextual factors on beliefs and practices. Section 2.3 discusses gaps identified from the literature review, which are to be addressed in the present research.

2.1 Performance Assessment

This section is composed of three subsections. Section 2.1.1 illustrates Assessment for Learning, which is contrasted with Assessment of Learning. Section 2.1.2 explains the background of the rise of performance assessment which aims at Assessment for Learning and provides its definition. In the last Section 2.1.3, the introduction and the current implementation of English performance assessment in Korea are elaborated.

2.1.1 Assessment for Learning

For years, the term “assessment” generally covered the processes of evaluating the outcome of teaching activities after the completion of the instruction (William, 2011). As Chappuis and Stiggins (2002) illustrated, it was commonly regarded that “assessment is as an index of school success rather than as the cause of that success” (p. 40). However, many researchers acknowledged the significant impact that assessment has on informing instruction, and it led to a clear distinction between two types of assessments: Assessment *for* Learning (AfL) and Assessment *of* Learning (AoL) (Assessment Reform Group, 1999; Winterbottom, Brindley, et al., 2008).

Assessment *for* Learning is any assessment whose design and practice prioritize promoting students’ learning (Black, Harrison, Lee, Marshall, & Wiliam, 2004). It is a set of evidence-informed classroom practices with the aim of enhancing learning and increasing standards of attainment (James & Pedder, 2006). In the context of student assessment, the difference between AfL and AoL is more evident. AoL is for grading and reporting and is generally designed for accountability, ranking, and certifying competency. In contrast, AfL is for teaching to promote learning. It provides feedback to teachers and students to assess themselves and one another, and this leads to modifying the learning and teaching activities they are engaged in (Black et al., 2004). Hence, true AfL takes place when teachers employ assessment as a part of the

instruction to facilitate students' learning instead of regarding it as extra work to a traditional lesson (Singh, 2000; Winterbottom, Brindley, et al., 2008).

Effective strategies of AfL involve (1) effective questioning, (2) sharing learning objectives and assessment criteria, (3) providing effective feedback about how to improve, and (4) self-assessment and peer assessment (Black et al., 2004; Winterbottom, Brinley, et al., 2008). The biggest benefit of AfL is that not only teachers but also students take another chance to contemplate the purposes of assessment (James & Pedder, 2006). Through AfL, they both appreciate more why they are doing what, and it leads to better learning of students.

In the search for innovative formative assessment aiming to benefit learning, this AfL approach has come to integrate formative and alternative assessment methods (Colby-Kelly & Turner, 2007). Alternative assessment plays a role as an alternative to traditional standardized testing and is distinguished from it (S. Lee, 2008). All the following assessments are included in the category of alternative assessment: performance assessment, authentic assessment, informal assessment, portfolio assessment, assessment by exhibition, and situated (or contextualized) assessment (Huerta-Macias, 2002). The present research will focus on performance assessment specifically with its role as AfL.

2.1.2 English Performance Assessment

Performance assessment aims to raise core abilities to respond to the demands of dynamically changing society nowadays. Memorizing pieces of knowledge is no longer the top priority when learning; what students can *do* with knowledge is what weighs (Darling-Hammond, Adamson, & Abedi, 2010). Hence, performance assessment evaluates learners by observing students use their existing knowledge to solve problems or perform tasks (Stiggins & Bridgeford, 1982). Instead of making a choice among pre-determined options, performance assessment asks students to construct an answer, perform an activity, or produce a product (Wren, 2009). From this view, it includes a broad spectrum of activities ranging from completing a sentence with several words to composing an extended essay. Through performance assessment, students no longer just identify a correct answer from a list provided; rather, they perform or construct original responses. This process helps to raise not only students' cognitive thinking or logic but also the competence to apply knowledge to solve meaningful and realistic problems (M. Lee, 2018; S. Lee, 2008). Furthermore, it enables teachers to collect information about what students can literally do with what they learn. Whether the skill being measured is speaking, writing, or listening, students actually perform tasks with these skills (McNamara, 1996), and the teacher rates the performance based on pre-determined criteria in performance assessment.

Since performance assessment is more appropriate for assessing and encouraging complex and high-level thinking skills, many countries have begun to mandate performance assessment since the late 1980s and early 1990s (Darling-Hammond et al., 2010). Currently, performance assessment maintains its essential role in educational systems in most high-achieving countries (Darling-Hammond et al., 2010). The situation in Korea is not different. In the pursuit of raising well-prepared global citizens to cope with today's dynamically changing world, performance assessment was introduced and has been given increasingly more weight (M. Lee, 2018; T. J. Park, 2016; Park & Chang, 2017).

2.1.3 English Performance Assessment in Korea

The introduction of English performance assessment methods was welcomed as authentic assessment in Korea with the hope of counting harsh criticism on standardized paper-based exams (M. Lee, 2018). Rather than studying merely *about* English, the desire and need to communicate in English has increased (Koh, 2015; Park & Chang, 2017). Therefore, performance assessment was introduced in 1999 (the 7th Education Curriculum reform) to correspond to the growing attention to raising communicative competency. Since English performance assessment 'involves oral and written production,

open-ended responses, integrated performance across skill areas, group performance, and other interactive tasks' (Brown & Abeywickrama, 2010, p. 16), it was expected to help to achieve the ultimate goals of language teaching: advancing fluency and communicative skills. The introduction of 2015 Revised National English Curriculum reinforced this educational reform with the emphasis on process-oriented performance assessment and student-centered learning (W. Lee, 2015). It was anticipated to boost communication skills and processing skills of information delivered in English with increased motivation and interest in learning English (S. H. Kim, 2017).

However, performance assessment has not been implemented in classrooms as it should be. Korean teachers are trapped in a situation where high-stakes examinations with norm-referenced results dominate the whole curriculum with the final objective of entering a prestigious university (W. Lee, 2015). This meritocratic system of Korean society made it difficult for teachers to implement performance assessment in the proper or desired way due to the loose link to the college entrance exam and doubts on reliability (T. J. Park, 2016).

Public English education should escape from the current situation. It can't be denied that Korea is facing a high-stakes examination environment where AoL is predominant and the main purpose of assessments is the accountability of students and schools. Nonetheless, establishing a

complementary relation between AoL and AfL is not impossible despite difficulties (Harlen, 2005; Wiliam, Lee, Harrison, & Black, 2004). In many schools, conducting performance assessments is considered additional burdens and irrelevant works to be done without connected influences on succeeding classes (S. Lee, 2008). However, Hargreaves (2004) argued that AfL, which true performance assessment aims for, can be realized without an extra struggle. On the contrary, it can sometimes be achieved with reduced effort. He continued that even though the beginning stages may demand work, there is a payoff afterward, for “AfL is a teaching strategy of very high leverage—working smarter, not harder” (Hargreaves, 2004, p. 24).

Many researchers advocated that assessment practice is significantly influenced by each individual teacher's own belief (Arkoudis & O’Loughlin, 2004; Davison, 2004; Griffiths et al., 2006; Remesal, 2011; Shim, 2015). Kahn (2000) pointed out that implementing new forms of assessment may be meaningless if teachers’ beliefs about assessment remain the same and unquestioned, especially if teachers, who are the agents of assessments, continue to be ignorant of their own beliefs. Therefore, to put performance assessments into practice as it was initially intended, the first fundamental step has to be the investigation of teachers’ beliefs.

2.2 Teachers' Beliefs about Assessment

In this section, empirical studies on teachers' beliefs about assessment are introduced. Section 2.2.1 elaborates the definition and roles of beliefs in assessments. Section 2.2.2 reviews studies investigating alignment between beliefs and practices in assessments, followed by the introduction of studies on Korean English teachers' beliefs in Section 2.2.3. Finally, section 2.2.4 unfolds the influence of contextual factors on teachers' beliefs and assessment practices.

2.2.1 Teachers' Beliefs and Assessment Practices

It has been widely advocated that teachers' beliefs about assessment have a direct impact on the design, implementation, and interpretation of their student assessments and evaluations. In the literature concerning teachers' beliefs, diverse terms such as values, attitudes, conceptions, and perceptions have been used. The present study chose the term *belief* under the definition proposed by Pajares (1992, p. 316): “an individual's judgment of the truth or falsity of a proposition.” In this study, this proposition comprises the purposes and types of performance assessment as identified by secondary school teachers in their English classrooms in Korea.

It is undeniable that the beliefs of teachers affect their perceptions and

consequently, influence their behaviors in classrooms (Pajares, 1992). Therefore, it is inferable that teachers with more positive or favorable beliefs about performance assessment are apt to accept assessment as a part of their teaching procedure and regularly employ various assessment procedures, while teachers with negative or less favorable beliefs are prone to approach the same activity less positively with a more restricted variety (Rogers, Cheng, & Hu, 2007). Hence, a positive correlation between beliefs and practices in overall school assessment has been empirically studied for many decades.

For instance, Gullickson (1984) surveyed 391 teachers in South Dakota and described that teachers with more positive beliefs about assessment had more agreement on the statement that the classroom tests influence on forming self-concept of students, enhance student interaction, increase student effort, and generally improve the learning environment. Green and Stager (1986) also conducted a longitudinal cross-sectional study with 555 teachers in Wyoming and found out that teachers with positive beliefs toward classroom testing showed a positive effect of testing such as motivating students, directing learning, and identifying problems of students. These results matched with the argument of Chester and Quilter (1998) who asserted that “Teachers must see value in it (assessment) if they are to buy into it” (p. 11). They concluded that developing positive beliefs of assessment was just as important as equipping teachers with the measuring and evaluating techniques.

The same is true in English assessment specifically. Assessing a language is multidimensional in that it demands assessing four main skills of English: reading, speaking, writing, and listening. Since each skill requires different assessment tools, techniques, and approaches, it is the teacher's job to determine when and how to employ specific tools and techniques (Serpil & Derin, 2017). In this complicated decision-making process of assessing a language skill, the influence of teachers' beliefs is magnificent. Therefore, studying English teachers' beliefs about assessment is necessary. Most of all, it allows policymakers and researchers to probe factors that may contribute to the improvement of assessment practices to utilize it as a channel of enhancing teaching and learning English (Elshawa et al., 2017). Furthermore, it is also helpful for English teachers themselves in that providing chances to examine their beliefs that shape practice leads to teachers' own professional development (James & Pedder, 2006).

2.2.2 Alignment between Beliefs and Assessment Practices

As illustrated in the previous section, the positive relationship between individual teachers' beliefs about assessment and their actual assessment practice has been attested by many researchers (Arkoudis & O'Loughlin, 2004; Davison, 2004; Griffiths et al., 2006; Remesal, 2011; Rogers et al., 2007; Shim, 2015). However, within the literature, researchers have been debating

extensively whether it is mandatory to change beliefs so as to change practices, and which change should precede. While Webb et al. (2004) insisted to focus directly on efforts to change practices because changing practices induces change in beliefs, Ajzen and Fishbein (1980) held a contrasting view that change in behavior is determined by changing beliefs first. They asserted that even though changes in beliefs are not enough to guarantee behavioral changes, they serve as the motive to act. What this debate implies is investigating teachers' beliefs about assessment alone is not sufficient. Both practices and beliefs should be scrutinized, and the degree or pattern of alignment between beliefs and practices should also be examined to find out which factor affects the relationship between beliefs and practices. Examining the relationship between beliefs and practices will help teachers to conduct assessment practices in the desired and meaningful way. Above all, inviting teachers to notice the degree of alignment between their beliefs and practices makes them choose their positions as either agents of 'change' or agents of 'conservatism' (James & Pedder, 2006). Therefore, several studies investigated the relationship between beliefs and perceived practices.

James and Pedder (2006) reported mismatches between values and practices of 558 teachers in England. They constructed a questionnaire to investigate how much teachers valued diverse assessment practices and how consistent their practices were with these values. The resulting data showed

that over a half of teachers thought they were carrying out assessments in pursuit of goals they certainly do not consider as valuable or desirable. Teachers seemed to dedicate themselves to the approach of Assessment for Learning, yet they confronted contradictions which needed to be resolved in order to actualize their values.

Winterbottom, Taber, et al. (2008) employed the identical questionnaire items of James and Pedder (2006) with 220 trainee teachers at the University of Cambridge, UK. They compared their results with those of qualified teachers of the study by James and Pedder (2006) and revealed that trainee teachers' gap between values and practices appeared larger than that of qualified teachers. This implied that they implemented practices to a less degree than they viewed suitable for promoting learning. They seemed to encounter more challenges and obstacles in realizing their values since they were still in the course of training and consequently had less autonomy or freedom.

The low degree of alignment between teachers' beliefs and actual practices was also noticeable in EFL/ESL classes. Muñoz et al. (2012) conducted a study of 62 English teachers' beliefs about assessment in general at a private university in Colombia and examined whether those beliefs matched with their practices or not. The instruments for this study were surveys, written reports, and interviews. The results showed a mismatch between what

they believed and what they practiced. Even though teachers held the conception of assessment for improvement, they lacked the ability to apply assessment results for improvement purposes. This outcome carried implications that opportunities for reflections and more guidance on conducting assessment practices were demanded.

Guadu and Boersma (2018) narrowed the scope of their research on formative assessment in writing and investigated beliefs and practices of EFL instructors in Ethiopia. Twenty-five EFL instructors revealed how they taught writing skills via a questionnaire of five-point Likert scale, and additional data were collected through semi-structured interviews. The statistical analysis of the questionnaire displayed a positive and significant correlation between their beliefs and actual practices. However, the qualitative data disclosed that their beliefs and practices were incongruent. Due to the time constraint and large class size, teachers confessed that they had a hard time presenting diverse types of tasks and providing timely feedback to shape learning.

To sum up, empirical studies on beliefs about assessment and its actual practices showed significant inconsistencies between them in many different educational settings outside Korea. In the next subsection, studies on assessment beliefs of Korean English teachers are revisited.

2.2.3 English Teachers' Beliefs in Korea

Korean English teachers' beliefs have drawn substantial research attention over the past decades (Choe, 2005; J. Choi, 2008; S. Choi, 2000; E. Kim, 1997). Among them, a few studies focused on teachers' beliefs about performance assessments. For instance, Kim and Yun (2015) investigated English teachers' perceptions and uses of assessment. They conducted a survey of 58 Korean English teachers across primary and secondary schools in Seoul. The questionnaire examined their perceptions of assessment and its applications. They proved that teachers who were conscious of the educational meaning of performance assessment tended to uphold performance assessment than those who were not.

The contrasting perspective between teachers and students was uncovered by G. H. Lee (2009) who explored teachers' and students' perspectives on middle school English performance tests by using a questionnaire and interview. Teachers and students expressed different opinions about satisfaction, utility, difficulty, and necessity of performance assessment. They also displayed complete inconsistency on the current problems and possible improvements. Based on the findings, this study suggested that detailed information about each type of performance assessment be provided to students and parents, and students' needs and interests be considered when designing performance assessment.

Others noticed diverse problems regarding the implementation of performance assessment. S. Lee (2008) conducted a study on how secondary English teachers perceived and carried out performance assessments. She surveyed 194 English teachers of middle and high schools in Incheon and Seoul. The findings showed that while teachers agreed on the need for performance assessment, several difficulties in implementing performance assessment were unavoidable. The major difficulties were the use of assessment methods that did not serve the purpose of performance assessment and lack of consistency in the process of assessment. Problems with reliability and contextual constraints were still to be resolved as well.

In-depth interviews with six experienced English teachers in different elementary schools disclosed further problems. Kim and Kim (2018) interrogated the perception of elementary school English teachers on the English performance assessment. In spite of the obvious benefits of English performance assessments, teachers confessed some difficulties: they employed test items such as multiple-choice questions along with de-contextualized and fragmented memory questions. Additionally, the frequency of executing performance assessment and providing individualized feedback were overwhelming for teachers since they had only two contact class hours per week with over a hundred students. These findings called for more in-service teacher training and material support.

As such, while investigating teachers' beliefs, most studies about performance assessment in Korea were interested in the actual situation or reality of performance assessment in the classrooms with a focal point on its difficulties or problems. Thus, the preceding research asks for a further study focusing on the relationship or degree of alignment between beliefs and practices in Korean secondary schools to contribute to the development of valid performance assessments.

2.2.4 Contextual Factors, Beliefs, and Practices

Numerous studies have shown that teachers' beliefs and practices are mutually informing. In this "symbiotic relationship" (Foss & Kleinsasser, 1996), it is proven that contextual factors play a crucial role in determining the degree of alignment between the beliefs and practices by teachers (Beach, 1994; Guadu & Boersma, 2018; Tabachnick & Zeichner, 1986). Language teachers' classroom practices are formed by a great extent of interacting and often conflicting factors, resulting in a situation where teachers' practices do not always mirror teachers' beliefs or personal theories (Borg, 2003). Figure 2.1 represents the interactive relationship between beliefs, classroom practices, and contextual factors. As illustrated, contextual factors affect both beliefs and classroom practices, whilst beliefs and practices interact with each other.

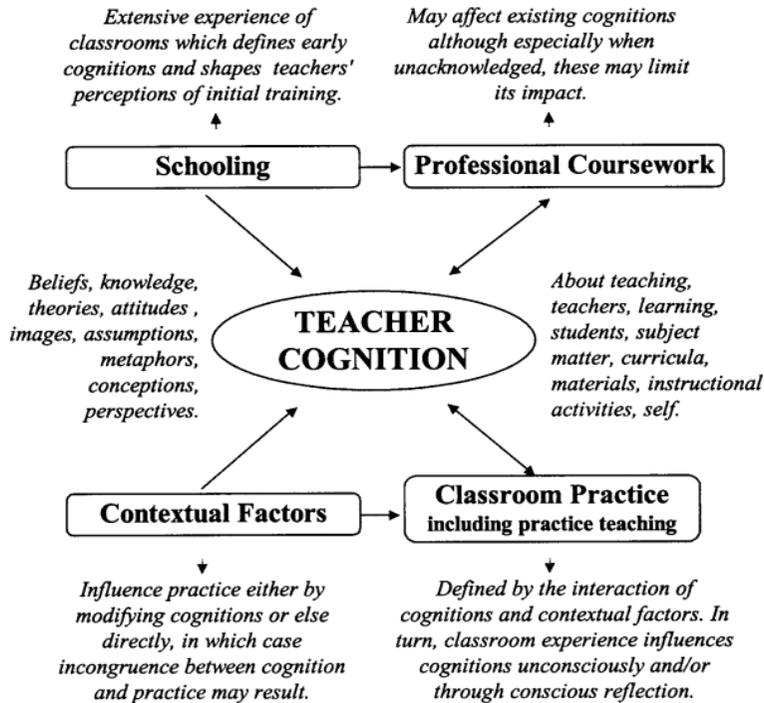


Figure 2.1 Teacher cognition, schooling, professional education, and classroom practice (Borg, 1997, as cited in Borg, 2003, p. 82)

Ample evidence supports this unsurprising power of contextual factors which range from parents, principal, and colleagues to curriculum mandates, standardized tests, and school policies. For instance, Crookes and Arakaki (1999) found that heavy workload affected teachers' pedagogical choices. A study by Richards and Pennington (1998) also revealed the impact of contextual factors which hindered language teachers from employing communicative method as they wanted.

Other researchers discovered the impact of contextual factors on assessment practices especially. Shih (2010) revealed that parental and student factors as well as social and school factors were considered in teachers' decisions on assessment practices. Sung and Jo (2015) examined the constructs of teacher-produced tests of English and disclosed the profound effect of contextual factors such as the high priority of preparing for college admission and test and policy regulations.

Since contextual factors shape what language teachers think and do in their classrooms, a deeper understanding of the contextual factors is necessary to delve into the relationship between beliefs and practices. Teachers built up different experiences regarding performance assessment in distinct schools, and this diversity in school experience is connected to the beliefs of teachers and their practices (Fulmer, Tan, & Lee, 2019). Thus, a mere investigation of beliefs and practices excluding the examination of contextual factors in which the beliefs and practices occur would inevitably yield a partial analysis.

However, relatively little prior research has paid attention to the wide range of possible influence of context on teachers' beliefs and practices of classroom assessment (Fulmer et al., 2019; Sung & Jo, 2015). Reflecting the importance of contextual factors, Fulmer et al. (2019) introduced a framework of the contextual influences on assessment practices by teachers to help researchers investigate the impact from various angles. This theoretical

framework adapted Kozma's (2003) approach to identify influences at three divided levels as illustrated in Figure 2.2.

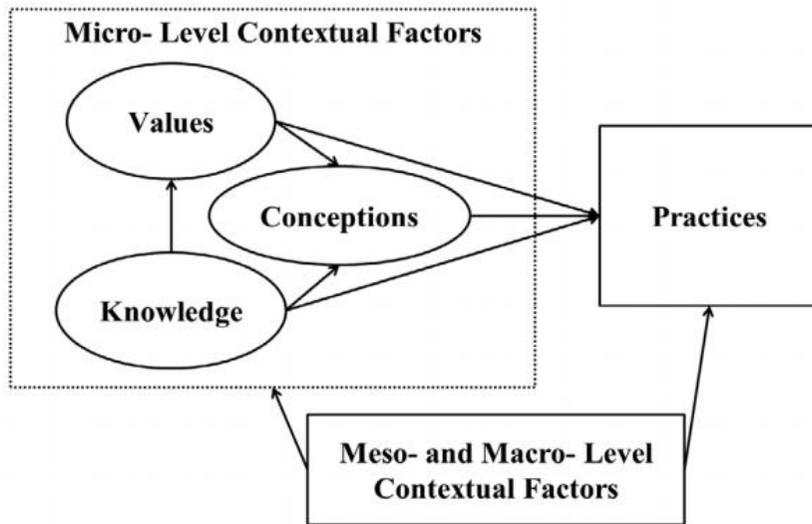


Figure 2.2 Conceptual framework connecting teacher's assessment practices with contextual factors at the micro, meso, and macro level (Fulmer et al., 2019, p. 168)

The framework divided contextual factors into three levels of micro, meso, and macro. The influences at the micro level include the effects of the individual teacher and their classroom, together with their values, conception, and knowledge. The factors at the meso level single out influence within the school like resources, facilities, school's climate, and leadership. Factors regarding the neighboring community or requests and expectations of parents

are also included in this level. The effects at the macro level refer to broad influences of the society, state, or nation. They encompass educational policies and standards, and cultural values. As assessment practice can be completely understood only when considering its contextual factors (Y. H. Kim, 2009), the present study also aimed to investigate contextual factors which might influence the degree of alignment between beliefs and practices of performance assessment.

The current study particularly focused on the investigation of the micro- and meso-level factors. The reason for excluding examination of contextual factors at the macro level is the existence of abundance of previous findings. A great proportion of studies regarding English performance assessment in Korea already claimed that the major reasons which prevent implementation of performance assessment in the initially intended and desired way are the current system of college entrance and the Korean national curriculum system (Bang & Chun, 2011; Kim & Yi, 2013; Koh, 2015; G. Lee, 2009; M. Lee, 2018; S. Lee, 2008; T. J. Park, 2016; Park & Chang, 2016, 2017). Since the system of college entrance and the national curriculum system remain generally the same despite a few changes, these prominent contextual factors at the macro level would be repeatedly mentioned without any new implication. Therefore, instead of reiterating the same results or implications revealed in former research, this study decided to concentrate on contextual

factors at the micro and meso levels.

2.3 Limitations of Previous Research

Extensive research was conducted on ESL/EFL teachers' beliefs and has revealed a substantial extent of the interplay between teachers' beliefs about assessment and their practices. Researchers in Korea also paid considerable attention to exploring Korean English teachers' perspectives and perceptions about assessment and about performance assessment specifically. However, many of them focused mainly on describing the current situation of implementing performance assessment in schools and addressed existing problems regarding performance assessment (M. Lee, 2018; S. Lee, 2008; Lee & Sung, 2017; T. J. Park, 2016; Park & Chang, 2016, 2017). Despite the well-documented significance of teachers' beliefs about assessments, a lack of understanding of Korean English teachers was still observable. When performance assessment aims to be conducted as a reflective cycle of teaching and assessment with high validity, examination of the Korean English teachers' beliefs about performance assessment is imperative (Kim & Kim, 2018). Therefore, the first step for valid and meaningful performance assessment should be a thorough examination of Korean English teachers' beliefs.

Furthermore, in spite of the breadth of existing literature, there remain

a lack of empirical studies on the relationship between beliefs and practices of Korean English teachers and gaps in our understanding of the influence of contextual factors that relate to the degree of alignment between beliefs and practices. It is vital to take into account the socio-cultural contexts where teachers work (Sung & Jo, 2015). Instead of just diagnosing current situations, contextual reasons which might bring differences in the degree of alignment between beliefs and practices should be discovered.

Considering the literature review, the purpose of the present study is to scrutinize teachers' beliefs about performance assessments, the relationship between their beliefs and practices, and the contextual factors which influence the degree of alignment between beliefs and practices. In order to help English teachers conduct performance assessment in a more meaningful and desirable way in the future, contextual factors at the individual- and school-level should be scrutinized through in-depth interviews.

CHAPTER 3.

METHODOLOGY

This study adopted a quantitative and a qualitative approach to examine the research questions. Section 3.1 explains specific information concerning the participants, and Section 3.2 describes instruments employed in the current study. Finally, Section 3.3 outlines the detailed procedures of the study, encompassing the process of data collection and analysis.

3.1 Participants

The mixed-methods approach in this study began with the collection and analysis of quantitative data from a questionnaire and proceeded with the collection and analysis of qualitative data from semi-structured interviews. 109 respondents were recruited for the questionnaire, among whom six teachers participated in the following individual semi-structured interview.

3.1.1 Respondents for Questionnaire

A total of 109 secondary school teachers in Korea were recruited from an online community of in-service English teachers for the online questionnaire. Through online postings, the participants were informed about the purpose of the questionnaire and agreed on participation. Teachers working in specialized schools such as vocational schools, science high schools, foreign language high schools, or autonomous private high schools were excluded from the data collection since the aim of this study is to investigate the beliefs and practices of Korean English teachers in ordinary secondary schools.

Among the total of 109 secondary teachers, 59 teachers were from middle schools and the remaining 50 teachers were from high schools. The detailed background information of the participants is presented in Table 3.1.

Table 3.1

Profiles of Teacher Participants in Questionnaire (n=109)

School level	Middle school 59(54.1%)		High school 50(45.9%)	
Gender	Male 9(8.3%)		Female 100(91.7%)	
Age	20s 8(7.3%)	30s 59(54.1%)	40s 31(28.4%)	50s 11(10.1%)
Degree	Bachelor 74(67.9%)		Master 35(32.1%)	

As Table 3.1 shows, the school level was almost evenly distributed. However, the gender distribution was highly unbalanced, which mirrored the severe shortage of male teachers in the school field. Over half of the teachers were in their thirties accounting for 54.1% whereas the percentage of teachers in their fifties was only 10.1%, possibly due to the survey conducted online. Teachers with master's degree accounted for 32.1 %.

3.1.2 Participants for Semi-Structured Interview

At the end of the questionnaire, a section was provided for teachers to show their willingness to participate in the subsequent individual semi-structured interview. Only those who agreed on the participation of the following interview were the candidates of participants. After analyzing the responses to the questionnaire, a total of six teachers were selected as the interviewees. The detailed information about the interviewees is provided in Section 4.3.

3.2 Instruments

The current study undertook two types of in-depth inquiry. First, a questionnaire survey of 52 items was carried out inquiring teachers' beliefs

about performance assessment and their perceived practice. Second, individual semi-structured interviews were conducted with the selected teachers. A detailed description of the questionnaire and interview questions is provided in the next subsections.

3.2.1 Questionnaire

The questionnaire used in this study was developed by adapting the one used in Elshawa et al. (2017). Since the original questionnaire aimed to investigate teachers' beliefs and practices regarding classroom assessment in the tertiary ESL/EFL context, a modification was necessary to fit the Korean EFL classroom setting at the secondary level. In order to guarantee the validity of the questionnaire for the Korean EFL setting, one professor who is an expert in English assessment and Applied Linguistics provided guidance, and the questionnaire was piloted with two in-service teachers. They offered constructive advice on “the wording of questions, the structure of questions, the response alternatives, and the ordering of questions” (Groves et al., 2009, p. 260). Thus, the modified questionnaire (Appendix 1) was confirmed as a usable and acceptable questionnaire for collecting required quantitative data.

The final version of the questionnaire consisted of 52 items in two sections. The first section of 26 items focused on teachers' assessment ‘beliefs.’

Using a four-point Likert scale, the range of response categories were from one (indicating “Strongly Disagree”) to four (“Strongly Agree”). Thus, higher mean scores indicated a more positive belief about the given aspect of performance assessment while lower mean scores signified less degree of English teachers’ agreement with the description.

The second section of 26 items focused on assessment ‘practices.’ Teachers expressed their opinions whether a particular practice was mostly true, often true, rarely true, or never true with reference to their own classroom assessment practices.

In order to compare and contrast teachers’ beliefs and practices, the content and order of questions in the first section (inquiring ‘beliefs’) and the second section (inquiring ‘practices’) were identical. This comparison was attested to be effective in diagnosing the relationship between present practice and beliefs of teachers (James & Pedder, 2006). Between the first and the second section, there was a part collecting the respondents’ demographic information. It was deliberately placed between the sections, rather than at the end of the questionnaire, to minimize the memory effect when answering the second section since it asks the same content as in the first section.

Twenty-six items in each section consisted of three subsections. Section A was designed for asking about purposes of performance assessment

(ten items). Next, in Section B, a total of eight items were devised to examine the methods and techniques of performance assessment. In the last part of Section C, teachers' feedback, grading, and reporting of grades were investigated with eight items.

Since the questionnaire depended on self-report of beliefs and practices, any possibility of an inconsistency between what teachers claim they do and what they actually carry out can't be denied. However, James and Pedder (2006) contended that the existence of values-practice gaps proved that teachers were prepared to be honest, and they were offering authentic replies about their values and their perceived practices. This study is also based upon this inference.

To remove a possible misinterpretation of the questionnaire items, the questionnaire was translated into Korean, and the Korean version was provided to teacher participants. Approximately ten minutes were required to complete the online questionnaire.

3.2.2 Semi-Structured Interview Questions

Since analyzing responses to a survey inevitably constrains the insights into individuals' experiences and thinking (Winterbottom, Taber, et al., 2008), an oral interview was chosen as an additional method for collecting data

in order to secure a richer and detailed dataset to probe the contextual factors which might influence the relationship between teachers' beliefs and their actual practices concerning performance assessment. Therefore, a total of six teachers were randomly selected to conduct the interview aiming to uncover influential contextual factors. Individual interviews were carried out in the Korean language to reduce the affective filter of participants.

This study applied a type of semi-structured interview to collect data. The interview questions were developed in reference to the Contextual Factors Questionnaire item (Fulmer et al., 2019). They were modified repeatedly via pilot study responses attained from three teachers who did not participate in the interview section. The final version of the interview questions is presented in Appendix 2.

3.3 Procedures

This section presents the detailed procedure of data collection. Section 3.3.1 elaborates the entire procedure of data collection, while Section 3.3.2 describes steps of data analysis.

3.3.1 Data Collection

After winning the approval from the Ethics Committee of the Institutional Review Board (IRB) of Seoul National University, the data collection started. 109 teacher participants were recruited through an online community of Korean English teachers, and a link to the Performance Assessment Beliefs & Practices Questionnaire was sent via messenger to the participants. A written description of this research was attached, and all the inquiries by the participant were answered. The participants were also notified that they have the right to discontinue participation in the study at any time. Information about the participants' background concerning gender, age, school level, degree, etc. was gathered through the questionnaire.

After analyzing responses to the questionnaire, a total of six teacher participants were chosen for the subsequent in-depth interview. Due to the COVID-19, reducing face-to-face contact was highly recommended, and hence, the interview was conducted by conversing through either online video conference or telephone, depending on the choice of the participants. The participants received an oral introduction to this study and were notified of their right of withdrawal from the interview. The participants received the consent form via e-mail and filled it. All the interviews were individually conducted, and each of them lasted approximately twenty minutes. All the interviews were audio-recorded and then transcribed.

3.3.2 Data Analysis

Quantitative and qualitative analyses were employed to examine the research questions. For quantitative methodology, SPSS 26 was used to conduct descriptive statistics and cluster analysis. For qualitative methodology, interviews were transcribed and analyzed based on thematic analysis.

3.3.2.1 Quantitative Analysis

The data gathered through the Performance Assessment Beliefs & Practices Questionnaire were analyzed. First, the responses about teachers' beliefs collected from item 1 to item 26 were descriptively analyzed. The response categories ranged from 1 (referring to "Strongly Disagree") to 4 ("Strongly Agree") using a four-point Likert scale. Their means and standard deviations were calculated on each item. Next, the responses about teachers' practices, which were examined through the questionnaire item 27 to 52, were descriptively analyzed. The response categories for assessment practices were also from 1 (referring to "Never True") to 4 ("Mostly True"), and the means and standard deviations of each item were measured.

Another item-specific level of analysis was conducted to diagnose the relationship between beliefs and practices of teachers. On each item, the mean score of practice was subtracted from the mean score of belief. As a result, a

positive number suggested that teachers valued a practice more than it was used while a negative number indicated that teachers used a practice even though their beliefs were against it.

Next, a cluster analysis using both teachers' beliefs factor and practices factor scores of the questionnaire was applied to unfold profiles of teachers' beliefs and practices of performance assessment. This helped explore variation amongst teachers more in detail. Different patterns of relationships between beliefs and practices were identified, and similarities and differences between clusters were discussed in the following chapter.

3.3.2.2 Qualitative Analysis

Interview data were analyzed qualitatively in order to uncover contextual factors which affected the degree of consistency between beliefs and practices. As a first step, recordings of participants' interviews were transcribed in verbatim. Since the interview was conducted in Korean, all verbal report was transcribed firstly in Korean and then translated into English. Then, guided by thematic analysis to detect themes which are the patterns of explicit and implicit content (Joffe, 2012), recurring themes were identified. Repeated analysis of the data redefined these themes and identified the final set of themes, which will be elaborated on in the next chapter.

CHAPTER 4.

RESULTS

This chapter presents the results of data analysis in the order of the research questions posed in Chapter 1. The first two sections display the results of quantitative analysis of the questionnaire data while the last section presents qualitative results from the data collected through interviews. First, Section 4.1 describes teachers' beliefs about the performance assessment with its three subsections. It further describes teachers' assessment practices and reports the relationship between teachers' beliefs and practices regarding performance assessment. Next, Section 4.2 presents the result of the cluster analysis identifying patterns of teachers' beliefs and practices. Finally, Section 4.3 reveals contextual factors affecting the degree of alignment between teachers' beliefs and practices at the micro and meso levels.

4.1 Teachers' Beliefs and Assessment Practices

Regarding the first and the second research question, Korean English teachers' beliefs were examined through the first half of the questionnaire

(from item 1 to 26) while their performance assessment practices were investigated via the second half of the questionnaire (from item 27 to 52). The extent of congruence between teachers' beliefs and practices was identified at an item-specific level of analysis comparing the mean scores of beliefs and practices of each questionnaire item. Before looking at the detailed item-level analysis, an overview of teachers' beliefs, their practices, and the difference between beliefs and practices are presented according to the three subsections of the questionnaire in Table 4.1.

Table 4.1

Mean Comparison of Three Subsections (n = 109)

Sec.	Content	Belief		Practice		Mean diff.
		Mean	SD	Mean	SD	
A	Instructional Purposes	3.45	0.61	3.02	0.77	0.43
B	Methods and Techniques	3.18	0.62	2.74	0.84	0.44
C	Feedback, Grading, and Reporting of Grades	3.26	0.61	2.95	0.74	0.31
	total	3.30	0.61	2.90	0.78	0.39

The greatest mean scores of both beliefs and practices were observed in Section A which investigated the instructional purposes of performance assessment. On the other hand, the lowest mean scores of beliefs and practices were in Section B asking about methods and techniques of performance assessment. While the scores of beliefs and practices in Section C, which

inquired feedback, grading, and reporting of grades, was the second-highest among the three sections, its mean difference between beliefs and practices recorded the lowest, meaning the highest degree of alignment between beliefs and practices among the three subsections in the questionnaire.

The specific description of teachers' beliefs, their practices, and the mean difference between beliefs and practices is presented with tables in the order of the three subsections of the questionnaire as in the following.

4.1.1. Instructional Purpose of Performance Assessment

The data from the Section A of the questionnaire showed whether teachers believed that performance assessment should be employed for diverse purposes. Their perceived practices and the mean difference between beliefs and practices were also analyzed. Table 4.2 shows means and standard deviations of English teachers' beliefs about the instructional purposes of assessment and their practices respectively. The difference between the mean scores of beliefs and those of practices is also included. Since survey item 6 demonstrates a negative statement, reverse coding is employed.

The descriptive analysis revealed that the majority of the teachers appeared to hold positive beliefs regarding using performance assessment to inform teaching and improve learning. This could be inferred through the

overall mean of 3.45 and standard deviation of 0.61 presented in Table 4.2.

Table 4.2

Instructional Purposes (n = 109)

Item #	Statement (Performance assessment...)	Belief		Practice		Mean diff.
		Mean	SD	Mean	SD	
1	helps to focus on teaching	3.55	0.60	2.91	0.77	0.64
2	helps to group students for instructional purposes.	2.73	0.88	2.84	0.78	-0.11
3	diagnoses strengths and weaknesses in students.	3.51	0.65	3.13	0.72	0.38
4	diagnoses strength and weaknesses in teaching.	2.86	0.85	2.78	0.85	0.08
5	provides information about students' progress.	3.62	0.52	3.14	0.70	0.48
6	creates competition among students. (reverse coded)	3.39	0.69	2.60	0.94	0.79
7	creates a valuable learning experience for my students.	3.80	0.40	3.34	0.67	0.46
8	motivates my students to learn.	3.72	0.45	3.19	0.76	0.53
9	provides feedback to students as they learn.	3.69	0.47	3.07	0.81	0.62
10	demonstrates the level of students' learning.	3.61	0.56	3.22	0.71	0.39
	total	3.45	0.61	3.02	0.77	0.43

The highest level of agreement was on the statement that performance assessment creates a valuable learning experience for the students (item 7) with the overall mean of 3.80 and standard deviation of 0.40. This statement also scored the highest mean score (3.34) and the lowest standard deviation score

(0.67) in assessment practices. The second-highest agreement (3.72) was on the belief that performance assessment motivates students to learn (item 8), which occupied the third highest mean scores in assessment practices (3.19) with the standard deviation of 0.76.

However, among the 109 participants, opinions regarding the purpose of performance assessment as helping to group students for instructional purposes (item 2) and diagnosing strength and weaknesses in teaching (item 4) varied recording low belief scores of 2.73 and 2.86 respectively. It appeared that although teachers believed that performance assessment should be employed for informing instruction in a way to improve students' learning, some of the teachers seemed skeptical about the role of performance assessment to provide feedback on their teaching practice and grouping students.

Looking at the difference between beliefs and practices, the range of mean difference in absolute value was from 0.08 to 0.79. To solely compare the size of the mean difference of each item, absolute values were used, and the interpretation of positive and negative number was added. The highest and the most noticeable difference was detected regarding the competition that performance assessment created (item 6) with the mean difference of 0.79. Even though many teachers believed that performance assessment should not create competition among students, teachers admitted that it did create

competition with the relatively high standard deviation of 0.94. This may reflect the highly competitive atmosphere in Korean secondary schools. The second-largest difference of means (0.64) was observed for the purpose of performance assessment as helping to focus on teaching (item 1). The difficulty to link performance assessment with teaching was revealed here. The third biggest mean difference (0.62) was noticed regarding students' learning. The practice of providing feedback to students as they learn (item 9) didn't meet the level of teachers' beliefs.

On the other hand, the least mean difference was detected in item 2 with negative 0.11 (helps to group students for instructional purposes) and item 4 with 0.08 (diagnoses strength and weaknesses in teaching). Interestingly, the mean scores of beliefs of these two items were the lowest among the ten items in Section A. Teachers' low level of beliefs seemed to have been transferred to the actual practices. In item 2 especially, the negative value of the mean difference might be interpreted that teachers might have failed to recognize their belief even though they were practicing it.

4.1.2 Methods and Techniques

This section presents the results of teachers' beliefs and practices about the appropriate and desired methods and techniques of performance

assessment, which was dealt with from item 11 to item 18 in Section B of the questionnaire. Table 4.3 provides the means and standard deviations of teachers' beliefs and practices. In the farthest to the right column, the difference between the mean scores of beliefs and those of practices is presented respectively. Reverse coding is employed for survey items 11 and 15, which demonstrated negative statements, for precise interpretation.

The descriptive analysis unveiled that teachers generally favored to employ various forms and types of performance assessment. This could be inferred from the overall mean of 3.18 and standard deviation of 0.62. However, their actual practices fell behind their beliefs with the overall mean of 2.74 and standard deviation of 0.84, recording the overall mean difference of 0.44.

Most of the teachers agreed strongly that sufficient time should be allowed to properly prepare performance assessments (item 14), recording the top mean score of beliefs (3.86) and the least standard deviation (0.35) throughout the whole questionnaire. Unfortunately, this same statement recorded the top mean difference between beliefs and practices as well. As many as 1.63 mean difference was observed, meaning only a few teachers were lucky to have sufficient preparation time for performance assessment in classrooms.

Table 4.3***Methods and Techniques (n = 109)***

Item #	Statement	Belief		Practice		Mean diff.
		Mean	SD	Mean	SD	
11	Paper and pencil assessment (e.g. dictation, filling in the blanks of conversation) is the best method in evaluating students' speaking and listening. (reverse coded)	2.54	0.76	2.78	1.06	-0.24
12	Performance assessment questions should reflect real life language use.	3.40	0.56	3.06	0.69	0.34
13	Language teachers need to use a variety of assessment methods to assess students.	3.57	0.52	3.12	0.73	0.45
14	Sufficient time should be allowed to properly prepare performance assessments.	3.86	0.35	2.23	0.89	1.63
15	Selected-response items (e.g. matching items, multiple-choice items, true - false items) are good methods of performance assessment. (reverse coded)	3.08	0.68	3.45	0.81	-0.37
16	Constructed-response items (e.g. journal entry, portfolio, short essay, sentence completion, reflective task) are good methods of performance assessment.	3.52	0.54	3.57	0.64	-0.05
17	Self-assessment by the student is a good method of performance assessment.	2.77	0.81	1.87	0.93	0.90
18	Peer-assessment is a good method of performance assessment.	2.71	0.77	1.84	0.93	0.87
	total	3.18	0.62	2.74	0.84	0.44

On the other hand, item 16 (constructed-response items are good methods of performance assessment) showed the third highest belief score

(3.52) and the highest practice score (3.57) resulting in the highest degree of alignment between beliefs and practices with the mean difference of negative 0.05 in Section B. It also ranked the second-highest place in the degree of alignment between beliefs and practices throughout the whole questionnaire. The high belief score was apparently transferred into classroom practice. Many schools seemed to employ various kinds of constructed-response items such as journal entry, portfolio, short essay, sentence completion, and reflective tasks, which were in line with teachers' beliefs.

Another interesting finding was the Korean English teachers' controversial beliefs about self-assessment (item 17) and peer-assessment (item 18) with the belief score of 2.77 and 2.71 respectively. It was disclosed that teachers held a skeptical view about employing these two methods in performance assessment, yielding an even lower rate of actual classroom practice, 1.87 for self-assessment and 1.84 for peer-assessment with the high standard deviation of 0.93 for both items. Regarding the degree of alignment between beliefs and practices, self-assessment (item 17) recorded the third-highest difference (0.90) while peer-assessment (item 18) marked the fourth-highest difference (0.87) among the 26 items in the whole questionnaire. The issue of self- and peer-assessment will be addressed in the discussion section.

4.1.3 Feedback, Grading, and Reporting of Grades

This section presents results from Section C of the questionnaire, which was related to the English teachers' beliefs about feedback, grading, and reporting of grades. Descriptive statistics about beliefs and practices are presented in Table 4.4. Mean differences between beliefs and practices are provided in the farthest to the right column.

The highest mean scores of both beliefs (3.83) and practices (3.86) in Section C were monitored in item 24, which stated that students should be informed about the marking criteria before being assessed. Obtaining high agreement and high rate of practice simultaneously, it recorded the highest degree of consistency between beliefs and practices throughout the questionnaire with the mean difference of negative 0.03.

The second-highest score in beliefs (3.74) and practices (3.83) in this subsection was about preparing a marking scheme before the assessment (item 20). This statement also won the fourth-highest degree of alignment between beliefs and practices in the whole questionnaire with the mean difference of negative 0.09.

Although teachers mostly agreed on when to prepare and inform marking criteria as stated above, they had divergent beliefs regarding students' position in the marking process (item 25). The score of that belief (2.20)

recorded the lowest throughout the whole questionnaire, while the score of the given practice (1.37) also had the same position. Even though ranking the lowest position in both belief and practice, the degree of alignment was also surprisingly low with the mean difference of 0.83. This result implied that Korean English teachers were unconvinced about involving students in designing the marking criteria, which led to the even fewer frequency of practice.

Table 4.4

Feedback, Grading, and Reporting of Grades (n = 109)

Item #	Statement	Belief		Practice		Mean diff.
		Mean	SD	Mean	SD	
19	Student effort is seen as important when assessing their learning.	3.45	0.57	3.18	0.80	0.27
20	A marking scheme should be prepared before assessment is given.	3.74	0.48	3.83	0.40	-0.09
21	Conferencing with students is a good way of giving feedback.	3.25	0.61	2.25	0.95	1.00
22	Criterion-referenced assessment is better than norm-referenced assessment.	3.26	0.76	3.40	0.94	-0.14
23	Students should be given feedback after performance assessment.	3.63	0.50	3.14	0.87	0.49
24	Students should be informed about the marking criteria before being assessed.	3.83	0.40	3.86	0.37	-0.03
25	Students should be involved in preparing the marking criteria.	2.20	0.71	1.37	0.65	0.83

26	Students should be given back their performance assessment results no later than a week after the assessment.	2.70	0.83	2.57	0.95	0.13
	total	3.26	0.61	2.95	0.74	0.31

Another noticeable mean difference between beliefs and practices was regarding conferencing with students as a way of giving feedback (item 21). This statement recorded the second biggest mean difference (1.00) in the whole questionnaire. Although many teachers agreed that conferencing with students was a good way of giving feedback on performance assessment with the belief score of 3.25 and standard deviation of 0.61, they replied that they rarely held a conference with students to give feedback in classrooms recording the practice score of 2.25 with a relatively high standard deviation of 0.95.

To investigate the obtained data more thoroughly, taking a broader view was necessary in addition to this item-level analysis. Thus, patterns of beliefs and practices were examined and analyzed in the next subsection to look at the bigger picture of the same data.

4.2 Patterns of Beliefs and Practices

To detect possible patterns of relationship between teachers' beliefs and practices guided under the third research question, cluster analysis was undertaken. Instead of using the difference values between beliefs and

practices for clustering, this research decided to use all the individual scores of teachers' beliefs and practices from item 1 to item 52 of the questionnaire. The reason for this choice was that the belief-practice difference values didn't convey the necessary information of desirableness about the beliefs or practice of performance assessment. Hence, teachers' beliefs and practices scores of the questionnaire were used for clustering, and this led to group individual teachers based on their similarities in both their beliefs and practices on each statement in the questionnaire.

Among many ways to compute the similarity of the clusters, Ward's linkage algorithm was opted to divide teachers into clusters. The minimization of within-cluster variance is the main criterion for forming clusters through this method, and its reliability was attested in reviews of different approaches to grouping cases (James & Pedder, 2006; Strauss & von Maltitz, 2017; Szekely & Rizzo, 2005). Not presuming any set number of clusters in advance, attempts were made to detect the proper number of clusters. Using SPSS 26, the dendrogram of hierarchical cluster analysis according to Ward's linkage was obtained, and the possible number of clusters were two, three, and four as observed in Figure 4.1.

Each possibility was examined closely. If dividing the participants into two clusters, the one cluster would be consisted of 70 teachers, whereas the other cluster would include only 39 teachers. On the other hand, if the

participants were divided into four clusters, the number of members in each cluster would be 48, 22, 14, and 25. These extremely unbalanced numbers of members in clusters would inhibit proper investigation of diverse pattern of beliefs and practices. Therefore, the best number of clusters evidenced by the dendrogram was decided to be three.

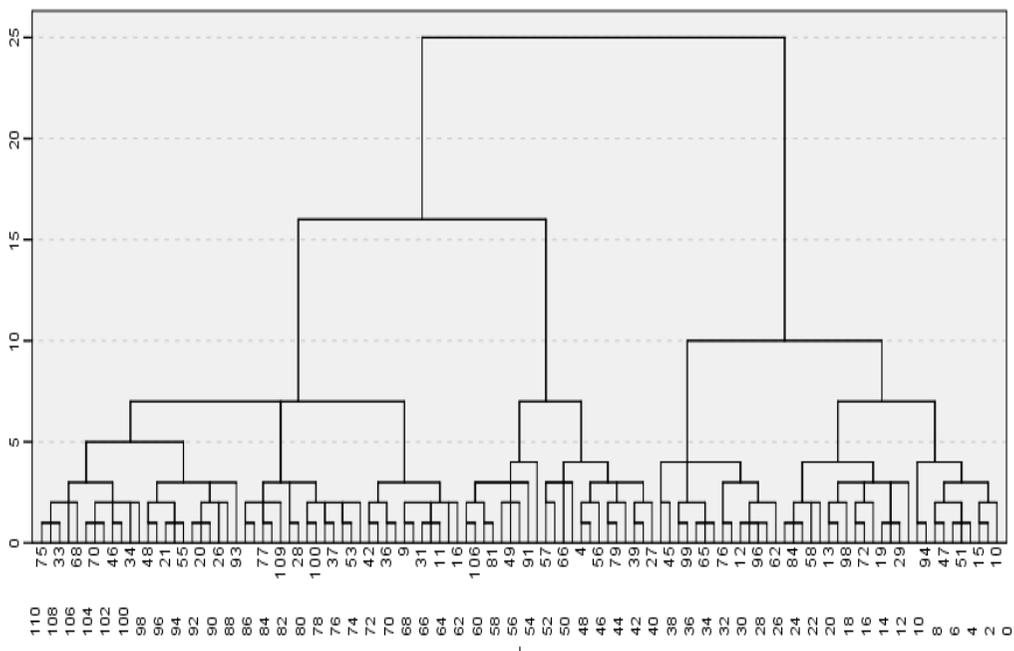


Figure 4.1 Dendrogram of cluster analysis

The three identified clusters were named ‘Typical Teachers,’ ‘Teachers in Dilemma,’ and ‘Progressive Teachers.’ Table 4.5 provides demographic information about each cluster: number of teachers in each cluster, their school level, age, degree and teaching experience are presented.

Table 4.5

Demographic Information of Clusters

Cluster	No.	School level		Gender		Age				Degree		Teaching experience (years)				
		Mid-dle	High	M	F	20s	30s	40s	50s	Bachelor	Master	- 5	5-10	11-15	16-20	20 -
Typical Teachers	48 (44)	31 (65)	17 (35)	2 (4)	46 (96)	4 (8)	25 (52)	14 (30)	5 (10)	35 (73)	13 (27)	7 (15)	17 (35)	15 (32)	4 (8)	5 (10)
Teachers in Dilemma	22 (20)	11 (50)	11 (50)	1 (5)	21 (95)	0 (0)	21 (95)	1 (5)	0 (0)	17 (77)	5 (23)	1 (5)	15 (67)	5 (23)	0 (0)	1 (5)
Progressive Teachers	39 (36)	17 (44)	22 (56)	6 (15)	33 (85)	4 (11)	13 (33)	16 (41)	6 (15)	22 (56)	17 (44)	5 (13)	13 (33)	7 (18)	9 (23)	5 (13)
Total	109 (100)	59 (54)	50 (45)	9 (8)	100 (92)	8 (7)	59 (54)	31 (28)	11 (11)	74 (68)	35 (32)	13 (12)	45 (41)	27 (25)	13 (12)	11 (10)

Note: The numbers in parenthesis indicate the percentage.

Next, Table 4.6 shows descriptive statistics of the identified clusters providing means and standard deviation of beliefs and practices of each cluster. Each cluster's mean difference between beliefs and practices is calculated in the farthest to the right column of the table.

Table 4.6

Descriptive Statistics of Clusters

Cluster	Number	Belief		Practice		Mean diff.
		Mean	SD	Mean	SD	
Typical Teachers	48 (44%)	3.26	0.51	2.96	0.70	0.30
Teachers in Dilemma	22 (20%)	3.13	0.49	2.39	0.64	0.74
Progressive Teachers	39 (36%)	3.46	0.44	3.16	0.56	0.30

The first cluster was labeled as 'Typical Teachers.' As Table 4.6 shows, they manifested a moderate degree of beliefs about performance assessment (Mean = 3.26) aligned with the Assessment for Learning approach, and their practice scores (Mean = 2.96) were also moderate compared to other clusters. The mean difference between beliefs and practices was 0.30, which was the same amount of difference as 'Progressive Teachers.' 'Typical Teachers' accounts for 44% of the respondents with 48 teachers, among whom are 31 middle school teachers and 17 high school teachers. The distribution of age and teaching experience is almost equivalent to that of total respondents without any noticeable difference. In contrast, the proportion of the number of

teachers with master's degrees falls short of that of total respondents as shown in Table 4.5.

The second cluster was named 'Teachers in Dilemma.' As shown in Table 4.6, their beliefs and practices scored the lowest, the mean of 3.13 and 2.39 respectively, compared to other clusters. At the same time, the mean difference between beliefs and practices was the largest among the three identified clusters with 0.74 of mean difference. Hence, this cluster with 20% of the respondents could be considered as 'Teachers in Dilemma' presenting the highest level of inconsistency between their beliefs and practices. Among the 22 teachers, half were working in middle schools, and the other half were working in high schools as presented in Table 4.5. Surprisingly, all the teachers in this cluster were female and were in their thirties except one male teacher who was the only one in his forties with more than twenty years of teaching experience. The proportion of the number of teachers with master's degrees was the lowest among the three clusters while showing a limited range of the years of teaching experience from 5 to 15 except two teachers one of whom was the above-mentioned male teacher.

The third and the last cluster was labeled as 'Progressive Teachers.' Even though the mean difference between beliefs and practices was identical as 'Typical Teachers,' it was distinguished from 'Typical Teachers' in that they showed the highest belief score (3.46) and practice score (3.16) for conducting

performance assessment for improving teaching and learning pursuing Assessment for Learning approach as shown in Table 4.6. Considering this fact, they could be called “Progressive Teachers” who endeavor to promote teaching and learning through performance assessment. Table 4.5 shows that thirty-six percent of the respondents belong to this cluster while recording the highest proportion of high school teachers compared to other clusters. The proportion of male teachers and teachers with master’s degrees also outnumbered that of the total participants. Looking at the distribution of age and years of teaching experience, ‘Progressive Teachers’ showed a slightly higher proportion of teachers over forties with more than 16 years of teaching experience.

To delve into the pattern of beliefs and practices of teachers in each cluster, graphs are used to visually present the necessary information at an item-specific level. Profiles for each three cluster of teachers were developed, and each cluster had their idiosyncratic profiles derived from their beliefs about performance assessment and their practices. In each graph, dividing lines are included to ease the comparison of patterns between the three subsections of the questionnaire, which are the purposes; methods and techniques; and teachers’ feedback, grading, and reporting of grades. The titles of subsections are provided in boxes at the top of each graph. All the three clusters’ beliefs, practices, and differences were compared directly in a single graph respectively to fully understand each cluster.

First, on the basis of the mean scores of beliefs of each cluster, Figure 4.2 is produced. As shown, all three clusters had a similar pattern of beliefs.

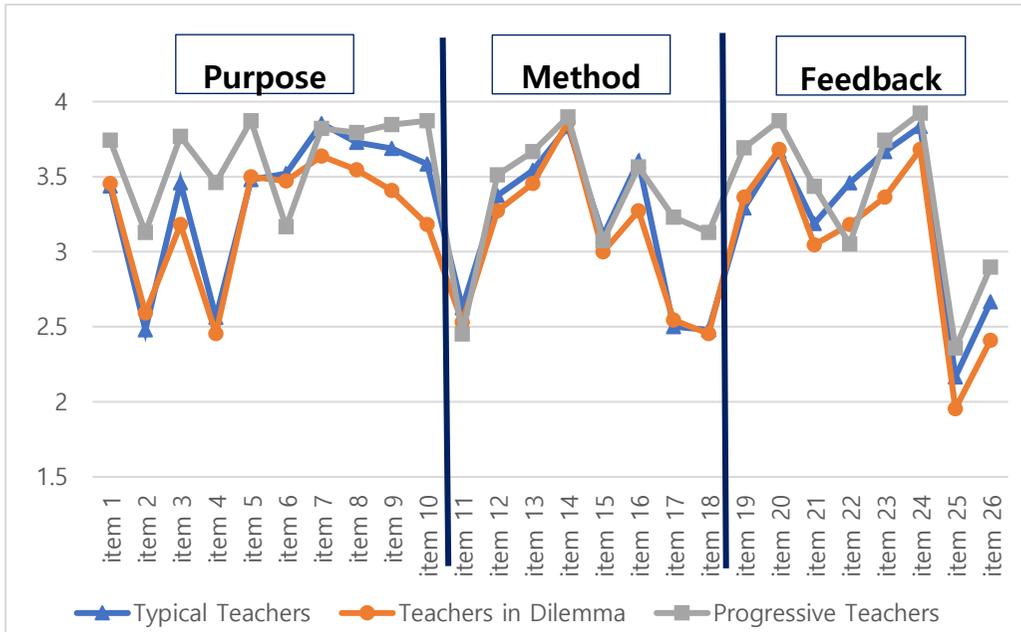


Figure 4.2 Comparing beliefs of clusters

However, the degree of their beliefs differed according to clusters. In most items, ‘Typical Teachers’ showed a moderate degree of beliefs while ‘Teachers in Dilemma’ demonstrated the lowest level of beliefs and ‘Progressive Teachers’ the highest level of beliefs. Especially ‘Progressive Teachers’ exhibited strong beliefs about performance assessment pursuing the goal of assessment for learning. They revealed extremely high agreement on the purpose of performance assessment as informing instruction (item 1, 2, 4, 5) as well as improving learning (item 3, 9, 10). Employing various methods

like self-assessment (item 17), peer-assessment (item 18), and conferencing with students as a way of giving feedback (item 21) were also considered desirable by ‘Progressive Teachers.’ Interestingly, however, they showed the least agreement with item 22, which stated that criterion-referenced assessment is better than norm-referenced assessment.

Next, the practice of each cluster was analyzed. Here, the differences between clusters were even more striking.

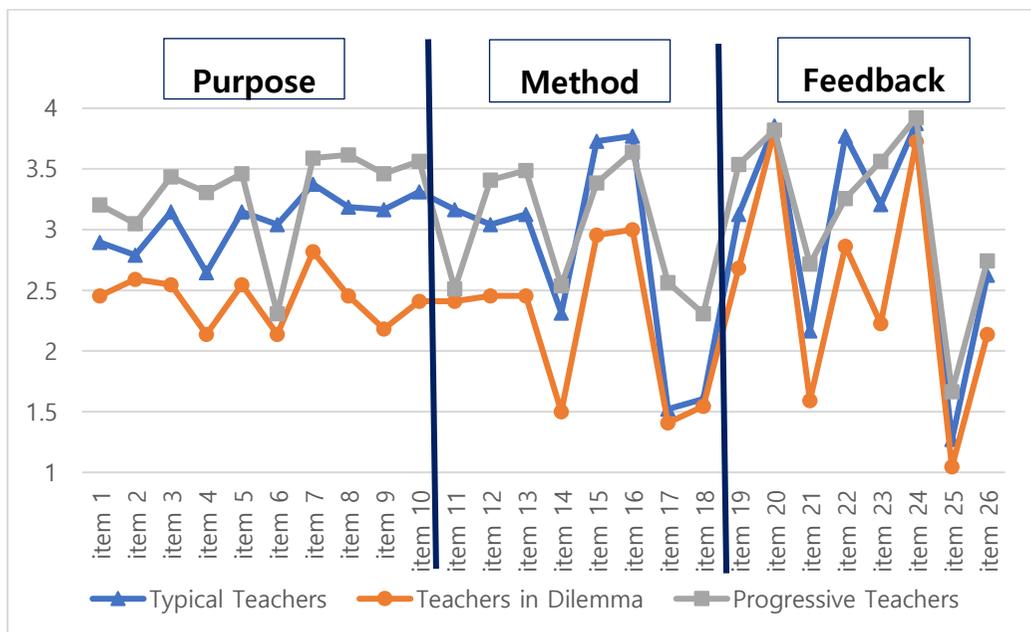


Figure 4.3 Comparing practices of clusters

Even though the ‘Typical Teachers’ and ‘Teachers in Dilemma’ held similar beliefs about performance assessment as Figure 4.2 showed, their actual practices in overall items were quite distinguishable in that ‘Teachers in

Dilemma' demonstrated the lowest level of practice regarding performance assessment for improving teaching and learning as displayed in Figure 4.3. In contrast with 'Teachers in Dilemma,' 'Progressive Teachers' conducted performance assessments to improve teaching and learning. They also conducted self- (item 17) and peer-assessment (item 18) markedly more than the other groups of teachers, which matched their beliefs as in Figure 4.2. They achieved the highest level of practice in areas which conform to the Assessment for Learning approach even though they used paper and pencil assessment in evaluating students' speaking and listening (item 11) and selected-response items (item 15). Another interesting point was that they somewhat created competition among students during performance assessment (item 6). Even though both 'Progressive Teachers' and 'Teachers in Dilemma' created competition among students (item 6) in classrooms, a difference was noticeable: 'Teachers in Dilemma' practiced it against their beliefs whereas 'Progressive Teachers' followed their beliefs and carried them into practice. This dissimilar level of alignment between beliefs and practices led to develop next figure which compared mean differences between clusters.

The size of mean difference between beliefs and practices was more clearly disclosed when looking at Figure 4.4. It directly compared the value of mean differences between beliefs and practices of each cluster in one graph.

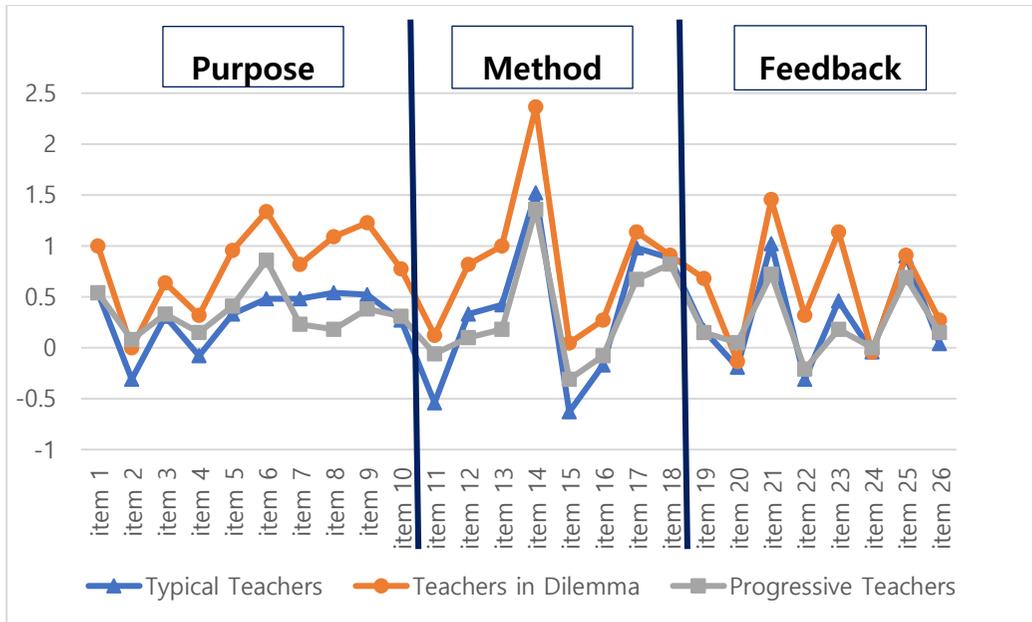


Figure 4.4 Comparing differences of clusters

Values below the zero point mean that they practiced the given aspect more than they believed to be right, whereas values above the zero point indicate that belief was transferred into practice less than they valued. The nearer the values are to the zero point, the smaller the mean difference between beliefs and practices is. The absolute value of each difference was the largest in ‘Teachers in Dilemma’ in most areas.

To compare beliefs and practices of each cluster more easily, the following figures (Figure 4.5, 4.6 & 4.7) are additionally presented which compared each cluster’s beliefs and practices in one graph. Comparing these three figures made it possible to intuitively notice the biggest difference

between beliefs and practices existing in 'Teachers in Dilemma.' The overall failure to achieve consistency between beliefs and practices for most of the areas in 'Teachers in Dilemma' was clearly observed in Figure 4.6, even though they were able to practice performance assessment in line with their beliefs in some areas such as in item 2 (Performance assessment helps to group students for instructional purposes), item 11 (Paper and pencil assessment is the best method in evaluating students' speaking and listening), item 15 (Selected-response items are good methods of performance), item 20 (A marking scheme is prepared before assessment is given), and item 24 (Students are informed about the marking criteria before being assessed). Interestingly, looking at all participants' mean differences between beliefs and practices of the above-mentioned items provided in tables in Section 4.1, all the mean differences were negative numbers ranging from -0.37 to -0.03, indicating that teachers practiced it more than they believed it right. Although 'Teachers in Dilemma' demonstrated quite a similar pattern and degree of beliefs as 'Typical Teachers' as observed in Figure 4.5 and 4.6, practice level of 'Teachers in Dilemma' showed a contrasting pattern compared to other clusters presenting the lowest degree of alignment between beliefs and practices.

In contrast, 'Progressive Teachers' displayed the highest degree of overall alignment between beliefs and practices as shown in Figure 4.7. However, some items displayed a significant mismatch between beliefs and

practices. These were item 6 (Performance assessment creates competition among students), item 14 (Sufficient time should be allowed to properly prepare performance assessments), item 17 (Self-assessment by the student is a good method of performance assessment), item 18 (Peer-assessment is a good method of performance assessment), item 21 (Conferencing with students is a good way of giving feedback), and item 25 (Students should be involved in preparing the marking criteria). The fact that the range of these six items' mean difference of all teacher participants was from positive 0.79 to 1.63 as presented in Section 4.1 means that not only 'Progressive Teachers' but also most Korean teachers tend to suffer the failure of putting these beliefs into practices.

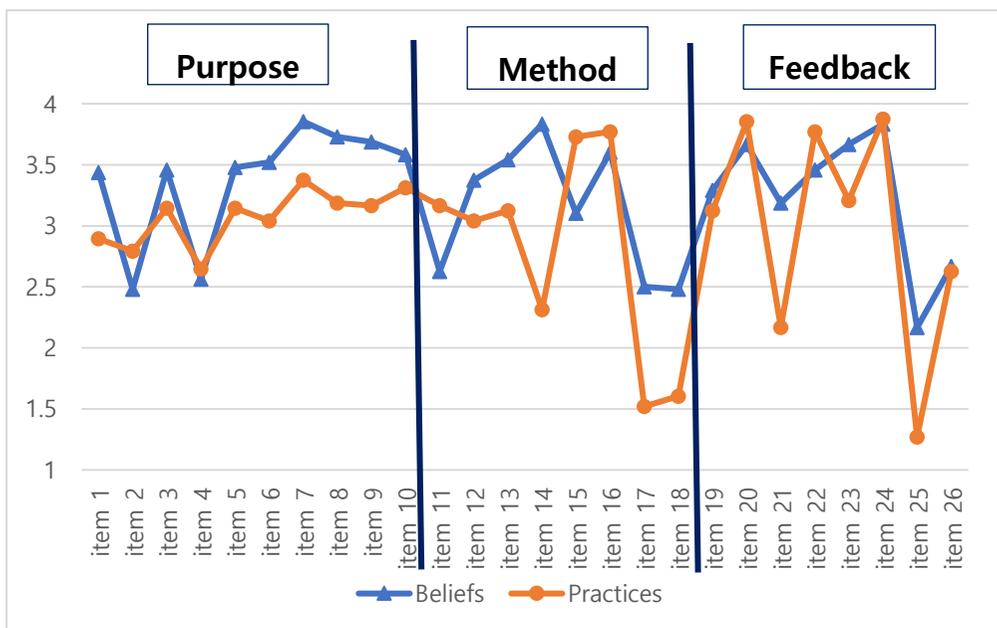


Figure 4.5 Comparing beliefs and practices of 'Typical Teachers'

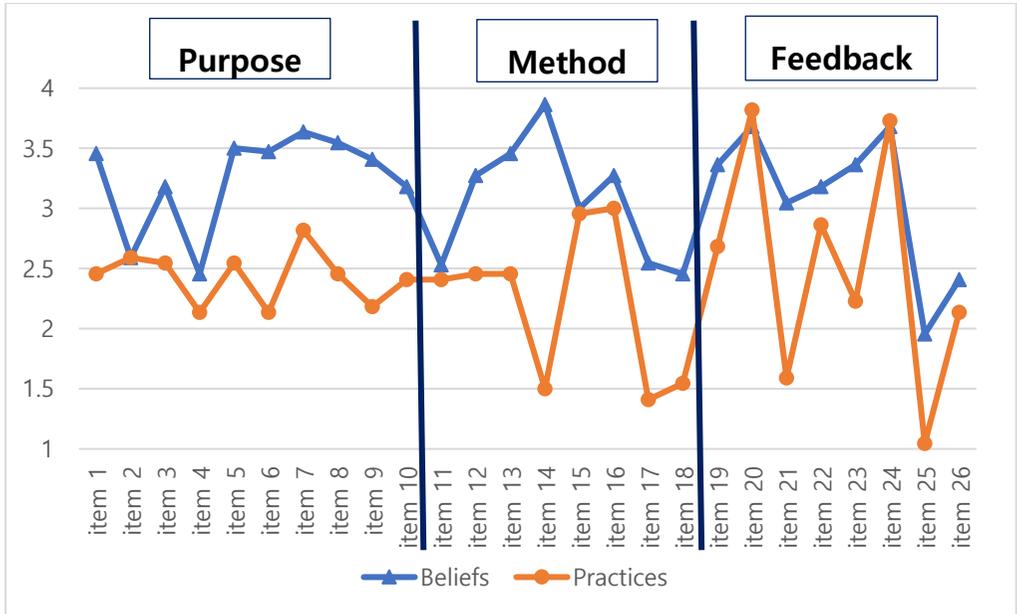


Figure 4.6 Comparing beliefs and practices of ‘Teachers in Dilemma’

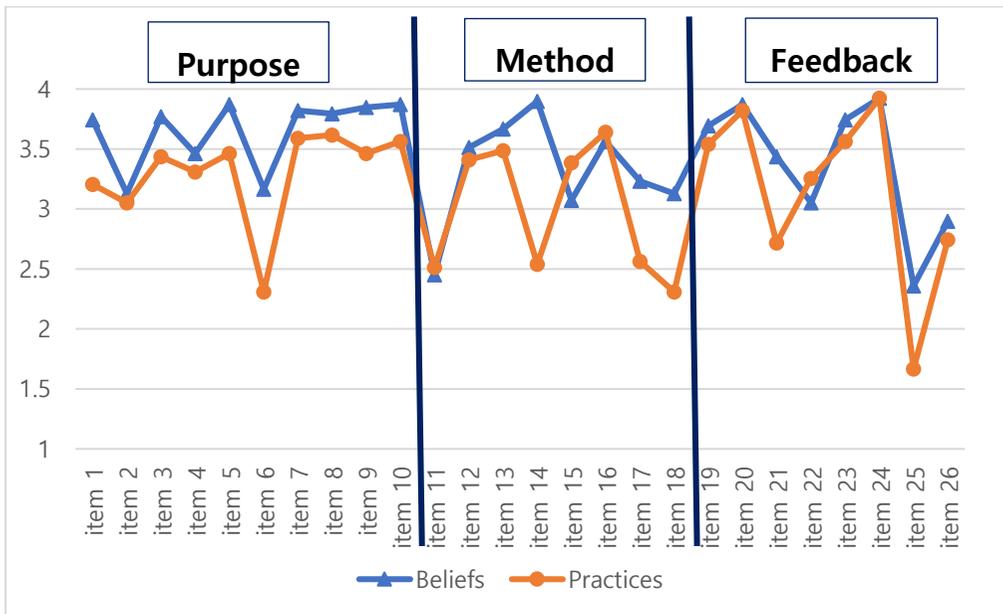


Figure 4.7 Comparing beliefs and practices of ‘Progressive Teachers’

This identification of clusters with different patterns stimulated further research questioning contextual factors that might affect the degree of alignment between beliefs and practices. To investigate the contextual factors, a qualitative method was employed whose result was elaborated on in the next section.

4.3 Contextual Factors

This section analyzed transcribed one-on-one interviews conducted with six teachers. According to the result of the cluster analysis of the data from the online questionnaire which showed intriguing patterns discussed in Section 4.2, two teachers from each cluster were randomly chosen as interviewees. The interviewees were selected only from those who showed agreement on the participation of the interview. Table 4.7 shows the percentage of agreement of participants on the interview.

Table 4.7

Agreement on Subsequent Interview (n=109)

Cluster	Agreement	Disagreement	Total number
Typical Teachers	26 (54%)	22 (46%)	48 (100%)
Teachers in Dilemma	10 (45%)	12 (55%)	22 (100%)
Progressive Teachers	14 (36%)	25 (64%)	39 (100%)
total	50 (46%)	59 (54%)	109(100%)

One middle school teacher and one high school teacher were randomly selected from each cluster to minimize the effect of school-level contextual differences. Since there were three clusters identified, a total of six teachers were contacted, and all the participants used pseudonyms in the study. Table 4.8 presents the specifications of the six participants in the interviews.

Table 4.8

Profiles of Teacher Participants in Interview (n=6)

Teacher	School	Gender	Age	Degree	Years of teaching experience
Typical Teacher A	middle	female	33	bachelor	7
Typical Teacher B	high	female	40	bachelor	17
Teacher in Dilemma A	middle	female	36	bachelor	8
Teacher in Dilemma B	high	female	39	bachelor	10
Progressive Teacher A	high	male	48	bachelor	17
Progressive Teacher B	middle	female	30	bachelor	6

Of the six participants, one participant was male, and five were female. Their age covered a range from 30 to 48, with the average age of 37.5. Their years of teaching English varied from six to 17 years, with the mean teaching year of 10.8. Two teachers were enrolled or had been enrolled in master's programs in English education at the time of data collection, but none of the participants received MAs in English Education.

Thematic analysis was conducted to examine contextual factors influencing the degree of alignment between beliefs and practices. The detected factors were divided into micro and meso levels following the framework of the contextual influences on assessment practices by teachers by Fulmer et al. (2019).

4.3.1 Contextual Factors at the Micro Level

The micro level incorporates the impact of the individual teacher and his or her classroom (Fulmer, Lee, & Tan, 2015). Training and coping style against students' feedback are included in this category.

4.3.1.1 Training

The number of participations in training related to all kinds of assessments varied across the six teacher interviewees. It ranged from one to five times according to the questionnaire. However, when teachers were asked about training related to performance assessment specifically, the number of participations decreased sharply compared to the number of participation related to all kinds of assessment. The interviewees answered that most of the training courses conducted in schools were about the regular pencil-and-paper exams such as mid-term or final-term exams, and relatively limited attention

was paid to performance assessment. As a result, both Typical Teacher A and B reported that they didn't participate in any training about performance assessment while Progressive Teacher A and B mentioned that they only received information about the overall system of performance assessment from time to time.

Surprisingly, the ones who participated in teacher training related to performance assessments exclusively were the 'Teachers in Dilemma' who showed the least level of beliefs and practices about performance assessment as Assessment for Learning and reported the biggest gap between their beliefs and practices. Teacher in Dilemma B had especially high aspiration and was a member of a professional learning community which aimed for Communicative Language Teaching, four Skills Integration, and Learner-centered Instruction.

Teacher in Dilemma B: A training about writing performance assessment was very impressive. That teacher made a scoring rubric with her students while providing some examples to students, and made students rate the writings. Of course, that school is told to be a school with very smart students. That was very memorable. Last year I participated in another training about performance assessment, and

the teacher shared her speaking performance assessment, (...)

As illustrated in the excerpt above, Teacher in Dilemma B voluntarily received diverse and practical training courses directly related to performance assessments, and she was eager to explain her training experiences. Within a short time, she enumerated as many as three training courses she was impressed by and shared her plan to invite one of the instructors of the training course to her school for her colleagues. She seemed very satisfied with the training courses in that the content of the training was neither superficial nor shallow but very informative with plenty of concrete examples. These seemingly ideal training courses, however, didn't appear to be absorbed in neither her belief nor her practice.

Teacher in Dilemma A also expressed satisfaction with her experience of participating in training about performance assessment particularly. She also explained that the training was practical in that it provided examples to apply in classrooms as shown in the following excerpt.

Teacher in Dilemma A: Most recently I participated in a training about speaking performance assessment conducted by Education Office. It was a one-time event, so it was a kind of workshop. They

informed me about the current trend of speaking performance assessment, what problems exist, what strength they have, and so on. A few teachers came and shared their lessons, and that was really good and impressive.

As observed in the case of Teacher in Dilemma B, Teacher in Dilemma A's participation in training didn't seem to be transferred into beliefs and actual practice as well. Since teacher training appeared not to have exerted a positive influence on teachers' beliefs and practices as observed in the analysis of the interview data, it could be argued, on the surface, that teacher training was not an influential contextual factor on the degree of alignment between beliefs and practices. However, the fact that two teachers in 'Teachers in Dilemma' were the most enthusiastic teachers in participating in training courses for performance assessment should be paid attention. There is a new possibility that their aspiration and enthusiasm toward teacher training about performance assessment might have adversely affected the severity of self-rating of assessment beliefs and practices since they might have a higher standard of performance assessment with the aim of Assessment for Learning which might have resulted from their participation in teacher training. Therefore, teacher training was pointed out as an influential contextual factor at the micro level and will be discussed in the next chapter.

4.3.1.2 Coping Style against Students' Feedback

Asked about the most satisfactory performance assessment, all the teachers chose one performance assessment task with pride and confidence. For instance, Typical Teacher B, Teacher in Dilemma A, and Progressive Teacher B selected essay writing as the most satisfactory performance assessment. To be specific, Teacher in Dilemma A and Progressive Teacher B expressed high satisfaction with the process writing, while Typical Teacher B emphasized writing an essay over 150 words to encourage students to express their own ideas logically. On the other hand, Typical Teacher A introduced dubbing, which she invented with her colleagues, as a speaking performance assessment where students prepared scripts according to an edited video clip and performed it in front of other classmates. Lastly, Teacher in Dilemma B and Progressive Teacher A chose extensive reading activity which led to performance assessments such as writing reading logs, group conference, and presentation.

All the performance assessments mentioned above were Assessment for Learning while evaluating the performance of four skills directly which we regard as authentic. Besides, there was one more thing to notice: all the performance assessments selected as the most satisfactory ones were those that were conducted most recently. What was impressive here was the teachers' effort to reach the current satisfactory performance assessment task. They

didn't stumble across the given task of performance assessment by chance. Instead, the assessment task conducted currently was the product of a combined effort to content both teachers and students. To do so, most teachers valued students' feedback about performance assessments, whether they were positive or negative, and tried to integrate this feedback into the implementation of performance assessment. For instance, Progressive Teacher A actively accepted students' negative feedback to improve his performance assessment task as in the following excerpt.

Progressive Teacher A: When I first implemented performance assessment related to extensive reading three years ago, it was the idea of my colleague, and I had difficulty implementing it. (...) I also asked my students about good points and bad points and tried to remedy the shortcomings. (...) When I collect feedback from students, I listen to them, and that helped me to decide on improvement direction or simplify the task.

The same adaptive coping style against students' feedback was also observable in Progressive Teacher B. She detected that the biggest problem of writing an essay when announcing the theme prior to the actual assessment

was that the students memorized sentences prepared at home or in private institutes individually. One or two times per semester, the problem of cheating, i.e., copying from a pre-written essay, occurred, and some students showed a negative attitude about it.

Progressive Teacher B: Some smart students even asked me if this is an 'essay' test. They were discontented about this type of performance assessment.

Progressive Teacher B took this negative feedback from students seriously. Her willingness to rectify this situation led to the implementation of process writing, which she is pleased with now.

The situation was the same in 'Typical Teachers.' They also expressed that they accepted students' feedback and conferred with colleagues to reach a more satisfactory type of performance assessment for both students and teachers.

However, a striking difference was noticeable in Teacher in Dilemma B. She was not able to cope with the negative feedback of students and decided to shut her ears to students' feedback.

Teacher in Dilemma B: *I don't care about students' reactions. They are too diverse, and some students like my lesson while others do not. Survey results are all different. So, recently I thought, I will just do what's fun to me. You guys (students) don't matter. I'm the only one who needs to have fun. (laughing)*

As the conversation deepened, the background for her choice was revealed. She had experiences of getting hurt and offended by students' comments and reactions. That made her build walls and protect herself from outside.

Teacher in Dilemma B: *I've often got hurt by students, so all I need is my satisfaction. No matter what you (students) say, I will conduct lessons as I want. Some students say it (reading class) is boring and ask why we are doing this. (...) These one or two comments are ringing in my mind. That breaks my motivation to teach, and I've concluded to do lessons that please me. (...) Just in case of complaints from students/parents, I know this is wrong, but I give full marks to high-level students because low-level students usually don't complain.*

Even though both Teacher in Dilemma B and Progressive Teacher A conducted the same performance assessment task, which was the task related to extensive reading, one teacher ignored students' feedback or reaction while the other accepted it as a stepping stone for improvement. Therefore, there is a strong probability that the detailed assessment practice and the level of students' satisfaction toward the performance assessment task may vary.

Teacher in Dilemma A also shared her experience of receiving negative feedback toward performance assessment tasks.

Teacher in Dilemma A: Students want authentic performance assessment. For example, some students directly ask me, "How come this is a speaking test?" They take some defiant attitudes. They are middle school students and express their opinions bluntly.

Even though Teacher in Dilemma A and B belonged to the same cluster, the coping style of Teacher in Dilemma A was different from that of Teacher in Dilemma B. Fortunately, Teacher in Dilemma A reported that she was able to reflect students' feedback and make performance assessment as authentic as possible. She attributed this success to the standards-based testing system and criterion-referenced evaluation in middle schools. She added that the reduced

burden of grading and ranking students made it possible to establish authentic situations with authentic materials for authentic performance assessment. This difference in the evaluation system was worth noticing. The coping style of Teacher in Dilemma B could not be fully interpreted without considering the fact that she worked in a high school where the norm-referenced evaluation was performed which emphasized discrimination and grading for the college entrance.

In sum, it could be inferred from excerpts above that the coping style against students' feedback is a contextual factor that brings the different size of inconsistency between beliefs and practices by teachers. However, this factor was not simple nor straightforward since it was observed with a reasonable possibility that the coping style might depend on the school level where teachers worked. The difference in the evaluation system between middle schools, which adopt criterion-referenced evaluation, and high schools, which adopt norm-referenced evaluation, might be interconnected with different coping styles, which will be discussed in the next chapter.

4.3.2 Contextual Factors at the Meso Level

The meso level embraces effects within the school, such as its culture, facilities, resources, and leadership. These factors are external to the classroom

even though they exert direct influence on it. Particularly, factors regarding the school itself or its neighboring community have an immediate impact (Fulmer et al., 2015; Owston, 2007). The analysis of the interview disclosed that the colleagues and complaints about grades from students or parents played a role in forming teachers' beliefs and practices resulting in different levels of inconsistency between them.

4.3.2.1 Colleagues

The analysis disclosed that colleagues exerted a strong influence in the implementation of performance assessment. Typically, in schools located in cities, more than two English teachers take responsibility for teaching the same grade. This means that cooperation and keeping balance in exchanging ideas and making decisions among English teachers of the same grade are extremely important. Based on the overall evaluation plan decided and announced in March, which is the beginning of a school year in Korea, teachers generally hold several meetings to decide detailed types of tasks and scoring criteria of performance assessment about a month prior to the actual performance assessment. Most of the teachers expressed that they felt free to exchange opinions and ideas regardless of their age or teaching experiences. However, both Teacher in Dilemma A and B exposed concerns for the reactions of colleagues.

Teacher in Dilemma A: *The flow of the meeting depends on the teachers' characters you're working with. It has nothing to do with the teaching experience. Some teachers react to my proposal asking if we really have to do that much. Some other teachers prefer doing exactly the same assessment as last year. Other teachers say to me that's a great idea and give more constructive ideas. I think it's fifty-fifty.*

This teacher continued claiming that the colleagues were the biggest obstacle in implementing performance assessment in her desired way.

Teacher in Dilemma A: *Dividing the same grade with other teachers is the biggest obstacle for me. That is, collaboration with other teachers sometimes helps to view things more objectively, whereas uncooperative teachers make performance assessment shallow and superficial.*

Teacher in Dilemma B expressed the same opinion as Teacher in Dilemma A in that the colleagues were the most demanding and challenging factor in the implementation of performance assessment.

Teacher in Dilemma B: *The most difficult part in implementing performance assessment is forming a consensus among colleagues. The process of a conference about performance assessment depends on the colleagues you are working with in the given year.*

The situation was strikingly contrasting with the one of 'Progressive Teachers.' The 'Progressive Teachers' didn't equate colleagues with hindrance. They rather affirmed that the colleagues were their overcoming strategy of difficulty in implementing performance assessment and additionally, the source of new ideas for performance assessment.

Progressive Teachers A: *When implementing performance assessment, there are always some unexpected problems. Whenever this happens, I talk with my colleagues and look up precedents. (...) New ideas for performance assessment come from my colleagues. I learn from my colleagues and create new assessments reflecting colleagues' methods.*

As shown in the excerpts above, the views on colleagues were contrasting. Some regarded colleagues as an impediment while others considered them as advisors. At first glance, one might conclude that this

difference of thinking might result from the different ability to cooperate with colleagues and hence affected the degree of alignment between beliefs and practices. However, peering into the data more closely, the contextual factor of colleagues at the meso level may be interconnected with the contextual factor of teacher training at the micro level. The hidden relationship between colleagues, training, and the degree of consistency between beliefs and practices will be discussed in the next chapter of discussion.

4.3.2.2 Complaints about Grade from Students/Parents

Since Korea is a meritocratic society, students and parents are very sensitive to the scores of every assessment (Park & Chang, 2017). It is not an exaggeration to say that university admission is the ultimate goal of most secondary school students. In Korea, the two kinds of requirements for university entrance are the CSAT (College Scholastic Ability Test) and student records throughout the high school years called *nae-shin*. Even though the records of middle school do not play a direct role in university admission and the newly introduced standards-based testing system reduced teachers' burden of grading and ranking students, middle school students cannot still escape from the competitive atmosphere since their records determine their admission into a prestigious high school, which is directly linked to the university

entrance. In other words, not only high school students but also middle school students pay excessive attention to grades and records, though the degree of attention may vary, and this undue attention and interest are extended to complaints and distrust about grading. Hence, none of the six teacher interviewees finished their interview without mentioning the word “complaint from students and/or parents about the grade.” The fact that performance assessment has a nonnegligible share of the final record but lacks a relatively objective form of grading (e.g. multiple-choice questions), complaints from students and/or parents about grades are one of the biggest stumbling blocks that hinder the ideal implementation of performance assessment. How much teachers care about complaints from students and/or parents about grades could be inferred through the following excerpts.

Typical Teacher A: What I take into consideration most when planning performance assessment is the complaints from students and/or parents about grades.

Teacher in Dilemma A: After an assessment, students talk about their grades, and whenever I hear that, I worry about complaints. That’s bothering, and I feel a lack of confidence. I try to be prepared for that.

In fact, many teachers were conscious about the issue of the reliability of grading and expressed worries about a fair evaluation when assessing students via performance assessment as in the following.

Progressive Teacher B: The most difficult part in performance assessment is how I grade and divide students. (...) When a student gets ten points and the other gets nine points, I'm confused if there really is a difference of one point between these two students.

Teacher in Dilemma B: It's difficult to evaluate fairly in performance assessment. (...) It's hard to follow a uniform criterion.

As a result, the outcome of the intensive preparation for possible complaints and teachers' apprehension about fair grades is the endeavor to establish flawless analytic scoring rubrics. The interview revealed that teachers dedicated a large portion of their preparation time to deciding the specific grading criteria and wished to be allowed to apply diverse types of scoring criteria in performance assessment.

Typical Teacher A: *To set clear grading criteria, I confer with my colleagues to decide specific criteria from a month before the actual performance assessment. (...) I usually conduct lessons directly related to performance assessment, and when there's a difference between the hours and the ways of conducting lessons in each class, the possibility of complaints arises. So, I plan it in detail in advance.*

Typical Teacher B: *I'm confident about setting scoring criteria and evaluating impartially, so I rarely get complaints about the grading results from students. (...) But the problem with my scoring criteria is that it's too quantified, and that makes qualitative assessment difficult.*

Teacher in Dilemma A: *I wish I could evaluate holistically. Or I wish reliable scoring criteria are available so that teachers are able to assess students without suffering from complaints about grades.*

The issue of objective scoring and grading illustrated above is due to the ranking practice for university entrance. Even in performance assessments, the need to rank students was inevitable regardless of whether they were middle or high school students though the degree may vary.

The interview data presented above clearly showed that no single teacher escaped from complaints from students or parents about grades. Therefore, this contextual factor didn't seem to affect the degree of alignment between beliefs and practices since all the teachers expressed the same worries about complaints about grades from students/parents. However, even though it did not relate to the degree of alignment, its huge impact on teachers' assessment practices was indisputable. Considering the magnitude of its influence, this study decided to address this contextual factor and include it at the meso level.

In addition, repeated readings of the transcribed interview with teachers discovered another interesting point: some teachers misunderstood the term *feedback*. Feedback is delivered information to students about how the state at this time relates to goals which are defined as specific standards, criteria, or targets in academic settings (Nicol & Macfarlane-Dick, 2006). Undoubtedly, feedback is a key strategy of performance assessment in that “formative feedback represents information communicated to the learner that is intended to modify the learner's thinking or behavior for the purpose of improving learning” (Shute, 2007, p. 1). However, some teachers confined feedback to providing explanations for losing points according to the criteria. This could be inferred through the following excerpts.

Typical Teacher A: *Although I can't give feedback to every student, I give specific feedback to students who visit and ask me to, and this will make an impression to all the students that this teacher is very careful and accurate when grading.*

Teacher in Dilemma B: *Feedback I give to students related to performance assessment is telling them this is your score. That's all.
(laugh)*

Teacher in Dilemma A: *All the feedback I give to students is their final grade. (...) When I give specific feedback to all students, then students will talk to each other. For example, that guy wrote that way and got a minus point, but I didn't get a minus point even though I wrote the same way. (...) There is a holistic scoring method, but students don't know about it, and they compare and complain... That hinders giving detailed feedback to students.*

This misconception about feedback was apparently related to the suffering from the complaints of students and/or parents. Some teacher participants misunderstood that providing feedback is explaining students

about the losing point, and thus it serves the role of a defensive tool against complaints about grade. Even though this misconception was not detectable across all interviewees, the drawback of the meritocratic system appeared to be severe and proper remedy seemed imperative.

CHAPTER 5.

DISCUSSION

In this chapter, the key findings of the current study presented in Chapter 4 are summarized and discussed according to research questions. Section 5.1 discusses the English teachers' beliefs about the performance assessment and the degree of alignment between these beliefs and perceived practices. Section 5.2 presents the contextual factors influencing the degree of alignment between beliefs and practices.

5.1 Beliefs and their Difference from Practices

The first research question posed for the current study was identifying Korean English teachers' beliefs about performance assessment. Regarding the beliefs about the performance assessment's purposes, the results presented convincing evidence of 3.45 of mean scores out of 4 indicating that the English teachers in Korean secondary schools tended to judge performance assessment to occupy an important role in improving teaching and learning. This was in line with abundant empirical research (Brown & Remesal, 2017; Elshawa et

al., 2017; Muñoz et al., 2012; Remesal & Brown, 2015) which presented teachers' high agreement with the purpose of assessment as serving the improvement of learning and teaching in the foreign language classrooms. Especially, teachers' strongest agreement with the role of performance assessment as providing a valuable learning experience and motivation to learn corresponded to the findings of N. Y. Kim (2017), Kim and Kim (2018), and Park and Chang (2016).

In terms of the methods and techniques of performance assessment, the results of the present study unveiled that teachers preferred to use a diversity of methods and techniques to evaluate students' language ability in their classrooms. This result corresponded with the finding of T. J. Park (2016) who highlighted the importance of performance assessment and recommended assessing language ability through various methods besides the traditional pencil and paper assessment.

Concerning feedback, grading, and reporting of grades, which are the key elements that enhance learning (Cheng & Wang, 2007), the result was in line with the findings of Cheng and Wang (2007) and Elshawa et al. (2017). Although these studies had contextual differences (Canada, China, Hong Kong, and Malaysia) and were conducted at the tertiary level in ESL/EFL setting, teachers' beliefs were identical in that most of the teachers tended to establish their marking criteria prior to the assessment and announced the scoring

criteria to their students in advance of the assessment. Through this process, the transparency of the assessment criteria was claimed to be increased.

Generally, most of the 109 teachers held beliefs that performance assessments should improve teaching and learning while employing diverse methods and providing constructive feedback. However, further analysis of the questionnaire responses exhibited some noticeable conflicting patterns of beliefs and practices for different aspects of performance assessment, which will be elaborated on in the next paragraphs.

The biggest gap of 1.63 between beliefs and practices was manifested in the amount of time needed for the full preparation of performance assessment. This result was in line with S. Lee (2008) who claimed that the contextual conditions should be improved for true performance assessment. She elaborated that teachers in the Korean school system often suffered from work overload with no room for preparing performance assessments. The problem of work overload and lack of time is probably connected to the school system. Changing the school system is certainly beyond the power of individual teachers, and that may be the reason for exhibiting the biggest difference between ideal and reality.

The second biggest mismatch between beliefs and practices was observable regarding feedback. Teachers strongly agreed that feedback should

be given while students learn and after performance assessment as well. As a good way of giving feedback, they valued conferencing with students. Unfortunately, these beliefs were not realized in classrooms exhibiting the mean differences between beliefs and practices which ranged from 0.49 to 1.00. This same mismatch was observed by N. Y. Kim (2017) who pointed out that most teachers appreciated the need and importance of feedback, but failed to apply it to their classrooms. Kim and Yun (2015) and G. H. Lee (2009) also arrived at the same conclusion. This lack of feedback in practice was claimed to have resulted from the number of students a teacher is responsible for. Typically, teachers in secondary schools take charge of multiple classes, and hence, providing feedback to all the students would be practically impossible (N. Y. Kim, 2017). In such a situation where individual feedback is difficult to be provided, teachers can't afford to think of conferencing with students. Therefore, teachers wouldn't dare to hold conferences with students to give feedback even though they acknowledge its value. This resulted in recording the second biggest degree of inconsistency, the difference value of 1.00, between beliefs and practices.

Instead of defending the status quo, other strategies to provide feedback should be pursued and implemented. Kim and Yun (2015) proposed devising diverse ways to provide feedback that can be easily modified by individual teachers, while N. Y. Kim (2017) suggested giving feedback in

groups or implementing peer-assessment. Talking about peer-assessment, however, Korean English teachers didn't have a favorable view toward it based on the evidence of the low belief score of 2.71 about peer-assessment and 2.77 about self-assessment in the questionnaire. In actual classrooms, the practice scores dropped even more sharply generating the third and fourth biggest mean difference between beliefs and practices. This supported the finding of G. H. Lee (2009) and N. Y. Kim (2017) which revealed that Korean English teachers attached comparatively limited value to peer- and self-assessment. However, as many researchers over decades have proved, useful formative feedback emanates not only from the teacher but also from self- or peer-assessment (Black & William, 1998; Fluckiger, Vigil, Pasco, & Danielson, 2010; Smith, 2007). Brown (2005) asserted that self- and peer-assessment are beneficial to students as these encourage students' metacognition (i.e., a means of learning about their own learning) and facilitate deep rather than surface learning. Furthermore, when properly managed, these methods save time for teachers, which Korean teachers crave for, and help students interpret criteria more thoroughly. Despite these advantages, many teachers are afraid of the low reliability and subjectivity of peer- and self-assessment (Hong, 2018; Park & Park, 2018). These shortcomings can be overcome by repeated training to get students accustomed to scoring criteria and processes (Hong, 2018). One of the ultimate goals for education is to raise self-learning ability, and its attainment

would be accelerated through self- and peer- assessment (Park & Park, 2018).

It is recommended that teachers be open-minded toward various kinds of assessment.

The fifth biggest mean difference between beliefs and practices was about the students' role in preparing the marking criteria. Based on the lowest belief score of 2.20 in the questionnaire, it could be inferred that most Korean teachers disagreed on including students when determining scoring criteria. This belief was transferred into practice resulting in the lowest practice score of 1.37. At the same time, the mean difference was as high as 0.83, indicating their negative belief and even lower degree of practice. This finding supported Cheng and Wang (2007) and Elshawa et al. (2017) who disclosed that teachers decided assessment "to be done on the students rather than with" the students (Cheng & Wang, 2007, p. 101). This same decision seemed to be pervasive among Korean English teachers in secondary schools as well. However, the benefit of involving students in negotiating or establishing the assessment criteria is evident in that it makes students fully understand what is anticipated of them (Brown, 2005). This is especially helpful when assessing group activities such as drama since assessment criteria can be designed to ensure an equal chance of achieving high grades for all students. Hence, expansion of views, thoughts, and experience about assessment is highly suggested. This may be challenging but fruitful and rewarding for teachers in the end.

The sixth greatest gap was observed regarding competition among students that performance assessment created. Teachers practiced it more than they valued resulting in the mean difference of 0.79. Even though most teachers in the survey told that they adopted criterion-referenced assessment as performance assessment as evidenced by the practice score of 3.40, many teachers could not deny that students were still in competition with other students. This finding was in line with the study of G. H. Lee (2009) illustrating that students recognized performance assessment as an influential examination since the scores are part of the final grades together with other paper-and-pencil tests, and this fact led students to compare and compete with others constantly. As Son (2006) clarified, the problem of the excessively competitive atmosphere of the formal testing regime that has been prevailing in secondary schools in Korea was born with the system of student records policy (*nae-shin*). Fundamental change is needed to cope with this problem.

The six biggest differences between beliefs and practices were discussed. Surely all these inconsistencies need to be addressed explicitly, and proper ways should be sought to resolve the current situation. One interesting thing to notice here was that some mismatches were due to the practical restraints requiring more fundamental innovations while others needed changes in both beliefs and practices. To be exact, among the six biggest differences stated above, the problems regarding time, feedback (conferencing

with students), and unnecessary competition seemed to be resulted from national-level practical limitations which demand structural changes to match practices with beliefs. Hence, possible solutions are likely to be beyond the power of individual teachers. In contrast, the rest of three differences, which were about self-assessment, peer-assessment, and students' role in marking criteria, needed changes in beliefs first. Teachers should escape from the traditional teacher-centered assessment and be open-minded and receptive of new forms of assessment. This kind of change seems to be within the capacity of teachers. Therefore, it is not impossible to start finding solutions for the inconsistencies between beliefs and practices. Teachers themselves should initiate the vital journey to arrive at valid and meaningful performance assessment.

5.2 Contextual Factors

The current study also aimed to investigate contextual factors affecting the degree of alignment between the teacher's beliefs about performance assessment and their actual practices. The identification of three clusters using both teachers' beliefs and practices scores of the questionnaire made it possible to examine the influential contextual factors which might differ among clusters. The three clusters were 'Typical Teachers', who showed a moderate degree of

beliefs and practices scores compared to other clusters, 'Teachers in Dilemma' with the lowest beliefs and practices scores and the largest difference between beliefs and practices, and 'Progressive Teachers' who displayed the highest level of beliefs and practices scores for conducting performance assessment for improving teaching and learning. Two teachers were randomly chosen as interviewees from each cluster, and their 20-minute semi-structured interviews were analyzed. For systematic analysis, Fulmer et al.'s (2019) conceptual framework linking assessment practices with contextual factors was used, and factors were divided into the micro and meso levels.

At the micro level, teacher training and coping style against students' feedback were noticed as influential contextual factors. Regarding teacher training, many researchers conducted investigations of its influence on their beliefs or cognition. In the majority of cases, it was confirmed that teacher training did exert an impact on teachers' cognition (Borg, 2003). However, this positive influence of teacher training on teachers' beliefs was not discovered in the present study. As elaborated in the previous chapter, the mean score of beliefs about performance assessment as Assessment for Learning was the highest in 'Progressive Teachers' followed by 'Typical Teachers.' 'Teachers in Dilemma' scored the lowest scores of beliefs. However, the interview revealed that the two teachers belonging to the 'Teachers in Dilemma' were the only ones who had chances of receiving training about performance assessment

specifically while showing satisfaction about the training they received. In other words, for these teachers, training didn't seem to have salient positive impact on their beliefs even though they were satisfied with the content of the training. This matched with the influential review by Kagan (1992) who suggested that the relationship between teacher education and teacher cognition was not significant. Empirical studies on the relationship between Korean English teachers' beliefs and teacher education programs also revealed that teacher education programs did not play any significant role in changing teachers' beliefs (E. Kim, 2008; H. J. Lee, 2006; Yook, 2010).

When we look at the practice factor, however, the influence of teacher training was claimed to be detected. It was especially worth noting that 'Teachers in Dilemma,' who had spontaneous enthusiasm toward training, displayed the lowest practice scores together with the lowest degree of alignment between beliefs and practices. The practice scores were derived from a questionnaire in a form of a self-report. Repeated participation in training might have raised their standards about their practices, and this high standard of performance assessment with the aim of Assessment for Learning might have made 'Teachers in Dilemma' self-rate their assessment practices more severely than teachers in other clusters. This possible influence of teacher training on the severity of scoring perceived practices was unexpected and surprising. More supporting evidence in the form of class observation would

help to understand the role of training more precisely.

Furthermore, the fact that ‘Typical Teachers’ and ‘Progressive Teachers’ had rarely a chance to participate in training about performance assessment exclusively provided an important implication for policymakers. This fact corresponded with the finding of T. J. Park (2016) who revealed that over 40% of Korean English teacher respondents hadn’t received any training about performance assessment. However, as many researchers suggested, long term and repeated in-service teacher training program should be provided (N. Y. Kim, 2017; Koh, 2015; McMunn, McColsky, & Butler, 2004; Wiliam & Thompson, 2008). In addition, instead of shifting the responsibility of application on individual teachers, feedback after training programs should also be presented to teachers (N. Y. Kim, 2017; Y. H. Kim, 2009). Even though individual developmental pathways appear essential to understand the influence of teacher education on language teachers’ beliefs (Borg, 2003), teachers should be consistently exposed to teacher training in order to enhance their professional development.

The other influential contextual factor at the micro level was the coping style against students’ feedback. When implementing performance assessment, various processes of decision making are involved where the teacher has to determine when and how to employ specific tools and techniques (Serpil & Derin, 2017). In this decision-making process, it is desirable to take

other people's ideas and concerns into account since this will make them participate in the given task more actively (Bressen, 2007). Hence, in the classrooms, teachers should have the compassion to listen to students' opinions and feedback for improvement of performance assessment. Like 'Typical Teachers' and 'Progressive Teachers,' acceptance of students' constructive feedback will lead to modification of performance assessment, and this will result in performance assessment with remarkable improvement.

Constructive feedback often takes the form of negative statements and is heart-breaking to listen to. However, when the same feedback is repeatedly provided and collected from students, there is an urgent need to address this feedback and try to solve the problem. Avoidance coping strategies, which Teacher in Dilemma B displayed in the interview, don't solve the problem at all and are associated with negative psychological wellbeing (Parsons, Frydenberg, & Poole, 1996). Especially, Teacher in Dilemma B employed a cognitive avoidance strategy by avoiding thinking about the stressor. She confronted the fundamental problem of coping with stakeholder's feedback which was diverse according to each personal test-taking experience and their individual needs (Shim, 2015), and instead of searching for a way to establish a good communication system, she just ignored it. This cognitive avoidance strategy is claimed to bring greater psychological distress than behavioral avoidance strategy (Parsons et al., 1996), and loss of chance of improvement.

Teacher in Dilemma B chose to show indifference to the recipients of one's service which are the students in this case. It was reported that teachers who are deeply frustrated and detached from students are unable to be effective in the classroom (Shukla & Trivedi, 2008), and this might have resulted in the biggest difference between beliefs and practices among the three identified clusters.

As Shukla and Trivedi (2008) put it, teachers are 'people workers' who are engaged in a large amount of interpersonal contact with students as their 'clients.' Therefore, teachers are unintentionally exposed to experience diverse psychological problems (Pishghadam et al., 2014). Rather than treating it as a problem which needs to be solved individually, increased attention to teachers' psychological well-being, which will affect students and classroom activities in a cyclic way, is requested.

The fact that two teachers in the cluster of 'Teachers in Dilemma' didn't show the same pattern relating to the coping style against students' feedback called for a deeper analysis of the data. When clustering the participants, the variable of middle and high school didn't affect the clustering result as shown in Table 4.5. The difference between middle and high school was not detected in quantitative data, but it was visible in the qualitative data of the interview. The interview data revealed that this difference of middle and high school did influence different coping styles even in the same cluster. The

less burden to grade and discriminate students in middle schools employing criterion-referenced assessment seemed to create more space for teachers to respond to students' feedback compared to high schools employing norm-referenced performance assessment. The introduction of achievement standards-based assessment, which resembles criterion-referenced assessment, brought greater autonomy to teachers in assessing students while aiming to measure students' skills more precisely and connecting assessment with teaching and learning simultaneously (S. H. Kim, 2017; Park, Cho, Hong, & Jung, 2017). That might have made Teacher in Dilemma A working in a middle school accept students' feedback more freely and positively compared to Teacher in Dilemma B who worked in a high school with norm-referenced assessment. The coping style of Teacher in Dilemma B might be another side effect of the high need to discriminate students for college entrance.

Next, two factors at the meso level will be discussed in detail. The first factor at the meso level detected in this study was colleagues. The analysis of the interview proved that colleagues exerted tremendous influence in the degree of alignment between beliefs and practices. Unlike 'Typical Teachers' and 'Progressive Teachers' who showed a comparatively high match between their beliefs and practices, only 'Teachers in Dilemma' with the least degree of alignment between beliefs and practices lamented the difficulty of communicating with colleagues. This substantial impact of colleagues was not

surprising since G. H. Lee (2009) revealed that 55.8% of teachers collected materials or tools of performance assessment through conferencing with colleagues. When dividing the same grade with other colleagues, the importance of collaboration increases dramatically, and the result of failure to meet a consensus with colleagues would be painful throughout the year. This was evident in the statement “it (performance assessment) depends on the colleagues” articulated by ‘Teachers in Dilemma.’ This finding lent support to Wiliam and Thompson’s (2008) acknowledgment of the importance of working and conferencing with colleagues when implementing formative assessment. They asserted problems could be remedied through conversations with colleagues. In the same vein, N. Y. Kim (2017) included the ability to collaborate with colleagues in evaluating secondary English teachers’ competencies in student assessment.

At this point, shifting the focus is needed to ponder what caused the adversity concerning communicating with colleagues by ‘Teachers in Dilemma.’ One plausible reason may be the teacher training that ‘Teachers in Dilemma’ willingly received. Through repeated training, they were likely to believe that they knew more about valid performance assessment. In fact, it is widely proven that teacher training and assessment literacy have a positive correlation (Levy-Vered & Nasser-Abu Alhija, 2015; Sato, Wei, & Darling-Hammond, 2008; Zhang & Burry-Stock, 2003). Equipped with high

assessment literacy, they seemed to be deeply discouraged and dismayed by colleagues who were unwilling to accept their opinions or plans. Hence, instead of concluding to link the inconsistency between beliefs and practices with the inability to collaborate with colleagues, focusing on reasons for this inability is highly suggested. This research chose teacher training as a possible reason and deduced that teacher training, colleagues, and the degree of alignment between beliefs and practices might be all closely intertwined with reasonable possibility.

The fact that most 'Teachers in Dilemma' were in their thirties with 5 to 15 years of teaching experiences (see Table 4.5) required additional attention. It is suggested that other factors such as their teaching experience, age hierarchy, or difference in school culture or school climate be taken into account when addressing troubles with colleagues. However, they were beyond the scope of this study and appear to be another area for future research. Structured observations of teacher meetings and more in-depth interviews are proposed as a more suitable method.

The second contextual factor at the meso level was complaints about grades from students and/or parents. Teachers were hypersensitive and responsive towards complaints about grades, and they invested enormous effort to forestall possible complaints. As a result, they developed scoring criteria as objectively as possible so that they could provide feedback to

students in the form of an unbiased explanation for the losing points. Teachers were conscious of the reduced reliability of the scoring of performance assessment and were well aware of how painful it would be being the target of complaint of students and/or parents. Hence, they struggled to avoid this terrible situation by implementing performance assessments in the most objective way. As Typical Teacher A clarified, they have “worries” and “fear” of being the object of students/parents’ complaints about grades.

This rampant obsession with students/parents’ complaints about grades frustrated teachers, and they were constrained from achieving their aspirations. No matter in what cluster they belonged, this was apparent in all the teacher interviewees. The result was in line with the finding by Sung and Jo (2015) that teachers were reluctant to include free-production items in assessments since these items imposed additional burdens such as dealing with complaints and grade appeals from students. It also supported the conclusion of Park and Chang (2016) who pointed out the complaints about grade from students and/or parents as the main factor that constricted speaking performance assessment in the school field. These existing findings were construed through a meso-level lens and listed here as the contextual factor at the meso level. However, the intricate connection between the macro, meso, and micro levels can’t be overemphasized (Fulmer et al., 2015). In other words, when considering meso-level factors, there is a need to consider both micro

and macro levels. As such, the complaints about grades from students and/or parents (meso-level) cannot be fully explained without mentioning the university entrance system of Korea (macro-level) and teachers' endeavor to develop reliable scoring criteria (micro-level). A more intentional examination of the potential impact of school and community variables will shed light on pursuing a complete understanding of teachers' assessment practices and their connection with beliefs.

CHAPTER 6.

CONCLUSION

This study sought to investigate beliefs about performance assessment of Korean English teachers and their practices. The contextual factors affecting the alignment between beliefs and practices were additionally explored. This chapter summarizes the major findings of the present study. In Section 6.1, a summary of the key findings and pedagogical implications are discussed. The chapter concludes with limitations and suggestions for further research in Section 6.2.

6.1 Major Findings and Implications

The primary objective of this research was to investigate beliefs about performance assessment of Korean English teachers in secondary schools and their practices and in particular, to examine the relationship between beliefs and perceived practices. The secondary aim of the study was to examine the contextual factors influencing the degree of alignment between beliefs and practices. A mixed-method research design of conducting a survey and an interview was adopted to collect both quantitative and qualitative data.

A preliminary finding from this research was that Korean English teachers generally believed that performance assessments should improve teaching and learning while employing various methods and providing constructive feedback. However, some inconsistency between beliefs and practices was observed. Some mismatch was due to the practical limitations requiring more fundamental innovations to match practice with belief while others needed changes in both beliefs and practices. The major contextual factors affecting the degree of alignment between teachers' beliefs about performance assessment and their actual practices were teacher training and coping style against students' feedback at the micro level, and colleagues and complaints about grade from students and/or parents at the meso level.

Pedagogical implications could be deduced from the findings of the present study. Firstly, mandatory in-service teacher training about performance assessment should be provided to many teachers. Many stakeholders tend to assume that the regular pencil-and-paper exams outweigh performance assessments. Therefore, substantial teachers focus more on regular pencil-and-paper exams and had rarely a chance of receiving training about performance assessment. However, performance assessment accounts for as much as 40% of the final grade in the case of Gyeonggi Province, and that implies that teachers should pay more attention to performance assessment. Besides, we shouldn't ignore the finding of this study that the most enthusiastic teachers in

participating in teacher training were the most discouraged ones when failing to achieve consensus among colleagues. One of the reasons that led to the failure of reaching an agreement might be the divergence in assessment literacy. If all teachers, regardless of their teaching experiences or age, receive training about performance assessment on a regular basis, these opinion gaps would diminish. The process of collaboration with colleagues would be smoother since they share similar knowledge about performance assessment. It has been over five years since the introduction of the 2015 Revised National English Curriculum, and there is no room left for blaming education policy only. Teachers themselves should be active and open-minded to make the right instructional choices. They need to constantly strive for personal advancement and step forward if they aim to be appreciated as a true assessor by the other stakeholders. Teacher training in good quality would enhance overall assessment literacy about performance assessment among teachers, and this will positively affect their beliefs and practices.

Secondly, the findings of the study provide suggestions for developing and providing specific assessment guidelines. Specific assessment guidelines should include various types of tasks of performance assessment for each of the four skills and their scoring rubrics according to each grade and possible classroom settings. Some guidelines are currently available; however, they are not diversified as they might look from the surface level, providing not much

help to teachers in the field. Concrete and applicable examples should be presented and feedback relating to applications in each school should also be repeatedly checked. Guidelines which can be easily customized will help teachers concentrate on the genuine purpose of performance assessment instead of arming themselves with unnecessary preparation for complaints from students and/or parents. In addition, to reduce burdens about giving individual feedback, diverse forms of delivering formative feedback should be presented and educated to teachers. The use of self-assessment, peer-assessment, and group assessment should be also considered and guided to teachers with practical application strategies. These will potentiate an overall improvement of performance assessment of English in secondary schools despite the restrictions imposed by a multitude of contextual factors at the macro level.

6.2 Limitations and Suggestions for Future Research

Several limitations existed in the current study. Firstly, a small sample size prevents a clear generalized statement about beliefs and practices regarding performance assessment in secondary schools in Korea. A larger sample from various regions across the country would allow a more applicable finding of the beliefs and practices of Korean English teachers.

Secondly, this research depended on self-report methods, which were an online questionnaire survey and interviews. It didn't involve the examination of conducting performance assessment in teachers' own classrooms. This signifies a meaningful space for future study to scrutinize the relationship between beliefs and classroom practices. Incorporating structured class observations and repeated in-depth interviews would offer an additional diagnostic insight into the practices of performance assessment.

In spite of these limitations, the study not only adds to the existing research of beliefs and practices regarding performance assessment but also addresses implications in assessment settings as to the question of how valid performance assessment should be implemented.

REFERENCES

- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Arkoudis, S., & O'Loughlin, K. (2004). Tensions between validity and outcomes: Teacher assessment of written work of recently arrived immigrant ESL students. *Language Testing, 21*(3), 284-304.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge: University of Cambridge School of Education.
- Bang, I. J., & Chun, S. Y. (2011). An analysis of high school English curriculum implementation by the high school record rating system. *CNU Journal of Educational Studies, 32*(1), 59-84.
- Beach, S. A. (1994). Teacher's theories and classroom practice: Beliefs, knowledge, or context? *Reading Psychology, 15*(3), 189-96.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*(1), 8-21.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*(2), 81-109.
- Bressen, T. (2007). Consensus decision making. In P. Holman, T. Devane, & S. Cady (Eds.). *The change handbook: The definitive resource on today's best methods for engaging whole systems* (pp. 212-217). San Francisco, California: Berrett-Koehler Publishers, Inc.
- Brown, G. T., & Remesal, A. (2017). Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian teachers. *Studies in Educational Evaluation, 55*, 68-74.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher*

- Education*, 1(1), 81-89.
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-44.
- Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360-389.
- Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly*, 4(1), 85-107.
- Chester, C., & Quilter, S. M. (1998). Inservice teachers' perceptions of educational assessment. *Journal for Research in Mathematics Education*, 33(2), 210-236.
- Choe, H. (2005). *Negotiation of status of Korean nonnative-English speaking teachers*. Doctoral dissertation. Indiana University, Bloomington.
- Choi, J. (2008). Teacher-learners' beliefs about proficiency goals and teaching methods for Korean secondary English education. *English Teaching*, 63(1), 3-27.
- Choi, S. (2000). Teachers' beliefs about communicative language teaching and their classroom teaching practices. *English Teaching*, 55(4), 3-32.
- Colby-Kelly, C., & Turner, C. E. (2007). AFL research in the L2 classroom and evidence of usefulness: Taking formative assessment to the next level. *Canadian Modern Language Review*, 64(1), 9-37.
- Crookes, G., & Arakaki, L. (1999). Teaching idea sources and work conditions in an ESL program. *TESOL Journal*, 8(1), 15-19.
- Darling-Hammond, L., Adamson, F., & Abedi, J. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford Center for Opportunity Policy in Education.
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334.
- Elshawa, N. R., Abdullah, A. N., & Rashid, S. M. (2017). Malaysian instructors' assessment beliefs in tertiary ESL classrooms. *International Journal of Education and Literacy Studies*, 5(2), 29-46.
- Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice:

- The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, 27(3), 243-256.
- Fluckiger, J., Vigil, Y. T. Y., Pasco, R., & Danielson, K. (2010). Formative feedback: Involving students as partners in assessment to enhance learning. *College Teaching*, 58(4), 136-140.
- Foss, D. H., & Kleinsasser, R. C. (1996). Preservice elementary teachers' views of pedagogical and mathematical content knowledge. *Teaching and Teacher Education*, 12(4), 429-42
- Fulmer, G. W., Lee, I. C., & Tan, K. H. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 475-494.
- Fulmer, G. W., Tan, K. H., & Lee, I. C. (2019). Relationships among Singaporean secondary teachers' conceptions of assessment and school and policy contextual factors. *Assessment in Education: Principles, Policy & Practice*, 26(2), 166-183.
- Gebril, A. (2017). Language teachers' conceptions of assessment: An Egyptian perspective. *Teacher Development*, 21(1), 81-100.
- Green, K. E., & Stager, S. F. (1986). Measuring attitudes of teachers toward testing. *Measurement and Evaluation in Counseling and Development*, 19(3), 141-150.
- Griffiths, T., Gore, J., & Ladwig, J. (2006, November). Teachers' fundamental beliefs, commitment to reform and the quality of pedagogy. In *Australian Association for Research in Education Annual Conference* (pp. 1-38).
- Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Hoboken, NJ: John Wiley & Sons.
- Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors' beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research*, 9(1), 42-50.
- Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *The Journal of Educational Research*, 77(4), 244-248.
- Hargreaves, D. H. (2004). *Personalising learning: Next steps in working laterally*. London: Specialist Schools Trust.

- Harlen, W. (2005). Teachers' summative practices and assessment for learning—tensions and synergies. *Curriculum Journal*, 16(2), 207-223.
- Hong, S. Y. (2018). The effect of self-assessment: A meta-analysis. *Journal of Educational Evaluation*, 31(1), 309-331.
- Huerta-Macias, A. (2002). Alternative assessment: Responses to commonly asked questions. In J. Richards & W. Renandya (Eds.), *Methodology in language teaching* (pp. 338-343). Cambridge, NY: Cambridge University Press.
- James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal*, 17(2), 109-138.
- Joffe, H. (2012). Thematic analysis. In D. Harper & A. Thompson (Eds.), *Qualitative research methods in mental health and psychotherapy* (pp. 209-223). Chichester: Wiley-Blackwell.
- Kagan, D. M. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, 62(2), 129-169.
- Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research*, 93(5), 276-286.
- Kim, E. (1997). A survey on effective English teaching methods in elementary schools. *English Teaching*, 52(2), 157-174.
- Kim, E. (2008). An exploratory study of curricular in English teacher (pre-service and inservice) training programs for primary and secondary schools in Korea. *English Teaching*, 63(2), 261-281.
- Kim, H. K., & Yun, H. O. (2015). Promoting teacher-based assessment: Korean English teachers' perceptions of assessment and its applications. *Teacher Education Research*, 54(2), 171-187.
- Kim, N. Y. (2017). *The development and validation of an assessment instrument for secondary English teachers' competencies in student assessment*. Doctoral dissertation. Seoul National University.
- Kim, N. Y., & Kim, J. R. (2018). A perception study of elementary school English teachers on the English performance assessment. *Teacher Education Research*, 57(4), 529-538.
- Kim, S. H. (2017). Analysis of changes in the 2015 revised National English Curriculum. *Modern English Education*, 18(3), 91-111.
- Kim, S. Y., & Yi, H. K. (2013). A case study of how English speaking and

- writing performance assessment works in middle school context. *Studies in English Education*, 18(2), 159-183.
- Kim, Y. H. (2009). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26(2), 187-217.
- Koh, H. W. (2015). Effects of background variables on secondary school English teachers' perceptions of usefulness and infeasibility of speaking performance assessment. *Studies in English Education*, 20(1), 187-208.
- Kozma, R. B. (2003). ICT and educational change: A global phenomenon. In R. Kozma (Ed.), *Technology, innovation, and educational change: A global perspective* (pp. 1-18). Eugene, OR: International Society for Technology in Education.
- Lee, G. H. (2009). *Teachers and students' perspectives of middle school English performance tests*. Unpublished master's thesis. Hankuk University of Foreign Studies, Seoul.
- Lee, H. J. (2006). A study on the impact of in-service teacher education on teachers' instructional change. *Foreign Languages Education*, 13(3), 283-320.
- Lee, M. (2018). A study of high school English teachers' teaching English speaking and performance assessment. *Journal of the Korea English Education Society*, 17(1), 107-126.
- Lee, S. (2008). A study on how secondary English teachers in Incheon and Seoul perceive and carry out performance assessment. *English Language Teaching*, 20(4), 305-327.
- Lee, W. (2015). Revisions of the national curriculum of English and challenges of English education. *English Teaching*, 70(5), 35-52.
- Lee, Y., & Sung, K. (2017). A study on middle school learners' perceptions of English evaluation. *Studies in Foreign Language Education*, 31(2), 49-87.
- Levy-Vered, A., & Nasser-Abu Alhija, F. (2015). Modelling beginning teachers' assessment literacy: The contribution of training, self-efficacy, and conceptions of assessment. *Educational Research and Evaluation*, 21(5-6), 378-406.
- McMunn, N., McColsky, W., & Butler, S. (2004). Building teacher capacity in classroom assessment to improve student learning. *International*

- Journal of Educational Policy Research and Practice*, 4, 25-48.
- McNamara, T. F. (1996). *Measuring second language performance*. London: Longman.
- Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile Issues in Teachers Professional Development*, 14(1), 143-158.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199-218.
- Owston, R. (2007). Contextual factors that sustain innovative pedagogical practice using technology: An international study. *Journal of Educational Change*, 8(1), 61-77.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307-332.
- Park, J., Cho, J., Hong, M., & Jung, H. (2017). Study on the status and recognition of school-based assessment in Korea. *Korea Institute for Curriculum and Evaluation*, 30(1), 29-52.
- Park, J., & Park, J. (2018). The current state and prospects of peer assessment. *Korean Journal of Cognitive Science*, 29(2), 85-104.
- Park, T. J. (2016). Examining how performance assessment in English is being implemented in high schools and how it is perceived by teachers and students. *Secondary English Education*, 9(3), 77-98.
- Park, T. J., & Chang, J. H. (2016). A study on high school teachers' opinions about issues of classroom assessment. *Secondary English Education*, 9(4), 95-115.
- Park, T. J., & Chang, J. H. (2017). A study on high school teachers' opinions about issues of communicative English teaching and assessment. *Korean Journal of English Language and Linguistics*, 17(4), 839-863.
- Parsons, A., Frydenberg, E., & Poole, C. (1996). Overachievement and coping strategies in adolescent males. *British Journal of Educational Psychology*, 66(1), 109-114.
- Pishghadam, R., Adamson, B., Sadafian, S. S., & Kan, F. L. (2014). Conceptions of assessment and teacher burnout. *Assessment in Education: Principles, Policy & Practice*, 21(1), 34-51.

- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472-482.
- Remesal, A., & Brown, G. T. (2015). Conceptions of assessment when the teaching context and learner population matter: Compulsory school versus non-compulsory adult education contexts. *European Journal of Psychology of Education*, 30(3), 331-347.
- Richards, J. C., & Pennington, M. (1998). The first year of teaching. In J. C. Richards (Ed.), *Beyond training* (pp. 173–90). Cambridge: Cambridge University Press.
- Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research*, 37(2), 179-182.
- Rogers, W. T., Cheng, L., & Hu, H. (2007). ESL/EFL instructors' beliefs about assessment and evaluation. *Comparative and International Education/Éducation Comparée et Internationale*, 36(1), 39-61.
- Sato, M., Wei, R. C., & Darling-Hammond, L. (2008). Improving teachers' assessment practices through professional development: The case of National Board Certification. *American Educational Research Journal*, 45, 669–700.
- Serpil, Ö. Z., & Derin, A. T. A. Y. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal*, 6(1), 25-44.
- Shih, C. M. (2010). The washback of the General English Proficiency Test on university policies: A Taiwan case study. *Language Assessment Quarterly*, 7(3), 234-254.
- Shim, K. N. (2015). Does performance assessment system operate as intended in a classroom-based assessment? *Primary English Education*, 21(4), 275-300.
- Shukla, A., & Trivedi, T. (2008). Burnout in Indian teachers. *Asia Pacific Education Review*, 9(3), 320-334.
- Shute, V. J. (2007). *Focus on formative feedback*. ETS Research Report RR-07-11. Princeton, NJ: Educational Testing service.
- Singh, B. (2000, September). *Formative assessment: Which way now?*. Paper presented at the British Educational Research Association Annual Conference, University of Sussex, Brighton, UK.

- Smith, G. (2007). How does student performance on formative assessments relate to learning assessed by exams?. *Journal of College Science Teaching*, 36(7), 28.
- Son, J. J. (2006). Sociological analysis of student record policy for admission in Japanese colonial period, 1930's. *Korean Journal of Sociology of Education*, 16(3), 135-165.
- Stiggins, R., & Bridgeford, N. (1982, March). The role of performance assessment in day to day classroom assessment and evaluation. In *Annual Meeting of the National Council for Measurement in Education*, New York.
- Strauss, T., & von Maltitz, M. J. (2017). Generalising Ward's method for use with Manhattan distances. *PloS one*, 12(1), 1-21.
- Sung, M., & Jo, V. H. (2015). Sociopolitical and contextual influences on teacher-produced achievement tests of English in Korean high schools. *The SNU Journal of Education Research*, 24, 115-148.
- Szekely, G. J., & Rizzo, M. L. (2005). Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method. *Journal of Classification*, 22(2), 151-184.
- Tabachnick, B. R., & Zeichner, K. M. (1986). Teacher beliefs and classroom behaviours: some teacher responses to inconsistency. In M. Ben-Peretz, R. Bromme & R. Halkes (Eds.), *Advances of research on teacher thinking* (pp. 84–96). Lisse, Netherlands: Swets and Zeitlinger.
- Webb, R., Vulliamy, G., Hämäläinen, S., Sarja, A., Kimonen, E., & Nevalainen, R. (2004). A comparative analysis of primary teacher professionalism in England and Finland. *Comparative Education*, 40(1), 83-107.
- Wiliam, D. (2011). What is assessment for learning?. *Studies in Educational Evaluation*, 37(1), 3-14.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49-65.
- Wiliam, D., & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–92). Hillsdale, NJ: Lawrence Erlbaum Associates
- Winterbottom, M., Brindley, S., Taber, K. S., Fisher, L. G., Finney, J., & Riga, F. (2008). Conceptions of assessment: Trainee teachers' practice and

- values. *The Curriculum Journal*, 19(3), 193-213.
- Winterbottom, M., Taber, K. S., Brindley, S., Fisher, L. G., Finney, J., & Riga, F. (2008). Understanding differences in trainee teachers' values and practice in relation to assessment. *Teacher Development*, 12(1), 15-35.
- Wren, D. G. (2009). *Performance assessment: A key component of balanced assessment system*. Virginia Beach: Virginia Beach City Public Schools, Department of Research, Evaluation, and Assessment.
- Yook, C. M. (2010). *Korean teachers' beliefs about English language education and their impacts upon the Ministry of Education-initiated reforms*. Doctoral dissertation. Georgia State University.
- Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16, 323-342.

APPENDICES

APPENDIX 1.	118
APPENDIX 2.	123

Appendix 1. Questionnaire

첫번째 설문은 당신의 “수행평가에 관한 신념”에 관하여 조사하는 설문입니다. 다음의 각 문장이 선생님이 평소에 수행평가에 대해서 갖고 계신 생각과 얼마나 일치하는 지에 따라서 응답을 해주시면 됩니다. 아래의 척도를 사용하여 응답해 주세요.

전혀 동의하지 않음(Strongly Disagree (SD)) 별로 동의하지 않음(Disagree (D)) 대체로 동의함 (Agree (A)) 매우 동의함(Strongly Agree (SA))

A 섹션: 수행평가의 목적

수행평가는	
1	교사가 수업에 집중할 수 있게 해주어야 한다. (helps to focus on teaching.)
2	교육적인 목적으로 학생들을 실력에 따라 그룹화하는데 도움을 주어야 한다. (helps to group students for instructional purposes.)
3	학생들의 강점과 부족한 점을 진단해 주어야 한다. (diagnoses strengths and weaknesses in students.)
4	나의 수업의 강점과 부족한 점을 진단해주어야 한다. (diagnoses strengths and weaknesses in teaching.)
5	학생의 발전상황에 대해 정보를 주어야 한다. (provides information about students' progress.)
6	학생들 사이에 경쟁을 유발해야 한다. (creates competition among students.)
7	학생들에게 배움의 경험을 제공해야 한다. (creates a valuable learning experience for students.)
8	학생들의 학업에 동기를 부여해야 한다. (motivates students to learn.)
9	수행평가를 통해 학생들은 배우는 과정 중에 피드백을 받아야 한다. (provides feedback to students as they learn.)
10	수행평가를 통해 학생들의 학습의 단계를 볼 수 있어야 한다. (demonstrates the level of students' learning.)

B 섹션: 수행평가의 방법 및 기술

11	지필 수행평가(예: 받아쓰기, 대화의 빈칸 채우기 등)는 학생들의 말하기와 듣기를 평가하는 좋은 방법이라고 생각한다. (Paper and pencil assessment (e.g. dictation, filling in the blanks of conversation) is the best method in evaluating students' speaking and listening.)
12	수행평가는 실제의 언어 사용을 반영해야 한다. (Performance assessment questions should reflect real life language use.)
13	교사는 학생들을 평가하기 위해 다양한 평가 방법을 사용할 필요가 있다. (Language teachers need to use a variety of assessment methods to assess students.)
14	교사가 수행평가를 적절하게 준비할 수 있도록 충분한 시간이 필요하다. (Sufficient time should be allowed to properly prepare performance assessments.)
15	선택적 응답 문항(예: 매칭 문항, 객관식 문항, 참 - 거짓 문항)은 수행평가에 있어서 좋은 방법이라고 생각한다. (Selected-response items (e.g. matching items, multiple-choice items, true - false items) are good methods of performance assessment.)
16	서술형 응답 항목(예: 일기, 포트폴리오, 짧은 에세이, 문장 완성, 생각하여 쓰기)은 수행 평가로 좋은 방법이라고 생각한다. (Constructed-response items (e.g. journal entry, portfolio, short essay, sentence completion, reflective task) are good methods of performance assessment.)
17	학생의 자기평가는 수행평가로 좋은 방법이라고 생각한다. (Self-assessment by the student is a good method of performance assessment.)
18	학생들끼리의 동료 평가는 수행 평가로 좋은 방법이라고 생각한다. (Peer-assessment is a good method of performance assessment.)

C 섹션: 피드백 및 점수내기

19	학생들의 노력은 그들의 학습을 평가할 때 중요한 요소라고 생각한다. (Student effort is seen as important when assessing their learning.)
20	평가를 실시하기 전에 채점방식을 준비해야 한다고 생각한다. (A marking scheme should be prepared before assessment is given.)
21	학생들과 회의를 하는 것은 피드백을 주는 좋은 방법이라고 생각한다. (Conferencing with students is a good way of giving feedback.)

22	절대평가가 상대평가보다 낫다고 생각한다. (Criterion-referenced assessment is better than norm-referenced assessment.)
23	수행평가 후 학생들에게 피드백을 제공해야 한다. (Students should be given feedback after performance assessment.)
24	학생들은 평가를 받기 전에 채점 기준에 대해 알아야 한다. (Students should be informed about the marking criteria before being assessed.)
25	학생들은 채점 기준 작성에 참여해야 한다. (Students should be involved in preparing the marking criteria.)
26	학생들은 평가 후 일주일 이내에 수행평가 결과를 돌려받아야 한다. (Students should be given back their performance assessment results no later than a week after the assessment.)

다음은 연구를 위해 필요한 정보들입니다. 기재해주세요.

인구통계학적 정보

1. 학교: 중학교 / 고등학교
2. 성별: 남 / 여
3. 나이: 20대 / 30대 / 40대 / 50대 / 60대
4. 학위: 학사 / 석사 / 박사
5. 교육 경력(년): 5년 이하 / 10년 이하 / 15년 이하 / 20년 이하 / 20년 이상
6. 주당 평균 수업시수:
7. 한 반의 평균 학생 수:
8. 평가관련 연수 수강 횟수:

두번째 설문은 “수행평가의 교실 실행 상황”에 관하여 조사하는 설문입니다. 다음의 각 문장이 선생님께서 실제로 교실에서 수행평가를 실행함에 있어 얼마나 일치하는 지에 따라서 응답을 해주시면 됩니다. 아래의 척도를 사용하여 응답해주세요.

<p>전혀 그렇지 않음 (Never True (NT))</p> <p>드물게 그러함 (Rarely True (RT))</p> <p>종종 그러함 (Often True (OT))</p> <p>대부분 그러함 (Mostly True (MT))</p>

A 섹션: 수행평가의 목적

1	교사가 수업에 집중하게 해주고 있다. (helps to focus on teaching.)
2	교육적인 목적으로 학생들을 실력에 따라 그룹화하는 것을 가능하게 해준다. (helps to group students for instructional purposes.)
3	학생들의 강점과 부족한 점을 진단해주고 있다. (diagnoses strengths and weaknesses in students.)
4	수행평가를 통해 나의 수업의 강점과 부족한 점을 파악하고 있다. (diagnoses strengths and weaknesses in teaching.)
5	수행평가를 통해 학생의 발전상황을 알 수 있다. (provides information about students' progress.)
6	학생들 사이에 경쟁을 유발하고 있다. (creates competition among students.)
7	학생들에게 배움의 경험을 제공하고 있다. (creates a valuable learning experience for students.)
8	학생들의 학업에 동기를 부여하고 있다. (motivates students to learn.)
9	수행평가를 통해 학생들은 배우는 과정 중에 피드백을 받고 있다. (provides feedback to students as they learn.)
10	수행평가를 통해 학생들의 학습단계를 볼 수 있다. (demonstrates the level of students' learning.)

B 섹션: 수행평가의 방법 및 기술

11	지필 수행평가(예: 받아쓰기, 대화의 빈칸 채우기 등)로 학생들의 말하기와 듣기를 평가하고 있다. (Paper and pencil assessment (e.g. dictation, filling in the blanks of conversation) is the best method in evaluating students' speaking and listening.)
12	수행평가가 실제의 언어 사용을 반영하고 있다. (Performance assessment questions reflect real life language use.)
13	교사는 학생들을 평가하기 위해 다양한 평가 방법을 사용하고 있다. (Teachers use a variety of assessment methods to assess students.)
14	교사가 수행평가를 적절하게 준비할 수 있는 시간이 충분하다. (Sufficient time is allowed to properly prepare performance assessments.)
15	선택적 응답 문항(예: 매칭 문항, 객관식 문항, 참 - 거짓 문항)을 수행평가로 쓰고 있다. (Selected-response items (e.g. matching items, multiple-choice items, true - false items) are good methods of performance

	assessment.)
16	서술형 응답 항목(예: 일기, 포트폴리오, 짧은 에세이, 문장 완성, 생각하여 쓰기)을 수행평가로 쓰고 있다. (Constructed-response items (e.g. journal entry, portfolio, short essay, sentence completion, reflective task) are good methods of performance assessment.)
17	학생의 자기평가를 수행평가의 방법으로 쓰고 있다. (Self-assessment by the student is a good method of performance assessment.)
18	학생들끼리의 동료 평가를 수행 평가의 방법으로 쓰고 있다. (Peer-assessment is a good method of performance assessment.)

C 섹션: 피드백 및 점수내기

19	학생들의 학습을 평가할 때 노력을 중요히 여기고 있다. (Student effort is seen as important when assessing their learning.)
20	평가를 실시하기 전에 채점방식을 준비하고 있다. (A marking scheme is prepared before assessment is given.)
21	학생들과 회의를 통해 피드백을 주고 있다. (Conferencing with students is a good way of giving feedback.)
22	수행평가에서 상대평가가 아닌 절대평가를 쓰고 있다. (Criterion-referenced assessment is better than norm-referenced assessment.)
23	수행평가 후 학생들에게 피드백을 제공하고 있다. (Students are given feedback after performance assessment.)
24	학생들에게 채점 기준을 평가 전에 알리고 있다. (Students are informed about the marking criteria before being assessed.)
25	학생들을 채점 기준 작성에 참여시키고 있다. (Students are involved in preparing the marking criteria.)
26	학생들은 평가 후 일주일 이내에 수행평가 결과를 돌려받고 있다. (Students are given back their performance assessment results no later than a week after the assessment.)

APPENDIX 2: Interview Questions

1. 선생님이 생각하시는 수행평가란 무엇일까요?
2. 평소에 수행평가에 대해 고민해본 점이나 걱정해본 부분이 있나요?
3. 수행평가에 대해 고민해본 결과로 개선된 점이 있나요?
4. 수행평가를 실행할 때 자신 있는 점과 자신 없는 점은 어떤 것이 있을까요?
5. 수행평가를 짤 때 가장 고려하는 부분은 어떤 부분일까요?
6. 수행평가는 언제 어떤 절차로 계획하시나요?
7. 수행평가를 시행함에 있어 매년 다르게 새로운 시도를 해 보시나요?
8. 새로운 수행평가 아이디어는 어디서 얻으시나요?
9. 아이디어 회의 시에 발언권이 보장되나요?
10. 수행평가를 실제로 실시하기 전에 어떠한 사전준비가 있을까요?
11. 동료평가나 자기평가 등의 다양한 평가방법을 시도해 보신 적이 있으신가요?
12. 수행평가를 적절하게 준비할 수 있는 시간이 더 충분하다면 수행평가의 유형이나 방향이 바뀔 것이라고 생각하시나요?
13. 교사가 수행평가를 적절하게 준비할 수 있는 시간이 부족할 때 어떤 방법으로 이를 해결하시나요?
14. 지금까지 본인이 했던 것 중 가장 만족스러웠던 수행평가는 어떤 것이었을까요?

15. 지금까지 본인이 생각하기에 가장 최악의 수행평가는 어떤 것이었을까요?
16. 본인이 생각한 바른 방향으로 수행평가를 실행하는데 있어서 제일 큰 걸림돌이 무엇인가요?
17. 그 걸림돌을 극복하기 위해서 기울인 노력에는 어떤 것이 있었을까요?
18. 평가 관련 연수 중 수행평가 관련된 연수는 몇 회 참여하셨나요?
19. 평가 관련 연수 중 실제 수행평가에 도움이 된 부분은 어떤 것이 있었을까요?
20. 수행평가 관련 연수에 바라는 사항이 있나요?
21. 채점은 어떤 방식으로 진행이 되나요?
22. 수행평가의 결과가 교사가 추후에 수업을 진행하는데 영향을 미칠까요?
23. 수행평가 관련한 피드백은 보통 어떻게 제공되나요?
24. 수행평가의 결과가 학생들을 파악하는데 도움이 되나요?

국 문 초 록

평가는 교수와 학습에 있어서 중심적인 역할을 수행한다. 한국에서는 1999년부터 높은 수준의 복잡한 사고력을 평가하고 장려하기 위해 수행평가를 시행하고 있지만, 이는 수능과의 낮은 연계성과 점수에 대한 낮은 신뢰성으로 인해 원래 의도한 방향대로 실행되고 있지 못하는 경우가 많다. 타당도 높은 수행평가의 성공적인 시행을 위해서는 실행에 영향을 미치는 교사의 신념을 면밀히 조사하고 그 신념과 실행과의 관계를 탐구할 것이 요구된다.

이 연구의 목적은 수행평가에 대한 교사의 신념, 이 신념과 실행의 관계, 그리고 신념과 실행의 일치 정도에 영향을 미치는 상황적 요인들을 탐색하는 것이다. 이를 통해 타당성 있는 수행평가를 보다 만족스럽게 실행할 수 방법을 고찰해 보고자 한다.

이를 위해 한국의 중등학교에 재학 중인 영어교사 109명이 52개 항목으로 이루어진 온라인 설문조사에 자발적으로 참여하였다. 설문지는 세 개의 하위 섹션으로 구성되었으며, 이는 각각 수행평가의 목적, 방법 및 기술, 교사의 피드백 및 점수내기에 관한 질문으로 이루어져 있다. 설문지 응답을 군집분석 하여 참여자들을 군집화 하였으며, 각 군집에서 무작위로 선발된 2명의 교사들은 추후에 진행되는 20분 정도 소요되는 반구조적 면접에 참여하였다. 군집 분석 결과 참가자들은 총 3개의 군집으로 나뉘어 졌으므로 총 6명의 교사가 개별 인터뷰에 참여하였다. 이들은 수행평가에 영향을 미치는 상황적 요인들에 대해 의견을 나누었고, 이는 교사/교실 수준과 학교/지역사회 수준의 요인들로 구별되어 분석되었다.

연구 결과에 의하면 대부분의 한국의 영어 교사들은 학습을 위한 평가(Assessment for Learning)로서의 수행 평가에 대해 긍정적인 신념

을 가지고 있었다. 교사들은 수행평가를 통해 교수와 학습이 향상되어야 한다고 생각하였으며, 이를 위해 다양한 방법을 고안하고, 적절한 피드백을 제공해야 한다고 생각하였다. 그러나, 몇몇 측면에서 이러한 믿음과는 상충되는 교실에서의 실행이 발견되었다. 이러한 교사의 신념과 실행의 일치 정도에 영향을 미치는 주요 상황적 요인 중 교사/교실 수준의 요인으로는 교사 훈련과 학생들의 피드백에 대한 대처 방식이 꼽혔고, 학교/지역사회 수준의 요인으로는 동료와 학생 및 학부모로부터의 성적에 대한 민원이 발견되었다.

표본 크기 및 방법론적 측면에서 일부 한계가 있었지만 이 연구는 수행 평가에 대한 신념과 실천, 그리고 이들 사이의 일치 정도에 영향을 미치는 다양한 상황적 요인을 파악하여 한국의 중등학교 교사들이 타당도 높은 수행평가를 보다 성공적으로 시행하는데 도움을 줄 것으로 기대된다.

주요어: 교사 신념, 신념과 실행, 수행평가, 상황적 요인, 한국의 영어교사

학 번: 2019-21002