Exploring Performance Assessment: Teacher Beliefs versus Classroom Practices in Korean Secondary English Education[*]

Kyungeun Min and Youngsoon So[**]

This study examines the beliefs and practices regarding performance assessment among Korean English teachers in secondary schools. A survey was conducted with 109 teachers using a 52-item online questionnaire to investigate the extent to which teachers' beliefs align with their instructional practices and to categorize them based on these criteria. Although teachers acknowledge the pedagogical advantages of performance assessment, a significant gap exists between their beliefs and actual implementation, primarily due to the considerable time required for its execution. Teachers with longer tenure and higher qualifications demonstrated a more positive and effective application of performance assessment. The findings highlight the necessity for targeted professional development to enhance teachers' use of performance assessment strategies, especially for those who are skeptical about their effectiveness. Such initiatives are essential for bridging the gap between beliefs and practices, fostering an environment that promotes the effective use of performance assessment, and ultimately improving teaching quality and student learning outcomes in Korean English education.

---

1. INTRODUCTION

Assessment plays a pivotal role in teaching and learning, significantly influencing student success and the efficacy of the curriculum (Cheng, Rogers, & Hu, 2004; Serpil & Derin, 2017). Teachers dedicate a considerable portion of their time to the development of assessment instruments and the analysis and reporting of results (Cheng et al., 2004). Rogers (1991) posits that up to one-third of instructional time may be allocated to activities related to assessment. Despite the crucial role of assessment and the substantial efforts of teachers in this area, there are instances in which assessments do not sufficiently mirror the instructional objectives (Serpil & Derin, 2017). This issue is especially evident in environments where school-based assessments are deeply influenced by high-stakes assessments, as is the case in South Korea. There, teachers commonly resort to teaching-to-the-test methodologies, which may lead to increases in scores that represent enhancements in test-taking strategies rather than actual gains in ability (Bang & Chun, 2011).

In response, Korea introduced performance assessment in 1999, emphasizing the importance of aligning assessment with learning and the necessity of accurately assessing English language skills (S. H. Kim, 2017). This emphasis has persisted in the Korean National English Curriculums including the most recent update in the 2022 National English Curriculum (Ministry of Education, 2022). However, performance assessments implemented in Korean secondary schools have not fully met these goals, with reports indicating a continued misalignment between assessment and learning (Lee & Sung, 2017).

Exploring teachers' beliefs about performance assessment is crucial for its successful implementation and the enhancement of assessment quality (Elshawa, Abdullah, & Rashid, 2017). Teachers' beliefs significantly impact classroom assessment practices, often more than teaching experience or socioeconomic contexts (Brown & Remesal, 2017). Their perceptions shape decision-making and classroom actions (Pishghadam, Adamson, Sadafian, & Kan, 2014), and critically examining these beliefs can contribute to professional development (James & Pedder, 2006).

Moreover, it is vital to examine teachers' assessment practices in conjunction with their beliefs. Current research indicates a misalignment between beliefs and practices, with teachers frequently encountering discrepancies between the two (Guadu & Boersma, 2018; James & Pedder, 2006; Muñoz, Palacio, & Escobar, 2012). A thorough understanding of the relationship between teachers' beliefs and practices is crucial for fostering professional growth in assessment (Winterbottom et al., 2008). Although the discrepancy between teacher beliefs and practices has been extensively explored across various aspects of language learning, research specifically addressing this issue within the context of Korean English education remains limited.

Comparing teachers' beliefs about assessment with their actual practices in Korean

secondary English education is particularly crucial for several reasons. Firstly, Korean English education is significantly influenced by high-stakes, norm-referenced exams linked to university admissions (W. Lee, 2015). Secondly, assessment results in secondary schools, particularly high schools, are considered in university admissions processes, highlighting concerns about the reliability of these assessments (Park, 2016). Ensuring scoring reliability in performance assessments is more challenging compared to multiple-choice questions, due to their constructed response formats. Despite these challenges, the Korean national curriculum endorses the use of performance assessments, recognizing their educational benefits. Given these contrasting aspects, exploring the alignment between teachers' beliefs and their actual assessment practices in Korean English education is of significant educational importance.

Therefore, the present study aims to investigate teachers' beliefs regarding performance assessments and examine the degree to which these beliefs are reflected in their teaching practices. The research is steered by the following specific questions:

1) What are Korean English teachers' beliefs regarding performance assessment?
2) How consistent are teachers' practices with their beliefs about performance assessment?
3) To what extent can English teachers be grouped into distinct clusters based on their beliefs and practices regarding performance assessment?

2. LITERATURE REVIEW

2.1. English Performance Assessment

Performance assessment is designed to enhance core competencies that meet the demands of today's dynamically changing society. Unlike traditional approaches that prioritize memorization, performance assessment emphasizes the application of knowledge in practical contexts (Darling-Hammond, Adamson, & Abedi, 2010). This method evaluates learners by observing how they use their existing knowledge to solve problems or perform tasks, moving beyond the selection of predetermined options to requiring students to construct answers, perform activities, or produce products (Wren, 2009). This approach not only promotes cognitive thinking and logic but also enhances students' abilities to apply knowledge to meaningful, real-world challenges (M. Lee, 2018; S. Lee, 2008).

Furthermore, performance assessment in English education provides teachers with insights into students' practical application of learned skills, particularly in speaking and writing (McNamara, 1996). Performances are evaluated against predetermined criteria. Given its proven effectiveness in fostering complex thinking skills, performance assessment

has been mandated in many countries since the late 1980s and early 1990s, continuing to play a pivotal role in the educational systems of these countries (Darling-Hammond et al., 2010). Consistent with this international trend, Korean education has continually emphasized performance assessments, linking them to English learning objectives since their incorporation into the national curriculum in 1999.

2.2. English Performance Assessment in Korea

The introduction of performance assessments in the Korean school assessment system in 1999 gave rise to the expectation that the prevalent over-reliance on traditional multiple-choice questions in English language classes would be addressed. Additionally, it was forecasted that English education in schools would progressively shift toward the primary objective of enhancing students' communicative competence, thereby improving the overall quality of English education within Korean schools (M. Lee, 2018). Despite the incorporation of performance assessments into the national curriculum, substantial transformations in the assessment of English in Korean secondary schools have not materialized. Further, existing literature indicates that assessments designated as performance assessments frequently fall short in effectively eliciting students' practical use of language (Park, 2016).

Two reasons may explain why English performance assessments in Korean secondary schools fail to meet the expectations set by the introduction of the new assessment policy. Firstly, the characteristics of performance assessments differ significantly from those of the College Scholastic Ability Test (CSAT), which is the most influential high-stakes examination for Korean high school students. Teachers feel pressured to assist students in achieving good scores on the CSAT, which in turn makes them reluctant to implement performance assessments that deviate significantly from the format of this examination (W. Lee, 2015). Secondly, the overall reliability of performance assessments tends to be lower compared to traditional written exams, further discouraging their implementation in secondary schools (Park, 2016). School records are expected to be as reliable as possible since they are considered in college admissions decisions.

The aforementioned reasons suggest that the performance assessments currently implemented in the Korean educational system may not fully realize the foundational goal of aligning with the curriculum. Ideally, an assessment that is aligned with curriculum objectives should help students progress in their learning, an approach encapsulated by the concept of 'assessment for learning' (AfL) (Black, Harrison, Lee, Marshall, & Wiliam, 2004). The revised 2022 Korean National English Curriculum addresses this concern by underscoring AfL and promoting a variety of assessment methods to support each student's developmental progress (Ministry of Education, 2022). Realizing the reforms proposed in

the National Curriculum requires a concerted effort to deepen English teachers' understanding of the features of performance assessment and to actualize the principles of AfL through the application of these assessments. Thus, examining teachers' perceptions of performance assessment from this perspective is of considerable importance.

2.3. Teachers' Beliefs and Assessment Practices

Extensive research has shown that teachers' beliefs have a significant impact on their assessment practices (Remesal, 2011). Kahn (2000) argues that these beliefs are critical to the successful adoption of new assessment methods in educational settings. Given that teachers are the ones who administer assessments, yet might not always be aware of their own belief systems, it is crucial to investigate these beliefs to ensure that performance assessments are implemented effectively and align with their designed objectives.

The interplay between teachers' beliefs and the design, implementation, and interpretation of assessments has been a subject of empirical research, revealing the multifaceted nature of this relationship (Gullickson, 1984). Studies often uncover discrepancies between teachers' beliefs and assessment practices across various educational settings, highlighting a consistent gap between teachers' assessment beliefs and their actual classroom practices, as evidenced by research in England (James & Pedder, 2006), Colombia (Muñoz et al., 2012), and Ethiopia (Guadu & Boersma, 2018). These studies pinpoint challenges like implementation difficulties and constraints such as time and class size and call for tailored reflection, guidance, and solutions to enhance assessment effectiveness within specific educational contexts.

Despite the extensive body of literature on the impact of teachers' beliefs on assessment practices, the specific beliefs of Korean English teachers remain less explored. Considering the pivotal role of performance assessments in advancing AfL objectives and thereby helping students to foster communicative competence in English, it is imperative to examine the beliefs of Korean English teachers concerning performance assessments (Kim & Kim, 2018). Therefore, a comprehensive investigation into the beliefs of Korean English teachers and how these beliefs inform their assessment strategies is essential to ensure the efficacy of performance assessments in English language education in Korea.

3. METHODOLOGY

3.1. Participants

A total of 109 secondary school teachers in Korea were recruited from an online

community of in-service English teachers for the online questionnaire. The detailed background information of the participants is presented in Table 1.

TABLE 1
Profiles of Teacher Participants in Questionnaire ($N = 109$)

School Level	Middle School		High School	
	59 (54.1%)		50 (45.9%)	
Gender	Male		Female	
	9 (8.3%)		100 (91.7%)	
Teaching Experience (years)	< 5	5-10	11-15	≥ 15
	13 (11.9%)	45 (41.3%)	27 (24.8%)	24 (22.0%)
Degree	Bachelor		Master	
	74 (67.9%)		35 (32.1%)	

As Table 1 illustrates, the distribution of teachers by school level is relatively balanced, with middle school teachers comprising 54.1% and high school teachers 45.9%. However, there is a significant gender imbalance, reflecting the broader trend of male teacher scarcity in schools, with males constituting only 8.3% of the sample. Teachers with teaching experience between five to ten years account for 41.3%, while only 11.9% are teachers with less than five years of teaching experience. Additionally, 32.1% of the teachers hold a master's degree.

3.2. Instrument

The questionnaire was adapted from one originally developed by Elshawa et al. (2017), which focused on teachers' beliefs and practices in tertiary English as a second or foreign language assessment. It was revised to specifically address performance assessment in the Korean secondary education context. Items that were irrelevant to the Korean setting (e.g., 'Students' final grades should be based on coursework only'), misaligned with the study's focus (e.g., 'Assessment items from published textbooks are a better source than those found on the internet'), or unrelated to performance assessment (e.g., 'The best means of assessing language is through summative assessment') were removed. Furthermore, items considered necessary but absent from the original questionnaire were added (e.g., 'Sufficient time should be allowed to properly prepare performance assessments').

The revised questionnaire consisted of 52 items, divided into two sections. The first section assessed teachers' beliefs about assessment, offering a response scale from one (Strongly Disagree) to four (Strongly Agree). The second section addressed assessment practices, with responses ranging from one (Never True) to four (True). Reverse coding was applied to three items that depicted practices contrary to those recommended for performance assessments, thus ensuring consistent interpretation of the results. These were

Item 6, regarding the competitive nature of performance assessments; Item 11, concerning the effectiveness of paper and pencil assessments for evaluating speaking and listening skills; and Item 15, relating to the appropriateness of selected-response items for performance assessments. Aligning the content and sequence of questions between the two sections enabled a comparative analysis of teachers' beliefs and practices, a method proven effective for elucidating the relationship between the two (James & Pedder, 2006).

Furthermore, the sections were subdivided into three subsections: Section A with ten items inquiring about assessment purposes; Section B with eight items examining methods and techniques; and Section C with eight items focused on teachers' feedback, grading, and reporting practices associated with performance assessments. The internal reliability of the questionnaire was assessed, yielding a Cronbach's alpha of 0.87. The measures for each of the three subsections were 0.83 for Section A, 0.68 for Section B, and 0.71 for Section C. Although the value for Section B is below the generally acceptable level of Cronbach's alpha, which is 0.70 or above (Nunnally, 1978), the questionnaire overall is deemed to have generally acceptable internal consistency.

3.3. Data Analysis

The data collected from the questionnaire were descriptively analyzed in the following order using SPSS 26. Initially, to explore teachers' beliefs about performance assessments, a descriptive analysis of Items 1 to 26 was performed. This phase involved calculating the mean and standard deviation for each item, as well as for the three subsections (see the Instrument section), and the composite section. Following this, Items 27 to 52 were descriptively analyzed to assess teachers' actual practices in performance assessments. The analysis provided the mean and standard deviation for each individual item, the three subsections, and the entirety of the practice section.

Subsequently, we explored the relationship between teachers' beliefs and their actual practices in performance assessments by comparing responses to matched items, each representing a belief and its corresponding practice. For instance, responses to Item 1, indicative of a specific belief, were contrasted with those to Item 27, which denotes the associated practice. We calculated a difference score for each matched pair by deducting the practice score from the belief score. A positive difference signified that the belief about a statement in the questionnaire was stronger than its actual implementation, while a negative value suggested that a practice was in use despite contrary beliefs. Descriptive statistical analyses of these difference scores were conducted for each item pair, across the three subsections, and for the questionnaire in its entirety.

In the final phase of analysis, hierarchical clustering was employed to categorize participating teachers based on their beliefs and practices in performance assessment. This

analysis aimed to identify patterns among the variations in teachers' beliefs and practices and to group them based on common characteristics. The Ward's linkage algorithm was chosen for its efficacy in minimizing within-cluster variance, thereby ensuring that the resultant clusters consisted of teachers with homogeneous profiles of beliefs and practices (James & Pedder, 2006). Subsequent chi-square tests were conducted to assess the relationships between the identified clusters and teacher characteristics. This procedure led to provide insights into the specific beliefs and practices regarding performance assessments among different teacher groups.

4. RESULTS

4.1. Teacher Beliefs and Practices in Performance

Table 2 provides an overview of the descriptive statistics for teachers' beliefs, their practices, and the difference between the two across three sections of the questionnaire.

TABLE 2
Descriptive Statistics by Sections ($N = 109$)

Section	Content	Belief		Practice		Belief-Practice Difference	
		M	SD	M	SD	M	SD
A	Instructional Purposes	3.45	0.61	3.02	0.77	0.43	0.86
B	Methods and Techniques	3.18	0.62	2.74	0.84	0.44	0.88
C	Feedback, Grading, and Reporting of Grades	3.26	0.61	2.95	0.74	0.31	0.78
	Total	3.30	0.61	2.90	0.78	0.39	0.84

Table 2 presents the following key observations. First of all, it is noted that the average belief score is higher than the average practice score across all three subsections, as well as across the entire questionnaire. This finding may indicate that English teachers' beliefs about performance assessment are not fully reflected in their practice. Second, in Section A, 'Instructional Purposes,' participants reported the highest belief scores ($M = 3.45$) on a scale of 4, indicating a notably strong agreement with the instructional purposes of performance assessment. Conversely, Section B, 'Methods and Techniques,' recorded both the lowest belief score ($M = 3.18$) and practice score ($M = 2.74$), resulting in the most pronounced discrepancy between beliefs and practices. This suggests a divergence between the endorsement of methods and techniques in performance assessment and their application.

Finally, the overall standard deviation values reveal that while teachers' beliefs are relatively consistent ($SD = 0.61$), there is greater variability in their reported practices ($SD = 0.78$). This implies a more extensive diversity in the application of performance assessment practices compared to the more uniform beliefs held by teachers.

Subsequent to the descriptive statistics of the subsections, the item-level analysis of teachers' beliefs, practices, and the differences between them is delineated. This analysis adheres to the sequence of the subsections in the questionnaire.

4.2. Instructional Purpose of Performance Assessment

Table 3 below summarizes the descriptive statistics regarding teachers' beliefs and practices pertaining to the instructional purposes of performance assessment.

TABLE 3
Instructional Purposes ($N = 109$)

Item	Statement (Performance assessment…)	Belief		Practice		Belief-Practice Difference	
		M	SD	M	SD	M	SD
1	helps to focus on teaching.	3.55	0.60	2.91	0.77	0.63	0.91
2	helps to group students for instructional purposes.	2.73	0.88	2.84	0.78	-0.11	1.06
3	diagnoses strengths and weaknesses in students.	3.51	0.65	3.13	0.72	0.39	0.83
4	diagnoses strength and weaknesses in teaching.	2.86	0.85	2.78	0.85	0.08	0.88
5	provides information about students' progress.	3.62	0.52	3.14	0.70	0.49	0.82
6	creates competition among students. (reverse coded)	3.39	0.69	2.60	0.94	0.80	1.03
7	creates a valuable learning experience for my students.	3.80	0.40	3.34	0.67	0.46	0.70
8	motivates my students to learn.	3.72	0.45	3.19	0.76	0.52	0.82
9	provides feedback to students as they learn.	3.69	0.47	3.07	0.81	0.61	0.82
10	demonstrates the level of students' learning.	3.61	0.56	3.22	0.71	0.39	0.72

The first notable finding is about Item 7, which addresses the role of performance assessment in fostering a valuable learning experience for students. This item garnered the highest belief score ($M = 3.80$, $SD = 0.40$) and a substantial practice score ($M = 3.34$, $SD = 0.67$), with the narrowest standard deviations indicating a consistent recognition of its importance among teachers. Additional items that reflected high belief scores—#8, #9, #5, and #10—suggest that these aspects of performance assessment are aligned with the AfL

principles, a view corroborated by the relatively high practice scores.

Conversely, skepticism was noted concerning the use of performance assessment for grouping students for instructional purposes (Item 2) and for diagnosing teaching strengths and weaknesses (Item 4), as reflected by the lower mean belief scores of 2.73 and 2.86, respectively. The high standard deviations for these items ($SD = 0.88$ for beliefs and $SD = 0.85$ for practices) point to diverse opinions among the teachers. Notably, beliefs regarding the diagnostic use of performance assessment to evaluate teaching effectiveness (Item 4) were less favorable compared to its use for assessing student abilities (Item 3), indicating a reluctance to apply performance assessment outcomes to teachers' self-reflection on teaching.

The mean difference between beliefs and practices spanned a range from a mere 0.08 (Item 4) to a substantial 0.80 (Item 6). The most pronounced divergence was observed in the context of performance assessment creating competition among students (Item 6). Considering the fact that this item was reverse-coded, a mean belief score of 3.39 ($SD = 0.69$) signals an adverse stance towards competition. Nevertheless, the practice scores, while also reverse-coded, suggest a deviation from these beliefs; it seems that competition is being induced in the process of the performance assessment. The second most significant discrepancy was in relation to performance assessment aiding the focus on teaching (Item 1), illustrating a disconnect between teachers' perceptions and their classroom practices.

Furthermore, the minimal belief-practice differences in Item 2 (negative 0.11) and Item 4 (0.08) coincided with the lowest belief scores in Subsection A, reflecting a transference of weak belief into practice. In particular, the negative belief-practice difference for Item 2 suggests a distinctive trend where practices exceed the low endorsement level.

4.3. Methods and Techniques

The descriptive statistics presented in Table 4 reveal teachers' beliefs regarding various forms and types of performance assessment as well as their actual classroom practices. The data indicate an overall inconsistency between the beliefs and practices of language teachers concerning performance assessment methods. Below are some noteworthy findings.

A significant disparity is noted in Item 14, which pertains to the requisite preparation time for performance assessment. The teachers strongly believed that ample preparation time is essential ($M = 3.86$, $SD = 0.35$). Contrary to this belief, the mean score for their reported practices was considerably lower ($M = 2.23$, $SD = 0.89$), suggesting a lack of adequate preparation time in reality. This belief-practice gap of 1.63 represents the largest among all items surveyed (refer to Tables 3 and 5 for comparison).

In contrast, Item 16 exhibits the most congruence between beliefs and practices concerning the use of constructed-response items in performance assessments, with a trivial

mean difference. The affirmative belief ($M = 3.52$, $SD = 0.54$) is well-reflected in practice ($M = 3.57$, $SD = 0.64$), indicating that teachers not only value but also actively implement this assessment type as performance assessment, which aligns with the principle of performance assessment by engaging students in language production.

TABLE 4
Methods and Techniques ($N = 109$)

Item	Statement	Belief		Practice		Belief-Practice Difference	
		M	*SD*	*M*	*SD*	*M*	*SD*
11	Paper and pencil assessment (e.g. dictation, filling in the blanks of conversation) is the best method in evaluating students' speaking and listening. (reverse coded)	2.54	0.76	2.78	1.06	-0.24	1.11
12	Performance assessment questions should reflect real life language use.	3.40	0.56	3.06	0.69	0.35	0.76
13	Language teachers need to use a variety of assessment methods to assess students.	3.57	0.52	3.12	0.73	0.45	0.80
14	Sufficient time should be allowed to properly prepare performance assessments.	3.86	0.35	2.23	0.89	1.63	1.01
15	Selected-response items (e.g. matching items, multiple-choice items, true - false items) are good methods of performance assessment. (reverse coded)	3.08	0.68	3.45	0.81	-0.37	0.89
16	Constructed-response items (e.g. journal entry, portfolio, short essay, sentence completion, reflective task) are good methods of performance assessment.	3.52	0.54	3.57	0.64	-0.05	0.74
17	Self-assessment by the student is a good method of performance assessment.	2.77	0.81	1.87	0.93	0.90	0.83
18	Peer-assessment is a good method of performance assessment.	2.71	0.77	1.84	0.93	0.86	0.93

Responses to Items 11 and 15, however, raise concerns. Teachers expressed a lower belief in the use of paper-and-pencil and selected-response items as performance assessment, as evidenced by low belief mean scores ($M = 2.54$, $SD = 0.76$ for Item 11; $M = 3.08$, $SD = 0.68$ for Item 15). The mean belief score for this perspective is pedagogically justified since these item types are typically considered less effective for eliciting performance in language from

students. Although the teachers' negative stance about using paper-and-pencil or selected-response items as performance assessment is theoretically warranted, the data suggest that these methods are still employed in classrooms, potentially more so than their perceived value, with the mean practice scores of 2.78 (SD = 1.06) for Item 11 and 3.45 (SD = 0.81) for Item 15. Another salient result highlights Korean English teachers' skepticism or undervaluation of peer-assessment (Item 18, M = 2.71, SD = 0.77) and self-assessment (Item 17, M = 2.77, SD = 0.81), both of which garnered even lower practice scores (M = 1.84, SD = 0.93 for peer-assessment and M = 1.87, SD = 0.93 for self-assessment). These findings imply that the teachers may not regard self- and peer-assessment as effective performance assessment strategies, thereby infrequently employing these methods in their English language classrooms.

4.4. Feedback, Grading, and Reporting of Grades

Table 5 presents findings on English teachers' beliefs and practices about Section C concerning feedback, grading, and reporting of grades.

TABLE 5

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

Item	Statement	Belief		Practice		Belief-Practice Difference	
		M	SD	M	SD	M	SD
19	Student effort is seen as important when assessing their learning.	3.45	0.57	3.18	0.80	0.27	0.85
20	A marking scheme should be prepared before assessment is given.	3.74	0.48	3.83	0.40	-0.09	0.52
21	Conferencing with students is a good way of giving feedback.	3.25	0.61	2.25	0.95	1.00	1.01
22	Criterion-referenced assessment is better than norm-referenced assessment.	3.26	0.76	3.40	0.94	-0.15	1.10
23	Students should be given feedback after performance assessment.	3.63	0.50	3.14	0.87	0.50	0.88
24	Students should be informed about the marking criteria before being assessed.	3.83	0.40	3.86	0.37	-0.03	0.35
25	Students should be involved in preparing the marking criteria.	2.20	0.71	1.37	0.65	0.83	0.66
26	Students should be given back their performance assessment results no later than a week after the assessment.	2.70	0.83	2.57	0.95	0.13	0.90

The items that garnered the highest belief mean scores concerned the preparation and communication of marking criteria prior to performance assessment. Specifically, Item 24, which pertains to informing students about marking criteria in advance of performance assessment, garnered a mean belief score of 3.83 ($SD = 0.40$), and Item 20, advocating for the preparation of a marking scheme before assessment, received a mean belief score of 3.74 ($SD = 0.48$). Their corresponding practice scores were marginally higher: 3.86 ($SD = 0.37$) for Item 24, and 3.83 ($SD = 0.40$) for Item 20. This indicates alignment between belief and practice, with practice slightly surpassing belief as indicated by negative mean differences.

Conversely, beliefs regarding student engagement in setting marking criteria (Item 25) and the prompt reporting of performance assessment results (Item 26) were lower, with mean scores of 2.20 ($SD = 0.71$) and 2.70 ($SD = 0.83$), respectively, and their enactment in practice was even less prevalent ($M = 1.37$, $SD = 0.65$ and $M = 2.57$, $SD = 0.95$, respectively). The disparity was especially pronounced for Item 25, which had a mean difference of 0.83, pointing to a significant divergence between belief and practice.

Another salient finding pertains to Item 21, which considers student conferencing as a feedback mechanism. A substantial mean difference of 1.00 found in this item indicates that the frequency of actual conferencing ($M = 2.25$, $SD = 0.95$) does not reflect the stronger belief in its value ($M = 3.25$, $SD = 0.61$). This may suggest that logistical challenges, such as time constraints, potentially inhibit the translation of this belief into consistent practice. This interpretation is corroborated by responses to Item 14, which emphasized the necessity for adequate preparation time for performance assessments (see Table 4). Teachers expressed a high belief in the importance of preparation time ($M = 3.86$, $SD = 0.35$), yet reported a lower incidence of practice ($M = 2.23$, $SD = 0.89$), resulting in a mean difference of 1.63. These two items – Items 21 and 14 – exhibit the most pronounced belief-practice discrepancies, implicating time as a pivotal factor in the correlation between teachers' beliefs and their classroom practices.

4.5. Identifying Patterns in Teachers' Beliefs and Practices Towards Performance Assessment: A Cluster Analysis Approach

Cluster analysis was conducted to examine the diversity within English teachers' beliefs and practices regarding performance assessment in Korean secondary schools. This method was used to determine whether the teachers constitute a homogeneous group or are segmented into distinct clusters characterized by their beliefs and practices. The objective was to identify inherent patterns and relationships in the data, aiming to identify subsets of teachers who exhibit similar profiles in terms of their beliefs and practices related to performance assessment. Without presupposing the number of clusters, the analysis sought to establish the appropriate categorization. The resulting dendrogram from the hierarchical

cluster analysis is presented in Figure 1.

FIGURE 1

Dendrogram of Cluster Analysis
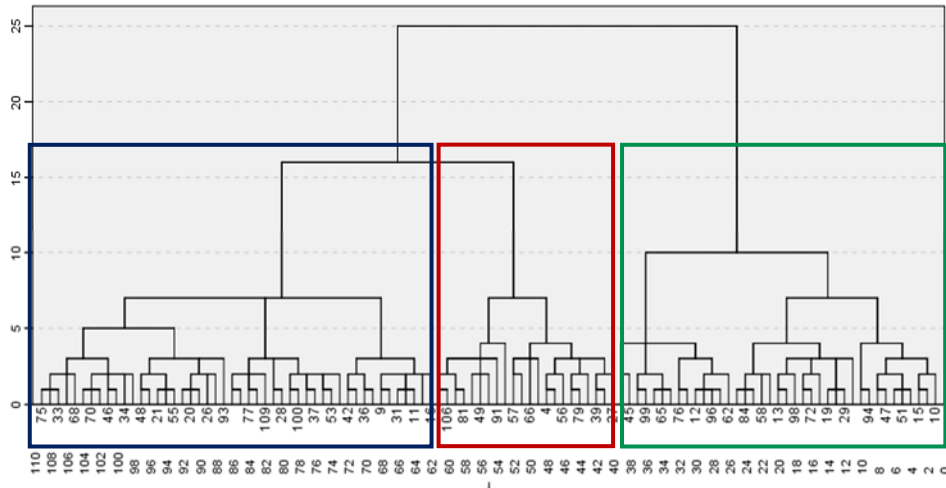
In determining the optimal number of clusters to represent English teachers' beliefs and practices, the dendrogram from the hierarchical cluster analysis was examined. Additionally, the relative sizes of the clusters were assessed to ensure that none were too small or large, facilitating meaningful comparisons and interpretations. These criteria led to the identification of three clusters that best explain the relationship between teachers' beliefs and practices as represented in the dataset. Figure 1 presents the dendrogram, delineating the three identified clusters, which are represented by blue, red, and green boxes, respectively.

To explore the characteristics of the clusters based on the teachers' beliefs held about performance assessment and the actual implementation practices, a descriptive statistical analysis was conducted for each cluster, as shown in Table 6.

TABLE 6

Descriptive Statistics of Clusters (N = 109)

Cluster	N	Belief		Practice		Belief-Practice Difference	
		M	*SD*	*M*	*SD*	*M*	*SD*
Cluster 1: Typical Teachers	48 (44%)	3.26	0.51	2.96	0.70	0.31	0.79
Cluster 2: Skeptical Practitioners	22 (20%)	3.13	0.49	2.39	0.64	0.75	0.90
Cluster 3: Aligned Advocates	39 (36%)	3.46	0.44	3.16	0.56	0.30	0.78

Table 6 presents the distribution of teachers' beliefs and their corresponding implementation practices regarding performance assessment, categorized into three distinct clusters. Cluster 1, comprising 44 percent of the participating teachers (N = 48), demonstrated moderate beliefs and implementation levels ($M = 3.26$, $SD = 0.51$ and $M = 2.96$, $SD = 0.70$, respectively), occupying an intermediate position relative to the other clusters. In contrast, Cluster 2, representing 20 percent of the sample (N = 22), contained teachers with the lowest belief scores regarding the efficacy of performance assessment in enhancing teaching and learning ($M = 3.13$, $SD = 0.49$). Additionally, the implementation of performance assessments was the lowest within the study ($M = 2.39$, $SD = 0.64$), resulting in the greatest disparity between beliefs and practices. Hence, the actual employment of performance assessments by these teachers was significantly lower than their belief scores. Cluster 3, accounting for the remaining 36 percent (N = 39), was distinguished by the highest mean scores for both beliefs ($M = 3.46$, $SD = 0.44$) and practices ($M = 3.16$, $SD = 0.56$), indicating a strong alignment with the AfL principles in their practices.

Upon reviewing the characteristics of each cluster concerning performance assessment, the clusters were labeled as 'Typical Teachers,' 'Skeptical Practitioners,' and 'Aligned Advocates,' respectively. Cluster 1 was termed 'Typical Teachers,' reflecting the predominant trends in beliefs and practices concerning performance assessment within the research sample and thereby establishing a reference point for comparison with the other groups. The label 'Skeptical Practitioners' for Cluster 2 implies their critical stance on the value of performance assessment, as evidenced by their low belief scores and even lower practice scores. Cluster 3, labeled 'Aligned Advocates,' captures the group's high belief and practice scores, suggesting their strong endorsement and application of performance assessment in line with AfL principles. This cluster analysis facilitated a classification of the participants, and their characteristics in each cluster are detailed in Table 7.

TABLE 7
Demographic Information of Clusters

		Cluster 1: Typical Practitioners	Cluster 2: Skeptical Practitioners	Cluster 3: Aligned Advocates	Total
School level	Middle	31	11	17	59
	High	17	11	22	50
Gender	Male	2	1	6	9
	Female	46	21	33	100
Teaching experience (years)	< 5	7	1	5	13
	5-10	17	15	13	45
	11-15	15	5	7	27
	≥ 15	9	1	14	24
Degree	Bachelor	35	17	22	74
	Master	13	5	17	35

Subsequent chi-square tests confirmed that teaching experience was the only characteristic significantly associated with the clusters, indicated by a chi-square statistic of $\chi^2(6, N = 109) = 14.45$, $p = .025$. No significant associations were detected with other characteristics such as school level, gender, and degree.

Table 8, which presents both observed and expected frequencies (with the numbers in parentheses indicating expected frequencies), reveals significant deviations between these frequencies. These deviations suggest a departure from the assumption of independence, thereby substantiating the significant results of the chi-square analysis. Notably, within the 'Skeptical Practitioners' cluster, the observed number of teachers with 5-10 years of experience (15) exceeded the expected frequency (9.08), while the number with more than 15 years of experience (1) was below the expectation (4.84). In contrast, the 'Aligned Advocates' cluster included a higher number of teachers with more than 15 years of teaching experience (14) compared to the expected frequency (8.59). These findings indicate that the 'Skeptical Practitioners' cluster predominantly consists of teachers with shorter teaching experiences, while the 'Aligned Advocates' cluster includes those with longer tenures in the teaching profession.

TABLE 8

Contingency Table for Cluster by Teaching Experience

	Cluster1: Typical Teachers	Cluster 2: Skeptical Practitioners	Cluster 3: Aligned Advocates	Row Total
Under 5 years	7 (5.72)	1 (2.62)	5 (4.65)	13
5 - 10 years	17 (19.82)	15 (9.08)	13 (16.10)	45
11 - 15 years	15 (11.89)	5 (5.45)	7 (9.66)	27
Over 15 years	9 (10.57)	1 (4.84)	14 (8.59)	24
Column Totals	48	22	39	109

Secondly, although the analysis did not achieve statistical significance ($\chi^2(3, N = 109) = 3.80$, $p = .149$), a higher proportion of teachers in the 'Aligned Advocates' cluster held master's degrees than expected; specifically, 17 out of 35 master's degree holders (49%) were in this cluster. Caution is advised in interpretation, yet this trend might suggest that advanced education and training could influence more positive attitudes towards performance assessment and its implementation in teaching practices.

5. DISCUSSION

The results presented above will be discussed according to the three research questions posed. In the first section, research questions #1 and #2 will be discussed together because a

more in-depth discussion is expected when comparing the results about teachers' beliefs or attitudes towards performance assessment with the actual implementation of performance assessment in classrooms. This will be followed by a section discussing research question #3, which examines the division of participating teachers according to the similarities or differences between their beliefs and practices regarding performance assessment.

5.1. English Teachers' Beliefs about Performance Assessment and their Assessment Practices

As reported in Table 2, teachers reported higher belief scores compared to practice scores across all sections, with the largest discrepancy observed in the 'Methods and Techniques' section. Despite the general agreement on the importance of performance assessment in all areas, there is a noticeable gap between teachers' beliefs and their reported practices. In the following paragraphs, major findings from three sections ('Instructional Purposes,' 'Methods and Techniques,' and 'Feedback, Grading, and Reporting of Grades') will be discussed.

The findings regarding teachers' beliefs of the purposes of performance assessment offer substantial evidence supporting their recognition of its significant contribution to improving teaching and learning. These findings are consistent with a body of empirical studies (Brown & Remesal, 2017; Elshawa et al., 2017; Muñoz et al., 2012), which reports consensus among language teachers on the value of performance assessment in foreign language instruction. Particularly notable is the teachers' high agreement that performance assessments facilitate a valuable learning experience (Item 7) and enhance motivation to learn (Item 8), corroborating the findings of N. Y. Kim (2017), and Kim and Kim (2018). Although these beliefs are reflected to some extent in classroom practices, the positive discrepancy between belief and actual implementation indicates that while teachers intend to utilize performance assessments to advance student learning, effectively integrating these assessments with instructional strategies remains a challenge.

Next, teachers' beliefs about performance assessment methods revealed a preference for diverse item methods and techniques to evaluate students' language proficiency in classrooms, with the notable exception of peer- and self-assessment, which they viewed less favorably. This aligns with Park's (2016) recommendation for assessing language ability through various methods beyond conventional paper-based tests. Despite this, traditional paper-and-pencil tests, specifically for speaking and listening assessment (Item 11), along with selected-response items (Item 15)—which are generally not recommended for performance assessment—were reported to be more widely used than what teachers' beliefs would predict.

Furthermore, it was observed that teachers' skepticism towards the efficacy of peer- and

self-assessment (M = 2.77, SD = 0.81 for self-assessment; M = 2.71, SD = 0.77 for peer-assessment) translated into a tendency to avoid these methods in practice (M = 1.87, SD = 0.93 for self-assessment; M = 1.84, SD = 0.93 for peer-assessment). This trend is consistent with observations made by N. Y. Kim (2017), who identified a reluctance among Korean English teachers to integrate these forms of assessment in their instruction. However, contrasting this reluctance, scholarly consensus suggests that formative feedback from peer- or self-assessment may yield numerous benefits, such as enhancing metacognitive processes, deepening student learning, improving autonomous learning competencies, saving instructional time, and improving students' comprehension of assessment criteria (Brown, 2005). These insights highlight the imperative for Korean English teachers to reassess the significance of self- and peer-assessment as integral components of formative performance assessment, and to consider strategies for their effective implementation in classroom settings. Consequently, there arises a need for specialized professional development programs designed to equip educators with the requisite knowledge and pedagogical competencies for the effective incorporation of a diverse array of performance assessment methodologies in their instruction.

In the domains of feedback, grading, and grade reporting—crucial for the assessment for learning (Black et al., 2004)—our findings concur with previous literature, including research conducted across varied educational contexts different from that of this study (Cheng & Wang, 2007; Elshawa et al., 2017). Despite contextual variances, a uniformity in teachers' beliefs was evident: they believed in the importance of establishing and communicating their evaluation criteria before conducting assessments (Items 24 and 20). On the other hand, teachers assigned minimal importance to involving students in the development of marking criteria (Item 25) and the prompt reporting of assessment outcomes within a week (Item 26); these low levels of belief were further underscored by an even lesser implementation of these practices in their assessment procedures.

Another salient finding pertains to Item 21, which considered the use of student conferences as a form of feedback. The data revealed that teachers possessed a strong belief in the value of student conferences (M = 3.25, SD = 0.61). Nonetheless, the actual implementation of such conferences occurred with a significantly lower frequency (M = 2.25, SD = 0.95), resulting in a marked discrepancy between belief and practice. This discrepancy between the teachers' valuing of student conferencing and its infrequent classroom application aligns with the findings from prior research conducted by N. Y. Kim (2017), and Kim and Yun (2015). This variance suggests that practical challenges, particularly time constraints, may inhibit the translation of this belief into consistent practice. Support for this inference was found in the responses to Item 14, which highlighted the necessity for adequate preparation time for performance assessments. While teachers recognized the significance of preparation time (M = 3.86, SD = 0.35), the frequency of their preparatory actions is less

(M = 2.23, SD = 0.89), yielding a mean difference of 1.63, the largest difference observed in this study.

Items 21 and 14 collectively demonstrate the most substantial belief-practice gaps, underscoring time as a crucial factor affecting the alignment of teachers' beliefs with their instructional behaviors. This discrepancy is further contextualized by the structure of Korean secondary education, wherein teachers commonly oversee several classes, leading to a higher student-to-teacher ratio (N.Y. Kim, 2017). Such a burden likely exacerbates time limitations (S. Lee, 2008), making the case for a strategic reduction in the number of students assigned to each teacher to alleviate these constraints (Park & Chang, 2017).

In summary of teachers' beliefs and practices of performance assessment, among the 109 teachers surveyed, there was a general tendency to acknowledge the pedagogical benefits of performance assessment as reflected by the questionnaire statements. Nevertheless, a discrepancy was evident between these positive beliefs and their practical application, with practices generally scoring lower than beliefs. Additionally, a pronounced skepticism was apparent in relation to certain desirable aspects of performance assessment, such as self- and peer-assessment and student engagement in developing marking criteria. This skepticism was mirrored in the limited incorporation of these strategies in their instructional practices. Given that a shift in beliefs is a prerequisite but not the sufficient factor for behavioral change (Ajzen & Fishbein, 1980), the data underscore specific domains that should be addressed within professional development programs focused on performance assessment. Furthermore, substantial gaps between belief and practice highlight that logistical challenges, such as time management, need to be addressed to effectively translate positive views on performance assessment into classroom practices.

5.2. Classifying English Teachers by Beliefs and Practices: Insights from Cluster Analysis

Cluster analysis revealed that English teachers in Korean secondary schools represent a non-homogeneous group with respect to their beliefs about performance assessment and the degree to which these beliefs are integrated into their pedagogical practices. The cluster analysis categorized participants into three distinct profiles: 'Typical Teachers' (44% of the sample), 'Skeptical Practitioners' (20%), and 'Aligned Advocates' (36%). The 'Typical Teachers' demonstrated moderate beliefs and practices scores relative to other groups. 'Skeptical Practitioners' were distinguished by the lowest scores in beliefs and practices and the greatest disparity between the two. In contrast, 'Aligned Advocates' showed the highest scores in both beliefs and practices related to performance assessment as a tool for enhancing teaching and learning. Despite similar discrepancies between beliefs and practices within the 'Typical Teachers' and 'Aligned Advocates' clusters, the latter exhibited higher belief and

practice mean scores, indicating that 'Aligned Advocates' possess and enact a more profound belief in the value of performance assessment.

Demographic analysis of three teacher clusters yielded notable findings. First, compared to the 'Typical Teachers' cluster, 'Skeptical Practitioners' generally had shorter teaching tenures, whereas 'Aligned Advocates' typically had longer ones. Furthermore, teachers in the 'Aligned Advocates' cluster were more likely to hold a master's degree than those in the other two clusters. Collectively, these observations suggest that greater teaching experience and advanced educational qualifications may correlate with more favorable attitudes towards performance assessment and its integration into teaching practices. Consequently, these results underscore the need for targeted teacher training and professional development programs aimed at enhancing the perception and implementation of performance assessment in educational settings (Borg, 2003). Specifically, for teachers in the 'Skeptical Practitioners' cluster who hold negative views on performance assessment, professional development should prioritize cultivating a more positive attitude. Once this mindset shift is achieved, teacher training should then focus on delivering practical guidelines for implementing performance assessments aligned with the principles of assessment for learning.

6. CONCLUSION

This study examined the beliefs and practices related to performance assessment among Korean English teachers in secondary schools. A total of 109 teachers participated by responding to an online questionnaire, which comprised 52 items split into two sections—one focusing on teachers' beliefs and the other on their instructional practices. The findings indicate that, although teachers generally acknowledge the benefits of performance assessment, there is a substantial gap in integrating these beliefs into actual classroom practices, especially when significant time investment is required. This discrepancy is largely due to pedagogical skepticism and practical challenges within the educational system.

Further analysis of the teachers' survey responses via cluster analysis indicated that the participants could be divided into three distinct clusters. These clusters, characterized by their varying beliefs and practices regarding performance assessment, suggest significant educational implications. Firstly, the findings underscore the necessity for in-depth, tailored professional development based on the characteristics specific to each cluster. For instance, teachers identified as 'Skeptical Practitioners' exhibit notably negative attitudes and are most resistant to adopting varied performance assessment methods. Professional development programs for these teachers should focus not only on positively shifting attitudes but also on equipping them with the skills required to effectively implement diverse assessment strategies.

Moreover, the results highlight the beneficial impact of teaching experience and educational qualifications on the adoption and effective utilization of performance assessment practices. Teachers in the 'Aligned Advocates' cluster, who generally possess longer tenures and higher academic degrees, demonstrate more positive attitudes and greater proficiency in applying these assessment methods.

Despite limitations such as the study's small sample size, the findings suggest a positive valuation of performance assessment among Korean English teachers, though its full integration remains a challenge. Targeted professional development and educational reforms are essential to address the logistical constraints faced by teachers, thereby fostering an environment conducive to the effective use of performance assessment to enhance teaching quality and student learning outcomes in Korea.

Applicable levels: Secondary

REFERENCES

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior.* Englewood Cliffs, NJ: Prentice Hall.

Bang, I. J., & Chun, S. Y. (2011). An analysis of high school English curriculum implementation by the high school record rating system. *CNU Journal of Educational Studies, 32*(1), 59–84.

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.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*(2), 81–109.

Brown, G. T., & Remesal, A. (2017). Teachers' conceptions of assessment: Comparing two inventories with Ecuadorian teachers. *Studies in Educational Evaluation, 55*(3), 68–74.

Brown, S. (2005). Assessment for learning. *Learning and Teaching in Higher Education, 1*(1), 81–89.

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

Cheng, L., & Wang, X. (2007). Grading, feedback, and reporting in ESL/EFL classrooms. *Language Assessment Quarterly, 4*(1), 85–107.

Darling-Hammond, L., Adamson, F., & Abedi, J. (2010*). Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning.* Stanford, CA: Stanford Center for Opportunity Policy in Education.

Elshawa, N. R., Abdullah, A. N., & Rashid, S. M. (2017). Malaysian instructors' assessment beliefs in tertiary ESL classrooms. *International Journal of Education and Literacy Studies, 5*(2), 29–46.

Guadu, Z. B., & Boersma, E. J. (2018). EFL instructors' beliefs and practices of formative assessment in teaching writing. *Journal of Language Teaching and Research, 9*(1), 42–50.

Gullickson, A. R. (1984). Teacher perspectives of their instructional use of tests. *The Journal of Educational Research, 77*(4), 244–248.

James, M., & Pedder, D. (2006). Beyond method: Assessment and learning practices and values. *The Curriculum Journal, 17*(2), 109–138.

Kahn, E. A. (2000). A case study of assessment in a grade 10 English course. *The Journal of Educational Research, 93*(5), 276–286.

Kim, H. K., & Yun, H. O. (2015). Promoting teacher-based assessment: Korean English teachers' perceptions of assessment and its applications. *Teacher Education Research, 54*(2), 171–187.

Kim, N. Y. (2017). *The development and validation of an assessment instrument for secondary English teachers' competencies in student assessment.* Unpublished doctoral dissertation, Seoul National University, Seoul.

Kim, N. Y., & Kim, J. R. (2018). A perception study of elementary school English teachers on the English performance assessment. *Teacher Education Research, 57*(4), 529–538.

Kim, S. H. (2017). Analysis of changes in the 2015 revised National English Curriculum. *Modern English Education, 18*(3), 91–111.

Lee, M. (2018). A study of high school English teachers' teaching English speaking and performance assessment. *Journal of the Korea English Education Society, 17*(1), 107–126.

Lee, S. (2008). A study on how secondary English teachers in Incheon and Seoul perceive and carry out performance assessment. *English Language Teaching, 20*(4), 305–327.

Lee, Y., & Sung, K. (2017). A study on middle school learners' perceptions of English evaluation. *Studies in Foreign Language Education, 31*(2), 49–87.

Lee, W. (2015). Revisions of the national curriculum of English and challenges of English education. *English Teaching, 70*(5), 35–52.

McNamara, T. F. (1996). *Measuring second language performance.* London: Longman.

Min, K. (2021). *Korean English teachers' beliefs about performance assessment and their practices.* Unpublished master's thesis, Seoul National University, Seoul, Korea.

Ministry of Education. (2022). *2022 revised National English Curriculum* (2022–33–14). Sejong, Korea: Ministry of Education.

Muñoz, A. P., Palacio, M., & Escobar, L. (2012). Teachers' beliefs about assessment in an EFL context in Colombia. *Profile Issues in Teachers Professional Development, 14*(1), 143–158.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Park, T. J. (2016). Examining how performance assessment in English is being implemented in high schools and how it is perceived by teachers and students. *Secondary English Education, 9*(3), 77–98.

Park, T. J., & Chang, J. H. (2017). A study on high school teachers' opinions about issues of communicative English teaching and assessment. *Korean Journal of English Language and Linguistics, 17*(4), 839–863.

Pishghadam, R., Adamson, B., Sadafian, S. S., & Kan, F. L. (2014). Conceptions of assessment and teacher burnout. *Assessment in Education: Principles, Policy & Practice, 21*(1), 34–51.

Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education, 27*(2), 472–482.

Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research, 37*(2), 179–182.

Serpil, Ö. Z., & Derin, A. T. A. Y. (2017). Turkish EFL instructors' in-class language assessment literacy: Perceptions and practices. *ELT Research Journal, 6*(1), 25–44.

Winterbottom, M., Brindley, S., Taber, K. S., Fisher, L. G., Finney, J., & Riga, F. (2008). Conceptions of assessment: Trainee teachers' practice and values. *The Curriculum Journal, 19*(3), 193–213.

Wren, D. G. (2009). *Performance assessment: A key component of balanced assessment system.* Virginia Beach, VA: Virginia Beach City Public Schools, Department of Research, Evaluation, and Assessment.