교육학박사학위논문

Exploring the Decision Making
of Korean High School Teachers of English
in Planning Performance Assessment

한국 고등학교 영어 교사들의 수행평가 계획 단계에서의
의사결정에 대한 탐구

2021년 8월

서울대학교 대학원
외국어교육과 영어전공

김 진 숙

Exploring the Decision Making
of Korean High School Teachers of English
in Planning Performance Assessment

한국 고등학교 영어 교사들의 수행평가 계획 단계에서의
의사결정에 대한 탐구

지도교수 소 영 순

이 논문을 교육학 박사 학위논문으로 제출함
2021년 3월

서울대학교 대학원
외국어교육과 영어전공
김 진 숙

김진숙의 교육학 박사 학위논문을 인준함
2021년 7월

위 원 장 _____

부위원장 _____

위   원 _____

위   원 _____

위   원 _____

ABSTRACT

Exploring the Decision Making of Korean High School Teachers of English

in Planning Performance Assessment

Jinsook Kim

Department of Foreign Language Education (English Major)

Graduate School of Seoul National University

The purpose of the present study is to explore how Korean high school teachers of English perceive and practice performance assessment, which is intended to be process-centered and *for* learning as presented in the assessment policies and guidelines of the 2015 revised national curriculum. To this end, the study analyzes the assessment methods and evaluation criteria teachers choose in planning performance assessments and the reasons why they make these decisions. This analysis is guided by the following two research questions: 1) How do Korean high school teachers of English plan their performance assessments and what assessment methods and evaluation criteria do they use? 2) How do they describe their choices of assessment methods and evaluation criteria?

Five high school teachers of English from two different schools (School X and School Y) in Gyeonggi Province participated in the study. The researcher conducted one-to-one 90-minute interviews with participants involving questions

formulated by reviewing recordings of the teachers' conferences for planning performance assessment and assessment plan documents. Ninety-minute group interviews followed in which participants were grouped by school. The researcher recorded and transcribed all the interviews and then analyzed the data. Also, the researcher examined the assessment planning documents of the two schools. The findings were interpreted in light of the concept of assessment *for* learning, which forms the core of the process-centered performance assessment presented in the 2015 revised national curriculum.

With respect to the first research question regarding the assessment methods and criteria chosen by the teachers of Schools X and Y, the findings demonstrated that performance assessments were conducted in a manner contradictory to the principles of assessment *for* learning. The teachers carried out writing, speaking, and listening assessments without actually teaching these skills. They did not give feedback during or after the assessments. Assessment methods for the speaking and listening assessments were not authentic and the assessment tasks for the speaking and class participation assessments largely drew on memorization. In addition, as with the class participation assessment, the teachers awarded grades to the students for completion and submission of their work rather than leading them to reflect on their learning.

The evaluation criteria also seemed to contradict the principles of assessment *for* learning. As with the essay writing and speaking assessments, the teachers assigned a heavier grade weight to task completion so as to judge the students' performances objectively, precluding criteria entailing subjective

judgment. The teachers prioritized ranking the students over promoting their learning and assigned heavier grade weights to the assessments emphasizing students' achievements. Moreover, the teachers did not clearly explain the criteria using concrete examples which could help students understand their learning goals.

With respect to the second research question, the findings showed that the teachers perceived the reasons for carrying out the assessments as "to comply with government policy," "to grade and report performance," "to fill out students' educational profiles," "to provide students an opportunity to practice speaking," and "to encourage students' active class participation." The teachers' perceptions of the reasons for carrying out the assessments appeared to be closely related with their selection of assessment methods and evaluation criteria. Furthermore, the teachers' perceived reasons for the assessments and decision making in selecting assessment methods and evaluation criteria were deeply affected by sociocultural factors, namely, bureaucratic pressures, such as government policy and reporting requirements, and considerations of university admission and external high-stakes assessment. Therefore, the present study found that sociocultural factors influenced teachers' perceptions of the reasons for implementing performance assessments, hampering the implementation of assessment *for* learning.

With these findings, the present study contributes to understanding how performance assessments are actually implemented in the classroom and provides pedagogical implications for successfully implementing and developing performance assessment *for* learning in a Korean EFL context.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1.

INTRODUCTION

The present study investigates how Korean high school teachers of English carry out performance assessment and why they carry it out in this way. Section 1.1 introduces the problems the present study addresses. Section 1.2 states the purpose of the study. Section 1.3 presents the two research questions. Section 1.4, the final section, describes the organization of the dissertation.

1.1. Statement of the Problem

Since the late 1990s, performance assessment has been widely used in Korean elementary and secondary-level school classrooms in the pursuit of alternative forms of assessment (e.g., Aschbacher, 1991; J. D. Brown & Hudson, 1998; Huerta-Macías, 1995; Norris, Brown, Hudson, & Yoshioka, 1998) to traditional standardized assessment (i.e., psychometric testing). Recently, with the growing popularity in international educational settings of the notion of assessment *for* learning (e.g., Black & Wiliam, 2009, 2010; E. Hargreaves, 2005; Hume & Coll, 2009; Kirton, Hallam, Peffers, Robertson, & Stobart, 2007; Marshall & Drummond, 2006; Wiliam, 2010), which proposes to promote (rather than simply

evaluate) students' learning, the Korean Ministry of Education (hereafter MOE) integrated *process-centered performance assessment* into the 2015 revised national curriculum (MOE & KICE, 2017). Based on assessment *for* learning, this new education policy opposes the assessment *of* learning merely for grading and reporting student performance thus emphasizing outcome over process. The new policy calls for English teachers to continuously assess students' performance during as well as at the end of teaching and learning, incorporating multiple forms of assessment such as short-answer/essay writing, presentations, peer and self-assessment, and logging individual student progress and development. As the policy emphasizes, process-centered performance assessment is meant to be *for* learning and diagnostic by providing students with feedback to promote their learning not only in cognitive but also affective domains and improve classroom pedagogy (KICE, 2017).

It seems that the MOE views process-centered performance assessment as a pedagogically desirable approach. Emphasizing achieving assessment *for* learning through process-centered performance assessment, research and studies by the Korea Institute for Curriculum and Evaluation (hereafter KICE) for the new 2015 revised national curriculum (KICE, 2018; MOE & KICE, 2017) stipulated that teachers should conduct process-centered performance in a manner according with its purpose, and to achieve this, teachers should follow the guidelines suggested by the Regional Office of Education to which the schools belong (i.e., Gyeonggido Office of Education in the case of high schools located in Gyeonggi Province) in implementing performance assessments. These guidelines (i.e., Gyeonggido

Office of Education, 2019) include particular types of performance assessment such as short-answer/essay writing assessment and stipulate the minimum grade weighting to be allocated for performance assessment.

Process-centered performance assessment has been implemented nationwide in the 10th grade since 2018, the 11th grade since 2019, and the 12th grade since 2020. In the meantime, the MOE has been actively training high school teachers to carry out process-centered performance assessment through online and offline programs.

Despite the government's enthusiasm and nearly 20-year policy support for performance assessment, process-centered performance assessment remains an underexplored area of research in the Korean EFL context. Only a few studies investigate perceptions regarding teachers' process-centered performance assessment in elementary or middle school settings (e.g., Ban, Kim, Park, & Kim, 2018), while high school settings remain unexplored. This omission may reflect the perceived lack of importance of process-centered performance assessment compared to regular paper-pencil assessments or external standardized testing such as the KSAT (Korean Scholastic Aptitude Test) in high schools. Meanwhile, the rising recognition of the merits of the newly introduced educational concept of assessment *for* learning has reinforced calls for performance assessment pursuing assessment *for* learning. In the interest of optimizing the implementation of performance assessment with the goal of assessment *for* learning in Korean EFL high school classrooms, then, it is necessary to explore the actual practice of performance assessment therein.

Besides this practical concern, there has been little theoretically or empirically informed research on process-centered performance assessment in the Korean EFL context. In particular, there is a need to apply a theoretical frame to understanding the new MOE policy of process-centered performance assessment in terms of its key concept, that is, assessment *for* learning, along with its characteristics, background, and goals. Such a theoretical approach is a prerequisite for understanding the policy in practice. In this respect, the present study is a theoretical and empirical attempt to examine whether performance assessment is being implemented in Korean EFL high school classrooms in a manner consistent with process-centered performance assessment as defined in the 2015 revised national curriculum.

1.2. Purpose of the Study

The present study aims to understand how the newly introduced MOE policy of process-centered performance assessment is implemented in high school classroom settings. Through a theoretical framework and empirical investigation, it particularly aims at exploring how Korean high school teachers of English perceive and practice this policy in classrooms.

In order to accomplish these objectives, the present study carefully reviews the key concept of process-centered performance assessment presented in the 2015 revised national curriculum and the core theory which the new policy draws on, assessment *for* learning. In addition, the study relies on the theory of

classroom-based teacher assessment, a form of performance assessment planned and carried out by the teacher in the classroom context. The literature relevant to classroom-based teacher assessment emphasizes exploring teacher's decision making in order to understand how teachers perceive and implement classroom-based teacher assessment (e.g., S. Clarke, 1998; Davison, 2008; Freeman, 2002; Hall, Webber, Varley, Young, & Dormant, 1997; Rea-Dickins, 2001). In classroom-based teacher assessment, teachers engage in a decision-making cycle of planning, consisting of implementing, monitoring, and recording (Rea-Dickins, 2001). Decisions made at the planning stage, such as identifying purposes, choosing assessment activities, and preparing students, have a significant impact on subsequent procedures. In particular, determining the assessment *purpose* is the most important factor affecting the other assessment concerns of the planning stage, such as selection of assessment methods and criteria, as well as the following stages (e.g., Breen et al., 1997; Calderhead, 1996; S. Clarke, 1998; Cumming, 2001). In planning assessment, as suggested in previous studies, teachers often perceive conflicts between summative and formative purposes (e.g., Brindley, 2007; Davison, 2004; Rea-Dickins, 2004). In particular, policy documents emphasizing accountability tend to lead teachers to adopt a summative rather than a formative purpose focusing on enhancing students' learning and their own teaching (e.g., Leung & Rea-Dickins, 2007).

Finally, many researchers have revealed that myriad variables affect teachers' decision making and assessment practices. Since teachers are social beings, their decision making is affected by various sociocultural factors such as

5

national policies and reporting requirements (e.g., Broadfoot, Osborn, Sharpe, & Planel, 2001; S. Clarke & Gipps, 2000; Gipps, 1994; McMillan & Nash, 2000; Stiggins, 1999b), sociocultural values (e.g., Brookhart, 2003; Carless, 2005, 2011; Hamp-Lyons, 2007), and personal beliefs about assessment (e.g., G. T. Brown, Lake, & Matters, 2011; Harlen, 2006; Harlen & Deakin Crick, 2003)

Based on the literature reviewed thus far, building on the theoretical framework of assessment *for* learning, the present study aims to examine implementation of performance assessment in a Korean EFL high school context, which is intended to be process-centered as instructed in the 2015 revised national curriculum.

After exploring teachers' decision making in selecting methods and evaluation criteria for performance assessment, the study seeks to reveal the reasons why teachers make these particular decisions. This will allow for a deeper understanding of how performance assessment is being implemented in Korean high school classrooms and how to improve this implementation. Thus, the findings of this study may provide stakeholders involved in Korean education with valuable insights for the successful design and development of performance assessment *for* learning.

For the qualitative data of the study, five Korean high school teachers of English from two different high schools in Gyeonggi Province, one located in Ansan City and the other in Pyeongtaek City, participated in individual and group interviews. The researcher then analyzed the interview data along with relevant documents related to the two research questions.

1.3. Research Questions

In order to examine teachers' perceptions and practices of performance assessment in Korean high school classrooms, keeping in mind that this performance assessment is intended to be process-centered and focus on improving learning as defined in the 2015 revised national curriculum, the present study investigates the assessment methods and evaluation criteria the teachers selected while planning assessment and the reasons why they made these decisions. The research questions are as follow.

1. How do Korean high school teachers of English plan their performance assessments? What assessment methods and evaluation criteria do they use?

2. How do they describe their choices of assessment methods and evaluation criteria?

1.4. Organization of the Study

The dissertation is organized into six chapters. Chapter 1 introduces the problem, purpose, and research questions. Chapter 2 introduces the language assessment in social perspective and examines teachers as social beings. It discusses the importance of teachers' perceptions and sociocultural context for their decision

making regarding the purpose of an assessment. Then, it reviews the meaning and practice of performance assessment and the definition of process-centered performance assessment presented in the 2015 revised national curriculum. Also, it presents a theoretical framework of assessment *for* learning and discusses the challenges of implementing assessment *for* learning in the Korean EFL context with respect to the concept of process-centered performance assessment. Chapter 3, pertaining to methodology, describes the participants, data collection procedures, and data analysis. Chapter 4 presents the findings with respect to the first research question and interprets the data in light of assessment *for* learning theory. Chapter 5 shows the findings regarding the second research question and discusses and analyzes them by drawing on previous studies. Finally, Chapter 6 evaluates the importance and significance of the study and suggests areas for further research.

CHAPTER 2.

LITERATURE REVIEW

This chapter first introduces a theoretical framework for language assessment as a social practice. In this context, teachers of classroom-based teacher assessment are social beings as well as agents, whose decision making is influenced by a variety of sociocultural factors. The second section explains the importance of teachers' decision making; specifically, it describes the importance of their perceptions of the purpose of classroom-based teacher assessment for planning and carrying out assessment. It also introduces the two purposes of classroom-based teacher assessment (summative versus formative) and discusses the contextual factors influencing teachers' determinations of the purpose. The last section introduces the meaning and practice of performance assessment in the Korean EFL context and the concept of process-centered performance assessment presented in the 2015 revised national curriculum. It then suggests the challenges of implementing assessment *for* learning in the Korean EFL context.

2.1. Language Assessment in Social Perspective

This section examines the social dimension of language assessment. Teachers are social beings as well as the agents of classroom-based teacher assessment, which means their decision making is affected by social context.

2.1.1. Theoretical Framework for Language Assessment as a Social Practice

Traditionally, language assessment emphasized validity (Bachman, 1990) and the achievement of language proficiency (Canale & Swain, 1980), which meant measuring cognitive differences between individuals using a psychometric approach drawing on linguistics and psychology. However, with the increasing understanding of the *social character* (McNamara, 2001) of both language assessment practices (Davies, 1997; Hawthorne, 1997) and constructs (Roever & McNamara, 2006), the traditional concepts of validity and language proficiency have been challenged. This turn toward the social character of language assessment has also been described in terms of abandoning the traditional positivist asocial paradigm of language assessment (e.g., Hamp-Lyons & Lynch, 1998; Lynch, 1997; Shohamy, 1998) in favor of sociocultural, constructivist, and cognitive perspectives, in other words, the interpretivist stance of postmodernism (Pennycook, 1994).

The most influential theory on the social dimension of language assessment is Samuel Messick's (1989) theory of test validity. Regarded as the preeminent theorist in this field, Messick insisted that all test constructs involve questions of values and developed the notion of consequential validity, namely, the consequences of score interpretation and test use. Figure 2.1 presents Messick's fundamental changes to the classic validity framework. The framework is read from left to right and top to bottom, with each cell successively encompassing the

concerns of the previous. In other words, the question of construct validity (upper left cell) entails consideration of utility (upper right cell), which entails consideration of values (bottom left cell), which finally entails consideration of social implications (bottom right cell). These last two concerns are of particular importance. *Value implications* refer to the value prioritized in a test context and how this value affects the interpretation of the test. *Social consequences* refer to the effects of a test's implementation.

FIGURE 2.1

Facets of Validity (Messick, 1989, p. 20)

	TEST INTERPRETATION	TEST USE
EVIDENTIAL BASIS	Construct validity	Construct validity + Relevance / utility
CONSEQUENTIAL BASIS	Value implications	Social consequences

Messick's theory prompted the Critical Language Testing movement (Lynch & Hamp-Lyons, 1996; Shohamy, 1998, 2001), which seeks to understand the ethics (Davies, 1997), political character (Khattri & Sweet, 1996), and impact of language assessment (Alderson, Clapham, & Wall, 1993). With researchers' interest in the social dimension of language assessment increasing, their attention has extended to the social character of language assessment in education systems,

including language assessments conducted in classroom contexts (e.g., Davison, 2004; Hall & Harding; 2002). Researchers have also begun to explore the social, cultural, and political values embedded in educational systems (e.g., Akiyama, 2004; Moore, 2005), government policy effects on standards-based assessments in classroom-based teacher assessment (Fulcher, 2004), and influence of the accountability concerns of educational providers on school curriculum and management (Evans & Hornberger, 2005).

In summary, since the turn toward the social dimension of language assessment practices and constructs, supported by Messick's (1989) theory emphasizing value implications and social consequences, researchers have come to question language assessment as an objective and value-free practice, instead focusing on the social values and consequences of language assessment. In this respect, researchers have investigated the social and political function of language assessment (e.g., Alderson, Clapham, & Wall, 1993; Davies, 1997) and paid more attention to social, cultural, and political influences on education systems and classroom-based teacher assessment (e.g., Abedi, 2004; Rogoff, 1990).

2.1.2. Teachers as Social Beings as well as Agents of Classroom-based Teacher Assessment

With regard to the social character of language assessment, some researchers have given special attention to the social dimension of *classroom-based teacher assessment* (e.g., Broadfoot et al., 2001; Hill & McNamara, 2012; Yin, 2010),

which refers to language assessment carried out in classrooms. Classroom-based teacher assessment has been increasingly promoted and adopted in the national educational policies of many countries including Australia, New Zealand, Canada and the United Kingdom (Davison & Leung, 2009; Leung, 2004; Leung & Rea-Dickins, 2007). Asian countries including Hong Kong and Singapore have also actively joined the mainstream (Fulmer, Tan, & Lee, 2019; Kennedy, Chan, Fok, & Yu, 2008; Tan, 2011).

The term classroom-based teacher assessment is interchangeably used with *teacher-assessment*, referring to assessment conducted by teachers in the classroom (e.g., Leung, 2005). More specifically, focusing on *who* is conducting the assessing, Hill and McNamara (2012) define classroom-based teacher assessment as "any reflection by teachers (and/or learners) on the qualities of a learner's (or group of learners') work and the use of that information by teachers (and/or learners) for teaching, learning (feedback), reporting, management or socialization purposes" (p. 397).

In classroom-based teacher assessment, teachers are the *agents* of assessment, which means they are directly involved in a constant process of decision making. They collect information about and evaluate students' learning and decide whether students have met the learning goals (e.g., Black & Wiliam, 2009; Cizek, 1996, 2000; C. M. Clark & Peterson, 1986; Hill & McNamara, 2012; Leung, 2005; McMillan, 2003). Emphasizing the importance of teachers' decision making in assessment, Davison (2004) asserts that the range and frequency of teachers' decisions determine the characteristics of classroom-based teacher

13

assessment. Similarly, Freeman (2002) claims that understanding how teachers make decisions in assessment is crucial to improving the assessment practice itself.

Considering the importance of teachers' decision making in classroom-based teacher assessment, many researchers assert the need to explore what factors affect teachers' assessment decision making. This is because, on the one hand, classrooms are not isolated or decontextualized spaces. Thus, classroom-based teacher assessment is practiced as a kind of language assessment that entails the imposition of values and a social impact, as Messick (1989) argues. On the other hand, as previous studies show, teachers, the agents of the assessment, are social beings whose assessment decisions are subjective, multi-dimensional, and context-dependent (e.g., Fulmer, Lee, & Tan, 2015; Broadfoot et al., 2001). In other words, much previous research argues that teachers' decision making is highly influenced by contextual factors (e.g., Leung & Teasdale, 1997; Rea-Dickins, 2001). In this respect, the present study will examine contextual factors influencing teachers' decision making as found in previous studies (Section 2.2.3).

In summary, the social dimension of language assessment has also been examined with respect to classroom-based teacher assessment. In classroom-based teacher assessment, teachers act as agents of assessment whose decision making significantly influences the character and practice of assessment. Because teachers are first and foremost human beings and classrooms are not separate from society, their decision making is significantly influenced by a variety of contextual factors (e.g., Freeman, 1989; Fulmer et al., 2015; A. Hargreaves, Earl, & Schmidt, 2002; Inbar-Lourie & Donitsa-Schmidt, 2009) (see Section 2.2.3).

2.2. The Purpose of Classroom-based Teacher Assessment

In this section, the study examines the importance of teachers' decision making in implementing classroom-based teacher assessment. In particular, it investigates the significance of teachers' perceptions of the purpose of assessment for the planning and other stages of the classroom-based teacher assessment cycle. Following that, it introduces the two purposes of classroom-based teacher assessment (summative versus formative) and the conflicts between them. Lastly, it reviews the literature pertaining to factors that affect teachers' decision making and perceptions of the purpose of assessment.

2.2.1. The Importance of Teacher's Perceptions of the Purpose of Assessment for Decision Making

As discussed above, teachers' decision making plays a significant role in planning and implementing classroom-based teacher assessment. In this regard, previous studies have attempted to discern the decision-making stages of the assessment cycle (Davison, 2008; Hall, Webber, Varley, Young, & Dormant, 1997; Rea-Dickins, 2001). For example, Rea-Dickins (2001) claims that classroom-based teacher assessment is composed of four main decision-making stages: planning, implementation, monitoring, and recording and dissemination. Figure 2.2 below provides examples of the priority concerns at each of these four stages.

FIGURE 2.2

The Four Decision-making Stages in Classroom-based Teacher Assessment

(Rea-Dickins, 2001, p. 435)

Stage 1: Planning

- Identifying the purpose for the assessment? (why?)
- Choosing the assessment activity (how)
- Preparing the learners for the assessment
- Who chooses/decides for each of the above?

Stage 2: Implementation

- Introducing the assessment (why, what, how)
- Scaffolding during assessment activity
- Learner self-& peer monitoring
- Feedback to learners (immediate)

Stage 3: Monitoring

- Recording evidence of achievement
- Interpreting evidence obtained from an assessment
- Revising teaching and learning plans
- Sharing findings with other teachers
- Feedback to learners (delayed)

Stage 4: Recording & Dissemination

- Recording & reporting progress towards NC
- Formal review for LEA or internal school purposes
- Strategies for dissemination of formal review of learners

While classroom-based teacher assessment does not necessarily entail the entirety of this cycle, the important point is that decisions made in the planning

16

stage especially influence subsequent procedures (Calderhead, 1996; Davison, 2008; Rea-Dickins, 2001; Shulman, 1986; Yin, 2010). In this regard, Yin (2010) describes assessment planning as "setting the table for subsequent assessment" (p. 185).

Teachers may consider a number of factors in planning classroom-based teacher assessment. Rea-Dickins (2001) suggests the following concerns of the planning stage: identifying the purpose, choosing the assessment activity, and preparing learners for the assessment. Meanwhile, Davison (2008) proposes that the planning stage involves identifying goals, establishing standards and criteria, and selecting appropriate assessment methods/schedule.

Among these, researchers agree that *purpose* is the most important factor affecting decision making at any stage of the assessment process, including the planning stage (Breen et al., 1997; Calderhead, 1996; Cumming, 2001; Rea-Dickins, 2001; Shulman, 1986). For example, S. Clarke (1998) found that teachers chose different strategies for planning, providing feedback, and reporting the assessment depending on whether its purpose was formative or summative. Rea-Dickins (2001) corroborates S. Clarke's (1998) findings with respect to the ESL/EAL context. By examining EAL (English as an additional language) class teachers and language support teachers in elementary schools in England and Wales, Rea-Dickins (2001) discovered that teachers selected distinct assessment activities and content in the assessment planning stage and employed dissimilar assessment strategies in subsequent stages depending on the purpose of the assessment. Breen et al. (1997) observed similar findings: Teachers in Australian

ESL schools adopted different assessment methods and practices depending on the purpose of the assessment. Finally, drawing on interviews with writing teachers in ESL/EFL classrooms across different countries, Cumming (2001) observed that teachers' writing assessment practices, task selection, and standards for achievement were closely related with the purpose of the assessment. For example, teachers aiming to assess English for specific purposes tended to prioritize assessing students' specific writing competencies, employing limited forms of assessment and limited criteria for evaluation. On the other hand, teachers aiming to assess English for general purposes tended to focus on assessing students' individual development, using diverse methods and broad criteria for evaluation.

In summary, teachers are the agents of assessment, involved at all stages of the assessment cycle, from planning to implementation and evaluating what needs improvement. As emphasized in existing research, decision making in the planning stage is important because it significantly affects the subsequent processes. In particular, teachers' determination of the purpose of the assessment in the planning stage substantially impacts their selection of assessment methods, standards or criteria, activities, and assessment strategies and practices.

2.2.2. Conflicts Regarding the Purposes of Classroom-based Teacher Assessment

Highlighting the importance of teachers' determination of assessment purpose, which significantly influences subsequent decision making in planning and other

procedures of classroom-based teacher assessment, most of the research in this field agrees that classroom-based teacher assessment plays two different assessment purposes, namely, assessment *of* learning, or summative assessment, and assessment *for* learning, or formative assessment.

On the one hand, when classroom-based assessment is used for summative assessment, the process of assessment tends to involve measuring students' achievement. The teacher thus collects certain data in order to assign the students a level (Leung & Dickins, 2007) on the basis of comparison with other students (norm-referenced) or pregiven standards (criterion-referenced). These results are then reported to parents, schools, and external authorities. This is the function of paper-pencil-based regular assessments (i.e., mid-term and final exams). In the words of the Qualifications and Curriculum Authority (QCA) (as cited in Leung and Mohan, 2004), classroom-based teacher assessment is a tool for assessment *of* learning, namely, summative assessment "carried out at the end of a unit or year or key stage or when a pupil is leaving the school to make judgments about pupils' performance in relation to national standards" (p. 337).

On the other hand, classroom-based teacher assessment can be *for* learning and formative in teaching and learning (e.g., Assessment Reform Group, 2002; Black, Harrison, Lee, Marshall, & Wiliam, 2004; Gipps, 1994). When the teacher perceives the function of assessment as *for* learning, the process of assessment tends to involve monitoring students' progress. Here, the teacher provides students with feedback and evidence of learning, which also informs the teacher with respect to improving the learning process (Assessment Reform Group, 2002; Rea-

Dickins, 2006). As the QCA (as cited in Leung, 2004) describes, assessment *for* learning in the classroom is chiefly characterized by "continual reflection and review about *progress*. When teachers and peers provide quality feedback, pupils are empowered to take the appropriate action. Teachers adjust their plans in response to formative assessment" (p. 22).

In general, teachers perceive the purposes of assessments according to their definition and characteristics. In other words, for classroom-based teacher assessment, which is intended to function as assessment *for* learning, teachers perceive its purpose as formative. And for assessment aimed at assessment *of* learning, teachers perceive its purpose as summative.

However, researchers have noticed that teachers perceive a tension between summative and formative purposes when implementing a classroom-based teacher assessment. Especially, they have found that with classroom-based teacher assessments officially meant for a formative purpose, teachers measure evidence of learning to report students' achievements rather than develop their learning and their own teaching (e.g., Arkoudis & O'Louphlin, 2004; Cheng, 2004; Gipps, 1994; Leung, 2004; McNamara, 2001).

Many studies have revealed that conflict between the purposes of classroom-based teacher assessment is inevitable because teachers are *social beings*, whose decision making is significantly affected by contextual factors (e.g., Brookhart, 2003; Cheng, 2004; Cheng, Rogers, & Hu, 2004, Davison, 2004; Freeman, 2002; Xiao, 2017; Xiaoju, 1990).

In summary, assessment *for* learning, or *formative assessment,* proposes to

promote students' learning, in contrast to assessment *of* learning, or *summative assessment*, which aims to grade and rank students. Assessment *for* learning requires students and teachers to continually reflect on the learning process. With the feedback acquired through the assessment activities, students are empowered and motivated to take the appropriate action to improve their learning. This ultimately leads students to be responsible for their learning. Meanwhile, the assessments enable teachers to adapt and modify their teaching to meet the particular learning needs of students. Although teachers generally perceive the purpose as intended, sometimes they discern conflicts between a formative and summative purpose, especially with respect to classroom-based teacher assessment, which is officially intended for formative assessment. Many previous studies have found that such conflicts are unavoidable because classroom teachers are social beings whose decision making is greatly influenced by contextual factors. In the following section, the present study thus examines the contextual factors influencing teachers' perceptions of the purpose of classroom-based teacher assessment.

2.2.3. The Sociocultural Contexts Influencing Teachers' Decision Making Regarding the Purpose of Assessment

As explained in the above sections, much research exploring classroom-based teacher assessment in social perspective has found that teachers' decision making is greatly influenced by internal and external contextual factors (e.g., Brookhart,

2003; Black & Wiliam, 1998; McMillan, 2003; Stiggins, 2001). Many studies (e.g., Allen, 1998; Stiggins, 1997; Wilson, 1996) have discovered that different levels of contextual factors affect teachers' decision making, from internal factors such as teachers' beliefs and attitudes to school or community context and national and sociocultural contextual factors. Most of these studies (e.g., Brindley, 2001; G. T. Brown, Lake, & Matters, 2008; G. T. Brown et al., 2011; Cizek, 2000, 2001; Kennedy et al., 2008; Xu & Liu, 2009) agree that, among such factors, in particular, national education polices and sociocultural values significantly affect teachers' decision making. This section thus reviews the literature on two sets of sociocultural factors affecting teachers decision making: national policies and sociocultural values.

National Policies and Reporting Requirements

Much previous research has found one of the most important factors affecting teachers' decision making in planning and implementing classroom-based teacher assessment to be national policies regarding assessment and curriculum (e.g., Black & Wiliam, 2004; Leung & Rea-Dickins, 2007; Leung & Scott, 2009; McKay, 2000). Most of these studies have discovered that teachers perceive a tension between summative and formative purposes due to the assessment frameworks in national educational policy (e.g., Black & Wiliam, 2004; Cizek, 2001) or curriculum (e.g., Bishop, 1997; Perfecto, 2012). More specifically, reporting requirements (e.g., Australian Certificate of Spoken and Written English

[Hagan et al., n.d.], the Canadian Language Benchmarks [Citizenship and Immigration Canada, 1993], and Common European Framework for languages [Trim, 1997]) typically featured in such assessment frameworks are often at odds with formative assessment. Even though these frameworks are presented as facilitating students' learning and improving the quality of teaching, the focus on "measuring learner performance summatively" and "reporting outcomes or evidence of students' progress and achievements" can interfere with assessment for learning (e.g., Cheng, 2004; Leung & Rea-Dickins, 2007; McNamara, 2001).

In most education systems, teachers and schools are responsible not only for teaching but also reporting their performances to an outside educational authority. This means that *accountability*, which is achieved through reporting assessment results, is an important part of a teacher's job (DeLuca & Johnson, 2017; Herman, 2007). In many countries, such as the United Kingdom, the United States, and Australia, national governments stipulate standards for learning outcomes, and teachers are expected to report on how and whether these standards are met through tests, report cards, etc. (e.g., Black, 1994). In this manner, reporting grades is not just about student but also teacher performance (Brindley, 1998, 2001; McKay, 2000). It is in this respect that Arkoudis and O'Loughlin (2004) observe the *administrative purpose* of assessment, referring to how governments document students' progress as a way of measuring educational accountability through assessment frameworks (i.e., English Curriculum and Standards Framework [CSF]) [Board of Studies, 2000b]) or education policy. Even though an assessment may be designed for a formative purpose, then, pressures with respect to

accountability can affect its implementation by teachers. Here, a key aspect of accountability is comparability (Gipps, 1994), or when concerns over accountability press teachers to ensure the conditions of assessment are consistent (i.e., comparable). For example, teachers should: present assessment tasks in a uniform manner; evaluate students' performance according to the same standards and rubric; and unvaryingly interpret assessment. In order to ensure such consistency, teachers frequently hold administrative meetings (Arkoudis & O'Loughlin, 2004; Brindley, 2001; Gipps, 1994). While carefully controlling assessment conditions, tasks, and criteria is important for summative assessment and accountability, these concerns can rather hinder the implementation of formative assessment. Indeed, calling for dynamic interactions between teachers and students, formative assessment entails contingency and spontaneity, precluding uniformity. In this respect, Gipps (1994) and Leung (2004) insist that formative and summative purposes are nearly incompatible. Likewise, the Assessment Reform Group (1999) claims that integrating grading or assigning levels into formative assessment is the main reason why assessment designed as formative is implemented as summative.

Nevertheless, there are some studies claiming that summative and formative assessment are in fact compatible (e.g., TGAT, 1988). According to this argument, assessment can be formative when carried out during the teaching and learning process but can also ultimately serve a summative purpose when scores are added up at the end of the course. However, other studies caution against such a mixed approach (Sadler, 1989). On the one hand, it may hinder summative assessment

by being too impressionistic (Leung & Teasdale, 1997). On the other hand, it may hinder formative assessment by producing in students the mindset that if an assignment does not contribute toward grade total it is not worth doing (Gipps, 1994).

Another problem in determining the purpose of assessment is the role of teachers. Formative assessment requires that teachers monitor students' language progress and provide feedback toward helping students improve. In other words, teachers act as a *facilitator* or *monitor* in formative assessment (Harlen & Winter, 2004). Under pressure to identify levels of achievement, however, teachers end up acting as a *rater* or *judge* (Leung & Rea-Dickins, 2007), the roles of the teacher in summative assessment. And under pressure to be objective and accurate, sometimes teachers just end up *ticking the boxes in a checklist* (Davison, 2004), that is, acting as *technician* by adopting *mechanistic criterion-based approaches* to the evaluation criteria (Davison, 2004; Davison & Leung, 2009; Leung & Rea-Dickins, 2007; Rea-Dickins, 2006).

Meanwhile, some studies investigating how assessment policies or national curriculums are carried out in local educational contexts show that assessment frameworks with a strong bureaucratic orientation can have problematic effects. In other words, under bureaucratic pressure, teachers can lose control of their assessment practices as well as teaching. Arkoudis and O'Loughlin (2004), for example, observed that the ESL companion, a state-mandated assessment framework in Australia, was often inappropriate for local education contexts. Meanwhile, ESL teachers in Australia lacked the authority to criticize or modify

the ESL companion document. The teachers were well aware of their responsibility to teach and assess students and were willing to develop more detailed assessment descriptors (i.e., related with criteria) and tailor them to their teaching context. However, they were unable to change the ESL companion because the Department of Education did not regard their suggestions as having a strong theoretical basis. Furthermore, the Department of Education denied the responsibility for reconciling assessment standards with pedagogical practice, asserting its task of delineating the use of the ESL companion for assessing students in principle.

In addition, many studies also argue that, contrary to their intention, mandated policies do not actually ensure reliability and consistency in classroom-based teacher assessment (e.g., Davison, 2004; Wiliam, 2001). While mandated policies are supposed to be both teacher- and context-free, studies show that their articulation can be ambiguous (Brindley, 2001), requiring teachers' interpretation. This can lead teachers to resort to assumptions (Davison, 2004).

In summary, teachers often perceive conflicts between the purposes of assessment that emerge from the competing demands of national educational policies and reporting requirements. Thus, even when assessments are officially designed with a formative purpose, teachers may end up implementing assessment with a summative purpose. Summative assessment involves measuring and assigning students' achievements a particular level in order to fulfill administrative and reporting duties (i.e., accountability), entailing consistency and uniformity in implementation. As mentioned above, teachers thus take on a role as a *rater* in

summative assessment. In contrast, formative assessment is characterized by contingency and spontaneity, requiring teachers to be *facilitators* or *monitors* interacting with students dynamically. Some studies also argue that national polices fail to consider specific classroom conditions, weakening teachers' authority over assessment and teaching, and require teachers to interpret vague statements relying on unexamined assumptions, thus undermining the very reliability and consistency they demand. Due to such policy inconsistencies, formative assessment is often displaced by summative assessment in the implementation process.

Sociocultural Values

Along with education policies, previous studies also regard sociocultural values and norms as significant factors affecting teacher's perceptions of assessment purpose (e.g., Cheah, 1998; Davison & Leung, 2009; Hamp-Lyons, 2007; Xiao, 2017). Socioculturally embedded meaning or values can shape or constrain teachers' decision making and perceptions of the purpose of assessment (e.g., Black & Wiliam, 2005; G. T. Brown et al., 2008; Kennedy et al., 2008). For example, examining assessment traditions in England, France, Germany, and the United States, Black and Wiliam (2005) revealed the complex interaction between shared sociocultural values and beliefs and teachers' teaching, decision making, and assessment and how these dynamics differed by society.

For the present study, research on the impact of Confucian-heritage

sociocultural settings on assessment are particularly noteworthy (e.g, Biggs, 1996; Carless, 2011; Li, 2010; Zhu, 1992). Carless and Lam (2014a) observe the intense prevalence of summative testing in Confucian-heritage societies such as China, Hong Kong, Singapore, Taiwan, and Japan and how this prevalence obstructs the acceptance of formative assessment. With a history going back as far as 2,000 years, competitive examinations originated in around 165 BC in Han-dynasty China to select government officials (Wright, 2001), and similar examination systems disseminated to the other Confucian-heritage countries of East Asia. Based on the idea that education leads to self-betterment, the examinations put special emphasis on mastery of the classics (Zhu, 1992). What encouraged the competitive examination system to take root was the *collectivist* nature of Confucian-heritage societies. In collectivist societies, since benefits tend to be distributed through personal relationships, there is a strong need for impersonal examinations to evaluate candidates for government office (Rohlen, 1983). Taking root through prolonged usage, testing came to be seen as a fair means of allocating and improving social status (Li, 2010). The ultimate significance of this context, with respect to the current study, is that in Confucian heritage contexts the main purpose of education has been perceived as passing examinations (Cheng, 1999).

In Confucian-heritage societies, both internal and external summative assessment strongly affect the education setting. According to Kennedy et al. (2008), internal summative assessment refers to the process by which teachers report grades and record students' achievements and progress (Harlen, 2007). External summative assessment, on the other hand, concerns testing administered

28

by institutions outside the school. These examinations influence the curriculum, pedagogy, and student learning strategies (Scott, 2001). Under such conditions, teachers particularly emphasize knowledge acquisition through memorization and practice. Furthermore, they tend to believe that achievement is a result of effort, regarding failure as a result of *laziness* rather than *lack of ability* (Biggs, 1996). Grades also serve as a tool for teachers to encourage students to strive for higher academic achievement or passing the examinations. Aware of the lasting social implications of exams on students' lives, it is quite difficult for teachers to abandon this results-oriented tradition.

This sociocultural context also influences students' learning dispositions. Students are under pressure to study hard and achieve high scores on exams. Accordingly, they do not seek to reflect on the knowledge they acquire in a manner beyond what is required for the exam (Carless, 2011; Carless & Lam, 2014b).

In Western cultures, by contrast, learning itself is viewed as the ultimate goal of education (Tweed & Lehman, 2002). According to Roos and Hamilton (2005), such essential cultural codes are not easily broken. It is thus understandable why formative assessment, which focuses on promoting rather than merely evaluating learning, is underappreciated in exam-centric Confucian-heritage societies compared with in Western societies. In this manner, emphasizing the importance of cultural effect on assessment, Hamp-Lyons (2007) divides cultures into *learning cultures* and *exam cultures*. She argues that learning cultures, where assessment is mostly used for learning and teaching, are better at implementing formative assessment, whereas exam cultures, where classroom-

based teacher assessment is used as preparation for external exams, are less adept in this regard. Likewise, Cheah (1998) and Kennedy et al. (2008) observe that countries with traditional exam-centric cultures, such as Hong Kong, Singapore, and China, experience difficulty in practicing meaningful formative assessment.

In summary, many studies (e.g., Black & Wiliam, 2005; Brookhart, 2003; G. T. Brown et al., 2008; Kennedy et al., 2008; Xiao, 2017) suggest that sociocultural values and norms influence teachers' decision making as heavily as national educational policies. With respect to East Asian countries of Confucian-heritage sociocultural settings, where the purpose of education is regarded as passing examinations and improving social status, some studies (e.g., Carless, 2005; Hamp-Lyons, 2007; Shen & Tam, 2008) argue that both internal and external summative assessment influences curriculum and pedagogy. Sociocultural values also affect teachers' beliefs about assessment and students' learning dispositions, leading them to treat learning as a means rather than an end. Naturally, such attitudes inhibit the implementation of formative assessment. Overall, unlike in Western cultures where learning itself is viewed as the ultimate goal of education, the values of exam-centric Confucian-heritage cultures hamper teachers' implementation of formative assessment meant to improve students' learning.

The Effect of Sociocultural Context on Teachers' Beliefs about Assessment

As reviewed thus far, previous research maintains that teachers' decision making is significantly affected by sociocultural context. This means that, viewed in a broader perspective, teachers' beliefs about assessment are substantially influenced by sociocultural context.

For example, Brindley (1989) and Stiggins (1999a) argue that some teachers believe assessment should be for learning for its own sake or the facilitation of learning, whereas others believe assessment should be for passing exams or getting higher scores in high-stakes assessments. Other studies find that teachers' perceptions of the goal of assessment substantially affect their decision making and implementation of the assessment in the classroom (Pollard, Triggs, Broadfoot, McNess, & Osborn, 2000; Reay & Wiliam, 1999). For example, Polland et al. (2000) propose that teachers who view the goal of assessment as helping students to pass an exam rather than to learn tend to prioritize coaching students to pass assessments over helping them to understand what it is they are actually learning. This leads students to adopt shallow learning strategies and avoid challenges (Ames & Archer, 1988; Benmansour, 1999; Crooks, 1988; Harlen & James, 1997).

Teachers' beliefs about what factors should be included in grading or rating assessment also influence their decision making with respect to the criteria for assessment and grading practices. For example, exploring the assessment practices of English language teachers in US high schools, Zoeckler (2007) concludes that some teachers believe that non-achievement factors such as effort and morality should be included when assessing students' language ability. Likewise, Pilcher-Carlton and Oosterhof (1993) maintain that, whereas some teachers prefer to

incorporate both achievement and effort in performing classroom-based teacher assessment and to allocate higher grades to students who work harder, some prefer to include only achievement in grading.

The notable point is that some studies have found that teachers' beliefs about assessment cannot be understood separate from sociocultural context. For example, G. T. Brown et al. (2011) found that secondary teachers in Finland and Queensland, unlike primary school teachers, spent much time preparing students for external high-stakes examinations, which secondary school students undergo prior to graduation. However, the teachers believed that the high stakes examinations were less valid, unreliable, and at odds with students' learning and their teaching (G. T. Brown, 2004; Shohamy, 2001). Conversely, it was found that Hong Kong (G. T. Brown, Kennedy, Fok, Chan, & Yu, 2009) and Singapore teachers (Noor, Muniandy, Krishnan & Mathai, 2010) believed that external examinations are an appropriate tool for measuring students' achievement and that ranking and competition motivate students to study harder. These findings demonstrate how the culturally shared values of a nation or region (e.g., East versus West), rather than the presence of external high-stakes examinations alone, can greatly affect teachers' beliefs, decision making, and assessment practices.

In sum, the literature has shown that teachers' beliefs about the purpose of assessment and grading and rating influence their decision making and assessment practices. At the same time, studies have found that teachers' beliefs about assessment are influenced by sociocultural values. In other words, teachers' beliefs cannot be understood apart from sociocultural context.

2.3. Process-centered Performance Assessment in the 2015 Revised National Curriculum

Based on the literature reviewed so far, this section focuses on process-centered performance assessment as presented in the 2015 revised national curriculum, which is the major object of the present study. First of all, to elucidate the distinguished features of performance assessment in the Korean EFL context, which are somewhat different from those commonly described in theory, the meaning and practice of performance assessment in the Korean EFL context is introduced in Section 2.3.1. Section 2.3.2 then introduces *process-centered assessment* and *process-centered performance assessment* based on the concept of assessment *for* learning as defined in the 2015 revised national curriculum. Following that, Section 2.3.3 presents a theoretical framework for assessment *for* learning and Section 2.3.4 discusses the challenges of implementing assessment *for* learning and critically reviews the literature on implementing assessment *for* learning in the Korean EFL context.

2.3.1. The Meaning and Practice of Performance Assessment in the Korean EFL Context

In the 1970s, traditional non-communicative assessment focusing on repeating mechanical words and sentences and producing preset question-and-answer patterns gave way to communicative assessment, or performance assessment,

which requires test takers to perform *real-life* language, namely, authentic, direct, and communicative language. Morrow (1977) defines performance assessment as evaluating students' use of spontaneous language in authentic activities and settings. J. L. Clark (1978) describes performance assessment as a *direct test* in which assessment tasks and procedures are as similar as possible to a real-life situation. Jones (1977) defines performance assessment as a test in which test takers demonstrate their ability to use functional language. As seen in these definitions, in performance assessment, students are expected to duplicate the language they would use in a non-test situation. In this respect, many researchers (e.g., J. D. Brown & Hudson, 1998; Shohamy, 1995; Wiggins, 2011) define performance assessment as assessment in which test takers are required to perform authentic tasks in a *real-life situation*.

Compared with traditional assessment, other characteristics of performance assessment include tapping into test takers' higher-level thinking and problem-solving skills, which are assessed in terms of process as well as end results. In terms of form, performance assessment can involve essay writing, interviews, problem-solving tasks, communicative pair-work tasks, discussions, and role-playing (e.g., Herman, Aschbacher, & Winters, 1992; Huerta-Macías, 1995).

In the Korean EFL context, performance assessment is used as an important measure of students' English ability in primary and secondary schools. In education policy (MOE & KICE, 2017), performance assessment is defined as assessment that directly tests students' knowledge, ability, and attitude as demonstrated in their performance. It contrasts this type of assessment with

traditional results-oriented testing, such as multiple-choice pencil-and-paper tests. The context should be similar to a real-life situation so students can demonstrate their performance ability in a real-life situation. Tasks should use a direct method such as writing or speaking. Finally, performance assessment should be a learning exercise and not purely evaluative.

In Korea, English performance assessment is cooperatively planned in English teachers' conferences held at the beginning of the semester by all the English teachers responsible for a given grade (i.e., year). The completed assessment plans are then submitted to the relevant office of education. After receiving feedback on the assessment plans from the educational offices, the teachers revise the plans if necessary. The revised plan is inspected in the committee for assessment in each school, and if there are no flaws, it is finally accepted. The final plan is then uploaded to the school website, allowing students, teachers, and parents to access information regarding performance assessment for the semester.

In practice, performance assessment has not been implemented in Korean classrooms completely in accordance with the policy description. For example, performance assessment has involved multiple-choice listening tests. In fact, the Gyeonggido Office of Education (2019) precludes using one-off, paper-pencil assessments for performance assessment for the reason that such practices invalidate the original objective of performance assessment, but it grants an exception with respect to assessing *listening ability* in English learning. Another example is the acceptance of *active class participation* as a form of performance

assessment. In other words, teachers count working hard in the classroom by responding to questions, completing tasks, or doing homework as performance assessment. Therefore, performance assessment tasks in Korean classrooms are not always authentic or direct, contradicting the definition of performance assessment in theory and in education policy. A better definition of performance assessment *in practice* in the Korean context, then, would be all assessments planned, implemented, and scored by the teacher for calculating semester grades other than midterm and final examinations.

In summary, the literature regarding performance assessment states that performance assessment should measure students' knowledge and communicative skills by employing authentic tasks such as writing essays that allow students to apply their knowledge in a real-world situation. In the Korean EFL context, although education policy defines performance assessment in accordance with theory, performance assessment in practice sometimes deviates from this definition. This means that performance assessment can be practically described in the Korean context as all assessments designed and formally conducted in classrooms by teachers contributing to a student's semester grade other than the midterm and final examinations.

2.3.2. Process-centered Assessment and Process-centered Performance Assessment in the 2015 Revised National Curriculum

The concept of *process-centered assessment* was introduced in the 2015 revised national curriculum. This new MOE policy pertains to all subjects (Chang et al., 2017; MOE & KICE, 2017). The KICE (2018) defines and describes the characteristics of process-centered assessment chiefly in terms of collecting information about students through classroom learning activities and providing feedback for the sake of students' development. Another key aspect is that the assessment is carried out by the same teacher who teaches the actual class content. This means that the teacher oversees the entire procedure of process-centered assessment (KICE, 2018, p. 165)

With respect to the background of the introduction of process-centered assessment, the KICE (2017) references the need to align with international trends, which means education emphasizing learning over accountability: "By collecting data regarding students' learning process rather than focusing on assessment for accountability, process-centered assessment should contribute to students' learning and development" (p. 15-16). [1] In other words, the KICE (2017) emphasizes that process-centered assessment should be *for* learning: "It is stressed that assessment is regarded as a tool *for* learning as part of the learning process, not as a means itself, or product-focused assessment, that is, assessment *of* learning" (p. 103).[2]

As shown thus far, the process-centered assessment in the 2015 revised national curriculum emphasizes assessment as *process*—as a part of learning. In

---

[1] All quotations from Korean sources are translated by the author.
[2] This passage is quoted from the English-language abstract of KICE (2017).

other words, the point of the new policy is that assessments in the classrooms be *for* learning, that is, a tool to promote students' learning.

The concept of process-centered assessment presented in the 2015 revised national curriculum has been carried out as a part of performance assessment in secondary schools. The MOE and KICE (2017) state that performance assessment should also be process-centered, reflecting the characteristics of process-centered assessment presented in the 2015 revised national curriculum. Admitting that even though performance assessment is already supposed to be process- as well as product-centered, the MOE and KICE (2017) acknowledge that the implementation of performance assessment thus far has not served to promote students' learning, focusing on product over process. Thus, the MOE and KICE (2017) assert that teachers should use performance assessment as a tool to promote students' learning and that performance assessment should be process-centered, which means teachers collecting various information and giving feedback to students during the assessment process toward aiding in students' learning and development. In this respect, the MOE and KICE (2017) list the characteristics of process-centered performance assessment as follow: 1) having a basis in the achievement standards of English learning; 2) carrying out assessment *during* (not after) the teaching process and connecting the assessment with teaching and learning; 3) measuring students' process in terms of knowledge, skills, and attitude; 4) collecting data from the assessments and providing feedback to the students in order to develop students' strength and reduce weaknesses; and 5) integrating teaching and learning and assessing and employing a variety of assessment

methods such as discussion, essay writing, projects, portfolios, and self- and peer assessment (p. 4-9).

In summary, the process-centered performance assessment presented in the 2015 revised national curriculum means that during the assessment *process*, teachers examine students' performance, collect data regarding their learning, and give feedback, motivating students to reflect on their learning and ultimately promoting students' learning. In this sense, assessment *for* learning, namely, assessment that improves students' learning, is the key characteristic of the process-centered performance assessment of the 2015 revised national curriculum. The new national educational policy is thus evidently designed not simply to evaluate but promote students' learning through performance assessment.

2.3.3. The Concept of Assessment *for* Learning

Assessment *for* learning is a concept that was introduced in the United Kingdom in an effort to develop national educational policy in the late 1980s. The Assessment Reform Group (1999), which played a major role in publicizing the concept, defines assessment *for* leaning as "any assessment for which the first priority in its design and practice is to serve the purpose of promoting students' learning" (Black et al., 2004, p. 10). The key point is the contrast with assessment *of* learning, which focuses on grading and ranking students rather than promoting their learning.

The Assessment Reform Group (2002) uses the term assessment *for* learning

synonymously with the term *formative assessment*. While this is generally the case in other studies, the Assessment Reform Group (2002) particularly stresses the *informing function* of assessment *for* learning, which helps the teacher decide and select what to teach in the next lesson and students understand what they have learned and what they need to learn next, as also shown in the following definition: "Assessment *for* Learning is the process of seeking and interpreting evidence for use by learners and their teachers to guide where the learners are in their learning, where they need to go and how best to get there" (Assessment Reform Group, 2002, p. 2). The Assessment Reform Group (1999) asserts that this informing function makes assessment *formative*, not only promoting students' learning but also improving their motivation and self-esteem, ultimately helping students to become responsible for their learning and engage in lifelong learning (Black & Wiliam, 1998; Day, 2002).

Assessment *for* learning is grounded in the *cognitive and constructivist* learning theory (Fensham, Gunstone, & White, 1994; Rogoff, 1990; Von Glasersfeld, 1987). In contrast to the behaviorist learning approach, drawing on the stimulus-response theory and focusing on rote learning, the cognitive and constructivist theory views students as analyzing and transforming new information drawing on what they already know, thus emphasizing learning contexts where students challenge their ideas (Iran-Nejad, 1995; Taylor, Fraser, & Fisher, 1997).

In accordance with the constructivist learning theory, assessment *for* learning treats students as active beings who can reflect on and improve their

learning with teacher guidance. Teachers are thus regarded as *supporters* or *facilitators* rather than *directors* of learning (Wood, Bruner, & Ross, 1976). Based on Vygotsky's (1962) concept of *scaffolding*, teachers are expected to examine students' performances and figure out their potential to advance in learning by pinpointing their *zone of proximal development* (Vygotsky, 1978), or the gap between what they can do on their own and what they can do with the help of others. The important point here is that some action must be taken to reduce the gap between actual and desired levels (Simpson, 2001). In this respect, feedback should be provided to the students indicating what needs to be done next (Sadler, 1989; Tunstall, & Gipps, 1996).

In assessment *for* learning, teachers' feedback plays an essential role (Black & Wiliam, 1998; Kluger & DeNisi, 1996; Sadler, 1989). Such feedback does not involve one-sided teacher-student interaction but discourse (Butler, 1988). This means students' involvement in discussion evoking thoughtful reflection on their learning (McNamara, 1998; Mohan, 1998). In addition, feedback should be constructive, addressing particular aspects of students' work and giving specific guidance regarding students' strengths and weaknesses in order for them to identify their own strengths and weaknesses (Black, 2001; A. L. Brown & Ferrara, 1985). Emphatically, such feedback does not mean grading or comparing students' work for the sake of ranking, which not only fails to offer guidance on improving students' learning but can also have a negative effect on students' motivation and self-esteem (Perrenoud, 1991; Pryor & Torrance, 1996). In other words, keeping in mind that feedback has an emotional impact, it should be constructive, sensitive,

and focus on students' work rather than the person, improving their enthusiasm and confidence (Dweck, 1986; Harlen, 2006; Perrin, 1991).

While feedback is important in assessment *for* learning, this does not mean students passively listening to the teacher (Fairbrother, 1995). Rather, they should actively incorporate feedback into their learning, taking responsibility for improving their learning (Andrews, Brown, & Mesher, 2018). In this respect, it is important that students be well aware of the goals or actual standards they are aiming toward (Sadler, 1989). Since it is often the case that learning goals are not explicit and require students' application in their implicit learning (Claxton, 1995), *criteria* can be useful to help students understand the goals teachers have in mind (Andrade, 2013; Zhang & Burry-Stock, 2003). Such criteria should be explained by the teachers in language the students can understand accompanied by examples of actual work seeking to fulfil the criteria (Frederiksen & Collins, 1989; Frederiksen & White, 1997). Only when students adequately understand learning goals may they begin to realize the disparities between their work and teachers' expectations, allowing them to accurately assess their own performance and consider how to improve (Stobart, 2006; Wiliam, 2001).

This process is known as *self-assessment*, another important feature of the constructivist learning approach, involving self-diagnosis, self-reflection, and planning ahead for improvement (Klenowski, 1995; Wiggins, 1992). This is also an important aspect of one of the overall goals of assessment *for* learning, which is teaching students to engage in *self-regulation* and *life-long learning* (Kitsantas, Robert, & Doster, 2004; Sadler & Good, 2006; Schunk, 1996).

42

Lastly, in assessment *for* learning, information teachers elicit by examining students' learning should also be used to modify and adjust teaching (Crooks, 1988; Dwyer, 1998). Although self-directed learning is a core characteristic of assessment *for* learning, the fact that teachers have to play a role in encouraging students to learn in this manner means they must adapt their teaching to changing circumstances in the classroom (Shepard, 2001). Only by being flexible and responsive can teachers ultimately promote students' learning in accordance with their needs (Black & Wiliam, 1998).

In sum, in assessment *for* learning, students should be responsible for their learning because the ultimate goal is for them to become *life-long learners*. For teachers, this means involving students in their learning, informing them of how well they are doing, guiding them towards next steps by providing constructive feedback, and responding to students needs by adjusting their own teaching.

2.3.4. The Challenges of Implementing Assessment *for* Learning in the Korean EFL Context

As illustrated thus far, assessment *for* learning is generally seen as desirable in Korean education policy, as reflected in the 2015 revised national curriculum's emphasis on integrating process-centered performance assessment into English-learning classrooms in order to improve students' learning and teachers' teaching, ultimately leading to students' life-long learning (Black & Wiliam, 2009, 2010; Bolhuis, 2003; Wiliam, 2010). However, as previous research indicates (e.g.,

Cheng, 1999; Morrison & Joan, 2002), changes in education practices often take longer than preferred by politicians and administrators. In fact, Morris, Fung-Lo, Chik, and Chan (2000) argue that assessment reform, especially when introducing progressive practices, is most resistant to reform. Moreover, the ultimate goal of assessment reform, namely, improving learning and teaching, is not easily attainable, requiring a prolonged period of time (Biggs, 1998; Yung, 2001).

Thus, previous studies argue that assessment reform is a complex and difficult endeavor (e.g., Fullan, 1999). Simultaneously, many studies assert that change in classroom-based teacher assessment is not achievable without teachers' active engagement (e.g., Day, 2002; Knight, 2002). They emphasize the important role of teachers as agents in classroom-based teacher assessment. Assessment innovation, then, should be preceded by changes in teachers' pertinent knowledge and attitudes (e.g., Carless, 2005; D. Clarke & Hollingsworth, 2002; Stiggins, 2001).

This fact suggests the importance of investigating teachers' perceptions of the newly introduced assessment policy in Korea. There have been some attempts to explore teachers' perceptions regarding process-centered assessment since its introduction. For example, Ban et al. (2018) surveyed primary and middle school teachers to investigate how they perceive *teacher by teacher process-centered assessment*, which is planned and implemented by teachers individually (i.e., without cooperation with other teachers). Teacher by teacher process-centered assessment is also in the 2015 revised national curriculum, even though, in practice, it has not yet been conducted in primary, middle, or high school settings.

44

Ban et al.'s (2018) study is meaningful in examining how primary and middle school teachers evaluate the potential of practicing teacher by teacher process-centered assessment. While the teachers perceived that teacher by teacher process-centered assessment would enhance students' creativity and comprehensive thinking ability, they were nonetheless concerned about the objectivity and fairness of the assessment, excessive workload for teachers, and lack of teacher training. However, because the study did not aim to investigate process-centered assessment in terms of its core meaning, it does not deal with process-centered assessment from the perspective of assessment *for* learning.

Ko (2019) also examined primary and middle school teachers' perceptions of teacher by teacher process-centered assessment. Although the teachers felt process-centered assessment was necessary to improve teaching and learning, they also thought that their lack of assessment expertise and insufficient understanding of process-centered assessment would eventually lead to a disparity between planning and actual practice. The study thus focused more on the meaning and character of process-centered assessment compared to Ban et al. (2018). In using only survey questionnaires to explore the factors inhibiting the practice of process-centered assessment and the support teachers would like to receive, however, it presented only quantitative data. In this respect, the study lacks qualitative data on how teachers perceive process-centered assessment and how the assessment is actually implemented.

Lastly, Shin, Ahn, and Kim (2017) tried to analyze the policy of process-centered assessment by exploring teachers' performance assessment practices at a

middle school in Seoul using surveys and interviews. The study showed that while policy documents emphasize process-centered assessments, in practice, process-centered assessment was rarely actually carried out. The authors thus called for teachers' active involvement in studying process-centered assessment. However, their study only addressed short-answer and essay writing assessment rather than performance assessment in general. Furthermore, the limited scope of the interview data could hardly reveal significant information on teachers' perceptions and implementation of process-centered performance assessment.

Therefore, compensating for the limitations of previous studies regarding process-centered performance assessment in Korean classrooms, first, requires exploring performance-centered performance assessment in light of its core concept, assessment *for* learning, based on a theoretical framework. When equipped with an effective theory, actual practice can be more accurately and meaningfully analyzed, the problems identified, and better solutions to these problems offered. Second, more in-depth data on teachers' perceptions and implementation of assessment is needed. In order to attain such qualified data, observing actual assessment in progress would be best. But if this is not possible, in-depth interviews and close examination of documents related to the assessment could help to understand teachers' perceptions and implementation of assessment. Third, rather than focusing only on teachers' perceptions of the assessment, analysis of the factors influencing those perceptions should be carried out. When such affecting factors are understood, more fundamental and effective solutions for the problems in assessment can be examined from a variety of perspectives.

Finally, studies targeting high school settings are urgently needed. As revealed through a review of the literature on the implementation of performance assessment in Korean classrooms, high school settings have been overlooked, which means that a more comprehensive picture of how performance assessment is being implemented in Korean classrooms is yet lacking.

In order to compensate for the limitations and gaps in previous studies, the present study aims to explore the practice of performance assessment focusing on teachers' selection of assessment methods and evaluation criteria and their perceptions regarding these choices. For this analysis, the study uses a theoretical frame and draws on the secondary literature on assessment *for* learning (the core concept of process-centered performance assessment), emphasizes the importance of teachers as social beings as well as agents of classroom-based teacher assessment, and uncovers the factors that influence teachers' perceptions of the purposes of assessment. It is hoped that this exploration may contribute to ensuring that process-centered performance assessment—more specifically, assessment *for* learning—becomes a fundamental practice in Korean high school classrooms.

CHAPTER 3.

METHODOLOGY

In Chapter 3, Section 3.1 describes the participants of the study. Section 3.2 explains the data collection procedures. Finally, Section 3.3 discusses the data analysis with respect to the research questions of the present study.

3.1. Participants

With the aim of understanding teachers' decision making regarding assessment methods and evaluation criteria in planning performance assessment and their perceptions with respect to these choices, the present study draws on qualitative research inquiry and analysis. For the qualitative data, the researcher recorded teachers' conferences, examined assessment plan documents, and conducted individual and group interviews for analysis. The study assumes that examination of teachers' selection of assessment methods and evaluation criteria in planning assessment can reveal information regarding actual implementation following the planning stage. As previous studies indicate, decision making in the planning stage significantly affects subsequent practice.

The participants of the study consisted of a total of six volunteers who were Korean teachers of English in high schools in Gyeonggi Province: three teachers were responsible for the 11th grade of School X located in Ansan City and three

teachers for the 10th grade of School Y in Pyeongtaek City. Toward completing the recordings of the teachers' conferences but prior to the interviews, one of the teachers from School Y dropped out of the study. This teacher expressed that she felt participation in an individual interview would be too burdensome. Nonetheless, she permitted the researcher to use the recordings of the teachers' conferences in which she had participated and the assessment plan documents she had produced.

Accordingly, a total of five teachers participated in the interviews. Information about the participants is as follows. In School X of Ansan City, Teacher A was female, 41 years old, and had 17 years of teaching experience. She was interested in content-based English learning. Teacher B was female, 30 years old, and had five years of teaching experience. School X was her first workplace. She was interested in teaching English through group projects. Finally, Teacher C, who was the most senior among three teachers, was male, 42 years old, and had 17 years of teaching experience. His main interests included English assessment and curriculum.

In School Y of Pyeongtaek City, Teacher D was female, 43 years old, and had 18 years of teaching experience. Her main interest was English assessment. Teacher E, who was senior among the two, was female, 52 years old, and had 26 years of teaching experience. When asked her area of interest with respect to English teaching, she stated that she had a general interest. Table 3.1 shows the participant teachers' information.

TABLE 3.1

Information about the Participants

Participant	School	Sex	Age	Teaching Experience (years)	Teaching Interest Area
Teacher A	X	Female	41	17	Content-based English learning
Teacher B	X	Female	30	5	Group-work learning
Teacher C	X	Male	42	17	English assessment and curriculum
Teacher D	Y	Female	43	18	English assessment
Teacher E	Y	Female	52	26	General

3.2. Data Collection Procedures

3.2.1. Listening to Recordings of the Teachers' Conferences

The data collection began with the start of the fall semester and progressed for four months (i.e., from August through November) in 2019. Before conducting interviews with the participants, it was necessary for the researcher to examine what the teachers decided in the teachers' conferences with respect to planning performance assessment and how they came to these decisions. It was believed that this data could help the researcher to understand the research context before meeting the participants in person, providing insight for constructing the interview questions.

While observing teachers' conferences was assumed to provide valuable information, there was the concern that direct observation might influence the

50

conference process. In this regard, with the participants' agreement, the teachers' conferences for planning performance assessment were recorded with the researcher absent.

A total of three teachers' conferences were recorded at School X. The first conference, lasting for 26 minutes, concerned the kinds of assessment to be administered, the percentage each assessment would account for in semester grades, and dates for the administration of each assessment. The second conference, lasting for 34 minutes, concerned essay writing assessments (including discussion of contents and methods) and administrative processes. The third conference, lasting for 41 minutes, concerned the evaluation criteria and grading methods related to all the assessments.

A total of three teachers' conferences were also recorded at School Y. The first conference, lasting for 20 minutes, concerned the kinds of assessment to be administered and the evaluation criteria and rubrics to be used. The second conference, lasting for 18 minutes, concerned essay writing assessments, including discussion of types, contents of the tasks, and evaluation criteria. The third conference, lasting for 25 minutes, concerned more detailed information about what was discussed in the second conference.

Table 3.2 below describes the contents discussed in each conference of School X and School Y.

TABLE 3.2

Contents Discussed in Teachers' Conferences in School X and School Y

School	Conference Session	Duration (min.)	Contents
X	1st	26	• Kinds of assessment to be administered • Percentage each assessment accounts for in semester grade • Dates for the administration of each assessment
	2nd	34	• Essay writing assessments (i.e., contents and methods of the assessments and administrative processes)
	3rd	41	• Evaluation criteria and grading methods related to all the assessments
Y	1st	20	• Kinds of assessment to be administered • Evaluation criteria and rubrics related to all the assessments
	2nd	18	• Essay writing assessments (i.e., types and contents of the tasks and evaluation criteria)
	3rd	25	• More detailed information about what was discussed in the second conference.

3.2.2. Examining Assessment Plan Documents

Before carrying out the interviews, in order to construct the interview questions, the researcher listened to recordings of the teachers' conferences for planning performance assessments at School X and School Y and examined the assessment plan documents of both schools.

3.2.3. Conducting Individual and Group Interviews

To collect data on the teachers' decisions in planning English performance assessment, individual interviews were conducted followed by group interviews in each school. Each individual and group interview was carried out for approximately 90 minutes in an empty classroom in each school. In the case of Teacher D of School Y, the individual interview proceeded for 148 minutes, prolonged by Teacher D's sharing of in-depth opinions on performance assessment and specific information about the conducting of performance assessment in School Y. With regard to all of the individual and focus group interviews, they were semi-structured interviews and conducted in Korean. The interview questions were devised by the researcher in advance based on the data examined up to that point. The researcher recorded all of the individual and focus group interviews, transcribing the recordings for analysis. Also, the researcher took notes throughout the interviews to capture the interviewees' ideas and revise or develop interview questions.

In the individual interviews, following the techniques for interviews of Alemu, Stevens, Ross, and Chandler (2017) and Creswell and Poth (2016), the researcher started with small talk, asking about personal information (e.g., age, teaching experience, and teaching interest area) to build rapport. As the interview proceeded, when more precise or in-depth information was required, the researcher asked participants to clarify details without disrupting the flow of the interview: "That sounds interesting. Could you tell me more about that?" Occasionally, the researcher improvised questions, hoping to elicit new or more valuable information.

The researcher also conducted two group interviews, one for School X and another for School Y, in which all the participants of each school participated. According to the literature (e.g., Carter, Bryant-Lukosius, DiCenso, Blythe, & Neville, 2014; Creswell & Poth, 2016), conducting group interviews after individual interviews allows the participants to hear each other's responses and offer additional comments they might not have thought of individually. This is because participation in a group interview encourages identification and the sharing of various perspectives with respect to a single topic. For this reason, each group interview was carried out in an empty classroom at each school and progressed for approximately 90 minutes.

3.3. Data Analysis

After finishing the individual and group interviews, the researcher transcribed the interview data. In addition to the interview data, the researcher examined the assessment documents of School X and School Y again. The documents pertained to performance assessment composition and grade weighting, specific information on each performance assessment, evaluation criteria and grading scores of the performance assessments, and guidelines for specific performance.

For the data analysis of the study, a thematic analysis method (Braun & Clarke, 2006) was employed. Thematic analysis is usually used to sort a set of qualitative data such as interview transcripts into broad and common themes, following a six-step process: familiarization, coding, generating themes,

reviewing themes, defining and naming themes, and writing up (Braun & Clarke, 2012). Thematic analysis is a popular method in analyzing qualitative data in many disciplines because it allows the researcher to identify patterns or meanings, which can be interpreted with respect to the researchers' judgment to seek answers to the research questions.

Drawing on thematic analysis, the interview data was coded using the software program NVivo version 12 and sorted according to two themes corresponding to the two research questions: "what assessment methods and evaluation criteria are used by Korean high school teachers of English" and "how they perceive their choices of the assessment methods and evaluation criteria." With respect to the first theme, the assessment methods and evaluation criteria of Schools X and Y were interpreted in light of the concept of assessment *for* learning, the objective of process-centered performance assessment presented in the 2015 revised national curriculum. Regarding the second theme, the teachers' perceptions, in particular, the reasons regarding their choices of assessment methods and evaluation criteria were interpreted to understand why they chose the specific assessment methods and evaluation criteria they did.

In the last phase of the data analysis, to improve credibility and validity, a member checking technique (Harper & Cole, 2012) was used. The analyzed data was shared with the participants by e-mail, requiring them to confirm the authenticity and accuracy of the interview data and to comment on it (Creswell, 2007; Lincoln & Guba, 1985).

CHAPTER 4.

MISALIGNMENT BETWEEN

TEACHING AND ASSESSMENT

This chapter presents the findings regarding the first research question, "How do Korean high school teachers of English plan their performance assessments and what assessment methods and evaluation criteria do they use?" In assessment theory, *assessment methods* refer to the ways, techniques, and instruments through which teachers collect evidence of student learning or achievement (Brookhart, 2003; McMillan, 2013). Assessment methods are closely related to the purpose or reason the assessment is carried out because the assessment purpose is embodied in the assessment methods (Breen et al., 1997; Calderhead, 1996). In this respect, in assessment *for* learning, which focuses on improving learning, assessment methods are expected to elicit information for both teachers and students indicating where students are in their learning, where they need to go, and how best to get there (Shepard, 2001).

In addition, *evaluation criteria*, which refer to the establishment of clear standards of achievement with respect to learning goals, not only inform teachers of the factors they should take into account when judging students' performance but also make clear to students what they are expected to do to demonstrate achievement of the learning goals (Frey & Schmitt, 2007; Stiggins, 1999b). In

assessment *for* learning, which aims at students becoming responsible for their learning, it is important that students be well aware of evaluation criteria (Brookhart, 2001). Taking into consideration the character of assessment methods and evaluation criteria required in assessment *for* learning, the present study explored teachers' decision making in planning performance assessment, namely, their selection of assessment methods and evaluation criteria.

The data revealed that their decision making regarding assessment methods and evaluation was incongruent with assessment *for* learning. This was evident in examining the assessment methods and evaluation criteria of four performance assessments (essay writing assessment, speaking assessment, listening assessment, and class participation assessment) carried out at Schools X and Y in terms of assessment *for* learning.

4.1. Overall Structure of the Assessment

In order to plan the implementation of performance assessments, the teachers of Schools X and Y each participated in three teachers' conferences. They also filled out a standardized performance assessment plan document known as the "performance assessment composition and grade weight" form, which elicits information regarding assessment types, activities, and grade weights for the semester.

In the case of School X, as shown in Table 4.1, performance assessment accounted for 60% of students' semester grade, with the midterm and final

examinations accounting for the other 40% (20% each). The performance assessment consisted of three parts: essay assessment (35%), which was composed of essay writing (20%) and an oral presentation (15%), class participation (15%), and listening assessment (10%).

TABLE 4.1

Performance Assessment Composition and Grade Weighting: School X

Category	Assessment Type		Assessment Activity (Task)	Grade Weight (%)
Regular Assessment	Midterm Exam		Multiple-choice Test	20
	Final Exam		Multiple-choice Test	20
Performance Assessment	Essay Assessment	Essay Writing	Essay Writing	20
		Oral Presentation	Oral Presentation	15
			Essay as Script	
	Class Participation		Quiz	15
			Homework	
	Listening Assessment		Multiple-choice Test	10

When the teachers of School X were asked in the group interview why they determined the performance assessment activities and grade weights in the manner they did, they emphasized policy requirements: The minimum grade weight for performance assessments should be 40% and 35% of a student's grade should be allotted to essay assessments. This is demonstrated in the testimonies of Teachers A and C.

Teacher C: It's due to the policy. We have to assign a grade weight of at least 40%

for performance assessments. At the same time, we have to assign a 35%

grade weight to essay assessments. These are mandatory requirements.

In addition, we thought that listening assessments and class

participation assessments should also be involved.

Teacher A: Yes, right. Following the two requirements and adding the listening

and class participation assessments easily brought the grade weight to

60%.

(Group interview at School X)

Teacher C also added that when the teachers of School X plan performance assessments, they first design essay assessments accounting for 35% of a student's semester grade and then configure the grade weights for the other assessments.

Teacher C: Above all, we first decided on administering essay assessment

accounting for 35% of a student's semester grade. Then we began to

think about other performance assessments.

(Individual interview with Teacher C)

According to the Gyeonggido Office of Education (2019), regarding assessments for high schools located in Gyeonggi Province, the 2015 revised national curriculum requires performance assessments to account for at least 40%

of a student's semester grade and mandatory essay assessment to account for 35% of a semester grade. The teachers explained that they assigned a 35% grade weight to essay assessment in accordance with the policy that "essay writing should account for 35% of a student's semester grade" (Gyeonggido Office of Education, 2019, p. 7).[3] And as shown in Table 4.1, the essay assessment (35%) of School X was composed of essay writing (20%) and oral presentation (15%). Generally, essay assessment refers to essay writing. However, defining *essay assessment* as "assessment in which students express their opinions or arguments in a logical way," the Gyeonggido Office of Education (2019) allows for oral presentations to replace essay writing if they maintain the original characteristics of the essay assessment, "expressing opinions or arguments in a logical way." Following this definition, the School X teachers decided to implement essay writing (20%) and oral presentations (15%) together accounting for 35% of a total grade.

The School X teachers' inclusion of an oral presentation in essay assessment, although aligning with policy guidelines, is an example of how teachers' choices can affect the implementation of assessment. If the teachers wanted, they could have carried out essay writing with a grade weight of 35%, not including oral presentation. When asked why they chose to include an oral presentation, the teachers of School X responded that they thought performance assessment should include assessment of all four skills, as shown in the following excerpts.

---

[3] The author translated the guidelines from the original Korean.

Teacher B: Regular examinations only assess students' reading ability. But other

than reading, English learning consists of three other skills: listening,

speaking, and writing. So, I think that these skills of listening, speaking,

and writing should be assessed.

(Individual interview with Teacher B)

Teacher C: Listening, speaking, and writing should be included in performance

assessments. They are all skills included in English learning.

(Individual interview with Teacher C)

While performance assessment in School X accounted for 60% of a student's semester grade, in School Y, performance assessment accounted for 40% of a student's semester grade, with the midterm examination covering 30% and final examination 30%. With respect to organization, performance assessment in School Y was divided into four parts: essay assessment (i.e., essay writing I and essay writing II), listening assessment, dialogue recitation assessment, and class participation assessment. Table 4.2 below shows the performance assessment composition and grade weighting of School Y.

TABLE 4.2

Performance Assessment Composition and Grade Weighting: School Y

Category	Assessment Type	Assessment Activity (Task)	Grade Weight (%)

Regular Assessment	Midterm Exam		Multiple-choice Test (21%) + Short-answer Test I (9%)	30
	Final Exam		Multiple-choice Test (21%) + Short-answer Test II (9%)	30
Performance Assessment	Essay Assessment	Essay writing I	Essay Writing	10
		Essay writing II	Essay Writing	10
	Listening Assessment		Multiple-choice Test	10
	Dialogue Recitation Assessment		Dialogue	5
	Class Participation		Completing Review Papers Exercising Vocabulary	5

As stated in the explanations of School X's performance assessment, the Gyeonggido Office of Education (2019) asserts that essay assessment should account for 35% of a student's semester grade. It adds that a *short-answer test* can also be regarded as a form of essay assessment, defining it as "assessment in which students summarize, explain, or describe a given topic (i.e., a concept, problem, or what they understand)" (p. 7).[4] Complying with the above guideline related to essay assessment, School Y teachers structured the essay assessment in terms of essay writing I (10%), essay writing II (10%), short-answer test I (9%), which was embedded in the midterm exam, and short-answer test II (9%), which was embedded in the final exam. This means that essay assessment made up 38% of a student's semester grade, three percent over the standard 35% allotted for essay assessment. This aspect was different from School X, which did not use not short-answer tests in either the midterm or final examinations.

---

[4] The author translated the guidelines from the original Korean.

When asked why they structured short-answer test I and short-answer test II as essay assessment, School Y teachers answered that the inclusion of a short-answer test as essay assessment was their choice. Teacher D also explained that it was because the students did not have the ability to write essays.

Teacher D: The students in my school are not prepared for writing English essays. They couldn't even spell the vocabulary in the textbook very well. For these students, English essay writing is too burdensome. That's the reason why we embedded short-answer test I in the midterm exam and short-answer test II in the final exam. We wanted to relieve the students' burden.

(Individual interview with Teacher D)

Teacher E also described the essay assessments as oppressing the students.

Teacher E: The essay writing is oppressive for the students of our school. Yet we have to do it as performance assessment.

(Individual interview with Teacher E)

According to the School Y teachers' explanations, short-answer tests are much easier compared to essay writing because students can complete the task by filling in the blanks from a choice of words or by rearranging words provided. For this reason, School Y teachers chose to structure short-answer tests I and II as

a type of essay assessment, administering them through the midterm and final exams, respectively. Examples of the short-answer test questions are as follows: "Explain what the underlined phrase 'his ideas' refers to (within seven words for each idea)"; "Rearrange the words below to make a sentence appropriate for blank (A) in the paragraph below." Figure 4.1 shows samples of School Y's short-answer test I.

FIGURE 4.1

Samples of Short-answer Test I: School Y[5]

1. Explain what the underlined phrase "his ideas" refers to (within seven words for each idea).

[4 points, 2 points each]

John Roebling, an expert at building suspension bridges, proposed the use of steel cables—instead of iron ones—that would be six times stronger than needed to support the Brooklyn bridge. In addition, he planned to build two large stone towers to hold up the bridge's road and enable people to walk across it. If 'his ideas' worked, the final result would be the longest, strongest suspension bridge ever built. John's ambition affected people, so construction began in 1869. However, he was involved in a ferry accident later that year and died of an infection not long after.

(1) _____   (2) _____

2. Rearrange the words below to make a sentence appropriate for blank (A) in the paragraph below.

[4 points, partial points possible]

Words like 'near' and 'far' can mean _____(A)_____. If you were at a zoo, then you might say you are 'near' an animal if you could reach out and touch it through the bars of its cage. Here the word 'near' means an arm's length away. If you were telling someone how to get to your local shop, you might call it 'near' if it was a five-minute walk away. Now the word 'near' means much longer than an arms' length away.

are / doing / depending / where / different / you / and / what / are / on / you / things

---

[5] The author translated the "Samples of Short-answer Test I: School Y" from Korean. The passages themselves were originally provided in English.

By completing a form on "performance assessment composition and grade weight," the School X teachers provided specific information on assessment, as shown in Table 4.3. The essay assessment consisted of writing an essay introducing an admirable person, an oral presentation where the essays were used as speech scripts, class participation (i.e., answering the teachers' questions), homework, and a listening assessment involving 20 multiple-choice questions.

TABLE 4.3

Specific Information on Each Performance Assessment: School X

Assessment Type		Assessment Method	Content	Administration Date
Essay Assessment	Essay Writing	• Write a 120-word essay within 50 min. • Follow the given conditions • No English dictionary allowed	• Introduce an admirable person and evaluate him or her in a logical way	2nd week of September
	Oral Presentation	• Give a 3-min. speech in front of the teacher alone • No referring to the script	• Introduce an admirable person and evaluate him or her in a logical way	4th week of September
Class Participation		• Answer the teacher's questions, do homework	• Contents learned in the classes	Throughout the semester

	Answer the 20 multiple-	Contents presented in the	September 26
Listening Test	choice questions of the EBS	2015 national curriculum	
	listening test		

Table 4.4 shows specific information on the assessment for School Y. The essay assessment was composed of essay writing I and II, listening assessment consisting of 20 multiple-choice questions, a dialogue recitation assessment in which students recited a dialogue within a given time limit, and class participation assessment involving a review handout and writing down words missed in the vocabulary test.

TABLE 4.4

Specific Information on Each Performance Assessment: School Y

Assessment Type	Assessment Method	Content	Administration Date
Essay Assessment — Essay Writing I	• Write a guided essay regarding three questions for 50 min. • Adhere to the word limit • No English dictionary allowed	• Introduce a person who overcame adversity and some lessons to be learned	3rd week of November
Essay Writing II	• Write a guided essay following the two requirements within 50 min. • Adhere to the word limit • No English dictionary allowed	• Write a summary and impressions of *The Elephant Man*	4th week of November
Listening Assessment	• Answer the 20 Multiple-choice questions of the EBS listening test	• Contents presented in the 2015 national curriculum	September 26

Dialogue Recitation Assessment	• Recite a randomly selected dialogue among three memorized dialogues within 90 sec. • Perform in front of the teacher in the hallway	• Three dialogues learned in class	First week of November
Class Participation	• Complete the review handout related to reading, vocabulary, and grammar • Write down the words missed in the vocabulary test	• Contents learned in class • Words missed in the vocabulary test	Multiple times throughout semester

In sum, the performance assessment of School X accounted for 60% of students' semester grade. The structure of the School X plan consisted of essay writing (40%), oral presentation (15%), listening assessment (10%), and class participation (15%). The performance assessment of School Y accounted for 40% of a students' semester grade. Its structure consisted of essay writing I and II (20%), dialogue recitation assessment (5%), listening assessment (10%), and class participation (5%). Following the requirements of the Gyeonggido Office of Education (2019) regarding performance assessment, both schools structured the performance assessments as accounting for at least 40% of a semester grade and implemented a mandatory essay assessment accounting for at least 35% of a semester grade.

4.2. Essay Assessment

As evident in the overall structure of the performance assessments of both Schools

X and Y, essay assessment was the most weighted form of performance assessment implemented. According to the explanations of the School X teachers, for the essay writing assessment, students were asked to introduce an admirable person and evaluate him or her in a logical way in at least 120 words. In addition to number of words, there were seven other requirements for the essay related to sentence construction and grammar. In order to help the students follow the guidelines, the teacher provided and explained the essay topic and requirements to the students along with a writing sample one week before posting the assignment in the class. According to the guidelines, students were required to prepare their writing in advance and write the essay within 50 minutes on a designated assessment date (see Figure 4.2).

FIGURE 4.2

Guidelines for Essay Writing: School X[6]

<div style="border:1px solid">

Guidelines for Essay Writing

Write an essay about an "admirable person" (not living) relevant to your desired future career. The essay should meet the following eight requirements.

<Essay Requirements>

1) Include a title.
2) Include an introduction describing the reason for your selection.
3) Include a main body describing the admirable person's interests, passions, and major achievements.
4) Include three pieces of factual information about the person using at least three grammatically correct sentences, each of which must include a subject and a verb.
5) Use relative clauses (i.e., relative pronouns or relative adverbs) at least two times (excluding in the title).
6) Include a conclusion evaluating the person and resolving the essay.
7) The essay must consist of at least 120 words.

<Sample>

Seok Jumyeong: A Famous Korean Butterfly Doctor	
I've dreamed of becoming a naturalist. While studying butterflies, I read an article about a Korean researcher. His name is Seok Jumyeong. Born in Pyeongyang in 1908, he was a world-famous biologist who researched butterflies in Korea during the Japanese colonial period.	Introduction (describing the reason for your selection)
While in high school, he began to take an interest in nature. After graduation, he worked as a teacher at his high school, where he devoted himself to butterfly research for 11 years. He traveled around the country to collect a total of 750,000 butterfly samples during his lifetime. By analyzing this vast number of samples, he corrected research errors made by Japanese scholars and reclassified Korean butterflies into 248 species—previously, they had been classified into more than 800 species.	Main body (describing the admirable person's interests, passions, and major achievements)
I think he is a pioneering scholar who brought scientific fame to Korea and he had such an enthusiastic attitude that noting could stop him from researching his field. Like him, I will do my best to fulfil my dream with passion and effort.	Conclusion (evaluating the person and resolving the essay)

</div>

---

[6] The author translated "Guidelines for Essay writing: School X," which was originally written in Korean. However, the sample "Seok Jumyeong: A Famous Korean Butterfly Doctor" was originally written in English.

When asked why the teachers of School X provided concrete guidelines and a writing sample in advance, Teacher C explained that this was because they do not teach writing in class.

Teacher C: We actually do not teach writing in class. However, we should assess writing. That is the policy. So, to help the students write essays, we presented a sample essay, provided specific guidelines, and posted the information in the classroom one week before the assessment date.

(Individual interview with Teacher C)

Also, when asked the reason why School X teachers selected the writing theme "an admirable person," Teacher C answered that the teachers thought this would be an easy topic for students because one of the reading lessons was also about "admirable people." Teacher C further explained how a writing activity regarding "an admirable person" was already suggested in that lesson.

Teacher C: In the teacher conference for the performance assessment, we selected the essay theme. The reading part of Lesson 6 was about "admirable people." So, we thought that after learning that reading part, students could think of their own admirable person. Also, the activity "Write an essay about an admirable person" was presented in the writing activity part of the text in Lesson 6.

(Individual interview with Teacher C)

A similar context was present at School Y. With respect to the question of how they taught writing, Teacher D explained that they did not teach writing in class. Instead, teachers made a sample assessment form including evaluation criteria at the bottom of the page, which was meant to help students with their writing. The form was posted in the classroom one week before the assignment handout was distributed. Both essay writing I and II were administered for 50 minutes.

Essay writing I was a guided writing exercise involving three questions. The requirement of the essay was to "introduce a person who overcame adversity and express what lessons you learned from that person." The requirement was followed by three questions: "What difficulties did he or she face?" (30-50 words); "What did he or she do to overcome the difficulties?" (40-60 words); and "What did *you* learn from the example?" (30-50 words). Essay writing II was also a guided writing exercise, this time involving one requirement: "Choose a Korean cultural heritage and discuss three characteristics of it (80-120 words)." Figure 4.3 shows essay writing I of School Y.

FIGURE 4.3

Essay Writing I: School Y[7]

Essay Writing I

- ● Write short answers to the questions in English.
- ● Fill in the blanks with the correct words.

Topic: "Introduce a person who overcame adversity and describe what lessons you learned from that person."

1) What difficulties did he or she face? (30-50 words)

```
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   10
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   20
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   30
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   40
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   50
```

2) What did he or she do to overcome the difficulties? (40-60 words)

```
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   10
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   20
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   30
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   40
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   50
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   60
```

3) What did *you* learn from the example? (30-50 words)

```
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   10
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   20
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   30
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   40
--------   --------   ---------   --------   --------   --------   --------   --------   --------   --------   50
```

Assessment Criteria	Task Completion
	Content and Construction
	Grammar, Spelling, Vocabulary, and Expressions

---

[7] The author translated "Essay Writing I: School Y" from the original Korean.

When asked why they selected the themes they did for essay writing I and II, Teacher D responded that the theme for essay writing I was related to what students were expected to learn in Lesson 3 and that the theme for essay writing II was related to what students were expected to learn in Lesson 6. In addition, when asked in the group interview why they selected a guided writing genre for essay writing I and II, School Y teachers D and E explained their belief that universities prefer information on a student's book reports when examining his or her educational profile for admission and that information diversity in an educational profile is advantageous for a student's university application. They had thus chosen a book report as a writing assessment task for the previous semester and, when planning the performance assessments for this semester, tried to implement diverse performance assessment tasks, even inspecting students' educational profiles from previous years and excluding those tasks the students had already performed.

Teacher D: When determining the performance assessment tasks, it is very important to consider what we should record in the "Specific Competencies and Specialties in English Learning" section in the educational profiles.

Teacher E: Yes, we cannot overlook that!

Teacher D: We tend to choose the writing task that might offer information the universities prefer. That's why we selected a book report task last semester.

Teacher E: Right, we cannot disregard which tasks are beneficial for students'
applications to university. In this respect, because the students already
performed a book report assessment last year, we did not choose this
task for this semester. If we chose a book report task again, information
on how a student did on a book report would be recorded again. This
results in overlapped information in the student's educational profiles.
We need new and diverse information.

Teacher D: Yes, so, we did not select a book report task this semester.

(Group interview at School Y)

Furthermore, Teachers D and E decided to conduct two essay writing assessments each accounting for 10% of a student's semester grade instead of one essay writing assessment accounting for 20%. Teacher D explained the rationale for this decision was that two different essays could offer two different sources of information for filling in the educational profiles.

Teacher D: In order to record diverse information in the "Specific Competencies
and Specialties in English Learning" section of a student's educational
profile, we determined to conduct two essay writing assessments with
two different themes.

(Group interview at School Y)

When examined with respect to the concept of the assessment *for* learning,

the data regarding essay assessment discussed thus far suggests the need to consider whether the teachers at both schools actually perceived and implemented the essay assessment as assessment *for* learning. The Assessment Reform Group (1999, 2002) asserts that for assessment to be *for* learning, it should be conducted in relation to teaching and learning. Describing the characteristics of process-centered performance assessment, MOE and KICE (2017) also emphasize the relationship with teaching and learning. Nevertheless, the data collected regarding both schools suggests that the teachers carried out essay writing assessments without conducting teaching and learning.

The theory of assessment *for* learning (Assessment Reform Group, 1999, 2002; Black & Wiliam, 1998) emphasizes that assessment *for* learning should focus on "how to improve students' learning." This involves the process of verifying students' capabilities and levels of learning so as to determine appropriate measures for their improvement. Teachers thus have to carefully observe students during the "learning and teaching *process*." In this respect, there are many different aspects of this process (i.e., tasks and questions that can demonstrate students' knowledge, understanding, and skills) that can be regarded as assessment (Assessment Reform Group, 2002).

In this sense, in Schools X and Y, where there is no teaching and learning involved in essay assessment, how could the teachers and students collect data to improve their learning and teaching? Just providing tests or conducting assessment separate from teaching could hardly be expected to increase students' learning. Even in the case where a writing activity with the same theme as the essay

assessment was present in the textbook, the teachers did not conduct this activity in class. In other words, the students did not learn how to write in class nor were they provided opportunities to practice writing. In this context, where students are not taught writing in class but somehow required to undergo assessment for writing, students likely end up learning writing through a private institute. Furthermore, since the writing themes were given a week in advance, this allowed for some students to seek external assistance in completing their essays, precluding any objectivity or fairness in the assessment.

In addition, as was evident in the case of School Y, when selecting the essay writing genre and number of assessments to be conducted, the teachers seemed more deeply concerned about filling out students' educational profiles with diverse information than with teaching and learning, although it is obvious that the theme and genre for essay writing should be determined in relation to the curriculum.

The data also shows how the students were not provided with constructive feedback that might lead to their improvement in essay writing. When asked if they gave feedback to the students regarding the essay writing, School Y teachers responded that they usually did not and School X teachers that they returned the score recorded on the writing assessment back with only grammatical mistakes highlighted. Regarding why they do not consider giving more concrete feedback to the students, Teacher B answered as follows:

Teacher B: In conducting the performance assessments, the scoring and grading kept me busiest.

Assessment *for* learning gives significant priority to *feedback*. Teachers should not provide only marks or remarks informing students of their success or failure; they should rather inform students "how to progress." This means dealing with the particular qualities and strengths and weaknesses of a students' work and how to improve or address them (Assessment Reform Group, 1999, 2002; Black & Wiliam, 2002). In spite of this fact, the teachers of Schools X and Y did not seem to significantly consider what kind of feedback to provide or how to provide opportunities for the students to develop their work.

The rubrics for the essay writing of Schools X and Y were as follow. As displayed in Table 4.5, the essay writing of School X consisted of the criteria *task completion*, *content and construction*, and *language use*.

TABLE 4.5

Rubric for the Essay Writing of School X

Category	Scoring Criteria	Score
Task Completion	The student fulfilled requirements 1), 2), 3), 6), and 7)[8].	5
	The student fulfilled most of the requirements (four among 1], 2], 3], 6] and 7])	4
	The students fulfilled some of the requirements (three among 1], 2], 3], 6] and 7])	3
	The students fulfilled few of the requirements (two among 1], 2], 3], 6] and 7])	2

---

[8] The seven requirements indicate the "essay requirements" described in the "Guidelines for Essay Writing: School X" (Figure 4.2).

	The students fulfilled almost no requirements (one among 1], 2], 3], 6] and 7])	1
	The student failed to fulfil the requirements 1), 2), 3), 6), and 7).	0
Content and Construction	The student fulfilled requirement 4) (three pieces of factual information).	4
	The student mostly fulfilled requirement 4) (two pieces of factual information).	3
	The student partially fulfilled requirement 4) (one piece of factual information).	2
	The student failed to fulfil requirement 4) (no factual information).	1
Language Use	The student fulfilled requirement 5), making few grammar errors and mostly using vocabulary appropriately.	4
	The student partially fulfilled requirement 5), making some grammar errors and sometimes using vocabulary inappropriately but still effectively communicating the main idea.	3
	The student poorly fulfilled requirement 5), making many grammar errors and often using vocabulary inappropriately, somewhat ineffectively communicating the main idea.	2
	The student failed to fulfill requirement 5), repeatedly making grammar errors and using vocabulary inappropriately, ineffectively communicating the main idea.	1

The criteria of School Y, similar to those of School X, consisted of *task completion*, *content*, and *grammar, spelling, vocabulary, and expressions*. In particular, it is noticeable that the criterion *language use* of School X was almost equivalent to the criterion *grammar, spelling, vocabulary, and expressions* of School Y.

TABLE 4.6

Rubric for Essay Writing I and II of School Y[9]

Category	Scoring Criteria	Score
Task Completion	All requirements met, complete and grammatically correct sentences.	4
	All requirements somewhat met, some incomplete and grammatically incorrect sentences.	3

---

[9] This rubric is for Essay Writing I (Figure 4.3) and Essay Writing II.

	Failed to meet all requirements, most sentences are incomplete and grammatically incorrect.	2
Content and Construction	Communicated the relevant idea in a coherent manner	3
	Communicated the relevant idea in a somewhat coherent manner	2
	Failed to communicate the relevant idea	1
Grammar, Spelling, Vocabulary, and Expressions	Few grammatical or spelling errors, appropriate use of vocabulary and expressions	3
	Some grammatical and spelling errors, inappropriate use of vocabulary and expressions not significantly disrupting communication of the main idea	2
	Many grammatical and spelling errors, inappropriate use of vocabulary and expressions significantly disrupting communication of the main idea	1

As for the reason for including the criteria of language use (School X) and grammar, spelling, vocabulary, and expressions (School Y), the teachers responded that these criteria helped them to distinguish between the students. They explained that high-proficiency students conformed to the grammar and vocabulary criterion—especially grammar—helping the teachers to distinguish students' proficiency levels.

Teacher E: Without "grammar" in the evaluation criteria, we could hardly rank the students. We should rank students precisely. The students should be ranked into nine levels.

<div align="right">

(Individual interview with Teacher E)

</div>

Teacher C: There could be disagreement regarding how harshly we should score grammatically wrong sentences. However, teachers would certainly agree that "grammar" is the most critical criterion to distinguish the

students' proficiency levels.

<div align="right">*(Group interview at School X)*</div>

The inclusion of the criterion *grammar and vocabulary* for essay assessment itself is not problematic. However, teachers' comments show that rather than considering the validity or usefulness of the criterion of *grammar* in light of teaching and learning, they seemed to prioritize *differentiating students*. With regard to this dynamic, Davison (2004) argues that when teachers are preoccupied with accuracy, they may act as *technicians*, taking a *mechanistic criterion-based approach*, rather than facilitating students' learning.

In addition, as demonstrated in the rubrics of both Schools X and Y, the criterion of *task completion* in the essay writing assessments accounted for a larger portion of a student's grade than *content and construction*. When asked the reason for this decision, the teachers answered that whereas the task completion criterion made it possible for the teachers to score the writing objectively, that is, according to specific predetermined conditions (e.g., write an essay at least 500 words in length), the content criterion required teachers' individual subjective judgment.

Teacher A: When it comes to the criterion of "task completion," we could score the essays by drawing on the specific conditions provided. It didn't require each teacher's subjective thinking.

<div align="right">*(Individual interview with teacher A)*</div>

Teacher C: Regarding the criterion of "content," it isn't easy for me to justify to students why I gave a low or high grade. This criterion is pretty subjective.

(Group interview at School X)

This shows that the teachers of Schools X and Y prioritized objectivity and fairness in pursuing grading performance. However, giving less weight to or excluding criteria entailing subjective judgment and giving more weight to criteria ensuring objectivity could lead teachers to regard their role as that of a *rater* rather than *facilitator*. Leung and Mohan (2007) argue that, in conducting formative assessment, the focus on grading objectively leads teachers to perceive their role as that of an assessor or judge rather than a supporter of language development or monitor of students' progress providing feedback.

Meanwhile, it seems contradictory that the teachers focused on objectivity and fairness in constructing rubrics for assessing writing. With respect to the essay writing, the teachers of both schools did not teach writing, merely informing the students of the writing guidelines and briefly explaining them. They also permitted students to seek out external help in completing their essay writing. Their assessment objectives reflected the aims of objectivity and fairness. Nonetheless, the teachers did not seem see this as problematic, only considering objectiveness and fairness in constructing the rubrics.

In sum, the teachers of Schools X and Y chose the essay writing assessment themes in relation to themes in the reading textbook. In particular, School Y

intended to exclude writing genres students had already took in order to record diverse information in their educational profiles. Both Schools X and Y selected *task completion, content and construction and language use* (School X) and *grammar, spelling, vocabulary, and expressions* (School Y) criteria. Both school teachers decided to assign a greater weight to the task completion criterion precluding subjective judgment and perceived the grammar and vocabulary criterion as a way to grade students in an accurate way to differentiate between them. Although the teachers did not teach the students writing or provide feedback for their writing work, they did not seem to feel this was problematic.

4.3. Speaking Assessment

With regard to speaking assessment, the School X students were also expected to give an oral presentation on an admirable person of their choice using the essays as scripts. When asked the reason for selecting the method "give a speech using the essay as a script," the teachers answered, as mentioned above (see explanation regarding constructing the performance assessment composition and grade weighting), that the Gyeonggido Office of Education (2019) permits the use of an oral presentation as essay writing assessment if it maintains the original characteristics of the essay assessment, "expressing opinions or arguments in a logical way" and accounting for at least 35% of a semester grade. In this regard, they explained that if they added an oral presentation worth 15% to the essay writing assessment already worth 20%, not only could they meet the grade weight

requirements for essay writing assessment but also implement speaking into the essay writing assessment.

After the students received their scored essays with grammatical mistakes highlighted by the teacher about one week after conducting the essay writing, they made corrections with the help of the teachers (if they asked), books, internet sources, etc. The students were expected to enhance the completeness of the essay by correcting all errors and meeting the requirements of the criteria of task completion, content and construction, and language use. The students then memorized the corrected script. The oral presentation was conducted in the hallway with only the teacher and presenting student present. The student was not allowed to refer to the script. Figure 4.4 below provides an example of the guidelines of the oral presentation of School X.

FIGURE 4.4

Guidelines for the Oral Presentation: School X[10]

Guidelines for the Oral Presentation

1. Assessment administration process

1) The order for the oral presentations shall be random.
2) The oral presentations shall be conducted in the hallway with only the teacher and student present.
3) Each oral presentation shall be completed within three minutes.
4) Students are not allowed to refer to their scripts during the presentation.
5) Each student shall submit his or her script to the teacher right before the presentation.

2. Evaluation criteria

1) Script completion: Teachers shall evaluate how well students improve their scripts based on the teachers' feedback.
2) Task completion: Teachers shall verify whether or not the presentation is completed within three minutes.
3) Pronunciation, intonation, and fluency: Teachers shall evaluate the degree to which student's pronunciation, intonation, and fluency are natural.

In School Y, students implemented a *dialogue recitation assessment* in which they chose one among three dialogues to memorize and recite in front of the teacher within 90 seconds. As shown in Figure 4.5, on average, each dialogue was composed of seven or eight lines (three or four lines per speaker), which the students had learned in the text. When reciting the dialogue, they were required to perform all the lines.

---

[10] The author translated "Guidelines for the Oral Presentation: School X" from the original Korean.

FIGURE 4.5

Three Dialogue Sets for the Dialogue Recitation Assessment of School Y

Set A	B: Look at Tina. She has a lot of friends. G: Yeah, people like her because she's very sociable and outgoing. B: I want to be as popular as she is, but I'm too quiet. I think I should be more like her. G: Well, you have your own merits. You always listen carefully to others when they are talking. B: Really? But I don't think that's such a great quality. G: Sure it is. Everyone needs a friend who is a good listener. B: Thank you for saying that. G: Try to always remember that it is important to be happy with who you are.
Set B	G: Hey, Danny. Have you heard about the big competition that's coming up? B: No. What kind of competition is it? G: It's for designing robots that can help in disasters. B: Oh… Do you mean the ones that enter dangerous areas people can't easily reach? G: Yes. Some robots can even fly over a disaster area and find people in need. B: That sounds cool! I think they're going to make rescue work much faster and safer. G: Definitely. I can't wait to see what ideas people come up with in the competition!
Set C	G: Look at these two people shaking hands. They look so happy to see each other. B: They do. But did you know that people didn't shake hands to be friendly in the past? G: Really? Then why did they do it? B: They would greet each other like this for safety reasons. G: What do you mean by that? How would shaking hands keep you safe? B: In the past, people shook hands with each other to see if the other person was hiding a weapon. G: Wow, that's so interesting. I had no idea how this custom started.

Figure 4.6 below shows the guidelines for the dialogue recitation, which was the speaking assessment of School Y. One week before administration of the assessment, the method, time limit, and criteria of the assessment were explained by the teachers and this information was posted in the classroom on a sheet of paper.

FIGURE 4.6

Guidelines for the Dialogue Recitation: School Y

Guidelines for Dialogue Recitation

● The students will randomly choose and recite one among the three dialogues.

● The assessment will take place in the hallway involving only the student and the teacher.

● The recitation should be completed within 90 seconds.

● Evaluation criteria: task completion, voice (audibility), pronunciation, and fluency

| Assessment Criteria | • Task Completion |
| | • Voice (audibility), Pronunciation, and Fluency |

With respect to the reason for selecting a dialogue recitation for the speaking assessment, teachers D and E explained that this method was effective because it could provide good opportunities for the students to practice speaking. They did not need to focus on making the speaking assessment authentic or reflect real-world situations.

Teacher E: I think this method is effective. I know that a meaningful speaking

assessment might not aim at memorizing whole dialogues like this.

However, this assessment method encourages students to read and

speak out loud. The method is satisfactory because it meets this

purpose.

<div align="right">

(Individual interview with Teacher E)

</div>

Teacher D: Although this method is unrelated to authentic communication in the

real world, I think this is okay. It encourages students to speak and

pronounce the words out loud. That's the point.

<div align="right">

(Individual interview with Teacher D)

</div>

The School X teachers also stated that speaking performance assessment, in particular, offered a valuable opportunity to students.

Teacher B: Students do not have chances to practice speaking. Usually, students

just read English passages silently. They do not speak out. However, in

speaking assessment, they have no option but to speak out.

<div align="right">

(Group interview at School X)

</div>

Teacher C: Due to the fact that students have hardly practiced speaking, I think

that speaking assessment can contribute to the improvement of their

*speaking ability more than any of the other parts of performance
assessment.*

<div align="right">

(Group interview at School X)

</div>

The School Y teachers expressed similar opinions. The following excerpts
show that the speaking assessment was significantly related to the School Y
students reading out loud in regular reading classes.

*Teacher E: If we do not assess speaking, the students only read the dialogues in
the texts silently. Even if I ask them to repeat after me, they don't.*

<div align="right">

(Individual interview with Teacher E)

</div>

*Teacher D: Students don't read out loud. The lower their proficiency, the less they
read out loud. Thus, conducting speaking performance assessments to
make students speak or read out loud is itself meaningful.*

<div align="right">

(Individual interview with Teacher D)

</div>

The School Y teachers explained that most of the English classes are
designed to teach reading, hardly involving writing or speaking. In this respect,
the teachers seemed to believe that since teaching speaking does not take place in
classrooms, speaking performance assessments could grant students opportunities
to practice speaking.

Teacher D: We do not teach speaking and writing. I feel sorry for that…. So, through carrying out speaking and writing in performance assessments, students can practice English language both orally and in writing. Actually, speaking, writing, and listening are all language skills we should teach in addition to reading. Even though teaching in the classes is preoccupied with reading, I think providing students an opportunity to practice [speaking] can help them to improve their language ability.

(Group interview at School Y)

Teacher E: If we do not assess, the students do not study. Since we teach reading in regular English classes and regular examinations are composed of reading and grammar questions, at the very least, performance assessments are a necessary measure. If we do not provide an opportunity to engage in speaking and writing in performance assessments, they would have no chance otherwise.

(Individual interview with Teacher E)

It was interesting to find that the teachers treated performance assessments as an opportunity to allow the students to "practice" skills they were not taught. In theory, assessment follows learning. According to Gipps (1994), classroom-based teacher assessment could be considered more valid than external forms of assessment in that it is directly related to content learned in class. Nonetheless, the

89

teachers did not appear to see anything problematic about implementing assessment without teaching the subject matter; they simply considered the benefits of providing students an opportunity to practice speaking.

In assessment *for* learning theory (Black & Wiliam, 1998), the ultimate goal of assessment is "to improve students' learning." To achieve this aim, students should actively involve their own learning and teachers should examine students' performance carefully and elicit evidence to cater to students' needs. Such evidence allows teachers not only to provide constructive feedback to students but also *adjust their teaching*. This is very important because this is how teachers reflect on their own teaching and seek ways to promote students' learning. This is why assessment *for* learning emphasizes *adjustment of teaching* as well as *students' active involvement* throughout assessment. Considering this fact, one could hardly expect enhanced teaching through assessment ultimately contributing to improving learning in Schools X and Y, where learning and teaching for speaking were not practiced.

The assessment methods of both Schools X and Y are also noteworthy. Considering the definition of performance assessment, dialogue recitation (School Y) and speech memorization (School X) do not seem appropriate as performance assessment methods.

Previous studies regarding performance assessment state that methods should involve productive skills and be as authentic as possible, performed in a real-world situation (e.g., J. D. Brown & Hudson, 1998; Shohamy, 1995). This principle is also expressed in KICE (2017, p. 16) and MOE and KICE (2017, p.

8). Authentic tasks allow teachers to examine the knowledge students can use in real-world situations, which permits teachers to assess true functional communicative ability. Moreover, according to Wiliam (2001), authentic tasks are not just for assessment but also learning. In other words, in the course of performing authentic tasks, students' learn. However, both schools employed a memorized dialogue or speech as the speaking assessment method. Teacher D's comment, "Although this method is unrelated to authentic communication in the real world," implied a deeper knowledge of how speaking assessment methods should be. Yet it appears that her knowledge did not translate into practice, as teachers prioritized providing opportunities to practice speaking by making the students memorize the dialogue and scripts.

Moreover, the assessment administration method at the two schools was also striking. First, there was the method of time limitation. School X's oral presentation was required to be completed within three minutes, while School Y's dialogue recitation was required to be completed within 90 seconds. Teacher E explained that they intended to provide students an opportunity to practice speaking by having students speak out loud. In this regard, they believed the time limit encouraged the students to practice more.

Teacher E: There should be a time limit. Providing a time limit makes students practice more.

(Individual interview with Teacher E)

In addition, the speaking assessments were administered in the hallway and not in front of the students in the classroom. The speaking assessments were thought not to require a student audience since the purpose was simply to practice speaking the dialogues.

Teacher E: This is not a real speech. We don't need to carry out the speaking assessment in front of all the students in the classroom.

(Individual interview with Teacher E)

The speaking assessment administration method of both schools, "conducted in the hallway with only the teacher and presenting student and with a time limit," does not seem appropriate in terms of the characteristics of assessment *for* learning as presented by the Assessment Reform Group (1999). First, as mentioned above, in theory, examinees are expected to perform an authentic task in a real-world situation. In the case of Schools X and Y, carrying out the task in the hallway with only the teacher as the audience is certainly far from a real-world situation.

Second, this administration method prohibited the students from sharing their performances with one another. Indicating the importance of peer assessment, the Assessment Reform Group (1999) claims that students can improve their learning by examining and discussing their performances, engaging in the kind of reflection that is the basis of learning. In addition, MOE and KICE (2017) state that in process-centered performance assessment, students can be assessors by

taking part in self- and peer assessment, which can ultimately improve their learning and increase assessment validity. If assessments are carried out isolated from a peer audience, then students are deprived of the opportunity to examine and assess others' performances (peer assessment). This also means missed opportunities for feedback from peers. In this respect, the speaking administration method of the two schools did not seem to support assessment *for* learning.

In fact, when the teachers asked how they provide feedback for the speech assessment, School Y teachers answered that they did not give feedback. Regarding the dialogue recitation assessment, Teacher D answered as follows.

Teacher D: I do not provide sufficient feedback for the performance. In that case, when the students' pronunciation is outstanding, I provide short comments, like, "You have very natural and good pronunciation." But when there are no notable features in the recitation, I just finish the test and record the score.

(Individual interview with Teacher D)

Teacher D seemed to provide feedback only to outstanding students. However, in assessment *for* learning, constructive feedback is most useful to low achievers and is designed to minimize achievement disparities, raising the achievement level of the class as whole (Black & Wiliam, 1998). In this respect, it is essential to provide feedback to all students, and especially poorer students.

Teacher C explained that he did not usually provide oral feedback but noted

some features of the students' speech like pronunciation and intonation for the purpose of recording it in the *Specific Competencies and Specialties in English Learning* section in students' educational profiles.

Teacher C: I think I always take into consideration the need to write something in the "Specific Competencies and Specialties in English Learning" section of the educational profiles. For example, I write down some characteristics of the students while they perform their tasks in oral assessments such as pronunciation. I then later record them in the "Specific Competencies and Specialties in English Learning" section.

(Individual interview with Teacher C)

However, when asked about providing students with feedback, Teacher C answered that students could read his comments in their educational profiles at the end of the year or the year after.

Teacher C: I do not give feedback right away to the student. I just record the comments in the educational profiles. Maybe they can see the comments at the end of the year or next year when all of the records for students' educational files are completed.

(Individual interview with Teacher C)

This feedback method would unlikely be helpful to the students. Assessment

for learning contends that teachers' comments can increase students' confidence and enthusiasm and contribute to learning by letting students know what to aim for and providing information for the skills and strategies they can employ to progress (Assessment Reform Group, 1999). In this respect, Teacher C was not well aware of how important feedback is in promoting students' learning.

With regard to the speaking evaluation criteria of Schools X and Y, the most noticeable thing was that, in both schools, the criterion of task completion was more weighted than those of pronunciation and intonation in School X (Table 4.7) and voice, pronunciation, and fluency in School Y (Table 4.8).

Table 4.7 shows that in the case of School X, while script completion and task completion were weighted with maximum scores of six and seven, respectively, pronunciation and intonation were weighted with a maximum score of two, a strikingly low number compared to those of script completion and task completion. In addition, by using a binary grading system, assigning either a score of zero or two, the teachers precluded the use of pronunciation and intonation to differentiate the students.

TABLE 4.7

Rubric for the Oral Presentation of School X

Category	Scoring Criteria	Score
Script Completion	Fulfillment regarding word total, content, and use of grammar and vocabulary.	6
	Partial fulfillment regarding word total, content, and use of grammar and vocabulary, but effective communication of the main idea.	4

	Poor fulfillment regarding word total, content, and use of grammar and vocabulary, somewhat disrupting communication of the main idea.	2
	Non-fulfillment regarding word total, content, and use of grammar and vocabulary, significantly disrupting communication of the main idea (or failure to submit a script).	0
Task Completion	Completed the entire script	7
	Completed most of the script (80% and above)	5
	Completed some of the script (50 to 80%)	3
	Completed a portion of the script (30 to 50%)	1
Pronunciation and Intonation	Clear pronunciation and intonation, no disruption of communication	2
	Unclear pronunciation and intonation, some disruption of communication	0

School Y was no different. The rubric for the dialogue recitation assessment of School Y shows that task completion was more weighted than voice, pronunciation, and fluency (Table 4.8). Moreover, while task completion was graded in terms of one-point intervals, voice, pronunciation, and fluency were graded in terms of 0.5-point intervals, precluding the differentiation of students through this criterion.

TABLE 4.8

Rubric for the Dialogue Recitation Assessment of School Y

Category	Scoring Criteria	Score
Task Completion	Completed most of the dialogue (more than 80%)	3
	Completed some of the dialogue (50 to 80%)	2
	Completed less than half the dialogue (less than 50%)	1

Voice (audibility), Pronunciation, and Fluency	Voice is audible, pronunciation clear with few pauses, communication not disrupted.	2
	Voice is generally audible, pronunciation generally clear with some pauses but communication not disrupted.	1.5
	Voice is inaudible, pronunciation poor with frequent pauses, communication disrupted.	1

The teachers explained that this was due to the preference for objectivity.

Teacher B: In speaking assessment, criteria such as "pronunciation and intonation" could be subjective. To rule out subjectivity as much as possible, we focus more on "task completion," I mean, whether the students finish within the time limit.

(Individual interview with Teacher B)

Teacher D: One of the ways to ensure objectivity is to underline the criterion of "task completion," assigning a higher weight to it than other criteria. I think that other criteria in speaking assessment such as "pronunciation" are judged differently according to the teacher.

(Individual interview with Teacher D)

According to the explanations of the teachers of Schools X and Y, they gave more weight to the task completion criterion in both the writing and speaking assessments, which was designed to eliminate any subjective judgment. However, for students to improve essay writing ability, although task completion is important, content and construction, or communicating the relevant idea in a coherent manner,

97

are also very important. Likewise, for students to improve speaking ability, in addition to task completion, clear and understandable pronunciation and intonation and fluency are significant. By assigning a greater weight to task completion to ensure objectivity, teachers may underestimate the importance of other criteria and pay less attention to those aspects of students' performance and ability. In other words, the preoccupation with task completion prevents teachers from comprehensively examining the aspects of performance related to the evaluation criteria, eventually hindering the collection of data to improve teaching and learning. Meanwhile, students themselves are likely to end up undervaluing the other criteria.

Another remarkable observation pertains to the rubric for the oral presentation of School X (Table 4.7). In the rubric, the *task completion* criterion was weighted by two-point intervals up to a maximum score of six. It is evident that, compared with the maximum grade weights given to other criteria, this weight was particularly heavy. The question is whether this criterion was appropriate for the oral presentation. When the scoring criteria for script completion were examined, it was found that they were identical to the criteria for essay writing, including task completion (i.e., word total), content, and use of grammar and vocabulary. This means that the teachers used the same criteria for both the speaking and writing assessments. Therefore, the students who achieved higher grades in the essay would naturally have also achieved higher grades in the speaking assessment. This finding could also be interpreted as indicating that the teachers did not significantly consider constructing a rubric completely

appropriate to speaking assessment.

In addition, even though the opportunity was provided to the students to correct the script, just as with the insufficient teaching of writing, the teachers' feedback was also insufficiently provided. Only when the students asked a specific question would teachers provide feedback. In a situation where the oral presentation was conducted only two weeks after conducting the essay writing (see Table 4.3), there was little guarantee that students would construct scripts much improved upon their essays. Once again, this demonstrates the difficulty of viewing a speaking assessment including a script completion criterion and maximum score of six as in fact a speaking assessment. Nevertheless, the teachers did not seem to see this as problematic.

Lastly, and most importantly, with respect to the evaluation criteria of the essay writing and speaking assessments, what deserves the most attention is the fact that the teachers did not adequately explain the criteria to the students. The teachers of both Schools X and Y merely let the students know the criteria before conducting the assessment by posting guidelines for the essay writing and oral presentation that included only the criteria categories at the bottom of the page without any description. In the case of School X, even though the guidelines for the essay writing included essay writing requirements, these were only explanations confined to *task completion* and *content and construction*.

According to Sadler (1989), students' understanding of their goals is essential to assessment with a formative purpose. Since learning goals can be ambiguous, however, implicit knowledge is important in this process (Claxton,

1995). Here, *criteria* can help students understand learning goals and what is required of them. Descriptions of criteria should thus be carefully worded and accompanied by examples of how the criteria will apply in practice. Teachers should also explain the criteria by discussing them with students and having them engage in peer and self-assessment. Through this process, students become responsible for their learning and no longer just passive recipients (Wiliam, 2010). Considering these points, it seems that the teachers of Schools X and Y need to more deeply consider ensuring that the students understand the evaluation criteria and, ultimately, learning goals.

In sum, regarding the speaking assessments, the School X teachers chose an oral presentation employing the students' previously written essays as scripts and the School Y teachers chose a memorized dialogue recitation. The teachers at both schools implemented the conditions of a time limit and a one-to-one (teacher-student) situation and assigned a greater weight to the task completion criterion, emphasizing objectivity. None of the teachers taught speaking, implemented authentic assessment methods reflecting a real-word situation, provided an opportunity for students to assess one another's performances, or shared the speaking assessment criteria, which might have helped the students understand their learning goals.

4.4. Listening Assessment

With regard to the listening assessment, the School X teachers adopted a

multiple-choice test. This method did not seem appropriate for performance assessment in theory. According to the teachers' responses, they were well aware of this. The Gyeonggido Office of Education (2019) also prohibits the use of one-off, paper-pencil assessments, arguing that this negates the original objective of performance assessment. Nonetheless, it states that the EBS multiple-choice listening test can be regarded as performance assessment (Gyeonggido Office of Education, 2019, p. 11). The School X teachers explained that because the policy permitted the use of multiple-choice questions as listening assessment, they did not see their use as problematic. They assessed students' listening ability with a listening test composed of 20 multiple-choice questions broadcast by EBS (Educational Broadcasting System in Korea). This test is broadcast via radio once per semester on a designated date and jointly administered by the sixteen regional offices of education. The difficulty is aimed at 11th-grade Korean students based on the contents of the 2015 revised national curriculum. The types and forms of assessment questions and answers are similar to those of the KSAT listening test. Figure 4.7 shows the actual EBS listening test implemented in School X.

FIGURE 4.7

EBS Listening Test Used for Listening Performance Assessment in School X

제2회 전국 16개 시·도교육청 공동 주관 영어듣기능력평가(고2)

2019. 9. 25. (수) 시행　　　제2학년 ＿＿＿반 ＿＿＿번 이름 ＿＿＿＿＿＿＿＿＿＿

1. 대화를 듣고, 남자의 마지막 말에 대한 여자의 응답으로 가장 적절한 것을 고르시오.
① Okay. Here's your public library card.
② No problem. We don't need your photo.
③ I'm sorry. I'll bring my photo right now.
④ Sure. The photo magazine section is over there.
⑤ Don't worry. We can take a picture of you here.

2. 대화를 듣고, 여자의 마지막 말에 대한 남자의 응답으로 가장 적절한 것을 고르시오.
① I sincerely wish you good luck.
② Thanks for taking your time with me.
③ Would you give me tips for the exam?
④ I also hope to have a good time with you.
⑤ How about buying a laptop instead of a desktop?

3. 대화를 듣고, 남자의 마지막 말에 대한 여자의 응답으로 가장 적절한 것을 고르시오.
① Please. That's really kind of you.
② Really? I think the office is too hot.
③ Thanks. I'll take medicine as you advised.
④ Right. We need a heater in the office right now.
⑤ Sure. I can wear my coat while the heater is off.

4. 다음을 듣고, 여자가 하는 말의 주제로 가장 적절한 것을 고르시오.
① 합성 섬유와 천연 섬유의 차이점 ② 효율적인 면직물 관리 요령
③ 다양한 면직물 가공 방법 ④ 한국 직물 산업의 역사
⑤ 면직물의 장점과 단점

5. 다음을 듣고, 남자가 하는 말의 목적으로 가장 적절한 것을 고르시오.
① 인공 눈물 신제품을 홍보하려고
② 콘택트렌즈 관리법을 안내하려고
③ 미세먼지의 위험성을 경고하려고
④ 유용한 건강 관리 앱을 소개하려고
⑤ 안구 건조 완화 요령을 알려주려고

6. 대화를 듣고, 여자의 주장으로 가장 적절한 것을 고르시오.
① 생명 존중 캠페인을 활성화시켜야 한다.
② 학교 시설 안전 점검을 정기적으로 시행해야 한다.
③ 학생들에게 심폐소생술(CPR) 실습 교육을 실시해야 한다.
④ 심폐소생술(CPR) 관련 영상 보급을 확대해야 한다.
⑤ 응급 구조 모범 사례를 발굴해야 한다.

7. 대화를 듣고, 남자의 의견으로 가장 적절한 것을 고르시오.
① 여행의 목적에 맞는 계획을 세울 수 있어야 한다.
② 정기적인 여행은 스트레스 해소에 도움이 된다.
③ 여행을 갈 때 모든 계획을 세울 필요는 없다.
④ 좋은 여행 프로그램은 의미 있는 순간을 제공해 준다.
⑤ 즐거운 여행을 하려면 마음이 맞는 사람과 함께 해야 한다.

8. 대화를 듣고, 두 사람의 관계를 가장 잘 나타낸 것을 고르시오.
① 부동산 중개인 - 고객 ② 요리사 - 촬영 감독
③ 호텔 직원 - 투숙객 ④ 사진작가 - 건축가
⑤ 교사 - 학부모

9. 다음 그림의 상황에 가장 적절한 대화를 고르시오.

① ② ③ ④ ⑤

10. 대화를 듣고, 남자가 여자에게 부탁한 일로 가장 적절한 것을 고르시오.
① 볼펜 찾아주기 ② 고민 상담해 주기
③ 상표 이름 알려주기 ④ 네임 스티커 주문해 주기
⑤ 온라인 쇼핑몰 주소 보내주기

11. 대화를 듣고, 남자가 여행 가방을 새로 구입하려는 이유로 가장 적절한 것을 고르시오.
① 가방 지퍼가 고장 나서 ② 가방 크기가 작아서
③ 가방이 구형이어서 ④ 가방을 분실해서
⑤ 가방이 더러워서

12. 대화를 듣고, 여자가 지불할 총 금액을 고르시오.
① $140 ② $170 ③ $230 ④ $260 ⑤ $300

13. 대화를 듣고, New Year's Concert에 관해 언급되지 않은 것을 고르시오.
① 주최 기관 ② 공연 날짜 ③ 초청 연주자
④ 입장료 ⑤ 예매 방법

14. Club Booth Exhibition에 관한 다음 내용을 듣고, 일치하지 않는 것을 고르시오.
① 운영 시간은 오후 2시부터 오후 5시까지이다.
② 동아리별 예산은 100달러이다.
③ 주제는 자유롭게 선택 가능하다.
④ 촬영된 영상은 학교 웹사이트에 게시된다.
⑤ 가장 인기 있는 동아리는 상을 받는다.

15. 다음 표를 보면서 대화를 듣고, 두 사람이 신청할 우유 배달 서비스를 고르시오.

Morning Milk Delivery Service

	Times/Week	Quantity	Flavor	Price/Month
①	6 times	500 ml	strawberry	$30
②	4 times	200 ml	plain	$15
③	4 times	1,000 ml	plain	$40
④	3 times	200 ml	strawberry	$10
⑤	3 times	500 ml	plain	$25

16. 대화를 듣고, 여자의 마지막 말에 대한 남자의 응답으로 가장 적절한 것을 고르시오.
① How about selling your vase next time?
② I'll try to arrive there as early as possible.
③ I don't usually buy things at the flea market.
④ I'll go to the flea market on Sunday afternoon.
⑤ Thank you for helping me start my own shop.

17. 대화를 듣고, 남자의 마지막 말에 대한 여자의 응답으로 가장 적절한 것을 고르시오.
① No problem. I'll make some food for you.
② No thanks. I don't drink coffee in the morning.
③ Sure. I'll follow your advice and see what happens.
④ Good. Thanks for recommending such a great doctor.
⑤ All right. I'll try your recipe for spicy pasta this time.

18. 다음 상황 설명을 듣고, Rachel이 Michael에게 할 말로 가장 적절한 것을 고르시오.
① I'd like to extend the time limit now.
② It's hard to choose a topic for debate.
③ Do you want to be the chairman of this debate?
④ You'd better not exceed the time limit on your turn.
⑤ I think we need more time to prepare for this debate.

【19-20】 다음을 듣고, 물음에 답하시오.

19. 여자가 하는 말의 주제로 가장 적절한 것을 고르시오.
① 건강에 좋은 식품의 종류와 효능
② 식품을 신선하게 보관하는 방법
③ 식품의 특성을 고려한 조리법
④ 균형 잡힌 식단의 중요성
⑤ 경제적인 식품 구매 요령

20. 언급된 식품이 아닌 것을 고르시오.
① meat ② cabbage ③ potato ④ tomato ⑤ mushroom

ⓒ 본 평가 문항의 저작권은 전국 16개 시·도교육청에 있습니다.

102

Listening assessment was administered in the same way in School Y, employing the EBS multiple-choice listening test. Here, the remarkable thing was that the School X teachers viewed the EBS multiple-choice question test for listening assessment as a good tool to assign grades and precisely differentiate between the students.

Teacher A: I know that multiple choice is not an appropriate task type for performance assessment. However, since multiple choice involves exactly correct answers, unlike with other performance-assessment methods such as essay writing or oral presentation, it makes it possible to distinguish precisely between students according to the number of correct answers obtained. We give scores for listening assessment by multiplying the number of correct answers by 0.3. This assessment allows us to accurately differentiate between the students. In order to rank the students' proficiency, we need this kind of method.

(Individual interview with Teacher A)

This excerpt demonstrates that the teachers intended to differentiate students through a multiple-choice question test. Thus, their use of this form of test could hardly be regarded as for assessment *for* learning.

In addition, feedback for the listening assessment was not provided. According to the teachers' responses, they did not provide any feedback whatsoever for the listening assessment, seeming more concerned with precisely

distinguishing between students through multiple-choice questions. In fact, if the focus is "how to improve students learning, " the ultimate aim of assessment *for* learning, in some respects performance assessment methods or types are arguably not all that important. The more important thing is providing constructive feedback, which can promote students' learning, after collecting data by examining students' assessment performance. In this respect, although the EBS multiple-choice question test was selected with a summative purpose (to distinguish students), this did not necessarily prevent teachers from giving constructive feedback. Such a course of action would align with the argument made in previous studies (Black et al., 2004; Wang, 2017) that summative assessment can also be used as assessment *for* learning by providing feedback after the test through question and answer or discussion. Carless (2011) also argues that since the test follow-up requires teachers to use test data to inform teaching and learning by analyzing students' test performance and designing remedial activities with the purpose of guiding students toward better learning, this follow-up strategy could be formative, namely, assessment *for* learning, albeit in a restricted sense. In this respect, for assessment *for* learning to take effect, what is chiefly important is providing feedback to students to improve their learning, as opposed to whether or not a performance assessment task is performance-based or not. In short, as previous studies argue, if teachers intend feedback to help students in their learning, summative assessment could also be sufficiently used for a formative purpose (Carless & Lam, 2014a; Kennedy et al., 2008; Xiao, 2017). Thus, classroom tasks should be justified in terms of the learning aims they serve; they support effective

learning if opportunities for students to communicate their evolving understanding are built into the planning.

Furthermore, as with the essay writing and speaking assessments, neither was listening ability taught in class. The School X teachers said they did not teach listening at all, while the School Y teachers explained that they let the students listen to dialogues and read them aloud, allocating one hour of class time per textbook chapter. In assessment *for* learning, assessment should not be a separate stage conduced at the end of class. Above all, for the assessment to be effective, there must be alignment between objectives, teaching and learning, and instruction (English, 1992; Harlen, 2007). In this sense, it would be difficult to evaluate the practices of Schools X and Y as effective, since assessment *for* learning was treated separately from learning.

The effort to design criteria for differentiating the students was also examined with respect to the rubric for the listening assessment. Scores were given according to the number of correct answers, as shown in Table 4.9.

TABLE 4.9

Evaluation Criteria and Grading System of the Listening Assessment: School X

Assessment Type	Evaluation Criteria	Grading System		
		Perfect Score (point)	Base Score (point)	Interval (point)
Listening Assessment	• Base score (4 points)			
	+ (The number of correct answers	10	4	N/A
	x 0.3 points)			

The criteria of School Y were basically similar (i.e., scores given according to the number of correct answers), although there were differences in the base score and coefficient in the scoring formulas, as shown in Table 4.10.

TABLE 4.10

Evaluation Criteria and Grading System of Listening Assessment: School Y

Assessment Type	Evaluation Criteria	Grading System		
		Perfect Score (point)	Base Score (point)	Interval (point)
Listening Assessment	• Base score (2 points) + (The number of correct answers x 0.4 points)	10	2	N/A

This evaluation criteria was also described as enabling the teachers to differentiate the students. However, what should be emphasized is that this criteria failed to provide information regarding the students' listening ability related to listening skills and goals such as *identifying the main idea* or *identifying a concrete idea*.

In sum, both Schools X and Y employed the EBS-multiple choice test for listening assessment with the intention to differentiate between students precisely. The evaluation criteria of both schools, rather than reflecting the goal of improving listening skills, revealed how students obtained scores according to the number of correct answers. Furthermore, as with the essay writing and speaking assessments, in both schools, the teachers did not actually teach listening or provide feedback after administering the test.

4.5. Class Participation Assessment

As for the last performance assessment of Schools X and Y, the *class participation assessment*, when asked why they selected this class participation assessment method, first, the School X teachers answered that they thought it would be helpful to encourage students' active class participation.

In School X, even though teachers were free to design activities for the class participation assessment, in general, they gave a stamp to students when they answered the teacher's questions regarding reading comprehension, vocabulary, and grammar or when they submitted homework in accordance with the teacher's instructions. The more stamps the students gained, the higher scores they obtained for the class participation assessment.

The School X teachers said that this encouraged students to actively take part in class, motivating them and even pushing them to be competitive.

Teacher B: When students answer the questions or work hard on group projects, I distribute up to three stamps. This is effective for making students participate in class. That is the purpose of making the "class participation assessment" one of the performance assessments.

(Individual interview with Teacher B)

Teacher A: There are some students who are not motivated in class. Observing that other students submit homework, make presentations, and receive

stamps from the teacher, the unmotivated students are encouraged or sometimes even become competitive.

(Individual interview with Teacher A)

Teacher B: Yes, I think it's [the class participation assessment] very effective to make students take part in class. I give questions related to reading comprehension in group activities. Only when all the group members complete the task do I give stamps. So, everyone should work hard. The students ask the other students of the group if they understand. The students know that if they don't work hard, this hurts the other students. I also assign vocabulary tests, group presentations, and individual presentations.

(Individual interview with Teacher B)

Second, the School Y teachers also said that the class participation assessment was aimed at encouraging students to actively participate and concentrate on the teacher's instructions in class.

Teacher E: For the "class participation assessment," students were required to complete the review comprehension questions at the end of class. If they could not answer the questions, the teacher would explain again. This encouraged the students to participate and actively engage in the class. I like the "class participation assessment." It's pretty useful for

involving students in class.

Teacher D: The "class participation assessment" is indispensable, considering that the students of this school have a low level of English proficiency. They cannot easily concentrate on the lesson. So, the "class participation assessment" plays an essential role in making them pay attention to the lesson.

Whereas in School X the class participation assessment allowed for teacher autonomy in selecting the activities for giving stamps, in School Y, the teachers uniformly implemented two activities for the class participation assessment, as shown in Figure 4.8. The first requirement was for students to complete a review assignment after the teacher finished instruction. At the end of class, the teacher handed out a review assignment. Students filled in the blanks related to vocabulary, content comprehension, and grammar, areas all learned in that class. The teacher gave some time for the students to individually complete the assignment before providing the answers. The students could thus fill in all the correct answers if they listened to and comprehended the teacher's explanations before handing in the assignment. The second requirement involved writing down English vocabulary words and memorizing them. After taking a vocabulary test, the students were asked to write down the words they had missed as a homework

assignment. They were thus allocated scores in terms of class participation

assessment according to their submission of these assignments.

FIGURE 4.8

A Sample for Class Participation Assessment: School Y

Check-up & Review   Student Number _____ Name _____

There was once a king **(A) [who / which]** wanted to know **(a)** <u>three things</u>: the right time to do everything, the most necessary people to pay attention to, and the most important thing to do. He thought that **(B) [know / knowing]** these things would ensure his success. He announced that he would reward anyone who could teach him these things. Many people traveled to his palace, but they all provided different answers to his questions. Regarding the first question, some said that **(b)** <u>the king should make a detailed schedule **so that** he could always know the perfect time to do everything</u>. Others claimed that he should take every situation into account and **(C) [wait / waited]** for the precise moment to act. Still others suggested that he should consult wise men.

<Vocabulary>

	English	Korean		English	Korean
1	pay attention to ~		7	make a schedule	
2		보증하다, 확실하게 하다	8	detailed	
3		발표/공표하다, 알리다	9		주장하다
4		보상(하다)	10	take A into account	
5		제공하다	11		정확한
6	regarding		12		제안하다

<Comprehension>

	Task	Answer
(a)	Why did the king want to know the three things?	
(b)	밑줄 친 (b) 해석하기 (*so that 주의)	

<Grammar>

	어법상 맞는 표현에 O표 하세요.	WHY??
(A)	**[who / which]**	
(B)	**[know / knowing]**	
(C)	**[wait / waited]**	

Lesson 5. What Matters Most   1st Period

The interview data shows that the teachers seemed satisfied with the

effectiveness of the class participation assessment, believing that the activities encouraged students' active class participation. Here, it is worth recalling that even though the teachers of Schools X and Y believed that rewarding students with grades was effective for encouraging and motivating them to work hard, many educational studies are skeptical about providing extrinsic rewards such as grades to improve motivation. For example, studies regarding the relationship between extrinsic rewards and intrinsic motivation in primary and second education settings (e.g., Deci, Koestner, & Ryan, 2001; Lepper, Greene, & Nisbett; 1973) argue that the expected rewards significantly undermine intrinsic motivation in the long term. Moreover, Black and Wiliam (1998) argue that emphasis on rewards or grades leads students to pursue better grades rather than improving their learning.

At Schools X and Y, the emphasis on grades as a motivating factor was apparent: "*It's pretty useful for involving students in class*"; "*Only when all the group members complete the tasks do I give stamps. So, everyone should work hard.*" However, it is precisely in this respect that the Assessment Reform Group (1999) and Black and Wiliam (1998) urge teachers to be cautious. They argue that when assessment is employed for classroom management, it can inhibit teachers' focus on students' learning needs. The Assessment Reform Group (2002) contends that rather than trying to manage the class with grades, teachers should help to motivate students. It argues that meaningful motivation can be preserved and enhanced by assessment methods that protect learners' autonomy, provide choice and constructive feedback, and create opportunities for self-direction.

For example, in the case of School Y, teachers seemed to believe that low-level students could concentrate on their studies by completing the review paper provided as class participation assessment. However, this raises the question of whether frequent testing can improve learning. In addition, this method does not require students' higher level thinking or problem-solving skills. In terms of assessment *for* learning, furthermore, teachers should rather prioritize identifying and improving upon students' weaknesses by providing meaningful feedback through teacher-student *interaction*. Nonetheless, when asked whether they gave feedback regarding the students' submissions, the School Y teachers answered that they did not. In this regard, the teachers should reconsider whether this assessment method is ultimately helpful for increasing students' intrinsic motivation and developing their learning.

The evaluation criteria of the class participation assessments of Schools X and Y reflect their intentions regarding class participation assessment. The evaluation criteria of School X was *the number of stamps corresponding to how many of the teacher's questions students answer and how much homework they submit*, as shown in Table 4.11.

TABLE 4.11

Evaluation Criteria and Grading System of Class Participation Assessment:

School X

Assessment Type	Evaluation Criteria	Grading System		
		Perfect Score (point)	Base Score (point)	Interval (point)
Class Participation	• Number of stamps corresponding to how many of the teacher's questions students answer and how much homework they submit	15	5	2

Similarly, the evaluation criteria of the class participation assessment of School Y was *the number of completed review papers and vocabulary exercise papers submitted*, as shown Table 4.12.

TABLE 4.12

Evaluation Criteria and Grading System of Class Participation Assessment:

School Y

Assessment Type	Evaluation Criteria	Grading System		
		Perfect Score (point)	Base Score (point)	Interval (point)
Class participation	• Number of submissions of completed review papers and vocabulary exercise papers	5	1	1

The evaluation criteria and grading systems for the class participation assessment of both Schools X and Y show that the grades were provided according

113

to the number of tasks students completed or assignments they submitted as directed by the teachers. The teachers of Schools X and Y explained that the criterion they emphasized for the assessment in the above plan was *effort* rather than *achievement*, as shown in the following excerpts.

Teacher D: In the "class participation assessment," we give scores to the students according to how many review comprehension questions they complete and submit and how much homework they submit. However, the real criterion is not the number of submissions. The fundamental criterion is "effort." Even if they couldn't complete the review comprehension material, it doesn't matter. This is because I let them know the answers after giving them some time. Only if they listened to me could they complete the review comprehension material and submit it. This means that I evaluate the "level of effort" rather than "level of achievement."

(Individual interview with Teacher D)

However, considering the low grade weights the teachers assigned for class participation, it is questionable as to whether their emphasis on effort was actually about improving students' learning. For example, in School Y, the class participation assessment accounted for only five percent of a student's semester grade. Teacher E said she was satisfied with this proportion and there was no need to increase it.

Teacher E: It isn't difficult for students to get the scores for the "class participation

assessment." By attending and concentrating in class, every student

can achieve a perfect score. Since anyone can do it, there is no need to

assign a greater weight to this assessment.

(Individual interview with Teacher E)

Teacher A of School X, where 15% of a student's semester grade was allocated to

the class participation assessment, expressed a similar opinion.

Teacher A: If the students make an effort, they can get a perfect score. That's not

tough work. The 15% weight is appropriate. No more is needed.

(Group interview at School X)

The teachers asserted that since they provided multiple opportunities for the

students to obtain scores through a variety of activities, any student could achieve

a perfect score if he or she made the effort.

Teacher B: If students make an effort in activities related to "class participation

assessment," they can gain a perfect score. The purpose of the "class

participation assessment" is "to actively engage students in class." So,

if a student does not get any stamps for activity A, I provide him or her

another opportunity in activity B. In this regard, the students can get a

perfect score if they only make the effort.

115

Teacher A: The "class participation assessment" gives a chance for every student to get a perfect score. For example, all of the students in a class I'm responsible for achieved a perfect score in the "class participation assessment" last semester.

According to the teachers' responses, since the class participation assessment was scored based on students' effort rather than achievement, which made achieving a perfect score relatively easy, they deemed it appropriate to assign it a low grade weight. This seems to imply that the teachers considered assessments that did not allow for differentiating or ranking less important. In other words, the teachers prioritized assessments designed to rank students over assessments designed to contribute to students' learning.

In sum, regarding the class participation assessment, the School X teachers implemented various activities including answering the teacher's questions and submitting completed homework assignments and the School Y teachers conducted activities including completing a review assignment and submitting vocabulary exercise papers. For the evaluation criteria, the School X teachers used the number of stamps a student obtained and the School Y teachers used the number of submissions of review papers and vocabulary exercise papers. The teachers at both schools believed that providing grades effectively motivated the

students to participate. While they admitted that these criteria valued effort over achievement, they ultimately assigned a low grade weight to the class participation assessment.

4.6. Summary of the Findings Regarding the Assessment Methods and Evaluation Criteria

In summary, when the performance assessment methods for Schools X and Y were examined, School X planned assessment methods as essay writing, oral presentation, listening assessment, and class participation and School Y planned them as essay writing I and II, dialogue recitation assessment, listening assessment, and class participation. Both schools established performance assessment as accounting for at least 40% of a semester grade and conducted mandatory essay assessment accounting for at least 35% of a semester grade, conforming to the requirements of the Gyeonggido Office of Education (2019). In both Schools X and Y, the essay writing assessment theme was chosen in relation to a theme of readings in the textbook. Also, School Y adopted writing tasks excluding tasks students had already performed with the intention of adding diverse information to their educational profiles. Regarding evaluation criteria, both Schools X and Y selected *task completion* and *content and construction* criteria. Furthermore, while School X also selected *language use* and School Y selected *grammar*, *spelling*, *vocabulary*, and *expressions*, these were equivalent in terms of performance of scoring. The teachers of both schools determined to assign a greater grade weight

to task completion over those criteria entailing teachers' subjective judgment. In addition, both schools selected the grammar and vocabulary criterion with the aim of grading students accurately.

With respect to the speaking assessment, School X conducted an oral presentation using the students' essays as scripts, while School Y carried out a memorized dialogue recitation assessment. Both schools implemented two identical conditions for these assessments, namely, a time limit and one-to-one (student and teacher) situation. As with the essay writing assessment, for the evaluation criteria of the speaking assessment, the teachers of both schools assigned more weight to the task completion criterion. Also, the teachers of both schools made sure that all criteria other than task completion did not contribute to differentiating the students: At School X, the teachers used a binary scoring system, assigning a grade of either zero or two, for pronunciation and intonation; at School Y, the teachers opted for a smaller grading interval scale for the criterion of voice, pronunciation, and fluency.

As for the listening assessment, both schools used the EBS multiple-choice test. The teachers explained that they wanted to differentiate students precisely with a multiple-choice test. Reflecting this intention, the evaluation criteria consisted of the number of correct answers, although the base score and coefficient in the scoring formula was different at each school.

Lastly, regarding the class participation assessment, the School X teachers decided to use a variety of activities such as answering the teachers' questions and doing homework. In the case of School Y, the activities included completing a

review assignment and submitting vocabulary exercise papers. The evaluation criteria included the number of stamps a student gained (School X) and the number of submissions of review papers and vocabulary exercise papers (School Y). Both schools' teachers perceived that rewarding students with grades was efficient to encourage students to participate in the classes. In addition, they thought that these criteria emphasized effort rather than achievement, but ultimately they decided to assign a low grade weight to the class participation assessment.

The present study examined the assessment methods and evaluation criteria of both Schools X and Y from the perspective of assessment *for* learning. With respect to the assessment methods, the most notable thing was that the writing, speaking, and listening assessments were conducted without any actual teaching. It is clear that the curriculum stipulates that teaching and assessment are indivisible (Wiliam, 2001). Drawing on assessment *for* learning theory, assessment should inform and drive teaching practice. The information gathered through assessment allows teachers to understand the gap between students existing knowledge and a desired outcome. By carefully analyzing and interpreting the evidence of students' accomplishments, teachers can help students meet the expectations for a course or lesson, and then revise and refine their teaching to meet students' needs, strengthening their teaching and ultimately leading to students' learning improvement. That is why adjusting teaching is the key factor to improving learning in assessment *for* learning (Assessment Reform Group, 1999).

The teachers' comments also revealed that they rarely provided feedback to the students with respect to all the assessments. In assessment *for* learning,

119

teachers' feedback based on the examination of students' performance is the essential factor for improving students' learning. The feedback should be constructive, informing not only where students are at in their learning but also how they might progress toward their learning goals. In addition, it should affect their confidence and enthusiasm.

Another noteworthy point is that the teachers did not employ authentic methods, as evident in the speaking assessment. The authenticity of tasks for performance assessment is important because this leads students to demonstrate their functional language ability in a real-life situation. Also, in the process of undertaking authentic tasks, students are not only assessed but also provided an opportunity to learn by addressing or solving tasks (Black, 2010; Wiliam, 2010). Assessment *for* learning requires students to be actively involved in learning through assessment, allowing teachers to elicit evidence of students' learning and students to reflect on their learning. In this respect, the speaking assessment tasks did not seem like they would have provoked the students' active involvement. Rather, required to recite a memorized dialogue, the students were treated as passive test-takers. Also, as evident in the speaking assessment administration methods, the use of a time limit and one-to-one situation would have deprived students of the opportunity to engage in peer assessment.

In spite of the importance of authenticity for assessment *for* learning, some previous studies also argue that even when assessment methods are inauthentic or are intended for grading students' performance rather than improving their learning, the focus should nevertheless be improving students' learning (e.g.,

Kennedy, 2007). In this respect, *teachers' constructive feedback*, rather than authenticity or the initial purpose in designing an assessment, would be most important element of assessment for learning.

Lastly, regarding the class participation assessment, the teachers used "stamps" or grades as rewards, review assignments, and lists of memorized words. Previous studies show that rewards only make students seek out the best marks or right answer rather than meeting their learning needs. Neither are review assignments or memorized word lists appropriate for assessment *for* learning. Such tasks fail to provoke students' higher-level thinking or problem solving skills or provide an opportunity for students to reflect on their learning.

Meanwhile, the evaluation criteria showed that the teachers were focused on differentiating between the students and ruling out any subjective judgment of their work. In assessment *for* learning, teachers should use evidence of students' learning obtained through assessment to promote teaching and learning. If teachers only focus on task completion, grammar, and vocabulary, they cannot monitor other aspects related to speaking and writing and gain sufficient information to improve teaching and learning. In this respect, as Leung and Mohan (2007) argue, they fail to act as a facilitator or supporter of learning who can give effective and constructive feedback to the students.

Another important point regarding the evaluation criteria is that the students were not made sufficiently aware of the evaluation criteria. In both Schools X and Y, the teachers did not provide explicit explanations for the scoring categories or performance standards. In assessment *for* learning, for students to become

responsible for their learning, it is essential for them to understand their learning goals. However, since criteria for evaluating the fulfillment of learning goals can be abstract, it is crucial to explain them through discussion with students in terms they understand accompanied by explicit examples.

It was also notable that the teachers prioritized effort in the class participation assessment, believing that this could improve students' learning. Nevertheless, by assigning a lower grade weight to the class participation assessment, they undermined their focus on effort. This finding suggests that the teachers did not espouse a firm belief that the ultimate purpose of their assessments was to promote students' learning.

This chapter discussed the research findings with respect to the first research question, "How do Korean high school teachers of English plan their performance assessments and what assessment methods and criteria do they use?" The decision making of the teachers of Schools X and Y in selecting assessment methods and evaluation criteria was analyzed and interpreted in light of assessment *for* learning, the principles of which form the basis for the process-centered performance assessment of the 2015 revised national curriculum. To review, the key aspects of assessment *for* learning (Black, 2010; Wiliam, 2010) are as follow: clearly explained evaluation criteria, activities eliciting evidence of learning, students actively engaging in peer and self-assessment, and constructive feedback advancing learning. The data of present study hardly revealed these key aspects of assessment *for* learning.

CHAPTER 5.

TEACHERS AS SOCIAL BEINGS

This chapter discusses the findings of the study in relation to the second research question, "How do the teachers describe their choices regarding the assessment methods and evaluation criteria?" In other words, this chapter explores the reasons why the teachers chose the specific methods and criteria they did and the factors underlying their thinking process behind this decision making.

The testimonies and documents provided by the teachers of Schools X and Y evince a decision-making process substantially affected by considerations of government policy, reporting requirements, university admission, and external high-stakes assessment. These findings reveal teachers not only as agents of their classroom-based assessments but also social beings whose decision making is significantly influenced by sociocultural context.

5.1. The Reasons Why the Teachers Chose the Assessment Methods and Evaluation Criteria

The factor the teachers referenced most often when asked the reasons for selecting the methods and criteria they did was *government policy*. In other words, they expressed the concern that we should "comply with government policy." This phrase, in one form or another, was repeated over and over in their testimonies.

Teacher A: The policy states that we should administer performance assessments

according to a designated grade weight. So, we just followed the policy.

(Group interview at School X)

Teacher D: Everything concerning performance assessments is planned to comply

with the requirements of the government policy.

(Individual interview with Teacher D)

Teacher C: Above all, I consider national policy when planning performance

assessments. I mean, what the guidelines of national policy state, for

example…what kind of performance assessments should be conducted

and how much of a student's grade should they account for…. And then

I consider the feasibility of each performance assessment.

(Group interview at School X)

The teachers seemed to feel that administering the performance assessments, in accordance with education policy, was an important duty. In particular, they were deeply concerned with making sure that the performance assessments accounted for the expected total in a student's grade, especially with respect to the essay writing assessment.

Teacher D: The most important thing I consider when planning performance

assessments is the national policy. This is the most significant frame to which I should conform. The policy also assigns the grade weight of performance assessments. For example, it states that we should administer essay writing assessments accounting for at least 35% of a student's semester grade. To observe the policy, we have to conduct essay performance assessment and make sure it accounts for 35% of a student's semester grade. In order to fulfill this requirement, we adjust the grade weights of other performance assessments.

<div align="right">(Group interview at School Y)</div>

Teacher A: When planning performance assessment, I do so according to the policy, for example, considering that essay assessment should account for a certain percentage, etc.

<div align="right">(Group interview at School X)</div>

The School Y teachers commented that although they had to comply with government policy in determining the grade weight for performance assessments, they were not content with this composition. The following comments made by Teacher E of School Y convey dissatisfaction with the total weight expected for performance assessment.

Teacher E: We decided to allocate 40% to performance assessments. This is the minimum required by the government policy. In our school, with

performance assessments accounting for 40% of a student's grade, regular assessments have come to make up the other 60%. In my view, since the class lessons mostly rely on reading, regular assessments (i.e., midterm and final exams), which are mainly composed of reading and some grammar questions, should be assigned a much higher grade weight than 60%. However, there is no choice whether or not to follow the policy requirements—we had to select 40% as the minimum grade weight for performance assessments. In any case, we could not assign a weight greater than 40% to performance assessments because most students in our school don't have the ability to perform assessments such as essay writing.

(Individual interview with Teacher E)

This excerpt demonstrates how Teacher E considered that since classes mostly consisted of teaching reading, assessment for reading should account for a much higher grade weight than 60%. Nevertheless, her own perspective was not well reflected in the total weight assignment to performance assessment due to the weight predetermined by the policy for performance assessment.

Teacher D also expressed discontent regarding the 35% weight given to essay assessments. She argued that most School Y students did not have the ability to write English essays.

Teacher D: The policy requires teachers to administer essay assessments worth

126

35%. This is a policy, I think, that does not take into account students'

English proficiency. Most students at our school don't have the ability

to write essays. They have difficulty even learning the English

vocabulary proposed in the textbooks. Nonetheless, even these students

are required by the policy to write English essays. I can't understand

this kind of performance assessments administered collectively without

considering students' differing proficiency depending on the school.

(Individual interview with Teacher D)

In short, although School Y teachers had their own opinion that assessment for reading should account for a greater grade weight because most of the classes were dedicated to reading and School Y students were not proficient enough to write English essays, it was not sufficiently taken into consideration in structuring the performance assessment composition and grade weighting of School Y because of the imperative to conform to policy. In other words, in planning performance assessment, the teachers' judgment with respect to the teaching context and students' writing ability was less important than policy requirements.

The teachers' testimonies illustrate that they felt compelled to administer performance assessment due to government policy, in particular, the guidelines for performance assessment stipulated in the 2015 revised national curriculum. Here, what is most remarkable is that even though the 2015 national curriculum describes its main focus as "assessment *for* learning," *improving students' learning*

seemed peripheral in teachers' implementation of the assessments. Rather, the data revealed that the teachers were significantly affected by obligations such as assigning grade weights and administering essay writing assessments. The important point is that when teachers prioritize an administrative purpose imposed by education policies in implementing performance assessment, this could lead to their losing control over their teaching (e.g., Brindley, 1998; McKay, 2000). For example, Arkoudis and O'Loughin (2004) argue that when the administrative purposes of assessment policy lead teachers away from focusing on teaching and learning, teachers are dispossessed of the authority to modify their assessment and control their teaching. Under such conditions, assessment *for* learning could hardly be expected.

With respect to their assessment plans, teachers responded that they thought they had a duty "to grade and report students' performance."

Teacher B: We conduct them [performance assessments] because we have to grade the students…. We have to give grades to the students by conducting them.

(Group interview at School X)

Teacher E: We are required to report the scores and grades in educational profiles. We have to administer the performance assessments because we have to report the grades.

(Individual interview with Teacher E)

128

Teacher B: We should allocate a grade to each student in the 11th grade. So, we have to be objective.

(Group interview at School X)

Teacher C: The essential matter for students in conducting performance assessments is, ultimately, grades. They want to achieve a higher ranking and go to a more prestigious university. So, I think that we should be fair in grading.

(Group interview at School X)

The participants explained that they should differentiate between the students based on grades. In Korean high schools, a student's semester score in English is determined by regular examinations (i.e., midterm and final exams) and performance assessments. Through the combined scores, the students are assigned a grade from one to nine, which is used to indicate their English proficiency when they apply to university. In this respect, the teachers said that even though the performance assessment itself was based on criteria, they were nonetheless always concerned about differentiating the students through grades.

Teacher D: The performance assessment scores are deeply related to the students' university admission. They make up at least 40% of a student's English semester grade. Whether the students apply to a university based on

their comprehensive profile or performance in a specific subject, it's a

pretty large proportion! We can't disregard it.

<div align="right">

(Group interview at School Y)

</div>

Teacher C: English scores are norm-referenced. We are required to assign a grade

of one to nine to the students. We teachers always take care about how

to discriminate between the students and how to assign different grades.

<div align="right">

(Group interview at School X)

</div>

Since scores were thought to be so important, the teachers also thought they should prioritize objectivity and fairness in implementing the performance assessments.

Teacher A: Since we have to provide students with grades, I think it's important

for us to be objective and fair.

<div align="right">

(Group interview at School X)

</div>

The teachers pursued objectivity and fairness in administering the performance assessments by making the conditions for testing and scoring consistent. Since the assessments were administered and rated by three different English teachers, it was imperative that the conditions of the performance assessments reflect unified methods and that the scoring of the assessments reflect unified criteria.

Teacher B: As much as possible, we three teachers try to be objective and fair in administering and scoring performance assessments. Although each teacher is responsible for different classes, the assessment methods and criteria should be identical. This guarantees objectivity and fairness.

(Individual interview with Teacher B)

The priorities of objectivity and fairness motivated the teachers to hold many conferences for planning, administering, and scoring the assessments. The teachers said that they do not inform the students of any of their decisions regarding the planning of performance assessment until the plans have been finalized.

Teacher C: Considering the fact that performance assessments are graded and recorded in students' educational profiles, uniformity of methods, criteria, and administration is most important.... To guarantee their uniformity, we have many teachers' conferences on making task formats, specific criteria, rating the assessments, and so on.... If these are not consistent, students complain. The students are so sensitive about their scores. So, we don't share any ideas with the students until we have a fixed and unified idea about the performance assessments.

(Group interview at School X)

Also in order to ensure fairness in administering the performance

assessment, in both Schools X and Y, the assignment handout was posted about one week before the due date in all of the classrooms of the same grade with guidelines on how to prepare the assessment, assessment administration process, and evaluation criteria. In addition, to guarantee objectivity they excluded criteria entailing subjective judgment (i.e., creativity) and included criteria ensuring objectivity. Furthermore, by adopting the multiple-choice method for listening assessment and including a *grammar* criterion, they tried to accurately differentiate students.

As the data shows, to grade and report performance, objectiveness and fairness were prioritized in choosing the assessment methods and criteria. The teachers tried to make the methods, criteria, and administration uniform to compensate for their implementation by different teachers. In order to do this, they held many teachers' conferences.

Here, it seems problematic that the teachers focused on objectiveness and fairness in administrating assessments without teaching writing. As mentioned above, no matter how teachers make the effort to administer assessments in a fair and objective way, conducting assessments without teaching and learning could bring about an unfair situation in which students depend on external help in order to prepare their assessments.

Teachers' focus on consistency is also observed in previous studies. Arkoudis and O'Loughlin (2004) and Gipps (1994), for example, show that the purpose of reporting accountability requires comparability, which requires consistency in turn. Teachers thus spend a great deal of effort in making sure

assessments are implemented in a standardized manner and that the criteria leave little room for differing interpretations. Naturally, this requires considerable coordination, leading to frequent teachers' conferences (Brindley, 2001).

The notable point here is that *consistency* is not a term associated with assessment *for* learning, namely, formative assessment. For an assessment to be *for* learning, it should be unplanned, contingent, interactive, and dynamic (Leung, 2004). Also, it requires that teaching and learning be integrated, which means that teachers respond to students' needs *during* the teaching and learning process. *During the teaching and learning process* is one of the key points of assessment *for* learning and process-centered performance assessment in the 2015 revised national curriculum. For assessment to be conducted during the teaching and learning process, the setting out in advance of the methods and criteria of an assessment could be an obstacle. Furthermore, consistency between different teachers is almost impossible. In this respect, as Teasdale and Leung (2000) point out, consistency is incompatible with formative assessment. This means that the reason "to grade and report performance," which led to an emphasis on consistency, jeopardized the implementation of assessment *for* learning in Schools X and Y.

Another reason why the teachers selected the assessment methods and criteria they did was "to fill out students' educational files." Teacher C explained that the national policy calls upon teachers to *report* the results and processes of performance assessments in the *Specific Competencies and Specialties in English Learning* section of students' educational profiles.

Teacher C: One thing to keep in mind is that we have to report information on the students such as performance process, attitudes, and results in the "Specific Competencies and Specialties in English Learning" section of the educational profiles. This is one of the points emphasized in the guidelines of the 2015 revised national curriculum.

(Individual interview with Teacher C)

Teacher E: We have to conduct performance assessments. Based on the students' performances in the assessments, we can extract some materials to fill out the "Specific Competencies and Specialties in English Learning" section of the educational profiles.

(Individual interview with Teacher E)

The 2015 revised national curriculum requires Korean high school English teachers to fill in the *Specific Competencies and Specialties in English Learning* section of students' educational profiles (e.g., KICE, 2018; MOE & KICE, 2017). The excerpt below demonstrates clearly that Teacher D felt that recording details for the *Specific Competencies and Specialties in English Learning* section was a duty.

Teacher D: I need specific information about students' ability or performance. Obtaining materials from the performance assessments to fill in the educational profiles is for the sake of the teachers as well as students.

I feel stressed when I have no material to record.

(Individual interview with Teacher D)

In addition, for university admission, universities rely on information recorded in the *Specific Competencies and Specialties in English Learning* section of students' educational profiles. In this regard, the English teachers seemed to feel a strong obligation to fill in the educational profiles with diverse and detailed information, which might eventually help the students to get into university.

Teacher D: The performance assessments play a significant role in providing information to fill out the "Specific Competencies and Specialties in English Learning" section of educational profiles. When a student applies to university through the admission type "comprehensive educational profile," the university evaluates the student by drawing on the content of the "Specific Competencies and Specialties in English Learning" section of his or her educational profile. So, we aspire to fill out as much as possible of that section. However, just writing "this student is diligent" is not convincing. Recording more specific information is advantageous for the students. In this respect, to fill out the section, we need a variety of distinct details regarding a student and his or her English ability.

(Individual interview with Teacher D)

Teacher E: I desire to write in as much diverse information as possible. I think that maybe most teachers do. This is beneficial when the students apply to university.

(Individual interview with Teacher E)

The teachers thus appeared to be preoccupied with procuring materials for filling out the *Specific Competencies and Specialties in English Learning* section of students' educational profiles. The following excerpt clearly shows how Teacher A recorded information regarding a student obtained through the performance assessments in the *Specific Competencies and Specialties in English Learning* section of the students' educational profiles.

Teacher A: I fill out the "Specific Competencies and Specialties in English Learning" section with information regarding a student's performance assessments. For example, I mention which theme the student selected in the essay writing assessment. The students often choose different themes. I think their selections reflect their identities and interests. I also briefly summarize the student's writing on that theme and record the feelings he or she expressed in the conclusion. I also sometimes register which relative pronoun or connective words the student can use in writing. By obtaining different information about each student in this way, I can fill out their "Specific Competencies and Specialties in English Learning" section. Finally, if there is still room in the

section even after recording the above information, I add how the

student actively participated in class, using information obtained from

the "class participation assessment" section.

(Individual interview with Teacher A)

Teacher C also said that the teachers of School X decided to include a speaking assessment in order to diversify the sources of information for the educational profiles.

Teacher C: If we conduct only writing assessments, we can only gain information

regarding students' writing skills. With the intention to record

information related to diverse skills, we decided to try an oral

presentation assessment for the first time this semester.

(Individual interview with Teacher C)

As the interview data shows, in terms of filling out the education profiles, the teachers were particularly concerned about providing information for the *Specific Competencies and Specialties in English Learning* section. This is a crucial document for students' university applications, and the teachers believed that the more abundant and diverse the information the more advantageous for students' university applications. This led the teachers to choose assessment tasks they considered would yield information universities would prefer (such as *book report*). For example, the School Y teachers intentionally examined the students'

previous educational profiles and precluded tasks they had already conducted in order to generate more diverse information.

The reason "to fill out students' educational profiles" showed that teachers conducted performance assessments to fulfill their obligation to report specific information about students' English ability and English scores in their educational profiles. This conforms to the requirement of the 2015 revised national curriculum that teachers report in detail the processes as well as results of performance assessment in educational profiles. The MOE distributes sample forms as references for teachers in making these records. Teacher D's comments were especially clear about how teachers feel obligated in this regard: "Obtaining materials from the performance assessments to fill in the educational profiles is for the sake of the teachers as well as students. I feel stressed when I have no material to record." In other words, filling in the educational profiles with information on students' English ability and scores is an MOE requirement, and teachers *had* to conduct performance assessments simply to gain materials to meet this requirement.

Another reason for selecting methods and criteria, according to the teachers, was "to provide students an opportunity to practice speaking." Teachers explained that they do not teach speaking and writing. For the reason of not teaching speaking and writing, the School Y teachers claimed that the students' low English proficiency hindered teachers from spending time teaching productive skills.

Teacher D: We could not deal with speaking and writing at all. First of all, we are

so busy teaching reading. We manage to complete the planned amount of reading passages. We definitely lack time…. Plus, the students' lack of proficiency wastes time. The students' English proficiency in our school is so low that they can hardly understand what I explain. They don't even know very basic vocabulary. I could rarely even consider teaching productive skills to the students, who are even distressed by learning reading.

<div align="right">

(Group interview at School Y)

</div>

Also, the teachers stated that because in-class teaching mostly consisted of reading instruction, students did not have an opportunity to practice speaking. In these conditions, they believed that in order for students to acquire the other three major skills of listening, speaking, and writing, they should at least provide opportunities to practice speaking.

Teacher D: Actually, other than reading, speaking, writing, and listening are all language skills we should teach. Even though our in-class teaching is preoccupied with reading, to provide students an opportunity to practice would help them to improve their language ability, I think.

<div align="right">

(Group interview at School Y)

</div>

Teacher E: In regular examinations, reading and grammar points are assessed. The regular examinations assess only a limited area. Accordingly, we

should assess speaking, listening, and writing in performance assessments.

(Individual interview with Teacher E)

Moreover, the School Y teachers believed that for the students who usually read passages or dialogues silently, the speaking assessment would encourage them to speak out loud and pronounce English words.

Teacher E: It [speaking assessment] makes students study. The students don't read out loud. They don't repeat after me. However, I find them reading out loud in the classroom on the morning of an assessment.

(Individual interview with Teacher E)

Teacher E added that if there is no assessment, teaching does not take place.

Teacher E: The students do not read out loud. They do not follow my instructions. Only through performing the speaking assessment do they speak out loud.

(Individual interview with Teacher E)

In this vein, the teachers of School Y seemed to think that it was enough for the students to have an opportunity to speak out loud through the memorized dialogue recitation. For the same reason, in assigning grade weights to the

assessment, the teachers prioritized the task completion criterion, which focused on whether a student completed the task within the allotted time, rather than pronunciation or other skills related with speaking ability. The excerpts below convey the teachers' views with respect to their speaking assessment methods and grade weighting for the criteria.

Teacher D: Whether a student completed the task, I mean, whether a student recited the dialogue from the first to last line, was the most essential point in administering this speaking assessment. So, we assigned a higher weight to the "task completion" criterion relative to others.

(Individual interview with Teacher D)

Teacher B also stated that the most important priority was for students to practice speaking, while pronunciation and fluency, for example, were less important. In this regard, Teacher B said that the teachers of School X prioritized task completion over pronunciation in the speaking assessment.

Teacher B: To practice speaking out loud is the major purpose of the speaking assessment, I think. Pronunciation is not a big deal. So, we assigned more weight to the "task completion" criterion when planning the performance assessments in the teachers' conferences.

(Individual interview with Teacher B)

The data reveals that teaching for productive skills was lacking in practice at Schools X and Y. In other words, the implementation of speaking assessment lacked a process of teaching and learning. It was rather the case that the assessments were used merely to report snapshots of achievements through the one-time assessment of productive skills. Furthermore, teachers were missing valuable opportunities provided in performing speaking assessments to develop students' productive skills through relevant instruction and specific feedback.

Also, the data indicates that, in implementing speaking assessments to provide students an opportunity to practice speaking, the teachers' selection of assessment methods, criteria, and administrative methods did not conform to the objectives and characteristics of the 2015 revised national curriculum stipulating that performance assessment is meant to be assessment *for* learning.

Lastly, as for the reason why the teachers chose to implement class participation assessments, they answered that they thought this would "encourage students to actively participate in the classes." This is conveyed in the following excerpts.

Teacher E: If I do not administer class participation assessment, many of the students just sleep. But if I administer a class participation assessment, they do not sleep or get distracted. Since they should complete the comprehension questions at the end of the class for the purpose of class participation assessment, most of the students, except for only a few, carry out the instructions.

Teacher D: Only when grades are given do students study enthusiastically. If grades are not given, they do not make an effort.

<div align="right">

(Individual interview with Teacher D)

</div>

Teacher B: When I suggest some activities in class, the students usually expect a reward. If I don't provide rewards for the activities, most of the students have little motivation to involve themselves. Students seem to take it for granted that they should be rewarded with stamps or scores through participating in the activities…. In this respect, we constructed the "class participation assessment." Through actively participating in the classes, the students can get stamps and obtain scores as a result.

<div align="right">

(Individual interview with Teacher B)

</div>

Teacher A: The students are so sensitive to the performance assessment scores that they actively participate in the activities related to the "class participation assessment." In the case of group activities, even if a student is unenthusiastic, he or she eventually participates in the activity. This is because the student knows that if he or she doesn't work hard, the group cannot gain stamps, and then the other members will blame him or her.

<div align="right">

(Individual interview with Teacher A)

</div>

As exemplified in the teachers' statements, given the fact that students are sensitive to rewards (i.e., grades), the teachers seemed to try to induce the students to actively participate in class by rewarding them with stamps. In fact, in School X, the stamps given to the students corresponded to scores, because the more stamps the students earned, the higher scores they achieved. Also, in School Y, scores were given according to the number of assignment submissions a student made.

The interview data shows that the teachers of Schools X and Y believed that these assessments effectively motivated students to attend class. However, many studies argue that using extrinsic rewards such as grading can hamper students' intrinsic motivation in the long term. Also, the Assessment Reform Group (1999) asserts that when classroom assessment takes on a management role, it can inhibit assessment *for* learning. In this respect, for assessment to be *for* learning, it is desirable for teachers to focus on examining what help the students need and provide constructive feedback rather than on making students submit the work in the name of participating in the class.

In addition, the data shows that the teachers valued effort over achievement as a criterion in the class participation assessment. Nonetheless, in deciding the grade weight for the class participation assessment, they assigned a low grade weight. This seems to indicate that even though they were willing to improve learning and teaching by conducting the class participation assessment, they still considered ranking students as more important.

144

In sum, the teachers perceived the reasons why they chose the assessment methods and evaluation criteria as "to comply with government policy," "to grade and report performance," "to fill out students' educational profiles," "to provide students an opportunity to practice speaking," and "to encourage students' active class participation." The data regarding the teachers' perceived reasons suggested that the teachers did not regard the purpose of performance assessment as assessment *for* learning.

5.2. Sociocultural Context and Teachers' Decision Making

Thus far, the teachers' descriptions of the reasons for their choices of assessment methods and evaluation criteria seem to indicate the strong influence of government policy and reporting requirements. This is significant because, as pointed out in the existing literature and emphasized in this study, government policy and reporting requirement pressure can negatively affect the implementation of assessment *for* leaning.

First, when asked to report the reasons for their choices of assessment methods and evaluation criteria, the teachers described reasons corresponding to an administrative purpose, appearing preoccupied with the duty to comply with education policy, that is, the requirements to implement essay writing assessment and report the results of students' work. In other words, during the planning stage the assessments came to reflect administrative, summative, and reporting purposes. This means that the performance assessments were *for* learning in theory but not

implemented as such in practice.

Such findings are not entirely unique to this study. Previous studies also show how bureaucratic pressure forces teachers to adopt a summative purpose for assessment, regardless of their intentions (e.g., Black, Harrison, & Lee, 2003; Cheng, 2004; Rea-Dickins, 2002). For example, Cheng (2004) discovered that although teachers of English in Canada and the USA view classroom-based teacher assessment as a beneficial tool to promote students' learning and their teaching, when they are under significant pressure from government policy and reporting requirements, they ultimately carry out assessment with a summative purpose.

Second, even though the 2015 revised national curriculum states that performance assessment should be assessment *for* learning, it also emphasizes "grading and reporting performance." More specifically, the policy calls upon teachers to integrate assessment with teaching and learning while simultaneously requiring them to report the process and product of students' performance in educational profiles. This reveals an assumption of the compatibility of formative and summative purposes. Reflecting this assumption, some have suggested that teachers can conduct formative assessment during teaching and learning by scoring pieces of work and then tallying these scores at the end of the course. However, Leung and Teasdale (1997) maintain that this style of teaching might be too impressionistic, that is, rely too much on teachers' subjective judgments. Also, Gipps (1994) insists that it may provoke in students the mindset that if an assignment does not contribute to their total score it is not worth doing. Moreover,

as discussed above, the interactive and dynamic characteristics of formative assessment are incompatible with summative assessment. In light of these considerations, the present study suggests the need to more deeply explore whether the 2015 revised national curriculum, which simultaneously emphasizes reporting requirements and assessment *for* learning, properly considers the feasibility of assessment in a classroom setting (see Chapter 6).

Third, the teachers' testimonies also suggested that the current national policy does not consider specific school conditions. Teachers D and E, for example, expressed unfavorable opinions regarding the mandatory implementation of essay writing assessment accounting for 35% of a student's grade. They claimed that the involvement of essay writing assessment in English as in other subjects is unreasonable given that Korea is an EFL context. They also asserted that the students of their school, who are unable even to accurately spell the vocabulary in the textbook, could not handle essay writing assessment. This implies that the nationwide mandating of essay writing by the Korean government fails to consider the specific context of English learning and students' English proficiency in each school. In this regard, it is unlikely that this policy will help to promote students' learning. As Brindley (1998) and McKay (2000) contend, state-mandated standardized forms such as the curriculum standards framework (CSF) are decontextualized, failing to consider specific conditions, including subject, task, language, and students (Brindley, 2002), which may inhibit teachers from gathering meaningful and accurate evidence through assessment that might reflect students' progress.

Fourth, the teachers also expressed frustration regarding the vagueness of education policy. Teacher B of School X, for example, expressed the difficulty of interpreting the vague statements of the national curriculum:

Teacher B: I don't understand the policies presented in the 2015 revised national curriculum. More specific and concrete explanation is needed, I think.

(Individual interview with Teacher B)

When national policies are written in abstract and generalized language, they require teachers to depend on their own knowledge and teaching experience. In this regard, Brindley (2001) and Davison (2004) point out how policy is rarely implemented completely in the manner intended. Also, Leung and Rea-Dickins (2007) maintain that national policies with ambiguous words frequently lead to teachers' interpretations in the process of implementation.

Finally, the emphasis on accountability and acting as a rater was also intelligible in the data. Teachers prioritized accountability, which led them to emphasize accuracy and objectivity in judging students' work. This meant teachers acting as *raters* of students' work rather than *supporters* of students' learning, which ultimately undermines the successful implementation of assessment *for* learning. This finding supports the claim that reporting requirements adversely affect the implementation of assessment *for* learning (e.g., Bishop, 1997; Black et al., 2003; G. T. Brown, 2011; Cizek, 2001; Leung & Rea-Dickins, 2007; Rea-Dickins, 2002).

In summary, the data shows that the pressure of government policy and reporting requirements significantly affected the teachers' selection of methods and evaluation criteria for the performance assessments. Even though the present government policy emphasizes assessment *for* learning, it also stipulates reporting requirements, and it was this latter factor that decisively influenced the teachers' selection of methods and evaluation criteria, ultimately hindering the implementation of assessment *for* learning. In other words, government policy and reporting requirements adversely affected the implementation of assessment *for* learning and the obtainment of meaningful information for students' development. As Black and Wiliam (2004) indicate, ironically, the bureaucratic demands imposed on teachers intended to improve education can actually function as the one of the biggest barriers to improving education.

Following the pressures of government policy and reporting requirements, the data showed that the teachers' selection of methods and evaluation criteria for the performance assessments was influenced by "university admission" and "internal and external testing."

In Korea, when students apply for university, their grades for English are reported in educational profiles along with their scores on the English language section of the KSAT. Moreover, for students applying to university through the *Comprehensive Educational Profile Admission* process, scores, grades, and information regarding ability recorded in the *Specific Competencies and Specialties in English Learning* section of their educational profile are immensely important.

In this context, university admission was an important factor affecting teachers' determinations in planning performance assessments. Keeping in mind that they should grade and report performance, the teachers focused on rating students and assigning grades objectively. In addition, they tended to fill in the *Specific Competencies and Specialties in English Learning* section in students' educational profiles with as much and as varied information as possible, believing diverse and detailed information advantageous for university admission. Thus, university admission, more specifically, "the importance of English scores and information regarding a student's English ability reported in educational profiles for university admission," significantly impacted the teachers' decision making.

Meanwhile, external high-stakes assessment was another significant influence on the teachers' decision making. In the interviews, teachers asserted that they do not have enough time to teach speaking and writing because their classes are mostly composed of teaching reading. When asked why they teach only reading, the teachers responded that this is because the regular examinations (i.e., internal testing consisting of midterm and final exams) are composed of mostly reading and grammar. In this respect, *internal testing*, which is mostly made up of reading questions, led teachers to decide to conduct performance assessments "to provide students an opportunity to practice speaking."

Teacher A: In classes, we spend most of the time teaching reading.... The reason why we only teach reading is...that regular examinations are largely composed of reading questions. For the students to prepare for the

150

reading examinations, we have to teach a minimum amount of reading

passages. We are so busy teaching reading that we have no time to

teach writing and speaking. It might sound like I'm making excuses….

But really we do not have enough time.

<div align="right">

(Group interview at School X)

</div>

The teachers also described how *external high-stakes assessment*, namely, the KSAT, was a significant factor affecting their perceptions and decision making in planning the assessments. Teacher D indicated that the inclusion of only reading questions in regular examinations is a consequence of the influence of the KSAT, which Korean high school students take to apply to university. She explained that because students' ultimate goal is to gain good scores on the KSAT and be admitted to university, teachers believe they should prepare the students to achieve a better score on the KSAT. In this respect, teachers consistently include reading questions similar to those of the KSAT on regular examinations, focusing teaching on training to answer such questions.

Teacher D: Students should prepare to take the KSAT. To get higher scores on the
KSAT is the ultimate goal of high school students. Accordingly, we
devise reading questions for the regular examinations similar to those
on the KSAT. And we teach students reading and make them practice
reading questions in class.

<div align="right">

(Individual interview with Teacher D)

</div>

It seems evident that "external high-stakes assessment mostly composed of reading questions" influenced teachers' determinations regarding planning performance assessment. Teachers explained that the students cannot learn speaking and writing in class because they spend most of the time learning reading, preparing for internal and external exams that are mostly composed of reading questions. This results in a situation where, since the students do not have time to learn speaking and writing, the teachers propose to give opportunities to practice speaking and writing through performance assessments—in other words, students are assessed with respect to areas they have not learned. The important point here is that the implementation of performance assessment with these methods and evaluation criteria can hardly lead to meaningful assessment *for* learning.

In Korea, it is not an exaggeration to say that the primary goal of high school students is to get into as prestigious and reputable a university as possible. When students apply to university, in general, two categories of information matter. The first includes their grades on regular internal testing, determined through midterm and final exams and performance assessment with respect to each subject over three years of high school, and information in the *Specific Competencies and Specialties* section of their educational profiles. The other includes their scores on the KSAT, which is a form of external testing. In short, in Korean society, students' university admission, which significantly affects their lives, is closely related with external summative examinations.

Thus, the present study shows that the Korean high school teachers' decision making for implementing performance assessments was largely affected

152

by government policy, reporting requirements, university admission, and external high-stakes assessment. Many previous studies note that external high-stakes examinations greatly affect teachers' decision making with respect to classroom-based teacher assessment (e.g., Black et al., 2012; G. T. Brown et al., 2011). For example, Cheng (2004) finds that the methods and procedures of internal testing in China mirror those of external testing, and that teachers feel obligated to coach their students for the tests. Similarly, Dweck (2002) and Shen and Tam (2008) show how Hong Kong students are trained to perform well on tests, which only limitedly or partially develops their understanding of the content being assessed.

Previous studies interpret the effects of government policy, reporting requirements, university admission, and external high-stakes assessment, altogether, as the result of sociocultural influence (e.g., Carless & Lam, 2014a; Kenney et al., 2008). For example, Kennedy et al. (2008) argue that teaching and learning in countries such as China, Hong Kong, and Taiwan are substantially affected by internal and external summative examinations. Also, teachers feel a strong obligation to report results, focusing on their accountability to the government. In these countries, the primary purpose of education is regarded as to prepare students to pass external public examinations, which are crucial for social mobility.

Especially, some previous studies conducted regarding exam-centric Confucian-heritage countries such as Hong Kong, China, and Singapore (e.g., Broadfoot & Black, 2004) also discovered that deeply rooted exam-centric values inhibit the implementation of formative assessment or assessment *for* learning. For

153

example, Chen, Kettle, Klenowski, and May (2013) found that even though teachers are required to implement formative assessment as national policy initiatives in China, the teachers in two Chinese universities focused on summative grades under the influence of exam-centric Confucian-heritage values, emphasizing product over process and the fact that English-language class scores significantly impact students' lives. Additionally, G. T. Brown et al. (2009) found that secondary school teachers in Hong Kong were greatly influenced by the emphasis on high-stakes assessment of a Confucian-heritage context and believed that examinations are an appropriate tool for evaluating competing students. Lastly, Fulmer et al. (2019) found that Singapore teachers discerned a tension between the high-stakes examination-oriented Confucian-heritage sociocultural environment and policy initiative emphasizing formative assessment. In this regard, Hamp-Lyons (2007) asserts that in an exam-centric culture, one can hardly expect that formative assessment will be implemented in compliance with the goal of learning for its own sake.

In sum, the present study demonstrates that the Korean high school teachers' decision making for conducting performance assessments was greatly affected by sociocultural context, mediated by government policy, reporting requirements, university admission, and external, high-stakes assessment.

5.3. Summary of the Findings Regarding the Reasons for the Reported Assessment Practices

The present study provides meaningful evidence that, as social beings, just like teachers of English in other exam-centric Confucian-heritage countries, the Korean high school teachers of English who participated in this study were also significantly influenced by sociocultural values, adversely impacting their implementation of assessment *for* learning in the classroom. In the individual and group interviews, the teachers of Schools X and Y frequently responded that the reason for their determination for selecting assessment methods and evaluation criteria was "to comply with government policy." They expressed their obligation to plan and conduct performance assessments in accordance with the 2015 revised national curriculum (KICE 2017, 2018; MOE & KICE, 2017) and guidelines of the Gyeonggido Office of Education (2019). For this reason, the teachers incorporated essay assessments into the performance assessments conforming to the requirement that essay assessments account for 35% of a student's semester grade. Furthermore, they respectively assigned 40% and 60% grade weights striving to meet the requirements with respect to the minimum grade weight for performance assessment.

Regarding why they selected the assessment methods and evaluation criteria they did, the teachers gave several reasons. First, they needed "to grade and report performance." They explained that they have to grade students through performance assessments and report the scores in their educational profiles. They

appeared to believe that their role to grade and report was important because performance assessments determine a large portion (minimum 40%) of a students' semester grade, and grades reported in students' educational profiles play a vital role when the students apply to university. Due to this concern, the teachers made considerable effort to ensure the performance assessments' objectivity and fairness and precisely distinguish between and rank the students. On the one hand, to make the assessments objective and fair, the teachers endeavored to make the assessment methods and criteria uniform by frequently administering teachers' conferences on planning the assessments. Meanwhile, they tended to exclude criteria requiring subjective judgment and include criteria enabling objective assessment. On the other hand, to accurately differentiate between the students, the teachers adopted assessment methods and criteria enabling them to make precise distinctions with respect to the students' proficiency levels.

Second, the English teachers needed "to fill out student's educational profiles," especially in terms of the *Specific Competencies and Specialties in English Learning* section. They seemed to feel responsible for filling in this section and hoped that conducting performance assessments would allow them to extract information on students' English ability or specialties. In addition, considering that the *Specific Competencies and Specialties in English Learning* section is crucial for students' university applications, they desired to add to it diverse and distinctive information, as preferred by universities. They also selected assessment tasks that could provide information advantageous for university applications and that allowed the teachers to obtain diverse and specific

156

information regarding each student.

Third, the teachers expressed that they need "to provide students an opportunity to practice speaking English." Even though the class teaching mostly consisted of reading, they thought that the students should nonetheless also acquire the skills of speaking, writing, and listening. To provide opportunities to speak, the School Y teachers viewed the speaking assessment as a means for allowing students to practice speaking, especially for those who normally avoid reading or speaking aloud. They thus adopted the method of reciting a memorized dialogue for the speaking assessment under the conditions of a time limit and one-to-one situation. The teachers also prioritized the task completion criterion of the speaking assessment, which stipulated that the students complete the task within the allotted time.

Fourth, the teachers explained that they desired "to encourage students' active class participation." For this reason, they employed a variety of activities related with teaching. Chiefly, they viewed the class participation assessment as an effective means of motivating students to participate. The teachers provided grades as rewards for students participation, which was defined as answering the teachers' questions and submitting completed homework assignments, review papers, and vocabulary exercises. The teachers asserted that they focused on effort over achievement when it came to students' participation. Nonetheless, because students could achieve a perfect score relatively easily in the class participation assessment, they also decided to assign it a low grade weight.

Overall, the reasons the teachers expressed for their chosen assessment

methods and evaluation criteria for the performance assessment were related to administrative, accountability, and reporting functions conforming to education policy and reporting requirements. Pertaining to these results, the notable fact is that the purposes reported by the teachers hardly accord with—in fact, they may even inhibit—assessment *for* learning as described in many previous studies. The administrative purpose of observing education policies can cause teachers to lose control of their teaching and assessment, preventing them from devising their own assessment to adapt to the specific conditions in which they are teaching (Arkoudis & O'Loughin, 2004). Furthermore, assessments with a summative purpose, focusing on comparability, consistency, objectivity, and accuracy (Brindley, 2001; Gipps; 1994; Leung & Rea-Dickins, 2007), cannot accommodate assessment *for* learning (Sadler, 1981; Teasdale & Leung, 1997), which is supposed to be interactive, dynamic, and integrated into everyday teaching and learning (Leung, 2004). Summative assessment leads teachers to regard their role as that of a *rater* of students' work rather than *supporter* or *facilitator* of students' learning process and development (Leung & Mohan, 2007). This hinders teachers from implementing assessment with the purpose of promoting learning (Assessment Reform Group, 1999; Sadler, 1981).

The purpose "to fill out students' educational profiles" led the teachers to select assessment methods and tasks advantageous for university applications rather than teaching and learning. This inhibited the assessment from serving an *informing function*, which is one of the most essential features of assessment *for* learning (Assessment Reform Group, 1999; Rea-Dickins, 2006). The purpose "to

provide students an opportunity to practice speaking" functioned to capture *snapshots* of student's achievements. This purpose could hardly facilitate the learning or performing process, and failed to contribute to informing learner progress and teaching pedagogy. The purpose "to encourage students' active class participation" led to a focus on giving grades as rewards, which can hinder the development of students' self-motivated learning. Meanwhile, although the teachers did not explicitly cite the purpose "to rank or grade students," their assignment of a relatively minor grade weight to the class participation assessment demonstrated their prioritization of differentiating between the students.

Altogether, the findings of the present study demonstrate that the reason for carrying out an assessment, as perceived by the teachers, significantly affected the teachers' selection of methods and criteria. This demonstrates the importance of teachers' perceptions in the implementation of assessment: Even though a policy may prescribe assessment *for* learning, teachers' perceptions of the purpose of the assessment may ensure that it becomes assessment *of* learning in practice. These findings suggest the need for more comprehensive research on the perceptions of Korean teachers of English in high schools regarding assessment in order to determine whether assessment *for* learning is being implemented in accordance with government policy at a more general level.

Finally, the findings support the claim that teachers' perceptions and decision making are significantly affected by government policy, reporting requirements, university admission, and external high-stakes assessment. These factors can be categorized as forms of bureaucratic pressure, which, in a broader

159

perspective, is ultimately rooted in sociocultural influence. Especially when drawing on previous studies (e.g., Chen et al., 2013; Fulmer et al., 2019), the findings of the present study can be understood in terms of the influence of exam-centric Confucian-heritage sociocultural values on Korean high school teachers' decision making for the implementation of performance assessment. In other words, the study demonstrates how teachers are social beings whose decision making is significantly affected by sociocultural values.

CHAPTER 6.

CONCLUSION

In this chapter, Section 6.1 summarizes the aims, methods, and key findings of the study. Section 6.2 considers the pedagogical implications of these findings. Section 6.3 discusses the limitations of the study and provides suggestions for future research.

6.1. Summary of Key Findings

6.1.1. Aims and Methods

The aim of the present study was to investigate how the performance assessment newly introduced to English education in Korea through the 2015 revised national curriculum, which emphasizes assessment *for* learning, is being implemented in Korean high school classrooms. In order to do this, the study drew upon previous research showing that the purpose of assessment determined in the assessment planning stage significantly influences subsequent decision making, including the selection of assessment methods and evaluation criteria, which are crucial factors in the actual implementation of assessment. In addition, viewing language assessment as a social practice, the study relied on research on the sociocultural factors that affect teacher's decision making in assessment planning. Above all,

taking into consideration that the process-centered performance assessment presented in the 2015 revised national curriculum emphasizes assessment *for* learning, namely, assessment to promote students' learning, the present study investigated the theoretical framework of assessment *for* learning. With the aim to explore the practice of performance assessment in light of the literature on assessment *for* learning, the study devised the following two research questions: 1) How do Korean high school teachers of English plan their assessment assessments and what assessment methods and evaluation criteria do they use? and 2) How do they describe their choices of the assessment methods and evaluation criteria?

To answer these research questions, individual and group interviews were carried out involving five teachers from two different high schools in Gyeonggi Province. The researcher then analyzed the interview data along with relevant documents for the implementation of performance assessment in the schools with respect to three themes of *assessment methods*, *evaluation criteria*, and *reasons for selecting the assessment methods and criteria*.

6.1.2. Key Findings

With respect to the first research question, assessment methods and evaluation criteria were examined in view of the concept of assessment *for* learning. As the process-centered performance assessment presented in the 2015 revised national curriculum maintains, *process* is the key factor allowing assessment to promote

learning. The reason why process is important is that *evidence* of students' learning is collected in the assessment process. In other words, in order to promote students' learning, during the assessment process of assessment, teachers should seek and interpret the evidence of students learning and decide where the students are in their learning, where they need to go, and how they can reach that destination. In the assessment process, students should actively involve themselves in their assessment, while what they say and do should be observed and judged by teachers in consideration of how students' learning can be improved. Teachers should provide students with constructive feedback about their strengths and weaknesses and how they can be addressed.

In addition, to promote students' learning, it is essential that students understand their learning goals. In this respect, criteria are a good tool for describing what students need to achieve. Teachers should thus share learning goals with students by providing and clearly explaining the criteria with concrete examples. This means that teachers should also provide an opportunity for students to engage in peer and self-assessment, which are the key skills allowing to students to take responsibility for their learning.

However, the findings of the study regarding the assessment methods and criteria selected by the teachers of Schools X and Y demonstrate the implementation of performance assessment in a manner contradicting the principles of assessment *for* learning. They conducted writing, speaking, and listening assessments without actually teaching these skills or providing feedback during or after the assessments. In addition, the assessment methods for the

speaking and listening assessments were inauthentic, having little to do with a real-life situation. The assessment tasks used for the speaking and class participation assessments did not utilize students' higher-order thinking skills, understanding, problem solving ability, or creativity but were rather based on memorization. Furthermore, as with the class participation assessment, the teachers provided grades as rewards for students' completion and submission of their work. In this regard, as Wiliam (2010) points out, this only leads students to seek out grades while avoiding more difficult tasks, hampering their reflection on their learning.

The teachers' selection of evaluation criteria also contradicted the principles of assessment *for* learning. With regard to the essay writing and speaking assessments, the teachers assigned a heavier grade weight to task completion with the intention of judging the students' performances objectively, excluding criteria requiring subjective judgment. Altogether, the teachers assigned heavier grade weights to the assessments focusing on students' achievements, rather than effort, prioritizing ranking the students over promoting their learning. Furthermore, the teachers did not share or clearly explain the criteria using concrete examples, ignoring another important principle of assessment *for* learning (Assessment Reform Group, 2002; Black, 2001; Black & Wiliam 1998; Wiliam, 2001).

The teachers' selection of assessment methods and evaluation criteria seemed closely related with their perceptions of the reasons for carrying out the assessments. As uncovered through the interview data, the teachers perceived the reasons for carrying out the assessments as "to comply with government policy," "to grade and report performance," "to fill out students' educational profiles," "to

provide students an opportunity to practice speaking," and "to encourage students' active class participation." These perceived reasons led the teachers to set the total grade weight of performance assessments at at least 40%, include a mandatory essay assessment accounting for 35% of a students' semester grade, and prioritize objectivity, fairness, consistency, and accuracy in assessment and administration. With respect to evaluation criteria, the perceived reasons drove the teachers to exclude criteria requiring subjective judgment and emphasize the grammar and vocabulary criterion, which helps to identify and distinguish high-proficiency students. They adopted essay writing themes and tasks they considered would be advantageous for students' university applications and adopted various tasks to record diverse information in the students' educational profiles. They also implemented a memorized speaking performance under the conditions of a one-to-one (teachers-student) situation and time limit. Furthermore, even though they emphasized effort over achievement in the class participation assessment, they assigned a lower grade weight to the class participation assessment because it could not contribute to ranking the students.

As mentioned above, the teachers' perceived reasons for the assessments and decision making in selecting assessment methods and evaluation criteria were significantly affected by bureaucratic pressure, consisting of government policy and reporting requirements. Overall, such bureaucratic pressure was found to be a negative influence on the implementation of assessment *for* learning. As shown in the findings, reporting requirements caused the teachers to prioritize objectivity, fairness, and accuracy in administering assessments and to regard their role as that

of a rater rather than supporter of students' learning, leading them to yield formative to summative purposes in implementing assessment.

In short, the findings of the present study show that while the objective of the 2015 revised national curriculum is to implement assessment *for* learning, because this is a national policy that involves reporting requirements, in practice, policies and reporting requirements tend to outweigh assessment *for* learning in teachers' decision making when it comes to actually implementing the assessment. This means that the requirements of the 2015 revised national curriculum may actually be inhibiting the fulfillment of its core objective of the implementation of assessment *for* learning.

Furthermore, the findings of the present study corroborate the premise that teachers' decision making is deeply affected by considerations of university admission and external high-stakes assessment. The teachers were concerned with filling educational profiles with advantageous information for university admission and focus most of their teaching on reading, preparing students for external summative testing, which ultimately affects their decision making in selecting assessment methods and evaluation criteria. This finding aligns with those of previous studies exploring exam-centric Confucian-heritage societies such as China, Hong Kong, and Taiwan, where classroom-based teacher assessment is usually used as preparation for external exams and teachers have difficulty in implementing assessment *for* learning as it is intended (e.g., Carless, 2011; Kennedy & Lee, 2008; Kennedy et al., 2008). In this respect, the findings of the present study point to the significance of sociocultural context for the

implementation of assessment *for* learning: Sociocultural factors shape the bureaucratic pressures influencing teachers' perceptions of the reasons for implementing assessment.

In conclusion, the present study found that, contrary to the objective of the 2015 revised national curriculum, which is to implement assessment *for* learning, the teachers perceived and carried out the performance assessments for administrative, accountability, and reporting purposes. In addition, it was found that the teachers' decision making in planning the assessments was significantly affected by sociocultural contextual values, which manifested in the form of bureaucratic pressures, namely, government policy, reporting requirements, university admission, and external, high-stakes assessment.

6.2. Pedagogical Implications

In the present study, the data showed that the teachers' selection of assessment methods and evaluation criteria did not accord with the principles of assessment *for* learning. They viewed performance assessment as having reporting and administrative purposes, and this perception significantly impacted their selection of assessment methods and evaluation criteria. These perceptions appeared to be rooted in exam-centric sociocultural values.

With respect to the findings, there are several points that have important implications for the improvement of performance assessments in Korean educational setting. First, interpreting the teachers' implementation of

performance assessment in light of the concept assessment *for* learning, evidence of assessment *for* learning was hardly detected. In other words, the important characteristics of assessment *for* learning—assessment embedded in teaching and learning, sharing learning goals or standards with students, students' involvement in peer and self-assessment, and providing constructive feedback—were rarely intelligible. The English teachers' perceived performance assessment in terms of administrative, grading, and reporting purposes, which substantially affected their selection of assessment methods and evaluation criteria. Besides the fact that the teachers did not implement assessment *for* learning, what is even more remarkable is that they conducted the assessments without actually teaching the relevant skills. Moreover, they did not appear to view this as problematic.

Second, the teachers' determinations of the purposes of the performance assessments were primarily affected by government policy and reporting requirements. As is often pointed out in previous research, national curriculums offer key stakeholders in the field of education advantages such as clear reporting requirements for communicating information on students' performance (S. Clarke & Gipps, 2000). However, they also tend to overlook pedagogical practice (Arkoudis & O'Loughlin, 2004; Leung & Rea-Dickins, 2007). This is also true for the 2015 revised national curriculum, which emphasizes assessment *for* learning (requiring teachers to focus on improving students' learning and teaching practice) but also requires teachers to report the products of students' performances. This means that the policy simultaneously emphasizes both summative as well as formative assessment. As shown in previous research, it is difficult to implement

assessment *for* learning and administrative purpose at the same time because they are often at cross purposes. As long as the purpose of *grading and reporting* is imposed on teachers, then, they might continue to prioritize summative over formative assessment.

Another problem with the guidelines for the performance assessment under the 2015 curriculum has to do with decontextualized requirements failing to consider specific conditions. In order to implement assessment *for* learning, teachers need to implement different assessment tasks according to their specific teaching context. Nonetheless, the guidelines of the Gyeonggido Office of Education (2019) affiliated with the 2015 revised national curriculum require teachers to implement essay writing assessment worth 35% of a student's grade and performance assessment worth at least 40% of a student's grade. With respect to the current study, this policy appeared not to account for the specific conditions in which assessments are carried out, restricting the teachers' authority over their teaching and implementation of performance assessment. This means that the teachers acted as *policy enforcers* (Leung & Rea-Dickins, 2007) rather than as teachers per se. Consequently, they appeared constrained in gaining meaningful and accurate information on teaching and learning that might lend support to their students' English learning development.

In addition to the government policy and reporting requirements, the teachers' decision making was also significantly affected by university admission and external high-stakes assessment. This means that exam-centric Confucian-heritage sociocultural values inhibited the practice of assessment *for* learning in

the classrooms (e.g., Brookhart & Durkin, 2003; G. T. Brown et al., 2009; Noor et al., 2010).

With these implications in mind, the study suggests three broad tracks for improving the implementation of assessment *for* learning in Korean classrooms with respect to three kinds of stakeholders in the field of English-language education: policymakers, assessment researchers, and teachers.

First, policymakers should carefully reconsider existing education policy. They need to develop a more specific model of assessment *for* learning that teachers can actually implement in classrooms while more actively informing teachers of the intended meaning of assessment *for* learning. The findings of the study indicate that the teachers were aware that the 2015 revised national curriculum emphasizes process-centered performance assessment. However, their responses in the interviews provided no evidence they were focusing on *process*, which enhances their teaching and learning. This may have been a consequence of their ignorance regarding how to practice process-centered performance assessment, which may lead them to fall back on more familiar mandatory guidelines and reporting requirements. To remedy this situation, policymakers can devise specialized teacher training for implementing process-centered assessment as assessment *for* learning. Furthermore, they may give teachers opportunities to monitor the quality of assessment procedures conducted in a variety of teaching contexts, leading them to create their own assessment tasks.

In addition, in terms of bureaucratic pressures, policymakers need to keep in mind the incompatibility of assessment *for* learning, the purported aim of the

policy, with administrative purposes. When teachers are preoccupied with reporting results, their focus is no longer on improving or facilitating learning but being objective and accurate, acting as *rater* rather than *facilitator* of learning, as was also the case in the present study. Reporting requirements should thus be de-emphasized in the policy. If there is no way around implementing mandatory standards and reporting requirements, a more effective policy would grant greater authority and flexibility to teachers, allowing them to adapt to the requirements based on their particular classroom conditions, rather than the current practice of stipulating a set minimum grade weight of 40% for performance assessment and 35% for essay writing assessment. This could eventually lead to more meaningful assessment *for* learning, improving in-class teaching and learning.

Furthermore, policymakers should consider the need for flexibility in teachers' assessment planning. As was the case in the present study, teachers normally plan performance assessment at the beginning of the semester and there is little opportunity to change this plan over the course of the semester. However, in assessment *for* learning, teachers' assessment plans should be adjusted in response to emerging ideas and skills (Assessment Reform Group, 2002). Also, planning should include information regarding how teachers will inform students of learning goals and criteria applied in the assessment and how they will provide *feedback* to students to help them progress. Neither was this dynamic present among the teachers' implementation of assessment in this study.

Second, assessment researchers need to pay attention to sociocultural context, suggesting reasonable expectations for policymakers regarding the

successful practice of assessment *for* learning. The prospects for this type of assessment are simply less than promising in the Korean EFL context due to a deeply embedded emphasis on summative assessment. While this does not mean giving up on assessment *for* learning, it does suggest the need to pursue a pragmatically formulated education policy. As Broadfoot (1999) remarks, "We need to recognize assessment as a social product in which the values and traditions of particular cultures and interests of specific groups within them combine to produce particular definitions of quality or merit" (p. 7). A pragmatic approach to assessment *for* learning, then, could actually entail implementation in a manner not entirely consistent with the Western or international literature. This means adjusting assessment *for* learning to the Korean cultural context, considering the priorities, expectations, and needs of students, teachers, and parents.

Previous research has explored the implementation of such a pragmatic approach in exam-centric cultures (e.g., Black et al., 2003; Carless, 2011; Carless & Lam, 2014b; Davison & Leung, 2009). For example, Broadfoot and Black (2004) argue that in an exam-centric culture where teachers and parents tend to view summative assessment positively and are reluctant to move away from it, the implementation of assessment *for* learning hinges on a positive synergy between formative and summative assessment. In practice, this means the use of summative tests in assessment *for* learning. In this respect, Black et al. (2003) coined the phrase "the formative use of summative tests" (p. 2). Another strategy, as outlined by Carless (2011), could be "test follow-up," where formative assessment is implemented after conducting summative tests to inform students of learning

problems, providing feedback according to the students' achievement levels, and recommending activities to facilitate students' learning. Meanwhile, Biggs (1996) discovered that in China, even though teachers are reluctant to move away from summative assessment preparing students for public examinations, if teachers employ summative assessment in a productive manner without excessively focusing on grading, helping students understand their learning by highlighting learning problems and providing learning strategies to overcome them, summative assessment can actually serve the same purpose as assessment *for* learning. Sitggins (2002) also found that in Confucian-heritage cultural contexts, such as Hong Kong, teachers of English can use dictation, typically regarded as a summative form of assessment, as a form of assessment *for* learning by providing *feedback* facilitating students' learning. In this sense, as Stiggins (2002) puts it, whether or not an assessment is formative or summative depends on a teacher's *mindset*. Such research suggests the important role that assessment researchers have to play in conceptualizing and formulating ways for policymakers to reconcile assessment *for* learning to the Korean EFL context.

Lastly, teachers also need to make more effort in implementing assessment *for* learning in classrooms. In the present study, the teachers certainly believed that all four skills of reading, listening, speaking, and writing should be assessed, meaning that they were well aware they should teach all four skills. Nonetheless, their actual teaching mostly consisted of reading. The teachers stated that this was because external high-stakes assessment (KSAT) consists of reading questions. Strictly speaking, however, the national curriculum states that teachers should

teach all four skills. The teachers' focus on reading was perhaps most apparent in the peculiar situation where they assessed speaking and writing without actually teaching them. In theory, assessment should only be administered with respect to something that has been taught. Indeed, the national curriculum states that teachers should assess what they teach.

The teachers also claimed that the national curriculum does not consider their specific teaching and learning conditions. However, even though the national curriculum specifies grade weights and requires essay writing, individual teachers can devise their own assessments in accordance with their students' proficiency or other conditions. In addition, although the School Y teachers maintained that policy requirements constrained their consideration of students' proficiency, in one respect, they did in fact adapt to the requirements in accordance with their students' proficiency by inserting short-answer tests into the midterm and final exams that served as essay assessment. It thus seemed that the teachers in the present study did not sufficiently understand or follow the basic guidelines of the national curriculum, instead fixating on the mandatory guidelines and reporting requirements. In this respect, the study suggested the need for teachers, just as much as policymakers and assessment researchers, to reflect on their performance assessment practices and endeavor to more comprehensively fulfill the basic guidelines for performance assessment provided in the national curriculum toward ultimately realizing assessment *for* learning.

6.3. Limitations and Suggestions for Further Research

The present study has provided insight into how performance assessment is implemented in Korean high school classrooms by examining the decision making Korean high school teachers of English in selecting assessment methods and evaluation criteria in planning performance assessment at two schools. The study found that the teachers' practice did not reflect the characteristics of assessment *for* learning. The teachers perceived the purposes of performance assessment in terms of administrative, grading, and reporting purposes, which affected their selection of assessment methods and criteria. The study interpreted these perceptions as primarily influenced by sociocultural contextual values.

Although these are meaningful findings that can help to improve the implementation of assessment *for* learning in the Korean context, the study nonetheless demonstrated several limitations. First, the investigation was limited to the planning stage of the assessments. Although rich data was attained through both individual and group interviews, each progressing for at least 90 minutes, this data did not extend to the entire process of the implementation of performance assessment, especially regarding *feedback*, which is a significant aspect of assessment *for* learning. The teachers explained how they provide feedback in the interviews, but this was insufficient. Thus, future studies might directly observe the entire process of the implementation of performance assessment. Finally, another limitation was the small number of participants (five from two different schools). Future research could incorporate a larger number of teachers across a

175

larger number of schools, allowing for more generalizable conclusions.

In spite of these limitations, the present study provides insight into how performance assessment is practiced in Korean EFL high school classrooms, how Korean high school teachers of English at large may perceive the purposes of performance assessment, and how this influences their selection of assessment methods and criteria when planning assessment. Above all, by suggesting a dissonance between the aim of the 2015 revised national curriculum and the way performance assessment is implemented in Korean high school classrooms, the current study is significant insofar as it alludes to the need for more in-depth research on how Korean high school teachers of English are implementing assessment *for* learning in classrooms. Policymakers and teachers as well as researchers need to consider the sociocultural values influencing teachers' perceptions of the purposes of assessment. Further knowledge of these processes may contribute to devising more effective performance assessment aimed at assessment *for* learning in Korean EFL high school settings.

REFERENCES

Abedi, J. (2004). The no child left behind act and English language learners: Assessment and accountability issues. *Educational researcher*, *33*(1), 4-14.

Akiyama, T. (2004). Introducing speaking tests into a Japanese senior high school entrance examination. Unpublished doctoral dissertation, University of Melbourne, Australia .

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Alemu, G., Stevens, B., Ross, P., & Chandler, J. (2017). The use of a constructivist grounded theory method to explore the role of socially-constructed metadata (Web 2.0) approaches. *Qualitative and Quantitative Methods in Libraries*, *4*(3), 517-540.

Allen, D. (1998). Introduction. In D. Allen (Ed.), *Assessing student learning: From grading to understanding* (pp. 1-17). New York: Teachers College Press.

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology*, *80*(3), 260-267.

Andrade, H. L. (2013). Classroom assessment in the context of learning theory and research. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 17-34). Thousand Oaks, CA: Sage.

Andrews, M., Brown, R., & Mesher, L. (2018). Engaging students with assessment

and feedback: Improving satisfaction and attainment with students as co-designers. *Practitioner Research in Higher Education*, *11*(1), 32-46.

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.

Aschbacher, P. R. (1991). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education*, *4*(4), 275-288.

Assessment Reform Group (1999). *Assessment for Learning: Beyond the Black Box [Pamphlet]*. Cambridge: University of Cambridge.

Assessment Reform Group (2002). *Assessment for Learning: 10 principles*. Cambridge: University of Cambridge.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford university press.

Ban, J. C., Kim, S., Park, C., & Kim, H. K. (2018). Teachers' perceptions on the teacher by teacher process fortified assessment policy. *Journal of Curriculum Evaluation*, *21*(3), 105-130.

Benmansour, N. (1999). Motivational orientations, self-efficacy, anxiety and strategy use in learning high school mathematics in Morocco. *Mediterranean Journal Educational Studies*, *4*, 1-15.

Biggs, J. (1996). Western misperceptions of the Confucian-heritage learning culture. In D. A. Watkins & J. B. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 45–67). Hong Kong & Melbourne: CERC & ACER.

Biggs, J. (1998). Assessment and classroom learning: A role for formative assessment*? Assessment in Education: Principles, Policy and Practice, 5*(1), 103–110.

Bishop, J. H. (1997). The effect of national standards and curriculum-based exams on achievement. *The American Economic Review*, *87*(2), 260-264.

Black, P. (2001). Formative assessment and curriculum consequences. In D. Scott (Ed.), *Curriculum and assessment* (Vol. 1) (pp. 7-24). Westport, Washington, DC: Greenwood Publishing Group.

Black, P. (2010). *Assessment of and for Learning: Improving the quality and achieving a positive interaction.* Paper presented to the meeting of representatives of the EU education ministers. Brussels: European Union.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74.

Black, P., & Wiliam, D. (2002). *Standards in public examinations*. London: King's College.

Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education*, *103*(2), 20-50.

Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *Curriculum Journal*, *16*(2), 249-261.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability (formerly: Journal*

of Personnel Evaluation in Education), *21*(1), 5.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81-90.

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, UK: Open University Press.

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., Burkhardt, H., Daro, P., Jones, I., Lappan, G., Pead, D., & Stephens, M. (2012). High-stakes examinations to support policy. *Educational Designer*, *2*(5), 1-32.

Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: a multidimensional perspective. *Learning and instruction*, *13*(3), 327-347.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, *3*(2), 77-101.

Braun, V., & Clarke, V. (2012). Thematic analysis. In H. Cooper (Ed.), *APA Handbook of Research Methods*: *Vol. 2, Research Designs* (p. 57-71). Washington, DC: American Psychological Association.

Breen, M. P., Barratt-Pugh, C., Derewianka, B., House, H., Hudson, C., Lumley, T., & Rohl, M. (1997). *Profiling ESL children: How teachers interpret and use national and state assessment frameworks (Vol. 1): Key issues & findings*. Canberra, AU: Commonwealth Department of Employment,

Education, Training and Youth Affairs.

Brindley, G. (1989). *Assessing achievement in the learner-centered curriculum.* Sydney, Australia: Macquarie University, National Center for English Language Teaching and Research.

Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programmes: A review of the issues. *Language Testing*, *15*(1), 45-85.

Brindley, G. (2001). Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing*, *18*(4), 393-407.

Brindley, G. (2002). Issues in language assessment. In R.B. Kaplan (Ed.), *The Oxford handbook of applied linguistics* (pp. 459-470). New York: Oxford University Press.

Brindley, G. (2007). Editorial. *Language Assessment Quarterly*, *4*(1), 1-5.

Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice, 11*(1), 7–26. doi:10.1080/0969594042000208976.

Broadfoot, P., Osborn, M., Sharpe, K., & Planel, C. (2001). Pupil assessment and classroom culture: A comparative study of the language of assessment in England and France. In D. Scott (Ed.), *Curriculum and assessment* (Vol. 1) (pp. 41-62). Westport, Washington, DC: Greenwood Publishing Group.

Brookhart, S. M. (2001). *The "Standards" and Classroom Assessment Research*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, Texas.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5-12.

Brown, A. L., & Ferrara, R. A. (1985). Diagnosing zones of proximal development. In J. V. Wertsch (Ed.), *Culture, communication, and cognition: Vygotskian perspectives* (p. 273-305). Cambridge: Cambridge University Press.

Brown, G. T. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice*, *11*(3), 301-318.

Brown, G. T. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, *3*, 45-70.

Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 347-363.

Brown, G. T., Lake, R., & Matters, G. (2008). New Zealand and Queensland teachers' conceptions of learning: Transforming more than reproducing. *Australian Journal of Educational & Developmental Psychology*, *8*, 1-14.

Brown, G. T., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education*, *27*(1), 210-220.

Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, *32*(4), 653-675.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1-14.

Calderhead, J. (1996). Teachers: Beliefs and knowledge. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 709–725). New York: macmillan.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, *1*(1), 1-47.

Carless, D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education: Principles, Policy & Practice*, *12*(1), 39-54.

Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. London: Routledge.

Carless, D., & Lam, R. (2014a). Developing assessment for productive learning in Confucian-influenced settings: Potentials and Challenges. In C. Wyatt-Smith, V. Klenowski, & P. Colbert (Eds.), *Designing assessment for quality learning* (pp. 167-182). Dordrecht, NL: Springer.

Carless, D., & Lam, R. (2014b). The examined life: Perspectives of lower primary school students in Hong Kong. *Education 3-13: International Journal of Primary*, *Elementary and Early Years of Education*, *42*(3), 313-329.

Carter, N., Bryant-Lukosius, D., DiCenso, A., Blythe, J., & Neville, A. (2014). The use of triangulation in qualitative research. *Oncol Nurs Forum, 41*(5), 545-547.

Chang, K.S., Lee, B. C., Jung, C. K., Hwang, J. B., Kim., K. J., Lee, H. J., Kim, H. M., Roh, Y. K., Jung, K. C., & Lim, N.G. (2017). A study of developing the assessment standards of the 2015 Revised National English Curriculum at the High School Level (Report No. CRC 2017-5-13). Seoul: Korea Institute for Curriculum and Evaluation.

Cheah, Y. M. (1998). The examination culture and its impact on literacy innovations: The case of Singapore. *Language and Education*, *12*(3), 192-209.

Chen, Q., Kettle, M., Klenowski, V., & May, L. (2013). Interpretations of formative assessment in the teaching of English at two Chinese universities: A sociocultural perspective. *Assessment & Evaluation in Higher Education*, *38*(7), 831-846.

Cheng, L. (1999). Changing assessment: Washback on teacher perceptions and actions. *Teaching and Teacher Education*, *15*(3), 253-271.

Cheng, L. (2004). The washback effect of a public examination change on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing* (pp. 147-170). Mahwah, NJ: Routledge

Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, *21*(3), 360-389.

Citizenship and Immigration Canada (1993). *Language benchmarks: English as a second language for adults*. Ottawa: Citizenship and Integration Policy

Division, Citizenship and Immigration Canada.

Cizek, G. J. (1996). Learning, achievement, and assessment: Constructs at a crossroads. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, Achievement, and Adjustment* (pp. 1-32). San Diego: Academic Press.

Cizek, G. J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice*, *19*(2), 16.

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, *20*(4), 19-27.

Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.) (pp. 255-296). London: Collier Macmillan.

Clark, J. L. (1978). *Direct testing of speaking proficiency: Theory and application*. Princeton, NJ: Educational Testing Service.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947-967.

Clarke, S. (1998). *Targeting assessment in the primary classroom.* London: Hodder & Stoughton.

Clarke, S., & Gipps, C. (2000). The role of teachers in teacher assessment in England 1996-1998. *Evaluation & Research in Education*, *14*(1), 38-52.

Claxton, R. (1995). Pupils as learners. In R. Fairbrother, P. J. Black, & P. Gill (Eds.), *Teachers assessing pupils: Lessons from science classrooms* (p. 105-120). Hatfield, UK: Association for Science Education.

Creswell, J. W. (2007). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. London: Sage.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*(4), 438-481.

Cumming, A. (2001). ESL/EFL instructors' practices for writing assessment: Specific purposes or general purposes? *Language Testing*, *18*(2), 207-224.

Davies, A. (1997). Special issue: Ethics in language testing. *Language Testing*, *14.*

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.

Davison, C. (2008). *Assessment for learning: Building inquiry-oriented assessment communities*. New York: Routledge.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, *43*(3), 393-415.

Day, C. (2002). *Developing teachers: The challenges of lifelong learning*. London: Routledge.

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*(1), 1-27.

DeLuca, C., & Johnson, S. (2017). Developing assessment capable teachers in this age of accountability. *Assessment in Education: Principles, Policy &*

Practice, *24*(2), 121-126.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*(10), 1040-1048.

Dweck, C. S. (2002). Messages that motivate. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 38–61). San Diego: Elsevier.

Dwyer, C. A. (1998). Assessment and classroom learning: Theory and practice. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 131-137.

English, F. W. (1992). *Deciding what to teach and test: Developing, aligning, and auditing the curriculum*. Newbury Park, CA: Corwin Press.

Evans, B., & Hornberger, N. H. (2005). No child left behind: Repealing and unpeeling federal language education policy in the United States. *Language Policy*, *4,* 87–106.

Fensham, P. J., Gunstone, R. F., & White, R. T. (Eds.). (1994). *The content of science: A constructivist approach to its teaching and learning*. London and Washington, D.C.: The Falmer Press.

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, *18*(9), 27-32.

Frederiksen, J. R., & White, B. J. (1997). *Reflective assessment of students' research within an inquiry-based middle school science curriculum.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Freeman, D. (1989). Teacher training, development, and decision making: A

model of teaching and related strategies for language teacher education. *TESOL Quarterly*, *23*(1), 27-45.

Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. A perspective from North American educational research on teacher education in English language teaching. *Language Teaching*, *35*(1), 1-13.

Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, *18*(3), 402-423.

Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, *1,* 253–266.

Fullan, M. G. (1999). *Change forces: The sequel*. London: Falmer.

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.

Gipps, C. (1994). *1994: Beyond testing: Towards a theory of educational measurement.* London: Falmer Press.

Gyeonggido Office of Education (2019). 2019 Hangnyeon-do godeung hakkyo hageop seongjeok gwalli sihaeng jichim. (GOE Publication No. 2019-3).

Hagan, P., Hood, S., Jackson, E., Jones, M., Joyce, H., & Manidis, M. (n.d.).

Certificates in spoken and written English I-Ii, Iii, Iv (3rd ed.) (Vols. 1-3). Sydney: NSW Adult Migrant English Service.

Hall, K., & Harding, A. (2002). Level descriptions and teacher assessment in England: Towards a community of assessment practice. *Educational Research, 44*(1), 1-15.

Hall, K., Webber, B., Varley, S., Young, V., & Dormant, P. (1997). A study of teacher assessment at Key Stage 1. *Cambridge Journal of Education*, *27*(1), 107-122.

Hamp-Lyons, L. (2007). The impact of testing practices on teaching. In J. Cummings & C. Davison (Eds.), *International handbook of English language teaching* (pp. 487-504). Boston, MA: Springer.

Hamp-Lyons, L., & Lynch, B. K. (1998). Perspectives on validity: A historical analysis of language testing conference abstracts. In A. J. Kunnan (Ed.), *Validation in language assessment* (pp. 267-290). Mahwah, NJ: Lawrence Erlbaum Associates.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, *39*(1), 69-95.

Hargreaves, E. (2005). Assessment for learning? Thinking outside the (black) box. *Cambridge Journal of Education*, *35*(2), 213-224.

Harlen, W. (2006). The role of assessment in developing motivation for learning. *Assessment and Learning*, 61-80.

Harlen, W. (2007). *Assessment of learning*. London: Sage.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning.

Assessment in Education: Principles, Policy & Practice, *10*(2), 169-207.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365-379.

Harlen, W., & Winter, J. (2004). The development of assessment for learning: Learning from the case of science and mathematics. *Language Testing*, *21*(3), 390-408.

Harper, M., & Cole, P. (2012). Member checking: Can benefits be gained similar to group therapy. *The Qualitative Report*, *17*(2), 510-517.

Hawthorne, L. (1997). The political dimension of English language testing in Australia. *Language Testing*, *14*(3), 248-260.

Herman, J. L. (2007). *Accountability and assessment: Is public interest in K-12 education being served?* (CRESST Report 728). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J. L., Aschbacher, P.R., & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Hill, K., & McNamara, T. (2012). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, *29*(3), 395-420.

Huerta-Macías, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5*(1), 8-11.

Hume, A., & Coll, R. K. (2009). Assessment of learning, for learning, and as

learning: New Zealand case studies. *Assessment in Education: Principles, Policy & Practice*, *16*(3), 269-290.

Inbar-Lourie, O., & Donitsa-Schmidt, S. (2009). Exploring classroom assessment practices: The case of teachers of English as a foreign language. *Assessment in Education: Principles, Policy & Practice*, *16*(2), 185-204.

Iran-Nejad, A. (1995). Constructivism as substitute for memorization in learning: Meaning is created by learner. *Education*, *116*(1), 16-32.

Jones, R. L. (1977). *Testing: A vital connection. The language connection: From the classroom to the world*. ACTFL Foreign Language Education Series (Vol. 9). New York: American Council on the Teaching of Foreign Languages.

Kennedy, K. J. (2007). *Barriers to innovative school practice: A socio-cultural framework for understanding assessment practices in Asia.* Paper Presented for the Symposium: "Student Assessment and Social and Cultural Contexts: How Teachers Respond to Assessment Reforms." Redesigning Pedagogy, Culture, Understanding and Practice Conference, Singapore.

Kennedy, K. J., & Lee, J. (2008). *Changing schools in Asia: Schools for the Knowledge society.* London: Routledge.

Kennedy, K. J., Chan, J. K. S., Fok, P. K., & Yu, W. M. (2008). Forms of assessment and their potential for enhancing learning: Conceptual and cultural issues. *Educational Research for Policy and Practice*, *7*(3), 197.

Khattri, N. & Sweet, D. (1996). Assessment reform: Promises and challenges. In M. B. Kane and R. Mitchell (Eds.), *Implementing performance assessment:*

Promises, problems and challenges (pp. 1-21). Mahwah, NJ: L. Erlbaum.

Kirton, A., Hallam, S., Peffers, J., Robertson, P., & Stobart, G. (2007). Revolution, evolution or a Trojan horse? Piloting assessment for learning in some Scottish primary schools. *British Educational Research Journal*, *33*(4), 605-627.

Kitsantas, A., Robert, A. R., & Doster, J. (2004). Developing self-regulated learners: Goal setting, self-evaluation, and organizational signals during acquisition of procedural skills. *The Journal of Experimental Education*, *72*(4), 269-287.

Klenowski, V. (1995). Student self-evaluation process in student-centered teaching and learning contexts of Australia and England. *Assessment in Education, 2,* 145-163.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254.

Knight, P. (2002). A systemic approach to professional development: Learning as practice. *Teaching and Teacher Education*, *18*(3), 229-241.

Ko, H. (2019). The study on the perception, actual condition, and support strategies of process-centered assessment by each teacher. *Journal of Learner-Centered Curriculum and Instruction*, *19*(9), 1137-1164.

KICE (Korea Institute for Curriculum and Evaluation) (2017). A study of Developing the Assessment Standards of the 2015 Revised National English Curriculum at the High School Level (Report No. CRC 2017-5-13).

KICE (Korea Institute for Curriculum and Evaluation) (2018). Development and Administration of Training Program for Process-Fortified Assessment by Teacher (Report No. CRE 2018-4).

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis. *Journal of Personality and Social Psychology*, *28*(1), 129-137.

Leung, C. (2004). Developing formative teacher assessment: Knowledge, practice, and change. *Language Assessment Quarterly: An International Journal*, *1*(1), 19-41.

Leung, C. (2005). Classroom teacher assessment of second language development: Construct as practice. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 893-912). New York: Routledge.

Leung, C., & Mohan, B. (2004). Teacher formative assessment and talk in classroom contexts: Assessment as discourse and assessment of discourse. *Language Testing*, *21*(3), 335-359.

Leung, C., & Rea-Dickins, P. (2007). Teacher assessment as policy instrument: Contradictions and capacities. *Language Assessment Quarterly*, *4*(1), 6-36.

Leung, C., & Scott, C. (2009). Formative assessment in language education policies: Emerging lessons from Wales and Scotland. *Annual Review of Applied Linguistics*, *29*, 64.

Leung, C., & Teasdale, T. (1997). What do teachers mean by speaking and listening: A contextualized study of assessment in the English National Curriculum. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.),

New contexts, goals and alternatives in language assessment (pp. 291-324). Jyväskyä, Finland: University of Jyväskyä.

Li, J. (2010). Learning to self-perfect: Chinese beliefs about learning. In C. K. K. Chan & N. Rao (Eds.), *Revisiting the Chinese learner: Changing context, changing education* (pp. 35-69). Dordrecht, NL: Springer.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.

Lynch, B. K. (1997). In search of the ethical test. *Language Testing*, *14*(3), 315-327.

Lynch, B. K., & Hamp-Lyons, L. (1996). *Positivistic versus alternative perspectives on validity within the LTRC*. Paper presented at the annual Language Testing Research Colloquium, Tampere, Finland.

Marshall, B., & Jane Drummond, M. (2006). How teachers engage with assessment for learning: Lessons from the classroom. *Research Papers in Education*, *21*(2), 133-149.

McKay, P. (2000). On ESL standards for school-age learners. *Language Testing*, *17*(2), 185-214.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice*, *22*(4), 34-43.

McMillan, J. H. (Ed.). (2013). *SAGE handbook of research on classroom assessment*. Los Angeles: Sage.

McMillan, J. H., & Nash, S. (2000). *Teacher classroom assessment and grading*

practices decision making. Paper presented at the Annual Meeting of the
National Council of Measurement in Education, New Orleans.

McNamara, T. (1998). Policy and social considerations in language assessment.
Annual Review of Applied Linguistics, *18*, 304-319.

McNamara, T. (2001). Language assessment as social practice: Challenges for
research. *Language Testing*, *18*(4), 333-349.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd
ed.) (pp. 13-103). New York: American Council on Education.

MOE & KICE (Ministry of Education & Korea Institute for Curriculum and
Evaluation) (2017). Gwajeong eul jungsi haneun suhaeng pyeongga
eotteokke halkkayo? (Report No. ORM 2017-19-2).

Mohan, B. (1998). Knowledge structures in oral proficiency interviews for
international teaching assistants. In R. Young & H. A. Weiyun (Eds.),
*Talking and testing: Discourse approaches to the assessment of oral
proficiency* (pp. 173-204). Amsterdam: John Benjamins.

Moore, H. M. (2005). Identifying "the target population": A genealogy of policy-
making for English as a Second Language (ESL) in Australian schools
(1947–1997). Unpublished doctoral dissertation, Ontario Institute for
Studies in Education, University of Toronto, Canada.

Morris, P. J. T. F., Fung-Lo, M. L., Chik, P. M. P., & Chan, K. K. (2000). One
function, two systems: Changing assessment in Hong Kong's primary
schools. In B. Adamson, T. Kwan, & K. K. Chan (Eds.), *Changing the
curriculum: The impact of reform on primary schooling in Hong Kong* (pp.

195-215). Hong Kong: Hong Kong University Press.

Morrison, K., & Joan, T. F. H. (2002). Testing to destruction: A problem in a small state. *Assessment in Education: Principles, Policy & Practice*, *9*(3), 289-317.

Morrow, K. (1977). *Techniques of evaluation for a notional syllabus*. London: Royal Society of Arts.

Noor, N. B. M., Muniandy, M. K., Shanmugam, S. K. K., & Mathai, E. J. (2010). Upper primary teachers' perceptions of PSLE English oral assessment. *English Language Teaching*, *3*, 142-151.

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. Honolulu, HI: University of Hawai'i Press.

Pennycook, A. (1994). Incommensurable discourses? *Applied Linguistics*, *15*(2), 115-138.

Perfecto, M. R. G. (2012). Contextual factors in teacher decision making: Extending the woods model. *Asia-Pacific Education Researcher (De La Salle University Manila)*, *21*(3), 474-483.

Perrenoud, P. (1991). Towards a pragmatic approach to formative evaluation. *Assessment of Pupils Achievement: Motivation and School Success*, 79-101.

Perrin, M. (1991). Summative evaluation and pupil motivation. In P. Weston (Ed.), *Assessment of pupil achievement: Motivation and school success* (pp. 169-173). Amsterdam, NL: Swets and Zeitlinger.

Pilcher-Carlton, J., & Oosterhof, A. C. (1993). *A case study analysis of parents',*

Teachers' and students' perceptions of the meaning of grades: Identification of discrepancies, their consequences, and obstacles to their resolution. Paper presented at the Annual meeting of the American Educational Research Association, Atlanta, GA.

Pollard, A., Triggs, P., Broadfoot, P., McNess, E., & Osborn, M. (2000). *What pupils say: Changing policy and practice in English primary education.* London: Cassell.

Pryor, J., & Torrance, H. (1996). Teacher-pupil interaction in formative assessment: Assessing the work or protecting the child? *The Curriculum Journal*, *7*(2), 205-226.

Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, *18*(4), 429-462.

Rea-Dickins, P. (2002). Exploring the educational potential of assessment with reference to learners with English as an additional language. In C. Leung (Ed.), *Language and additional/second language issues for school education: A reader for teachers* (pp. 81-93). York: York Publishing for NALDIC.

Rea-Dickins, P. (2004). Editorial. Understanding teachers as agents of assessment. *Language Testing, 21*(3), 249-258.

Rea-Dickins, P. (2006). Currents and eddies in the discourse of assessment: A learning-focused interpretation 1. *International Journal of Applied Linguistics*, *16*(2), 163-188.

Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': Structure, agency and the

construction of identity through assessment. *British Educational Research Journal*, *25*(3), 343-354.

Roever, C., & McNamara, T. (2006). Language testing: The social dimension. *International Journal of Applied Linguistics*, *16*(2), 242-258.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford: Oxford university press.

Rohlen, T. P. (1983). *Japan's high schools*. California: University of California Press.

Roos, B., & Hamilton, D. (2005). Formative assessment: A cybernetic viewpoint. *Assessment in Education: Principles, Policy & Practice*, *12*(1), 7-20.

Sadler, D. R. (1981). Intuitive data processing as a potential source of bias in naturalistic evaluations. *Educational Evaluation and Policy Analysis*, *3*(4), 25-31.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119-144.

Sadler, P. M., & Good, E. (2006). The impact of self-and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31.

Shulman, L. S. (1986). Paradigms and research programs in the study of teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 3-36). New York: Macmillan Publishing.

Schunk, D. H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, *33*(2), 359-382.

Scott, D. (Ed.). (2001). *Curriculum and assessment* (Vol. 1). Westport, Washington, DC: Greenwood Publishing Group.

Shen, C., & Tam, H. P. (2008). The paradoxical relationship between student achievement and self-perception: A cross-national analysis based on three waves of TIMSS data. *Educational Research and Evaluation*, *14*(1), 87-100.

Shepard, L. (2001). *The role of classroom assessment in teaching and learning*. California: Center for Research on Education, University of California.

Shin, H., Ahn, S., & Kim, Y. (2017). A policy analysis on the process-based evaluation-focusing on middle school teachers in Seoul. *Journal of Curriculum Evaluation*, *20*(2), 135-162.

Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics*, *15*, 188-211.

Shohamy, E. (1998). Applying a multiplism approach. In E. Li & J. Gregory (Eds.), *Testing and evaluation in second language education* (pp. 99-114)*.* Hong Kong: Language Centre, The Hong Kong University of Science and Technology.

Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Harlow, UK: Pearson Education.

Simpson, M. (2001). Differentiation and assessment. In D. Scott (Ed.), *Curriculum and assessment* (Vol. 1) (pp. 25-40). Westport, Washington, DC: Greenwood Publishing Group.

Stiggins, R. J. (1997) *Student-centered classroom assessment*. Englewood Cliffs, NJ: Prentice-Hall.

Stiggins, R. J. (1999a). Assessment, student confidence, and school success. *The Phi Delta Kappan*, *81*(3), 191-198.

Stiggins, R. J. (1999b). Evaluating classroom assessment training in teacher education programs. *Educational measurement: Issues and practice*, *18*(1), 23-27.

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, *20*(3), 5-15.

Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10), 758-765.

Stobart, G. (2006). The validity of formative assessment. *Assessment and Learning*, 133-146.

Tan, K. (2011). Assessment for learning reform in Singapore–quality, sustainable or threshold? In R. Berry & R. Adamson (Eds.), *Assessment reform in education* (pp. 75-87). Dordrecht, NL: Springer.

Taylor, P. C., Fraser, B. J., & Fisher, D. L. (1997). Monitoring constructivist classroom learning environments. *International Journal of Educational Research*, *27*(4), 293-302.

Teasdale, A., & Leung, C. (1997). *Teacher assessment at key stage 1: Teacher and scale meanings in speaking and listening.* Paper presented at the British Educational Research Conference, York.

Teasdale, A., & Leung, C. (2000). Teacher assessment and psychometric theory: A case of paradigm crossing? *Language Testing*, *17*(2), 163-184.

TGAT (Task Group on Assessment and Testing) (1988). *National curriculum: A*

report. London: Department of Education and Science and the Welsh Office.

Trim, J. L. M. (1997). The proposed Common European Framework for the description of language learning, teaching and assessment. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment* (pp. 415-421). Jyväskylä, Finland: University of Jyväskylä and University of Tampere.

Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, *22*(4), 389-404.

Tweed, R. G., & Lehman, D. R. (2002). Learning considered within a cultural context: Confucian and Socratic approaches. *American Psychologist*, *57*(2), 89.

Von Glasersfeld, E. (1987). Learning as a constructive activity. *Problems of Representation in the Teaching and Learning of Mathematics*, *3*(17), 83-90.

Vygotsky, L. S. (1962). *Thought and language*. New York: Wiley.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wang, X. (2017). A Chinese EFL teacher's classroom assessment practices. *Language Assessment Quarterly*, *14*(4), 312-327.

Wiggins, G. (1992). Creating tests worth taking. *Educational Leadership, 49*(8), 26-33.

Wiggins, G. (2011). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, *92*(7), 81-93.

Wiliam, D. (2001). An overview of the relationship between assessment and the curriculum. In D. Scott (Ed.), *Curriculum and assessment* (pp. 165-81). Westport, Washington, DC: Greenwood Publishing Group.

Wiliam, D. (2010). The role of formative assessment in effective learning environments. In D. Hanna, I. David, & B. Francisco (Eds.), *The nature of learning: Using research to inspire practice* (pp. 135-155). Paris: OECD Publishing.

Wilson, R. J. (1996). *Assessing students in the classrooms and schools*. Scarborough, Canada: Allyn & Bacon.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89-100.

Wright, D. C. (2001). *The history of China*. Westport: Greenwood Press.

Xiao, Y. (2017). Formative assessment in a test-dominated context: How test practice can become more productive. *Language Assessment Quarterly*, *14*(4), 295-311.

Xiaoju, L. (1990). How powerful can a language test be? The MET in China. *Journal of Multilingual & Multicultural Development*, *11*(5), 393-404.

Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, *43*(3), 492-513.

Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Language Assessment Quarterly*, *7*(2), 175-194.

Yung, B. H. (2001). Examiner, policeman or students' companion: Teachers'

perceptions of their role in an assessment reform. *Educational Review*, *53*(3), 251-260.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, *16*(4), 323-342.

Zhu, W. (1992). Confucius and traditional Chinese education: An assessment. In R. Hayhoe (Ed.), *Education and modernization: The Chinese experience* (pp. 3-22). Oxford: Pergamon Press.

Zoeckler, L. G. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American secondary education*, *35*(2), 83-102.

APPENDIX

Interview Questions

1. Why do you think you conduct performance assessment in general?

2. What do you prioritize most in planning performance assessment?

3. Which purposes elicit the greatest amount of discussion, hence more teacher conferences?

4. Why did you compose the performance assessment plan as consisting of essay assessment, class participation, and listening assessment (in the case of School X)? Why did you compose the performance assessment plan as consisting of essay assessment, listening assessment, dialogue recitation assessment, and class participation assessment (in the case of School Y)?

5. What are the reasons each performance assessment in your assessment plan is included? Could you state the purpose of each performance assessment? Or, why do you think you conduct each performance assessment?

6. Why does performance assessment account for 60% of students' semester grade (in the case of School X)? Why does performance assessment account for 40% of students' semester grade (in the case of School Y)?

7. Why do you assign a different grade weight for each performance assessment?

8. How do you determine which method(s) (e.g., writing, speaking, listening) to apply for each specific assessment? Could you describe this process in detail?

9. How do you determine the criteria for each assessment as shown in the

assessment plan?

10. Why do you prioritize criterion A over criterion B?

11. How did you determine the decision interval grade?

12. Why do you carry out short-answer tests as part of the mid-term and final examinations?

13. You said "_____" in the _____ teacher conference. Why did you say this?

14. Could you explain about _____ in the performance assessment?

15. How do you give students feedback in each performance assessment?

16. Do you think the purpose of the performance assessment is well fulfilled?

17. Is the performance assessment connected to what the students learn in classes?

18. Why do not you teach writing or speaking skills in your classes?

19. Do you think performance assessment helps to improve students' learning or teaching instruction?

국문 초록

본 연구는 2015 개정교육과정에서 수행평가가 과정에 초점을 맞추고 학생들의 학습을 향상시킬 수 있는 평가가 되어야 한다는 것을 강조하고 있음을 염두에 두고, 한국의 고등학교 영어교사들이 수행평가를 어떻게 인식하고 시행하고 있는지를 알아보는 것을 목적으로 한다. 이를 위해 본 연구는 연구문제로서 첫 번째, '한국의 영어교사들은 어떻게 그들의 수행평가를 계획하는가? 교사들이 사용하는 평가 방법과 평가 기준은 무엇인가?' 두 번째, '교사들은 자신들이 선택한 평가 방법과 평가 기준에 대해 어떻게 설명하는가?'를 정하고, 교사들이 수행평가를 계획하면서 선택하는 평가 방법과 평가 기준을 살펴보고, 이러한 의사결정에 대한 이유를 알아보자 한다.

두 개의 경기도 소재 고등학교에서 근무하는 총 5명의 영어 교사들이 본 연구에 참여하였다. 연구자는 먼저, 각 학교에서 학기 초 이루어진 수행평가 계획을 위한 협의회 녹음자료와 수행평가 계획서들을 살펴봄으로써 인터뷰를 위한 질문을 마련하였다. 연구자는 이 인터뷰 질문들을 가지고 일대일로 개인 인터뷰를 개인당 약 90분 동안 진행하였으며, 이후, 각 학교의 교사들이 모두 참여하는 그룹 인터뷰를 각 학교에서 약 90분 동안 진행하였다. 이러한 인터뷰 자료는 연구자에 의해 기록되고 전사되었으며, 두 학교의 평가계획과 관련한 문서들과 함께 모두 분석되었다. 이 결과들은 교사들이 계획하고 실시하는 수행평가가 2015개정교육과정에서 제시하고 있는 학습을 *위한* 평가인가의 관점에서 검토되었다.

첫 번째 연구질문과 관련하여, 두 학교의 교사들이 선택한 평가 방법과 평가 기준에 관련한 결과는 두 학교의 수행평가가 학습을 *위한* 평가라고 보기 어렵다는 것을 보여주었다. 무엇보다도 교사들은 쓰기, 말하기, 듣기에 대한 기술들을 가르치

지 않고, 이 기술들에 대한 수행평가를 실시하였다. 교사들은 평가 동안 또는 평가 후에 피드백을 학생들에게 제공하지 않았으며, 말하기와 듣기 평가에 대한 평가 방법이 진정성을 갖추지 못하고 학생들의 암기력에 주로 의존하고 있었다. 학생들의 수업참여 평가에서도, 교사들은 학생들이 평가를 통해 자신의 학습을 돌아보도록 독려하기 보다는, 학생들이 주어진 과업을 완수하고 제출한 것에 대해 점수를 부여하는 것에 초점을 맞추었다.

평가 기준도 또한 학습을 *위한* 평가의 원칙에 맞지는 않았다. 쓰기평가와 말하기 평가에 있어서, 교사들은 학생의 수행을 객관적으로 평가하기 위해 '과제완성'의 평가 기준에 더 큰 점수를 부여하였고, 교사들의 주관적 판단을 요구하는 평가 기준은 배제하는 경향이 있었다. 교사들은 학생들의 학습을 증진시키기 보다는 정확한 성적을 매기는 것을 우선시하였고, 학생들의 성취도에 초점을 두고 점수를 부여하였다. 또한, 교사들은 학생들이 자신의 학습 목표를 이해할 수 있도록 도울 수 있는 평가 기준을 설명한다거나 구체적 예시를 활용하고 있지는 않았다.

두 번째 연구질문과 관련하여, 본 연구의 연구결과는 교사들이 수행평가를 실시하는 이유를 '정부의 정책에 따르기 위해', '학생들의 수행을 점수화하고 보고하기 위해', '학교생활기록부를 기록하기 위해', '학생들에게 말하기를 연습할 기회를 주기 위해', 그리고 '학생들의 적극적인 교실 수업 참여를 이끌기 위해' 라고 인식하고 있음을 보여주었다. 교사들이 인식하고 있는 수행평가를 실시하는 이유 및 목적은 교사들의 수행평가를 위한 평가 방법과 평가 기준 선택과 밀접한 관련이 있는 것처럼 보였다. 또한, 수행평가를 실시하는 것에 대해 교사들이 인식하고 있는 이유와 평가 방법과 평가 기준을 선택하는 데 있어서의 교사들의 의사결정은 정부의 정책, 보고에 대한 요구와 같은 사회의 관료주의적 압박과 대학 입시, 수학능력시험과 같은 요인에 의해 크게 영향을 받고 있었다. 즉, 본 연구는 사회관료적 압박감과 같은 사회

적 요인이 교사들이 수행평가를 실시하는 이유에 대한 인식에 영향을 미치면서, 수행평가가 학습을 *위한* 평가가 되는데 있어서의 장애로 작용한다는 것을 발견하였다.

　　　이러한 본 연구의 결과는 한국의 고등학교 교실에서 영어 수행평가가 어떻게 실제로 시행되고 있는지를 이해하는데 크게 기여할 수 있다. 동시에, 한국의 EFL 상황에서 수행평가가 과정에 초점을 두고 또한 학습을 *위한* 수행평가가 되기 위해 어떠한 노력이 필요한 지에 대한 제언을 제공한다.

주요어: 수행평가, 과정 중심 수행 평가, 학습을 위한 평가, 교사의 의사 결정, 사회
　　　　문화적 맥락