Examining Student-completed Teacher Evaluation and

Stakeholder Perception in a Korean EFL University Context

Submitted by Kyung Rhan Kim to the University of Exeter
as a thesis for the degree of
Doctor of Education in TESOL
In April 2021.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

Signature ……

ABSTRACT

Student-completed teacher evaluation forms (TEFs) are used by many educational institutions as an efficient way to assess teaching performance. However, both researchers and practitioners have raised doubts and concerns about the way they are used. This includes the validity of student-completed TEFs and the possible adverse effects they have on teachers due to their high-stakes nature.

The aim of this thesis is to identify and examine underlying assumptions about the student-completed TEF used to evaluate native English speaking teacher (NEST) performance within a Korean EFL university context. These assumptions are considered from three stakeholder perspectives: students, NESTs, and administrators. This study seeks to gain insight on the similarities and differences in stakeholder perceptions regarding the TEF in terms of its general purpose, usefulness, and preferred use, as well as opinions on the best way to evaluate NESTs. Questionnaires, semi-structured interviews, and focus groups are conducted with four students, four NESTs, and four administrators.

This study suggests that the current, student-completed TEF used to evaluate NESTs is considered to be insufficient and inadequate by all stakeholders. Using a grounded approach to analyze the qualitative data, findings show that all stakeholder groups perceive a) a gap between the TEF's intended and actual purpose b) a lack of TEF usefulness and validity c) a strong preference for TEF formative use and d) a need for a more tailored TEF to evaluate NESTs. These views suggest the creation and implementation of a comprehensive NEST evaluation system rooted in the needs of both students and NESTs. In addition, certain critical, context-specific, and neoliberal factors appear to augment the negative effects of student-completed TEFs on NESTs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

On the surface, teacher evaluation appears to be a fairly straightforward concept. It is a process or system that educational institutions around the world engage in to assess teachers and to make necessary improvements and decisions within their contexts. Teacher evaluation is conducted at all levels and fields of education through a variety of methods, but one of the most commonly used ways to review teacher performance is through the use of student-completed teacher evaluation forms (TEFs), especially within higher education (HE).

Student-completed TEFs have long been used to assess teaching practices in HE classrooms. Most consist of generic survey forms distributed during and/or at the end of the semester as a way for students to reflect and give feedback on what they learned in their classes. The resulting quantitative and qualitative data are analyzed and generally used for formative and summative purposes. Teachers can use TEF data and feedback in a formative way to improve future classroom practice, while administrators generally use the TEF in a summative manner to evaluate and rank overall teacher performance.

Although the purpose of student-completed TEFs may seem relatively simple, the implementation and use of such forms might be more complicated than we assume. For instance, the content and purpose of each TEF item, the way each item is understood or perceived by stakeholders, and the context in which the TEF is conducted are all factors that can significantly affect the resulting data. In particular, factors influenced by cultural contexts and different languages could potentially lead to critical misunderstandings about student-completed TEFs. For these reasons, as a native English speaking teacher (NEST) at a Korean university, I have come to question the purpose of student-completed TEFs and, more specifically, how they are used to evaluate NESTs within the HE EFL context in which I am currently situated.

In this chapter, I will provide the study context, the rationale that explains why I find it important to examine student-completed TEFs within my professional context, the research

questions of this study, and a brief summary outline of the remainder of this thesis.

1.1 Study context

In Korea, educational attainment has been generally linked to occupational and social status as well as high income earning (Robinson, 1994; Lee & Brinton, 1996). It is the main motivation for Koreans to pursue higher education (HE) (Kang, 2015). This became especially so after the International Monetary Fund (IMF) financial crisis in 1997 that deeply impacted Korea's education reform and economy. During this time, 'globalization' became a buzzword that sent Korean parents into a flurry of action to help their children achieve better social status through higher education. It marked the beginning of what many call 'education fever,' which is highly pervasive in Korean society (Lee, 2005; Lee, Lee & Jang, 2010). This can be seen in the dramatic increase in Korea's participation rate in higher education, which went from 6% in 1960 to 65% in 2007 (Korean Educational Development Institute, as cited in Chae & Hong, 2009).

English education fever, in particular, quickly spread when the use of English was equated with globalization and paired with educational reform, economic growth, and status recognition (Park, 2009). This could be seen both within academic and social settings. For instance, English is one of the main subjects in the national College Scholastic Aptitude Test (CSAT). Students receive rigorous CSAT-centered English education in high school, while their parents invest in private English lessons to help their children receive high CSAT scores in order to attend the best university possible.

In addition to the "premium" of having a four-year college degree (Lee & Brinton, 1996, p. 181), the level of one's English has also been viewed as high premium and a measure of one's competence or performance indicator at many Korean companies (Koo, 2007). This perception and the great desire to gain employment upon graduation led to a sharp increase in investment regarding English education as Koreans noticed that one's English proficiency could potentially yield a 20% income difference (Koo, 2007). As a result, because English

was perceived as social, economic, and global cultural capital, English education in Korea became a US$10 billion per year industry by 2007 (Ministry of Education (MoE), as cited in the 2013 Statistics Korea report). Indeed, the use of English is second only to Korean in South Korea (Jo, 2008; Jeon, 2009), and at the tertiary level, there has been an increasing trend for English as a medium of instruction (MoI) across all subjects (Lee, 2010; Cho, 2012; Kang, 2012).

However, despite this rapid expansion of higher education and perceived need for English, Korea has recently begun experiencing a steady decrease in university student enrollment resulting in the lowest rate of admissions ever (Kim, 2008). According to the MoE (2005), enrollment in HE institutions will drop from approximately 3,278,000 in 2000 to 2,336,000 in 2020 (as cited in Kim, 2008). These numbers are projected to fall even further to 1,511,000 by 2030, which could lead universities that are highly dependent on student tuition fees to go into survival mode. Because of these financial reasons and still-high demand for quality English education, it seems that many Korean universities and postgraduate schools hire an unusually high proportion of foreign faculty members (Kim, 2008). Expatriate native English speaking teachers (NESTs) are also recruited as a “core globalization strategy” to attract both local and international students in an attempt to improve enrollment and global ranking for these universities (Cho, 2012, p. 19).

As English has been mostly taught in preparation for the CSAT rather than as a means of communication, many Koreans still have difficulty speaking English (Kim, 2001; Chun, 2014). Therefore, in addition to hiring NESTs for globalization and financial reasons, the Korean MoE began recruiting NESTs in order to improve students’ English proficiency (Kwon, 2000). It should be noted, however, that the majority of these NESTs were hired regardless of having any previous teaching experience or proper qualifications (Han, 2005; Chun, 2014; Howard, 2019). This is because NESTs are automatically viewed by Koreans as competent and ideal English language teachers (Park, 2006; Wang & Lin, 2013; Howard, 2019).

My university has also employed a substantial number of NESTs to teach English language classes. In order to ensure that NESTs are performing to standard, administrators mainly use student-completed TEFs to assess classroom teaching practices and rate overall performance. While this is a commonly used method to evaluate teachers at all levels of education, I have a number of reservations regarding the use of student-completed TEFs specifically in relation to NESTs within a Korean university EFL context.

1.2 Study rationale

My interest and concern in student-completed TEFs are based on existing literature as well as observations and experiences within my own professional practice as both a NEST and the head of the Professional Development Committee (PDC) for the General English Program (GEP) at my university. Numerous studies have been conducted on the need for and use of reliable teaching evaluation methods. Student-completed TEFs in particular have been extensively examined by many researchers as they are one of the most commonly used tools to measure and assess teacher performance, especially in higher education (Penny, 2003; Wines & Lau, 2006; Burden, 2008a, 2008b, 2010; Thomas & Yang, 2013; Śliwa & Johansson, 2014). Data gathered from these TEFs can be used in either a formative or summative manner.

Summative evaluation is generally related to administrative matters, such as job, personnel, or school status decisions (Darling-Hammond et al., 1983, as cited in Bailey, 2009), while the formative use of TEFs can help teachers to improve their pedagogical practices and curriculum quality. By giving teachers the opportunity to professionally develop themselves and their classroom practices through the use of evaluation data, this can have a positive effect on student learning outcomes (Danielson & McGreal, 2000; Yoon, Duncan, Lee & Shapley, 2008; King, 2014). Formative assessment data can also help to identify program characteristics that can inform decisions on improving or further developing the program's effectiveness (Patton, 2008). Therefore, the formative use of student-completed TEFs can

be beneficial for not only teachers but also students and administrators.

However, many studies have called attention to the risks of using student-completed TEFs especially with regard to faculty evaluation at university or college level. For instance, in their study, Wines and Lau (2006) found that the vast majority of the questions contained on student evaluation forms were unclear or unrelated to classroom teaching performance. This brought to light the undesirable effects of using non-validated evaluation questions, such as teachers tailoring their speech and pedagogy to maintain high evaluation scores for the purpose of obtaining promotion or tenure. Further effects involved reduced teaching quality, lack of professional development, and subsequently, the possibility of diminished student learning outcomes. Other negative outcomes of using student-completed TEFs include fear and anxiety that teachers experience due to a perceived lack of power in such situations as TEF scores are frequently linked to job security (Burden, 2008a). In addition to administrators making decisions related to hiring and contract renewal/termination, TEF scores are also used to give bonus monetary incentives (Pennington & Young, 1989). However, despite the body of literature criticizing this narrow use of student-completed TEFs, most universities around the world continue using them as the core or sole criterion for summative evaluation.

Like this, the high-stakes use of student-completed TEFs to evaluate one's teaching competence and performance can be observed at my university as well. As head of the PDC, my responsibilities include creating opportunities for GEP NESTs to work collaboratively in order to communicate in a more effective and productive manner, reflect positive changes in the curricula, and thus help improve student performance. One way to achieve these goals is to find ways for NESTs to engage in reflexivity both on an individual level and within a group dynamic regarding our teaching practices. At minimum, this requires the formative use of evaluation feedback.

Unfortunately, the only feedback GEP NESTs are given or have access to are a set of

student-completed TEF scores and two qualitative comments on the form. This is particularly problematic for NESTs because of the lack of transparency regarding teacher evaluation criteria in conjunction with language barrier issues. That is, the TEF cannot be effectively used in a formative way to help the majority of NESTs because all of the content is available only in Korean. Since most of the qualitative comments are also written in Korean, this creates further difficulty for NESTs in terms of understanding how students feel about the curriculum and their experiences within the classroom. These language issues and the accompanying lack of information, therefore, restrict the opportunity for NESTs to effectively use the TEF.

In addition, the constant pressure from university administrators during faculty meetings to improve TEF scores rather than pedagogical practices could lead many GEP NESTs to feel a lack of legitimacy, agency, and autonomy, which are important factors associated with being professional educators (Nixon, 1996; Beck &Young, 2005; Jeon, 2009). For these reasons, it seems that at my university, teacher evaluation and professional development demonstrate a somewhat mutually exclusive relationship rather than one that is constructive and reciprocal.

All of this led me to consider a number of questions about the current TEF and its use in relation to GEP NEST faculty members:

- Why is a student-completed TEF the only method used to evaluate teachers/NESTs? Is this sole criterion sufficient to determine the value and worth of teacher knowledge and pedagogical practice? Are there other evaluation methods of which NESTs are not aware?
- Why are NESTs encouraged to improve their TEF scores yet not provided with the necessary information to understand the form?
- How do students understand the purpose of the TEF and its items? Do they believe that the TEF adequately addresses their language learning goals?

- How do NESTs view the TEF? In what ways to do they find the TEF useful or not useful in improving their teaching practices?
- How do administrators perceive the TEF? Why do they use only the final numerical value of a high-stakes TEF to determine the effectiveness of NESTs whom students have had for only one semester?
- Are there any particular factors or circumstances that influence these stakeholders' understanding or perceptions of the TEF?

In addition to the personal and professional reasons I have stated, I believe it is important to address the gap in literature regarding empirical studies focusing on student-completed TEFs and NESTs within HE EFL contexts. Despite the large body of literature about teacher evaluations, much of this research pertains to literature review or observations of teacher evaluation in general (Pennington, 1989; Theall & Franklin, 1991; Rea-Dickens, 1994; Theall & Franklin, 2001), within ESL contexts (Master, 1983; Pennington & Young, 1989; Wennerstrom & Heiser, 1992), within EFL contexts but at a primary or secondary level (Al-Mutawa, 1997; Al-Mutawa & Al-Dabbous, 1997), or involves differences in student perceptions of NESTs and non-native English speaking teachers (NNESTs) (Chun, 2014; Kasai, Lee & Kim, 2011; Śliwa & Johansson, 2014).

In particular, I find it necessary for more studies of student-completed TEFs and NESTs to be conducted within a Korean university EFL context. One reason is because most universities use student-completed TEFs to evaluate all of their teachers regardless of the subject matter. Another reason involves the large and active presence of an English education industry in Korea, which is mostly geared towards preparing high school students for competitive college/university entrance exams and subsequent tertiary level English classes (Sorensen, 1994; Koo, 2007). This market has led to the propensity of HE institutions hiring more and more NESTS to teach EFL classes. Employing NESTs as English language teachers is considered advantageous because all NESTS are seen as

competent and more effective than NNESTs despite the lack of rationale for this belief (Han, 2005; Park, 2006). However, it is rather troubling to see how NESTs are generally perceived and regarded in Korea because “expertise, linguistic awareness, and cultural awareness should be prerequisites for being a competent English teacher” (Wang & Lin, 2013), yet there is little evidence to show that NESTs are being hired based on these qualities.

It is also puzzling to see a significant lack of studies evaluating NESTs and their teaching performance or effectiveness through the use of multiple and valid assessment methods considering how much emphasis is placed on recruiting NESTs. There are some qualitative studies examining Korean student perceptions of effective EFL teacher characteristics (Park & Lee, 2006; Kasai et al., 2011; Barnes & Lock, 2013; Chun, 2014) and student-completed TEF studies conducted in other East Asian EFL university settings such as Japan and Taiwan (Burden, 2008a, 2008b, 2010; Thomas & Yang, 2013). However, I have not been able to find any specific or significant empirical or theoretical research related to student-completed TEFs and NESTs within a higher education Korean EFL context.

1.3 Research questions

This study therefore aims to examine and explain underlying assumptions about the student-completed TEF used to evaluate NEST performance at a Korean university from three stakeholder perspectives: students, NESTs, and administrators. In order to gain insight on how these stakeholders view the TEF and its use, this study will be guided by the following primary question:

- What are the similarities and differences among stakeholder group perceptions of using a student-completed TEF to evaluate NEST performance and the factors influencing these perceptions?

In order to gain insight on this primary question, study participants are asked to share their perceptions on the following research questions:

1. What do they perceive to be the general purpose of the TEF and its items?
2. In what way do they consider the current TEF useful or not useful?
3. What is their most preferred way to use the TEF?
4. What do they consider to be the best or ideal way to evaluate NESTs?

1.4 Thesis structure

This thesis is presented in six chapters. Chapter 2 will provide a review of literature that outlines how my primary question relates to what is already known in the field of teacher evaluation. Chapter 2 also includes literature summary charts to help better recognize gaps in this area of research. Main concepts and relevant keywords for this study were established first. Then empirical and theoretical studies related to these constructs were located and organized with a focus on context to describe current trends in teacher evaluation within higher education EFL, its methods, and the stakeholders involved. Chapter 3 describes and justifies my chosen methodology and outlines the data collection and analysis methods I used with reference to appropriate research methodology literature. It also considers the ethical implications of each method. Chapter 4 presents the study findings and consists of a detailed write up of my data analysis. Chapter 5 provides a discussion of these findings. It explains how the findings answer the research questions and how they relate to existing literature. Finally, in Chapter 6, I will give the conclusion of my thesis that includes the main implications of the findings, its contribution to knowledge, the limitations of the study, and recommendations for further research, practice, and policy. Throughout this chapter, I will also reflect on how I was affected by the study on both a personal and professional level.

CHAPTER 2: LITERATURE REVIEW

Evaluating teachers has long been a difficult and complex matter for stakeholders at all levels of education (Pennington & Young, 1989; Bailey, 2009; Isoré, 2009). Numerous studies have been conducted to measure and assess the state of educators and their teaching competencies at primary and secondary levels (Al-Mutawa, 1997; Al-Mutawa & Al-Dabbous, 1997; Isoré, 2009), tertiary level (Wennerstrom & Heiser, 1992; Wines & Lau, 2006; Burden, 2008a, 2008b, 2010; Thomas & Yang, 2013; Śliwa & Johansson, 2014), and across all fields of education and contexts (Al-Mutawa, 1997; Danielson & McGreal, 2000; Kang & Hong, 2008; Isoré, 2009; Looney, 2011; Alamoudi & Troudi, 2017) . These studies show a need for a sound teacher evaluation system involving the collection and interpretation of constructive information that can provide opportunities for teachers to professionally develop themselves and their curriculum.

They also suggest that the existence of a well-implemented teacher evaluation system can have a positive influence on student learning outcomes as well as teacher practice, professional development, and administrative matters. It is, however, also important to consider underlying factors and policies that can affect such teacher evaluation methods and systems, particularly student-completed teacher evaluations. These factors include how the field of education is currently perceived and how educational institutions are operated in a more neoliberal global and business-oriented manner.

This chapter will review in detail teacher evaluation and explore the role of neoliberalism in English language education. Each section of this chapter defines and then states the purpose, issues, and effects of these concepts with regard to this study. In addition, teacher evaluation and neoliberal influences are discussed within a Korean university EFL context in relation to the primary question: what are the similarities and differences among stakeholder group perceptions (i.e. students, NESTs, administrators) of using student-completed teacher

evaluation forms (TEFs) to assess NEST performance, and what factors influence these perceptions?

2.1 Teacher evaluation

2.1.1 Purpose and definition of teacher evaluation

Teaching is a “personal, social, complex activity” (Howard & Donahue, 2015, p. 1) and as such, the proper assessment of one’s teaching quality and competency is a multifaceted process. Teaching quality is a seemingly obvious yet vague term often used interchangeably with “teaching performance” or “teaching effectiveness” with regard to student-completed surveys or questionnaires. Factors associated with “good” teaching quality and its evaluation in higher education include providing feedback, clear goals, appropriate workload, appropriate assessment, positive learning outcomes, organization and presentation, interaction, and motivation (Ramsden, 1991; Guolla, 1999; Kember, Leung & Kwan, 2002; Ginns, Prosser & Barrie, 2007).

Examining teacher evaluation is extremely important because its goals and procedures help to define the standards of the changing profession of ESL (Pennington & Young, 1989) and language teaching in general (Bailey, 2009). Some perceive teacher evaluation as a process or system that assesses the quality and effectiveness of a teacher’s performance by placing value upon each aspect that is examined (Looney, 2011). Patton (2008) described teacher evaluation as a “systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding” (p.38).

Spooren, Brockx and Morelmans (2013) further comprehensively defined the purpose of teacher evaluations as “(a) improving teaching quality, (b) providing input for appraisal exercises (e.g. tenure/promotion decisions) and (c) providing evidence for institutional

accountability (e.g. demonstrating the presence of adequate procedures for ensuring teaching quality; Kember et al., 2002)" (p. 599). Here, accountability can also be understood as "the answerability of staff to others for the quality of work" (Weir and Roberts, 1994, as cited in Burden, 2008a, p. 479). Therefore, based on existing literature and for the purpose of this study, teacher evaluation is understood as the comprehensive and systematic process of assessing teacher knowledge, performance, and effectiveness with the aim of improving teaching quality and demonstrating accountability.

Teacher evaluation is generally used for formative and summative purposes (Danielson & McGreal, 2000; Wines & Lau, 2006; Patton, 2008; Isoré, 2009; Bailey, 2009; Looney, 2011; Alamoudi & Troudi, 2017). While formative assessments are intended to provide teachers opportunities to professionally develop themselves and their classroom practices (Nunan & Lamb, 1996; Danielson & McGreal, 2000; Alamoudi & Troudi, 2017), summative evaluations are important for administrators because "educational institutions around the world are under increasing pressure to adopt reliable and cost effective measures and instruments to assess and evaluate teacher's knowledge and competence in the classroom" (Troudi, 2009, p.60). That is, teacher evaluation is a system involving individual teacher development such as skill and knowledge improvement as well as administrative matters related to job, personnel, or school status decisions (Darling-Hammond et al., 1983, as cited in Bailey, 2009).

From this, one can infer that the purpose of teacher evaluation and the needs it should serve is two-fold: improvement and accountability (Danielson & McGreal, 2000; Penny, 2003; Burden, 2008a). It is expected that these two aims could have a positive effect on student learning outcomes because the intent behind implementing well-designed teacher evaluation systems is to provide opportunities of professional growth that encourage individual teacher development which meets student needs. This, in turn, can raise student achievement, and could thereby increase the overall quality of educational institutions.

2.1.2 Assessment tools and methods

Ideally, teacher evaluations involve the use of multiple assessment tools and methods for both improvement (formative) and accountability (summative) reasons. Formative methods include self-evaluation and reflection (Delandshere & Petrosky, 1994; Al-Mutawa, 1997; Ropers-Huilman, 1999; Troudi, 2009; Howard & Donaghue, 2015; Almamoudi & Troudi, 2017), collaborative work such as peer evaluation (Delandshere & Petrosky, 1994; Ropers-Huilman, 1999; Almamoudi & Troudi, 2017), and feedback meetings (Troudi, 2009; Howard & Donaghue, 2015).

In addition to such qualitative methods, teacher evaluation systems can include external assessment tools and quantitative methods for summative purposes, such as classroom observations (Master, 1983; Pennington & Young, 1989; Troudi, 2009; Howard & Donaghue, 2015), teaching competency indices (Al-Mutawa & Al-Dabbous, 1997), competency tests (Pennington & Young, 1989), student achievement (Pennington & Young, 1989; Looney, 2011), and student-completed evaluations (Theall & Franklin, 1991; Wennerstrom & Heiser, 1992; Theall & Franklin, 2001; Penny, 2003; Burden, 2008a, 2008b, 2010; Alamoudi & Troudi, 2017) at school, regional, or national levels.

With regard to teacher evaluation, “[a]ny system for teaching evaluation needs to be tied to a clear set of standards and competences” (Looney, 2011, p. 441) as well as fair and reliable criteria (Isoré, 2009). However, such systems are difficult to achieve for a number of reasons. For example, most educational institutions show a tendency to value ease and efficiency when collecting teacher evaluation data for administrative purposes, which include ranking teachers for promotion or job retention (Penny, 2003; Corbin & Strauss, 2015). Therefore, there appears to be a general propensity to use a single, student-completed, generic questionnaire or survey form that mainly consists of scaled, cross-curricular items. These items fall in with what are considered good teaching qualities, and students are generally required to select scaled, numerical values that best correspond with their course experience. However, it is questionable as to whether one’s teaching quality or performance

can indeed be accurately measured or evaluated in this manner. It is important to consider that issues stemming from the sole use of a TEF can be compounded when taking into consideration influencing factors such as context and language.

2.1.3 Issues and effects of teacher evaluation

Teacher evaluation has long been a complex and contentious topic (Penny, 2003; Bailey, 2009) and “controversial and ill-defined process in education” (Pennington & Young, 1989, p.619). One of the main perceived issues with teacher evaluation include the use of metric forms to assess one’s teaching quality because many of these student-completed questionnaires are specifically designed with a focus on faculty appraisal rather than for formative purposes or helping educators feel that their teaching is valued (Kember et al., 2002). In addition, while some studies show insignificant impacts of gender or age on student evaluation of teaching effectiveness (Tran & Do, 2020), others strongly suggest that these factors, along with race and/or ethnicity, could have a bias effect towards teachers. For example, studies indicate that young male professors are rated more favorably than their older male or female colleagues (Arbuckle & Williams, 2003; Macnell, Driscoll & Hunt, 2014; Joye & Wilson, 2015) and that females and persons of color receive lower scores than white male instructors (Chávez & Mitchell, 2020).

Other weaknesses with teacher evaluation methods and policies include the lack of consistency regarding timely formative and summative assessments (Nunan & Lamb, 1996; Danielson & McGreal, 2000; Isoré, 2009; Looney, 2011), inadequate evaluation systems (Newton & Braithwaite, 1998; Looney, 2011), ineffective evaluation methods and tools (Looney, 2011), and the disparity of perceived value between formative and summative evaluation. Although each type serves its own fundamental purpose, it is difficult to see find studies on a single system that could satisfy both aims of teacher evaluation. Educational institutions seldom place equal value on formative and summative assessments; rather, administrators seem to apply integrated models of evaluation where summative evaluation

often carries more weight and is “a final assessment, a make-or-break decision at the end of a project or funding period” (Bailey, 2007, as cited in Alamoudi & Troudi, 2017, p. 33). The use of student-completed TEFs in particular in this manner by administrators is particularly problematic for NESTs for a number of reasons. Its high-stakes nature, lack of transparency regarding the criteria, and sociocultural factors such as language and cultural barriers could significantly affect the way NESTs accept and understand it. As a result, critical questions have been raised regarding the validity and reliability of student-completed TEFs and their effects on teacher professional development.

Validity and reliability concerns

One of the major concerns regarding student-completed TEFs involves the validity and reliability of student responses, particularly with regard to the use of metric forms to evaluate teaching performance. Penny (2003) observed “strong empirical support for the validity of student ratings as one indicator of teaching quality” with the use of multiple “well-constructed and validated instruments” (p. 401), but that student ratings and evaluation scores can be no more valid than the instrument used to collect the information. However, many researchers and educational institutions appear to equate student opinions with factual knowledge and thus regard student-completed TEFs as valid sources of data that can accurately and effectively assess teaching performance (Spooren et al., 2013).

For instance, Guolla (1999) discussed the use of such metrics to evaluate teaching performance in close relation to student satisfaction and saw it as a “fruitful method” of acquiring pertinent information to improve one’s teaching (p. 95). Ramsden (1991) stated that teaching quality could be effectively evaluated through performance indicator (PI) metrics, such as the Course Experience Questionnaire (CEQ), due to the strength of the CEQ scales. He further stated that potential bias in student-completed evaluations could be minimized if “exacting controls over methods of administration are imposed” and if “students are asked only about those aspects of teaching which they are qualified to comment on” (pp. 131-132).

Like this, while some researchers do not observe any issues with the use of scaled metrics to assess one's teaching quality or performance, others such as Ginns, Prosser and Barrie (2007) supported the use of a modified CEQ as a PI, but they emphasized that the form was not designed to gather specific feedback about certain individual subjects or teachers and their practices. Instead, they supported the use of such scores at the faculty or administrative level (e.g. competitive performance-based funding, administrators initiating meaningful dialogue with departments or programs in order to address issues).

In addition, Kember et al. (2002) identified a number of concerns with this evaluation method. In their study based on student perceptions, they failed to find evidence that student-completed feedback questionnaires helped improve the overall quality of teaching and learning. They stressed that a general correlation between student feedback and teaching quality improvement did not imply causality unless there was a "[statistically] significant rise in ratings and when the feedback is accompanied by counseling or improvement activities" (p. 413), which could only be achieved through a thorough examination of scores over an extended period of time across the entire university.

Kember et al. (2002) and Tran and Do (2020) also pointed out the now-common requirement for universities to conduct faculty appraisal quality reviews in order to show that there are sufficient measures in place to ensure an acceptable level of teaching quality (i.e. as a way to demonstrate accountability). Therefore, the assumption that student feedback opinions are accurate, measurable sources of knowledge is cause for concern because rarely are TEFs created with validity and reliability in mind; rather, studies have suggested that university administration's main consideration is the time and cost efficiency in which the TEF data can be collected and analyzed (Penny, 2003; Troudi, 2009; Corbin & Strauss, 2015).

A critical issue with teacher evaluation is the lack of consistency, and thus reliability, when conducting them. Reliability involves the "dependability, consistency, replicability over time,

over instruments, and over groups of respondents" (Cohen et al., 2007, p. 146). However, despite the general consensus regarding the need to assess and evaluate teacher performance both in a formative and summative manner, this does not seem to occur with the expected frequency that accompanies such administrative policies. Studies show wide discrepancies between teachers' beliefs and perceptions of teacher evaluation and their actual evaluation experiences (Newton & Braithwaite, 1988; Al-Mutawa, 1997) and significant differences between student and teacher views of student-completed teacher evaluations (Sojka, Gupta & Deeter-Schmeiz, 2002), as well as a gap between stated and actual aims (Wines & Lau, 2006; OECD, 2005a, 2009a, as cited in Looney, 2011).

For instance, Nunan and Lamb (1996) observed that although teacher evaluation and "external supervision" are "mandatory aspects" for many teachers with regard to their employment, some are not formally evaluated at all, much less by their peers or colleagues, in any collaborative fashion (p.238). In addition, teachers in many countries report that individual teacher evaluations are not conducted systematically, evaluators are untrained, methods and tools are ineffective, and that these evaluations may not provide timely feedback or detailed information on classroom practices (OECD 2005, 2009, as cited in Isoré, 2009). Such inadequate and disappointing evaluation experiences indicate a high probability of both flawed reliability and validity in most teacher evaluation methods.

Validity generally refers to how accurately a method or instrument measures what it is intended to measure (Spooren et al., 2013). Building on previous conceptual models, Onwuegbuzie et al. (2009) developed a meta-validity framework model for the purpose of assessing construct, content, and criterion-related validity (as cited in Spooren et al., 2013). They described each type of validity as:

> "[Construct validity is] the extent to which an instrument can be seen as a meaningful measure of a given characteristic… [content validity is] the extent to which the items of an instrument are appropriate representations of the content being measured… [Criterion validity is] the extent to which scores are

> related to another independent and external variable that can serve as a direct measure of the underlying characteristic" (p. 601).

For the purpose of this study, TEF validity is more holistically regarded as the correctness of the findings based on how accurately the TEF items reflect what they purport to measure.

In addition to the concern that the use of metric forms cannot accurately assess teaching quality or performance, there are a number of other validity concerns regarding the use of student-completed teacher evaluations. While some studies and researchers seem to support the focus on students and student satisfaction levels when evaluating teachers and teacher effectiveness in higher education (Nasser & Fresko, 2002; Penny, 2003), Wines and Lau (2006) called attention to the "folly of using student evaluations of college teaching for faculty evaluation" because research indicated that "approximately 80% of the questions contained on student evaluation forms either are unclear, subjective, or ambiguous or are unrelated to classroom teaching performance" (pp.177-178). It is therefore highly plausible that such vague or confusing wording of TEF items, as well as their lack of relevance to effective teaching, can lead to a high degree of variance in interpretations and inaccurate responses (Block, 1998; Penny, 2003; Spooren et al., 2013), thus leading to TEF validity issues.

In addition to validity concerns related to lack of clarity and relevance of TEF wording, external variables can negatively affect TEF responses. For instance, studies have shown that critical factors such as fear and anonymity can influence TEF student feedback accuracy and quality (Gordon & Stuecher, 1992; Svinicki, 2001). Such factors are based on critical perspectives of power, knowledge, and language. (Agger, 1991). They relate to the whole structure of society regarding education and the "access, power, disparity, desire, difference" within (Pennycook, 2001, p. 6) and emphasize the need for collaborative reflexivity and meaningful dialogue in order to overcome these issues (McLaren et al., 2009; Pennycook, 2001). From this perspective, knowledge is viewed as fluid and dynamic and

shaped or contextualized by social, cultural, historical, and political values (Agger, 1991; Ponterotto, 2005). The discursive resources that study participants draw on in terms of their social and/or cultural positioning and responses (Agger, 1991) can also affect the validity of TEF results, so it is essential to consider such critical and contextual aspects when reviewing TEF data.

Moreover, a lack of knowledge or guidance on how to complete the TEF in an effectual manner, as well as why it is important to do so, can also have a damaging effect on all stakeholder groups. For example, NEST participants in a study conducted at a Japanese university expressed concern that such absence of awareness in knowledge could lead to students providing “non-committal responses further denuding the value of [TEF] feedback” (Burden, 2008a, p. 484). That is, failing to provide students with necessary information about the TEF in a detailed and systematic way can lead to them giving feedback that is low in both quantity and quality, which can lead to poor TEF validity. This aligns with observations made by Dunegan and Hrivnak (2003) that failing to present students with sufficient guidance could lead to “mindless” evaluation behavior where students respond in an automatic sense rather than a reflective and meaningful manner.

Furthermore, in a study conducted by Simpson and Siguaw (2000), they observed that without proper guidance on how to complete the TEF, some students might perceive it as a tool for revenge or payback against the teacher. Responses based on these perceptions would not correspond with the TEF’s intent and therefore would lead to flawed validity. This lack of guidance and knowledge could also adversely affect administrators. For instance, Franklin and Theall (1989) found that administrators responsible for using student-completed TEFs to make personnel decisions lacked relevant knowledge of the instrument. This led to them frequently making critical errors in interpreting the data, which was cause for great concern due to the TEF’s high-stakes nature and potentially serious effects on teachers.

There are additional resulting adverse effects from using non-validated evaluation questions, failing to take into account the distinctive characteristics of language teaching in specific contexts (Borg, 2006), or not considering other external influencing factors when assessing teachers (Neumann, 2001; Lee, 2010). Wines and Lau (2006) observed a "near-obsession with student teaching evaluations [that] has led [college] faculty to carefully tailor their pedagogy towards maintaining evaluations that continue employment rather than educate students" (p.209) and thus "cause faculty members to tailor their speech to obtain promotion and tenure" (p.180). Such situations could lead to a further decrease in teaching quality and student learning. This unfortunate observation brings attention to another critical issue with using student-completed TEFs: the negative effects on teacher professional development.

Effects on professional development

It is important to understand professional development in relation to teacher evaluation and how it can influence one's practice. There are many studies trying to determine how professional development can be defined to encompass most educational contexts and situations. Essential keywords related to this construct include legitimacy, agency, autonomy, professional learning, collaborative effort, and reflection (Delandshere & Petrosky, 1994; Crookes et al., 1995; Nixon, 1996; Johnson, 2000; Beck & Young, 2005; Jeon, 2009; Tang & Choi, 2009; Troudi, 2009; Murray, 2010; Looney, 2011; Sim, 2014; Howard & Donaghue, 2015).

Professional development can be understood as a set of practices involving "knowledge and skills and understanding of individuals or groups in learning contexts that may be identified by themselves or institutions" (Nicholls, 2000, p. 371) with much collaborative effort (Ropers-Huilman, 1999; Troudi, 2009; Looney, 2011). Kelchtermans (2004) summarized this concept as "a learning process resulting from meaningful interaction with the context (both in time and space) and eventually leading to changes in teachers' professional practice (action) and their thinking about their practice" (p. 220). For the purpose of this study, professional development will be understood as a temporal, spatial, and evolving learning process in

which educators attempt to enhance the quality and standards of their practice. This practice is characterized by autonomy and meaningful collaborative effort within their communities of practice.

In previous years, the concept of professional development involved teachers who taught for the sake of education itself and who had pedagogical freedom and autonomy over curriculum development and decision-making without being hampered by external constraints and pressures by their educational institutions or the government (Hargreaves, 2000). Researchers have discussed professionalism as occupational values worth working for and that are marked by factors such as autonomy (Nixon, 1996; Hargreaves, 2000; Evetts, 2012) and trust and confidence between and among practitioners and employers (Helsby & McCulloch, 1997, as cited in Hargreaves, 2000; Beck & Young, 2005; Evetts, 2012). Understanding how these values relate to practice is key to developing one's teaching abilities and beliefs. This can only occur if all professional development elements are properly aligned and good policies are set in place within an encouraging atmosphere in which teachers feel brave enough to implement innovative practices and take risks that can positively contribute to English teaching and learning (Chisman & Crandall, 2007). It therefore seems only natural that one's teaching performance should be evaluated on such a premise.

Unfortunately, there are a number of issues and challenges that teachers face regarding this matter. Researchers and policymakers acknowledge the necessity of professional development and the evaluation of its impact (King, 2014). Numerous studies suggest positive causal links between teacher professional development, teaching effectiveness, and student outcomes (Yoon et al., 2008; Murray, 2010; Looney, 2011; Çelik, Arikan & Caner, 2013; King, 2014). In addition, there is a connection between inadequate teacher evaluation systems and professional development because the lack of evaluation, formative or otherwise, can lead to weak teaching management (Nunan & Lamb, 1996). However, despite all this, there is often little clarity as to how a balanced teacher evaluation and

professional development system can be achieved in a methodical way, especially at HE institutions. This can be attributed to the absence of research, systematic support, and teacher awareness regarding professional development, which results in a lack of sense in legitimacy, voice, agency, and autonomy that are needed in being recognized and functioning as professional educators (Nixon, 1996; Harvey, 2001; Kelchtermans, 2004; Beck &Young, 2005; Jeon, 2009).

One leading criticism of teacher evaluation in relation to professional development is the school administration's sole reliance on student-completed summative assessment as a means of determining and recognizing teacher performance and competency (Pennington, 1991; Wines & Lau, 2006; Bailey, 2009; Isoré, 2009; Troudi, 2009), rather than concentrating on formative measures for teacher development and quality control. For example, in their study on teacher perspectives on teacher evaluation, Newton and Braithwaite (1988) observed that "teachers were most concerned that assessment should lead to feedback and improvement in their own performance and perceived the existing system to be sadly lacking in this area" and that "there was no clear purpose" of the system (p.285). One possible explanation for the emphasis on summative assessment could be policy-related in that while summative teacher evaluations are required by local and state mandates, formative evaluations are not (Danielson & McGreal, 2000). Such use of student ratings of teaching performance tend to neglect the potential usefulness to both teachers and administrators (Penny, 2003).

There are grave concerns that follow the singular use of summative evaluation, particularly within higher education. For instance, qualitative studies reflect recurring themes of stress and anxiety (Danielson & McGreal, 2000) that many teachers experience due to the way TEF scores are used in decision-making regarding job security, promotions, and pay (Pennington, 1991; Penny, 2003; Bailey, 2009). Teachers also perceive a lack of power in such situations, which could make them feel "afraid… worried… scared… or threatened" (Burden, 2008a, p. 485). Nunan and Lamb (1996) stated "[e]xternal evaluation, particularly

when it is for purposes of certification or continued employment, can be extremely threatening. In fact, it may well be the most anxiety-creating situation the teacher is ever likely to face." (p. 238). Teachers may experience stress to the point where it can lead to "feelings of alienation" (Tang & Choi, 2009, p. 14), "de-humanizing effect" (p. 1), and "performativity culture" where "professional development is more likely to become externally-driven" (p. 15) rather than self-directed through teacher agency. These issues might be heightened for those situated within an EFL context due to context-specific influences.

2.1.5 Teacher evaluation in Korean EFL context

There is much literature related to teacher evaluation in various countries. However, critical or empirical teacher evaluation studies conducted on EFL teachers or NESTs within the Korean context are scarce. There is research on student and teacher perceptions regarding characteristics of effective English teachers (Park & Lee, 2006; Barnes & Lock, 2013) but not on overall performance assessment or evaluation methods. There are also studies on teacher quality, hiring standards, and top-down teacher evaluation policies on a national level for Korean teachers in fields such as mathematics, but the same cannot be said for those teaching EFL (Kang & Hong, 2008). Dickey (2006) pointed out "[i]n the field of teaching English as a foreign language (TEFL), none of the three major international societies (TESOL Inc., IATEFL, and AsiaTEFL) have established ethical guidelines for teachers" (p. 17). The absence of any professional standards or guidelines for NESTs teaching EFL is concerning because it seems to be in direct conflict with the high-stakes nature of student-completed TEFs that are used by most universities. If NEST teaching practices are to be assessed in a summative manner that is commonly associated with employment status, it only stands to reason that specific professional requirements, as well as recruiting and teaching criteria, are established by administrators and upheld by NESTs.

However, there is a significant absence of such professional standards for NESTs in Korea. For example, many NESTs do not come from a teaching background, and 48% of them hold a degree unrelated to education (Smith, 2010, as cited in Howard, 2019). This context-

specific problem could be partially attributed to the strongly misplaced trust that Koreans exhibit toward all NESTs as being capable teachers where language competence is favored over proven teaching ability (Seol, 2012). Park (2006) also noted that the “public faith” in NESTs in Korea has “grown so strong” despite there being no rationale provided for this (p. 128).

This misperception can be observed in the government-sponsored English Program in Korea (EPIK) that hires and places NESTs throughout the public school system. Wang and Lin (2013) found in their study that while there were general rules EPIK NESTs should follow, there was a conspicuous lack of specific NEST recruitment guidelines other than the native English speaking requirement. They also found that at EPIK, “NESTs have to conduct team teaching with local non-native English speaking teachers (NNESTs) and provide professional development for NNESTs” (p. 9). This implies that NESTs are experts who are somehow better than NNESTs and are in automatic possession of high teaching proficiency and professional development methods despite the fact that only 26% of EPIK participants hold teaching certificates (Ahn, Park & Ono, 1998, as cited in Wang & Lin, 2013).

Howard (2019) repeatedly mentioned this “paucity of research directly investigating NEST professional identity in Korea” (p. 1481) in her research. These types of studies suggest that educational institutions do not find professional development necessary for NESTs because their ability to speak English with native fluency is indicative of impeccable teaching qualifications and competency that do not require further improvement. This, perhaps, could explain why it is difficult to find Korean EFL studies on appropriate NEST-specific evaluation methods as there is an absence of professional teaching standards that NESTs are supposed to adhere to and improve upon. However, this is not to say that other countries similar to Korea (i.e. East Asian, heavily invested in EFL education) demonstrate more transparent and structured regulations, policies, or professional standards related to student-completed teacher evaluations.

For instance, in Taiwan, Thomas and Yang (2013) examined HE institutions and went so far as to state that top-down evaluation led to educators being “forced into survival mode to save themselves from punitive measures that may threaten their employment. This Darwinistic evaluation results in a dumbing down of education in Taiwan and decreases faculty morale” (p.117). This echoes observations made by Wines and Lau (2006) where they found that student-completed evaluations steered university faculty members towards the problematic behavior of shaping their speech and pedagogy around maintaining high enough evaluations for continued employment rather than focusing on student learning improvement.

Burden (2008a, 2008b, 2010) and his studies brought to light NEST perceptions of student-completed TEFs at a Japanese university. These TEFs were administered at the end of the semester and were used as the sole criterion for summative evaluation. Results indicated that NEST participants in these studies were generally unaware of the purpose of the evaluation, nor were they asked to provide any input, which pointed at their lack of knowledge and voice in the matter. They also questioned the validity in the TEF for a number of reasons, such as the cross-curricular nature of the TEF, relevance of items to teaching performance, student feedback quality and various external factors such as class size, student attitudes, teacher personality etc. (Burden, 2008a, 2010). Other issues influencing NEST views of the TEF included top-down reform by the Japanese Ministry of Education, managerial views of “institutional accountability,” job security concerns, and “student-as-consumer” views (Burden, 2008a, p. 489).

In addition to the need for NEST-specific TEFs or other assessment methods, it is also important to consider context-specific factors that influence one’s knowledge of student-completed teacher evaluations. For example, Barnes and Lock (2013) stressed the importance of NESTs needing to be constantly aware of less-obvious factors that might cause discord between them and students, such as Korean EFL learners’ resistance to participate in class and their passive learning style. Han (2005) reflected on how Korean

students fear to make mistakes in order to save face in terms of their lack of general conversation or presentation skills. Such factors could have an effect on how students assess NESTs on the TEF, and both teachers and administrators should take this into account when analyzing and using TEF data.

There are also other indications of sociocultural factors affecting NESTs within their context that could, in turn, affect the way they perceive student-completed TEFs and the stakeholders involved. For example, Howard (2019) observed in her study on constructing professional identities that NEST participants at a Korean university experienced feelings of isolation and detachment from both the faculty and its administrators. The NESTs attributed this to feeling “continually conscious of their status as foreigners and [struggling] with language barriers” (p. 1482) as well as the “rigid hierarchical nature of Korean society, which also pervades private and public-sector institutions” where “management rarely listens to the foreign teachers” (p. 1488). It therefore seems plausible that issues stemming from such challenges can prevent NESTs from fully understanding or accepting the criteria used to assess their teaching performance.

Through these studies, we can see that there are systemic issues related to student-completed teacher evaluations. Therefore, there is a need to examine teacher evaluation in a more thorough and systematic way in order to help bridge the gap in teacher evaluation literature for NESTs, especially for those teaching in Korea. In order to do this, it is important to acknowledge and examine other underlying influences affecting NESTs and the use of student-completed TEFs. This includes top-down reform policies, job security fears, and viewing students as consumers within a global English education market. These factors are all part of a market-driven concept called neoliberalism, which many argue is now fully integrated within the field of EFL and especially at a higher education level. The present study will examine this perception of neoliberal influences on EFL and student-completed TEFs for support.

Table 2.1

Summary of empirical teacher evaluation studies

Author(s)	Context	Teacher evaluation focus	Participants	Instruments
Al-Mutawa, (1997)	Primary/EFL	Evaluation of competencies of EFL primary school teachers	Teachers	Evaluation forms
Borg (2006)	Various educational levels/EFL	Distinctive characteristics of foreign language teachers with reference to specific contexts	Students Teachers Subject-specialists	Questionnaires Group discussions Observations Essays
Burden (2008a, 2008b, 2010)	HE/EFL	ELT teacher views on student-completed, end-of-semester evaluation forms for teacher development	NESTs Non-NESTs	Interviews
Dunegan & Hrivnak (2003)	HE	Characteristics of student evaluations of teachers	Students	Questionnaires Evaluation forms
Franklin & Theall (1989)	HE	Assessing knowledge of users of student ratings of instruction	Teachers	Questionnaire
Kember et al. (2002)	HE	Student feedback questionnaires and the effect on teaching quality from student and faculty perspectives	Students Teachers	Questionnaires
Lee (2010)	HE/EFL	Japanese student perceptions of EFL teachers	Students	Questionnaires
Nasser & Fresko (2002)	HE	Faculty perceptions on student evaluations	Teachers	Questionnaires
Newton & Braithwaite (1988)	Primary/ secondary	Teacher perceptions on evaluating teachers	Teachers	Questionnaires
Park & Lee (2006)	Secondary/ EFL	Student and teacher perceptions of effective English teacher characteristics	Students Teachers	Questionnaires
Simpson & Siguaw (2000)	HE	Faculty response to student evaluations of teaching	Teachers	Questionnaires
Sojka et al. (2002)	HE	Similarities and differences in student and faculty perceptions regarding student evaluations of teaching	Students Teachers	Questionnaires
Tran & Do (2020)	HE	Student evaluation of teaching and the effects of age, gender, seniority	Students	Questionnaires

Table 2.2

Summary of theoretical teacher evaluation studies

Author(s)	Context	Teacher evaluation focus
Alamoudi & Troudi (2017)	EFL	Issues in the evaluation of English language teachers
Danielson & McGreal (2000)	HE	Teacher evaluation to enhance professional development
Delandshere & Petrosky (1994)	General	Teacher performance and knowledge within an educational assessment context
Gordon & Stuecher (1992)	HE	Effects of anonymity and accountability on teacher evaluations
Isoré, M. (2009)	Various	Current teacher evaluation practices in OECD countries
Johnson (2000)	HE	Implications of student evaluation questionnaires on teaching and professional development
Kang & Hong (2008)	General	Teacher evaluation policies within South Korean education context
Looney (2011)	General	Teacher evaluation for improvement, characteristics of effective teachers, and policy suggestions
Neumann (2001)	HE	Differences in teaching methods across disciplines in higher education
Nunan & Lamb (1996)	General	Issues with teacher evaluation use
Patton (2008)	General	Moving towards more useful evaluations
Penny (2003)	HE	Shortcomings of SRT (student ratings of teaching)
Pennington (1989, 1991)	ESL/EFL	Issues and recommendations for the implementation of faculty evaluation within EFL/ESL contexts
Pennington & Young (1989)	ESL/TESOL	Approaches to faculty evaluation for ESL
Spooren et al. (2013)	Various	Validity of student evaluation of teaching
Svinicki (2001)	HE	Understanding student feedback on teacher evaluation
Theall & Franklin (1991, 2001)	HE	Issues with student-completed teacher evaluation and use of its data
Troudi (2009)	EFL/TESOL	Teacher evaluations and recognizing teacher contributions
Wennerstrom & Heiser (1992)	HE/ESL	Student bias in instructional evaluation with ESL context
Wines & Lau (2006)	HE	Implications of using student evaluations for faculty evaluation/personnel decisions

2.2 Neoliberalism

2.2.1 What is neoliberalism?

Neoliberalism can mean different things to different people (Bernstein, Hellmich, Katznelson, Shin, & Vinall, 2015). However, within language education, neoliberalism reflects an economic ideology that views the use of English as an imperative construct of global competitiveness (Piller & Cho, 2013). In this market-like domain, English is seen as a valuable and essential commodity that can lead to greater power and influence. Many EFL learners around the world acknowledge and accept that learning and using English is a must if they wish to develop their future careers and social status because English is viewed as the key to global success (Troudi, 2009; Kang, 2012; Pederson, 2012; Price, 2014; Ennser-Kananen, Escobar, & Bigelow, 2016). As Price (2014) observed, "[s]ince the cultural capital of English competence acts as a gatekeeper to higher education and employment markets, students have little choice but to learn English WELL" (p.570). As such, English language learners, especially those in countries like Korea, are now expected to be "autonomous student-consumer[s] who [are] responsible for managing his or her own lifelong creative capital development" (Abelmann, Park, & Kim, 2009, p.232) and to make themselves more marketable (Park, 2010) on both a local and global level. Because of these expectations, English is sometimes perceived as a panacea for most socioeconomic problems; however, as indicated in the following, it might actually cause harm.

Globalization, competitiveness, global/university ranking, efficiency, accountability, and consumer values are regularly used terms in the current English education industry, and they have been implemented in many countries over the past four decades (Giroux, 2002; Kauppi & Erkkilä, 2011; Bernstein et al., 2015). In higher education, the neoliberal idea of a university is to provide English education based on a corporate business model that delivers goods and services within a global market. Here, English is viewed as a job skill and commodity, while language learners are seen as consumers/customers/clients, and language teachers are expendable skill/service- providers or contract-workers (Bernstein et

al., 2015). Such institutions tend to quantify students and teachers through high-stakes tests and student-completed TEFs, respectively, in an attempt to increase efficiency and accountability.

This neoliberal perception of English has affected NESTs in more subtle ways as well. For example, Barratt and Kontra (2000) examined studies that surveyed “the clients of [NESTs] - their students” in order to “get a consumer’s view of what works and what does not for a visiting [NEST] in a strange land” (p. 19). Their choice of vocabulary when describing NESTs and students in their article demonstrated the effects of neoliberal factors and policies in EFL as well as the influence these factors can have on student-completed NEST performance evaluations. Although there are some countries, such as Saudi Arabia, that seem to support the neoliberalization of English and/or higher education (Elyas & Al-Sadi, 2013; Le Ha & Barnawi, 2015), the vast majority of studies conducted within HE contexts show how neoliberal discourse and subjectivity embedded in administrative policies and teacher evaluation systems trickle down to teachers to have a direct and negative effect on their practice and student learning outcomes (Beck & Young, 2005; Ramanathan & Morgan, 2007; Tang & Choi, 2009; Piller & Cho, 2013; Thomas & Yang, 2013; Price, 2014; Bernstein et al., 2015; Ennser-Kananen et al., 2016). The following sections will examine these effects in detail.

2.2.2 Issues and effects of neoliberalism

Effects on teacher evaluation

Hargreaves (2000) said that “market principles have become embraced so strongly by many governments that school and other public institutions have been rationalized… made more economically efficient, and set in competition against one another for ‘clients’” (p.168). Because of this fierce competition stemming from the normalization of market principles, there is much strain and conflict within the education field and tension between agency and structure (Coldron & Smith, 1999).

The perceived high-stakes nature of teacher evaluations, especially student-completed TEFs, can be largely attributed to such neoliberal discourse and policies that have become an integral part of education systems around the world. For instance, rather than using TEFs for formative purposes with regard to improving teaching quality and student learning outcomes, they are used as an efficient means to quickly assess and rank teaching performance based on a set of numerical scores (Theall & Franklin, 2001). TEFs are viewed as a method of accountability that demonstrates the presence of adequate procedures to ensure teaching quality (Kember et al., 2002). The use of student-completed TEFs is also seen as a "key indicator in quality monitoring" (Penny, 2003, p. 400). However, whether this method of evaluation indeed quickly and accurately assesses teaching performance and accountability is questionable.

Because of the need to efficiently manage administrative matters such as measuring teacher performance (e.g. did the teacher adequately provide the English language service?) while considering student opinions and input (e.g. was the student satisfied with the service provided?), most teacher evaluation methods come in the form of questionnaires with rating scales. These forms generally consist of items that cover most academic subjects in broad terms, rather than subject or context-specific questions. Unfortunately, as mentioned earlier, it is common to see TEFs being used as the sole evaluation criterion, especially for teachers in higher education. Although using the same student-completed, cross-curricular TEF guarantees that it will be unfair for both students and teachers (Emery, Kramer & Tian, 2003), Thomas and Yang (2013) observed that the "one size fits all and efficiency models prevail" because "[m]arket driven solutions are seen as essential" (p.112) and are the most efficient use of time and resources.

Within these business-oriented HE contexts, students are viewed as consumers of a product (Harvey, 2001). Therefore, administrators seem to prefer the use of student-completed TEFs and numerical scores as a way to measure student satisfaction levels. Some researchers

and practitioners attribute this troubling view to the increasing social, economic, and political pressures universities face these days (Penny, 2003) where the survival of private universities desperately depends on student tuition fees (MoE, 2005; Burden, 2008a, 2008b). As Thomas and Yang (2013) observed, “schools that are in a financially critical position tend to treat students as customers because student tuition is the major financial resources that schools need to survive…the schools will please their current students by lowering the standard of academic achievement” (p.117).

Because of this, many teachers find themselves fixating on their final, end-of-the-semester TEF scores to the point where “collusion between neoliberal interests, [and] an obsession with evaluation… are wreaking havoc” on education systems (Thomas & Yang, 2013, p. 107). It is perhaps because of this perceived neoliberal influence on education and growing neoliberal interest in controlling teachers’ actions (Raaper, 2017) that leads to the failure of students, teachers, and administrators alike to value the importance of conducting effective, meaningful teacher evaluations. Therefore, it is not far-fetched to think that such attitudes can have a negative effect on student-completed TEF results and, consequently, TEF validity.

Effects on professional development

Professionalism and professional development in education has changed from “occupational and normative” values (Evetts, 2012) that “improve quality and standards of practice” (Hargreaves, 2000) characterized by autonomy to a concept that is now highly affected by external neoliberal factors such as managerialism (Archer, 2008) and marketization (Hargreaves, 2000; Beck & Young, 2005) that limit the development of said values. Significant and meaningful professional educator components such as autonomy, trust, and confidence have become obscure values in a somewhat hostile teaching climate affected by neoliberal regulations that were shaped by flourishing economies and business-oriented organizational structures adopted by HE institutions.

Due to the high-stakes nature of teacher evaluation influenced by neoliberal factors and the effect it can have on the livelihoods of teachers, their pedagogical practices, and their perceived worth, we can infer that such evaluations can have a harmful impact on their professional development. Alamoudi and Troudi (2017) stated that “formative evaluation can be used to feed professional development decisions” (p.33), but some teachers might develop a skewed perception of professional development as seen in a year-long qualitative study conducted by Crookes et al. (1995). Although this study was based within a primary and secondary ESL context, the findings illustrate neoliberal influences on ESL teachers’ attitudes towards professional development. For example, study participants refused to attend any professional development events or in-service workshops and only reiterated the need for useful and appropriate classroom materials. These demands stemmed from a standardized-test driven curriculum in which teacher accountability towards students’ high-stakes, standardized test scores and rankings were directly related to the amount of funding their schools could receive. This perpetuated the ESL teachers’ desire to get “quick fixes” rather than engaging in reflective professional development and acquiring collaborative knowledge over time (Crookes et al., 1995, p. 192).

This quick fix attitude can also be observed in the use of high-stakes student-completed TEFs where the NESTs’ focus is on quickly improving their final evaluation scores and student satisfaction levels for job security purposes, rather than applying it in a more systematic, formative manner. As mentioned before, HE context studies have shown that using student-completed TEFs for efficient, summative purposes lead to teachers concentrating on increasing their evaluation scores by specifically adapting their curriculum to fit evaluation criteria that continue employment rather than focusing on teaching quality or performance (Simpson & Siguaw, 2000; Wines & Lau, 2006), which would be a great disservice to students.

2.2.3 Neoliberalism in Korean EFL context

Neoliberalism has had a profound effect in Korea and, more specifically, in Korean EFL education. In 1995, the Korean government implemented the Educational Reform for the 21st Century that emphasized autonomy and accountability of educational institutions and introduced performance-based funding that encouraged excessive competition among these HE institutions (Chae & Hong, 2009). However, Korea's neoliberal turn was triggered in earnest by the International Monetary Fund (IMF) crisis in 1997. This disastrous event led to a near-collapse of the Korean economy and "intensified privatization, individuation, and globalization related transformations" where "individuals became self-managers under the demands of neoliberalism" (Abelmann, Park, & Kim, 2009, p.231). This need to become more globally recognized (Sung, 2012) has led to some significant shifts and changes in English education, such as the increase in NEST employment and enforcement of English as the sole method of instruction (MoI) at a number of HE level institutions. However, many have come to question the wisdom in these changes after witnessing a series of negative effects on the stakeholders involved.

Effects on Korean EFL learners

Koreans are generally motivated to pursue higher education because of its perceived correlation to status and high income earnings (Robinson, 1994; Lee & Brinton, 1996; Kang, 2015). For example, as of 2005, 97% of 18 year olds graduated from high school, while 82% of them went on to university or other HE institutions (KEDI, OECD, as cited in Kim, 2008). Therefore, the most substantial part of the Korean education system involves a high-stakes university entrance exam; namely, the national College Scholastic Aptitude Test (CSAT) which is administered only once a year. With the extraordinarily high value placed on entering university and obtaining a degree, the CSAT-oriented Korean education system has earned undesirable monikers such as "testocracy" (Sorensen, 1994, p. 17), "examination hell," (Seth, 2002, p. 140), and "life and death" (Card, 2005) due to its exceedingly competitive nature (Koo, 2007). This could be explained through the CSAT's intended purpose: to provide valid, reliable objective data for selecting students into colleges and

universities with the goal of developing “autonomous and creative Koreans who will lead the age of globalization… in the 21st century” (Korea Institute of Curriculum and Evaluation, 2001). In this entrance exam, English is one of the core sections that both students and universities focus heavily on.

However, there are a number of concerns arising from such a high-stakes exam and particularly with regard to English education. The CSAT is used to determine student admissions to university, but its preparation process has significant long-term effects on high school students’ social and academic lives (Kim, 2006; Jung, 2008). Some researchers and practitioners observe that much of CSAT preparation relies on rote memorization rather than discussions or debates, which in turn produces “robot” students who are ill-equipped to manage the rigors of tertiary education and later have difficulty filling their roles in society where they are expected to contribute both nationally and globally (Lee, 2013). These problems become more evident in relation to English education and the use of English in the classroom at a university level. For example, due to CSAT-oriented school practices, Korean high school students are rarely given opportunities to speak or use English in the classroom. As a result, when these students enter university, they tend to demonstrate an overall lack of general English conversation, discussion, and presentation skills (Han, 2005).

One critical example that further illustrates the gravity of neoliberal influences on English education in Korea involves the disturbing number of student suicides that occurred at Korea Advanced Institute of Science and Technology (KAIST), an elite university in Korea (Piller & Cho, 2013). A number of major local and international news outlets attributed these suicides to a specific policy implemented by university administration in an effort to become more globally recognized: the exclusive use of English as MoI across the entire curriculum (Chang & Park, 2011; McDonald, 2011). This purportedly led to a tremendous amount of stress and burden for ill-prepared students who could not follow the English-only lecture content. Some blamed the effects of intense competition while others pointed at factors such as social inequality in Korea (Robinson, 1994) and the lack of educational opportunities afforded to

students of lower socioeconomic status who could not afford private English education to aid in their studies (Koo, 2007). While Piller and Cho (2013) did not presume to draw a distinct connection between English as MoI and the suicides in their study, they did argue that "to understand the spread of English – despite its obvious costs – one has to look outside language and link language explicitly to the socio-economic order" (p. 24). In other words, they suggested taking a neoliberal perspective in understanding the role of English and English education in Korea.

The unfortunate incidents that occurred at KAIST can be seen as an extreme case of neoliberal English education gone wrong. However, it is not unreasonable to assume that, despite the perceived requirement and need to use English, Korean EFL learners cannot help but exhibit negative attitudes towards learning English. This could be due to the amount of stress and anxiety they endured within the aforementioned intense CSAT-oriented English education system; in turn, these previous experiences could have an effect on how university students currently view the use of English. Other factors might include being caught between the Korean government's neoliberal view of English and teachers' goals for English education within the classroom (Shin, 2007). In the middle of this conflict are EFL teachers, which include a significant number of NESTs at every level of English education.

Effects on NESTs

Although English is "widely perceived to be the cause of immense social suffering in Korean society… this linguistic burden is simultaneously embraced as natural and incontestable" (Piller & Cho, 2013, p.24) because Koreans believe that the use of English is inextricably linked to successful careers, high social standing, and global recognition (Kang, 2012; Pederson, 2012). As a result, there has been a major focus on English education in Korea and an increasing number of NESTs employed by educational institutions at all levels (i.e. primary, secondary, and tertiary) in order to keep pace with global competition (Alptekin & Alptekin, 1984; Kim, 2001; Jo, 2008; Kang, 2012; Kim, 2012; Wang & Lin, 2013).

Within HE institutions marked by managerialism and marketization, many NESTs find themselves fixated on improving their TEF scores while shouldering the burden of improving both local and global school rankings (Tang & Choi, 2009). However, despite the increase in NEST employment and the perceived need to use English on a global level, little research has been conducted on NESTs in Korea regarding their professional teaching standards or teaching performance evaluations, especially at a university level. This is a perplexing issue that should be examined because it is in stark contrast to the notion of autonomy and professionalism afforded to Korean teachers and guaranteed by the government (Yeom & Ginsburg, 2007). That is, while there are studies on teacher quality, hiring standards, and teacher evaluation policies on a national level for Korean teachers (Kang & Hong, 2008), the same cannot be said for NESTs. This seemingly contradicts the Korean government's desire for high quality, effective EFL instruction by NESTs in order to cultivate citizens who can compete on a global level (Pill, 2005; Pederson, 2012; Wang & Lin, 2013). It also differs from other countries and governments that "despite global recession… continue to invest in teachers' professional development as a means of enhancing pupil outcomes" (King, 2014, p. 89).

Effects on administrators

Perhaps out of all the stakeholders, HE administrators are the most directly affected by neoliberal influences within the Korean education system. Despite the neoliberal market-oriented structure of Korean HE policies, the role of the Ministry of Education (MoE) remains the same because it still has direct control over both public and private sectors as well as the authority to fund private HE institutions and to regulate their operations (Kim, 2008; Chae & Hong, 2009). This can be seen through the increasing number of Korean universities seeking to improve their performance and funding through the increase in NEST hiring and cross-curricular use of English as a MoI that has become a major criterion by which they are evaluated (Lee, 2010; Kang, 2012). While this educational policy aims to "increase research and teaching productivity in rankings of world-class institutions of higher learning" (Kang, 2012, p. 30), it also "almost forces the universities in Korea to offer ECMs (English-mediated

courses) in a core curriculum" (Lee, 2010, p. 470). As a result of such global competitiveness and top-down government reform, it seems there has been a decrease in university autonomy as most universities have shifted their attention from 'true education' and enriching student learning to improving their global ranking.

In addition to a growing focus on performance at a global level, university administrators have also been concentrating on efficiently assessing teaching performance and ranking, increasing student satisfaction levels, and proving institutional accountability (Theall & Franklin, 2001; Chae & Hong, 2009). This has led to a "great interest in controlling teachers' actions" (Thomas & Yang, 2013, p. 112), and the use of generic student-completed TEFs to accomplish all of the above through a single, efficient instrument. As a result, HE administrators within EFL contexts likely perceive the use of general, student-completed TEFs to evaluate NESTs to be adequate for immediate ranking and summative purposes, no matter how grossly unfair this might seem to the teachers (Theall & Franklin, 2001). Unfortunately, there seems to be an overall lack of literature that involves administrator perspectives of EFL education and teacher evaluations that can further shed light on this view, much less those concerning NESTs within a Korean context.

Table 2.3

Summary of neoliberalism studies

Author(s)	Context	Neoliberalism focus
Abelmann et al. (2009)	HE	Effects of neoliberalism and structural inequality on Korean university students
Bernstein et al. (2015)	EFL	Critical examination on neoliberalism in relation to second and foreign language education
Elyas & Al-Sadi (2013)	HE	Neoliberalism and the globalization of the Saudi higher education system
Ennser-Kananen et al. (2016)	EFL	Empirical study on neoliberal reasons for foreign language learning
Kang & Hong (2008)	General	Structural inequality in Korean educational contexts
Kang, H.D. (2012)	Primary/EFL	Primary English education in Korea with regard to policy, practice, issues, and possible solutions
Kauppi & Erkkilä (2011)	HE	Rapid restructuration of higher education at global level
Le Ha & Barnawi (2015)	HE/EFL	English, neoliberalism, and internationalization in Saudi higher education
Park (2010)	EFL	Impact of neoliberalism on English learning in Korea
Pederson (2012)	EFL	Effects of neoliberal views on EFL fostering a critical appropriation of language and culture
Penny (2003)	HE	Shortcomings of SRT (student ratings of teaching) and neoliberal influences
Piller & Cho (2013)	HE/EFL	Neoliberal language policy effects in Korean higher education
Price (2014)	EFL	Neoliberalism and globalization in English education
Raaper (2017)	HE	Neoliberalism and fair student assessment policy
Ramanathan & Morgan (2007)	TESOL	Language policy and the neoliberal agenda
Shin (2007)	EFL	Globalization and English language teaching in Korea
Thomas & Yang (2013)	HE	Neoliberalism, globalization, and evaluation
Tang & Choi (2009)	General/ EFL	Neoliberal influences on teacher professional development
Troudi (2009)	EFL/TESOL	Neoliberal views of English and recognizing teacher contributions within this context

2.3 Summary of teacher evaluation and neoliberalism

Examining teacher evaluations and the neoliberal influence on English and English education on various stakeholders within the Korean EFL context leads to a number of perplexing questions. For example, with regard to students, there is heavy investment and importance placed on English, as we can see in competitive, high-stakes exams such as the CSAT. This shows that even at secondary level education, English is perceived as “an important social ladder to ascend to the high road to better social class” (Kim, 2006, p. 167) and that Korean EFL learners see English as the key to a successful future in today’s globalized society. This leads to the reasonable assumption that students would be willing to engage in activities related to the improvement of their English learning, such as providing detailed and constructive TEF feedback that NESTs can use and later implement in the curriculum. However, this does not necessarily seem to be the case based on teacher evaluation studies in which participants lament the lack of quantity and quality of student-completed TEF data and feedback.

Based on the literature, one can perhaps speculate that, at least for Korean university students, the role of English is limited to mostly that of a major component of an extremely stressful and high-stakes college entrance exam that helped them to attain their immediate goal of entering university and obtaining a degree. The need to invest in English learning could diminish once this immediate goal is achieved. Also, the high level pressure and competition involved with taking the CSAT could have had a detrimental effect on their investment in English as well, which in turn could affect the quality of TEF responses related to their university EFL courses and teachers.

Another puzzling question involves NESTs and their seeming lack of knowledge regarding TEFs used to assess their teaching practices, as seen in Burden (2008a, 2008b), as well as an overall lack of such research. This seems particularly so within my own professional context. Considering the value and significance placed on TEF scores and feedback for job security reasons, it is unclear as to why NESTs demonstrate an absence of knowledge or

have difficulty understanding the TEF, its purpose, items, results, and feedback. We can perhaps surmise that language barriers NESTs experience within their contexts could explain their struggle to access TEF-related knowledge (Howard, 2019). Interestingly, however, none of the NEST participants in studies conducted by Burden (2008a, 2008b, 2010) within a Japanese EFL context mentioned any language difficulties regarding the TEF. Therefore, further examination and identification of what factors affect NESTs and their knowledge of the TEF is needed in order to make effective use of it and its data.

Finally, there is some confusion involving administrators and what information they actually wish to glean from student-completed TEFs. On the surface, it appears that these TEFs are intended to improve teaching quality and increase student learning outcomes while efficiently and effectively rating NESTs. However, in order to achieve this goal, it is necessary for both students and NESTs to have a common understanding of the TEF. Administrators are expected to provide guidance in this regard, but many do not appear to do so. As administrators create the TEF, they have the responsibility of creating clear and relevant items that aid with improvement. Unfortunately, many TEFs appear to lack this clarity and focus. For example, Burden (2008a) found in his study that NESTs were encouraged by administrators to engage in collaborative and communicative English teaching, but the student-completed TEF used to evaluate their performance consisted of items related to lecture-style teaching. An extensive literature review done by Spooren et al. (2013) also indicated contradictions between stated and intended aims of TEFs and their perceived uses.

For these reasons, there is a need to critically examine the use and perceptions of student-completed TEFs because of the significant effects they can have on all stakeholder groups regarding student learning outcomes, teaching performance, and overall school performance. This need is particularly strong concerning the use of student-completed TEFs to assess NESTs within Korean EFL contexts as they might experience more acute effects regarding their teaching practices within a highly neoliberal English education system. In

addition, a significant lack of studies regarding all stakeholder perceptions (i.e. students, teachers, administrators) of NEST teaching standards and evaluations methods within EFL contexts also calls for this manner of research. It is therefore my hope that the findings in the present study might help fill this gap in literature.

CHAPTER 3: METHODOLOGY

The purpose of this study was to critically examine and understand the underlying assumptions and knowledge that students, native English speaking teachers (NESTs), and administrators (i.e. stakeholder groups) have about the student-completed teacher evaluation form (TEF) used to evaluate NEST performance at a Korean university. The primary question guiding the study was:

- What are the similarities and differences among stakeholder group perceptions of using a student-completed TEF to evaluate NEST performance and the factors influencing these perceptions?

To gain insight on stakeholder perceptions about the primary question, study participants were asked to share their views on the following research questions:

1. What do they perceive to be the general purpose of the TEF and its items?
2. In what way do they consider the current TEF useful or not useful?
3. What is their most preferred way to use the TEF?
4. What do they consider to be the best or ideal way to evaluate NESTs?

To best explore and explain the similarities and differences among stakeholder perceptions of the student-completed TEF and their influencing factors, I took a grounded approach to this qualitative study. As mentioned in the introduction chapter, I had already possessed some general knowledge of NEST opinions of the TEF and was aware of possible responses that could emerge because these “guiding interests” (Charmaz, 2014, p. 31) were what originally motivated me to conduct the study. However, I was not at all sure what to expect regarding student and administrator perceptions of the TEF within their own situated worlds as I never had the opportunity to discuss such matters with them.

In order to provide a full picture of everyone's assumptions regarding the TEF, I thought it necessary to follow culturally sensitive procedures that could show how "logic and emotion combine to influence how persons respond to events" and reveal the participants' underlying beliefs and meanings in those responses (Corbin & Strauss, 2015, p. 11; Creswell & Creswell, 2018). For these reasons, I considered it best to take a grounded approach that would require simultaneously collecting and analyzing data sets from study participants, each informing and focusing the other throughout an interactive research process that merged both past interactions and current interests into the study (Charmaz, 2005; Corbin & Strauss, 2015; Schwandt, 2015). This approach allowed me to apply a set of flexible analytic methods and strategies to gather rich and sufficient data. This data was then analyzed through a rigorous series of cumulative coding cycles and constant comparison that could ultimately lead to the development of a theory rooted in the original data (Cohen, Manion & Morrison, 2007; Saldaña, 2009; Charmaz, 2014; Schwandt, 2015).

Based on this approach, this chapter discusses both theoretical and practical matters regarding the present study and consists of the following:

1. Introduction of study setting and its participants
2. Outline of the data collection methods with reference to appropriate literature
3. Outline of analysis methods and procedure with reference to appropriate literature and a worked example
4. Ethical considerations

3.1 Setting and participants

3.1.1 Setting

The study was conducted at a four-year, all-female university in Seoul, Korea. The university's General English Program (GEP) provides students with compulsory English language classes that cover the four basic skills: speaking, listening, reading, and writing.

GEP classes are mandatory elective courses taught only by NESTs over the course of 16 weeks. Mandatory elective classes are non-major classes that students must take in order to fulfill their credit hours. Per university policy, all students must attend and pass at least two GEP classes (e.g. Discussion & Presentation, Reading & Writing) as part of their graduation requirement. The GEP classes are leveled based on English proficiency, and students are required to take a level placement test in order to sign up. In addition to these GEP classes, NESTs are also responsible for leading various non-credit English Clubs for students who wish to further practice speaking and interacting in English outside of the classroom.

NESTs are evaluated twice during a regular 16 week semester: the midterm TEF is given during Week 4, while the final TEF is provided to students during Week 12. Both forms are student-completed and designed by university administrators. The final TEF is used to evaluate and rank all teachers on campus, including GEP NESTs, and consists of both quantitative and qualitative items. This same TEF is applied to all majors and subjects, and the items/questions are provided in Korean (Appendix A). However, the midterm TEF greatly differs from the final in a number of distinctive ways. First, it consists of only a few questions that do not wholly reflect the items in the final TEF (Appendix B). Another difference is how little attention is called for students to complete it or for teachers to review the midterm TEF results, unlike the final. In addition, only the final TEF results are examined and used by the university for administrative purposes.

Completing the TEF is voluntary for students. However, the university tries to encourage students to fill out the final TEF by providing them with minor incentives, such as allowing students to access their final grades a few days earlier than those who do not complete the form. The perceived benefit of this is the extra time allotted for students to review their scores and to contact their teachers if they a) detect any calculation errors or b) wish to contest the final grade they received prior to the end of the grading period.

3.1.2 Participants

Questionnaires were used as a tool to recruit participants through a simple sampling process and then selected based on their prior, relevant knowledge and experience of the TEF. Relevance was a matter of choosing those who could provide responses that were critical to understanding the primary question as well as shed light on issues pertaining to any influencing factors (Schwandt, 2015). As a result, the study participants consisted of members from each of the three stakeholder groups involved or affected by the TEF: four students, four NESTs, and four administrators.

Student participants were chosen through a process that involved asking my NEST colleagues for permission to distribute short, voluntary, anonymous questionnaires to students in their GEP classes and English Clubs. After briefing students about the study, explaining the purpose of questionnaire, its role in the study, and assuring them that the questionnaire had no bearing on their current studies, the forms were distributed among classes comprised of students from different grade levels (e.g. first, second, third, fourth year) and class levels (e.g. Level 1, 2, 3). A total of 100 students were asked to fill out the questionnaire and then invited to contact me via email if they were interested in volunteering for the study. For those who responded, I made individual inquiries, via email, about the amount of experience they had with the TEF (i.e. the frequency in which they either completed or reviewed the TEF) and their field of study. From this process, I was able to select four participants whom I hoped could help explain how students perceived the TEF.

A similar process was followed for NEST participants. Questionnaires were distributed to 25 GEP NESTs at the end of a faculty meeting. This allowed me to address everyone simultaneously and explain the purpose of the study. I then asked the NESTs to return their anonymously completed questionnaire forms to my office mailbox. NEST participants for the study were chosen from a list of volunteers who had completed the questionnaire and who had also privately contacted me to express their interest in the study. Based on the length of their employment at the university, as well as their personal and professional backgrounds, I

was able to choose four NESTs whom I thought could best provide a good representation of responses regarding the TEF from a teacher's perspective. When recruiting administrator participants, I had initially approached 10 administrators associated with the GEP and contacted them individually in person as well as via email. Of the 10, four responded and volunteered to take part in the study.

Students

Jina, Yoon, Yena, and Mia were former GEP students whom I had never taught or had contact with prior to the study. They were also all members of different English Clubs when recruited for the study. Jina and Yoon were third year students; Yena and Mia were both fourth year seniors, but Mia was scheduled to graduate at the time of the study while Yena had one more semester left. The students came from various fields of study, had completed the TEF five or more times, and brought diverse perspectives of the TEF based on their backgrounds and past experiences. For instance, Mia had been the university student body president during the previous year. This position and the responsibilities that followed gave Mia opportunities to interact with main administrators about a wide range of topics and thus serve as a bridge between them and the student body. She was therefore seen as a participant who could perhaps provide a more detailed and contextual student perspective of the TEF in a holistic and enlightening way.

Table 3.1

Student participants

Participant	Years of study	TEF completion frequency
Jina	3	Once per semester (final only: 5 times)
Yoon	3	Once or twice per semester (midterm and final: total 7-8 times)
Yena	4	Once per semester (final only: 7 times)
Mia	4	Once per semester (final only: 8 times)

NESTs

I selected four experienced NESTs (Kelly, Mark, John, Rob) who came from a wide range of personal and working backgrounds and held various positions within the GEP in order to gain richer data from a range of perspectives. For the purpose of this study, experienced NESTs refer to teachers who have taught at the university for at least five years. This was to ensure a higher probability of the participants being familiar with the TEF and providing relevant and constructive responses to help explain NEST assumptions of the TEF, how they perceived its use in evaluating their teaching performance, and what factors affected their views.

For example, at the time of the study, Kelly was the GEP head teacher, and Mark was the head curriculum coordinator who also had much teaching experience within the American public school system. Rob had been the GEP head teacher in previous years and but had also worked in corporate human resources prior to teaching, while John was a member of the GEP Professional Development Committee and the only bilingual participant in the

study. John's ability to speak both Korean and English was a substantial factor that I took under consideration when choosing participants because of his potential to offer more insight on possible language-related issues regarding TEF knowledge and use.

Table 3.2

NEST participants

Participant	Years of teaching experience	TEF review frequency
Kelly	7	Twice a semester
Mark	7.5	Almost never
John	7.5	At least once a semester (midterm review only when reminded)
Rob	14	At least once a semester

Administrators

Siri, Hyuna, Hana, and Jin were GEP administrators who worked closely with NESTs at the time of this study. Hana, Hyuna, and Siri were administrative staff members who interacted with GEP NESTs on a daily basis on various matters ranging from managing class schedules to coordinating monthly faculty meetings. Unlike these three, Jin held the title of supervisor (과장/*gwajang)* within the General Elective College of which the GEP became a part. Many NESTs assumed that Jin's position and status as a supervisor afforded her more opportunities to communicate with those in main administration.

All four GEP administrators were involved with the TEF in terms of identifying NESTs with below average scores, translating qualitative student feedback for those NESTs, and then

reporting this information to main administrators. It should be noted that both NESTs and GEP administrators have always perceived people working within main administration as being higher in position and having more authority or agency within the university system. Therefore, for the purpose of this study, administrator participants are referred to specifically as ‘GEP administrators.’ The distinction between GEP administrators and main administrators was made throughout the study.

Every effort was made to recruit main administrators directly involved with the TEF because the data collected from these individuals could have proven to be quite valuable for the study. Unfortunately, this was extremely difficult to do for a number of reasons related to the university system and communication issues. For example, the GEP was originally part of the Department of English Language and Literature, but for reasons unknown, it became a separate, yet non-autonomous program, that was no longer associated with any particular department or college within the university. At the beginning of the study, the university underwent a series of major changes that included significant restructuring of departments and programs. This included the GEP, which went from a floating English language program to becoming a part of the university’s General Elective College. As a result of this restructuring, the GEP no longer had a head director, whose main roles included leading the program, interacting with NESTS, using the TEF data for administrative purposes, and serving as a bridge between NESTs and main administration in terms of communication. Therefore, despite the support I received from Jin to gain access to main administrators and distribute more questionnaires, it was not possible to communicate my request through official channels without the help of someone in an even higher position.

Table 3.3

GEP administrator participants

Participant	Years of administrative work experience	Core GEP administrative duties
Siri	One semester	• Material prep for GEP-related matters • Translating qualitative TEF feedback and other documents for main administration
Hyuna	1.5	• Coordinating GEP extracurricular programs between students and NESTs (e.g. English Clubs) • Translating qualitative TEF feedback • Processing quantitative TEF data
Hana	3	• Organizing and managing NEST teaching schedules • Translating qualitative TEF feedback for main administration • Processing quantitative TEF data for teacher awards
Jin	4 (supervisor for 2 years)	• Supporting GEP NESTs • Managing GEP classes and other general elective classes • Handling NEST contract renewals • Reporting NEST TEF scores to main administration

3.2 Data collection methods

In order to acquire rich and sufficient interpretive data from this study and to fully explore the similarities and differences among stakeholder group perceptions of the student-completed TEF and the associated influencing factors, I made an effort to adhere to guidelines suggested by Charmaz (2014). They included asking myself the following questions:

- Am I collecting enough background data about persons, processes, and settings to understand and portray the full range of contexts of the study?
- Are the data sufficient to reveal changes over time?
- Have I gained multiple views of the participants' range of actions?
- Am I gathering data that enable me to develop analytic categories?
- What kinds of comparisons can I make between data? How can these comparisons generate and inform my ideas?

As Charmaz (2014) also advised, I tried to let the study's primary question shape the methods I chose in order to best 'see' the data and to help develop my emerging ideas. As a result, this study was conducted in three stages based on a "constellation" of data collection methods: questionnaires, semi-structured interviews, and focus group discussions. That is, rather than viewing these methods as independent tools, they were seen as part of a collaborative system of helpful strategies for collecting, managing, and analyzing qualitative data within an iterative process (Charmaz, 2014). In addition to data from questionnaires, interviews, and focus group discussions, I also wrote and used analytic memos to better reflect upon and analyze the codes and categories that emerged from the data. Concepts that emerged from this process were "compared with more empirical indicators and with each other to sharpen the definition of the concept and its properties" (Schwandt, 2015, p. 63).

Prior to discussing the data collection methods in detail, it is important to note that the participants' English proficiency levels did not affect the quality or quantity of their responses

during any stage of the study. As a proficient bilingual, I used my knowledge and experience of the Korean language and culture within my professional context to “dig beneath the surface” and understand the significance of certain responses more quickly (Corbin & Strauss, 2015, p. 78). I therefore gave participants the choice to use Korean, English, or both during all stages of data collection in order to acquire in-depth and detailed data from all “relevant situational and social contexts” (Charmaz, 2014, p. 18). This is because context not only “grounds concepts” but also “minimizes the chances of distorting or misinterpreting intent” and thereby helps the researcher to stay true to the data (Corbin & Strauss, 2015, p. 70). However, while I felt that my position as a bilingual GEP NEST afforded me better opportunities to elicit more contextual data, I also found it essential, from an ethical standpoint, to engage in constant reflexivity as the researcher in this study. The strengths and weaknesses of my dual identity as both NEST and researcher will be discussed later in this chapter as well as in the conclusion.

3.2.1 Questionnaires

Questionnaires are helpful in that they take a general purpose and turn it into more “concrete, researchable fields about which actual data can be gathered” (Cohen et al., 2007, p. 318). Therefore, questionnaires were used in the beginning of the study in order to a) recruit study participants b) use the emerging data to familiarize myself with the participants’ views about the TEF and c) to better prepare myself for the following interviews.

The questionnaire served as an opportunity to invite students, NESTs, and administrators to take part in my study. It also provided potential participants with enough information to understand the gist of the study that allowed them to make an informed decision on whether they wished to participate in it. Initial results showed that there were respondents who could provide relevant insight on TEF knowledge. They also revealed inconsistencies in how the TEF was perceived by and within each stakeholder group. Criteria used to narrow down the pool of participants included not only their general interest in the study and availability but

also the amount of experience they had with the TEF.

Questionnaire items were based on the study's research questions and aimed to elicit answers regarding the participants' general understanding of the current student-completed TEF. When creating the questionnaire, I strove to follow the suggestions made by Cohen et al. (2007) to ensure that it:

- Was clear on its purposes
- Was clear on what needs to be covered in order to meet the purposes
- Asked the most appropriate kinds of questions
- Elicited the most appropriate kinds of data to answer the research purposes and sub-questions

As a result, the questionnaire consisted of three open-ended questions to allow for flexibility and contextualization and two closed questions, which were all related to the current TEF at the time of the study (Appendix C). Open-ended questions were asked because there were too many possible categories of responses that would lead to an otherwise very long list of closed question options. I also thought that open-ended questions gave participants the freedom to provide as much detail as possible in their answers, which could lead to unanticipated yet useful ways of viewing the data (Cohen et al., 2007).

In addition to recruiting study participants, the information gathered from the questionnaires also helped me to create interview sub-questions to be used for clarification purposes and to better identify influencing factors affecting stakeholder perceptions of the TEF with relation to NEST performance evaluation (see section 3.2.2). This then allowed me to refine and finalize the interview questions for the second stage of data collection.

3.2.2 Semi-structured interviews

Corbin and Strauss (2015) stated that interviews are preferred as a method of data collection in studies that take a grounded approach, and Cohen et al. (2007) mentioned a number of useful purposes of conducting interviews for such qualitative studies. These include gathering information directly related to the research objectives, using them to help identify variables and relationships, or delving deeper into participant responses to determine the meaning of what they said. In other words, interviews allow participants to “discuss their interpretations of the world in which they live” and to express how they feel about certain situations from their own perspective, which in turn, can generate knowledge about the research topic (Cohen et al., 2007, p. 350). Corbin and Strauss (2015) recommended conducting unstructured interviews as they yield the richest data and likely cover every aspect of the research topic the participant wishes to share, whereas a major weakness in semi-structured interviews is the difficulty in ascertaining whether issues that are relevant to the participants have been sufficiently discussed and covered.

However, despite this critique, I felt that semi-structured interviews were more appropriate for my study. One reason was because although I would initially guide the interview in a certain direction, the “emergent nature of interviews” would eventually “shift control to the participant” (Corbin & Morse, as cited in Charmaz, 2014, p. 71). I therefore thought that semi-structured interviews could still give the participants enough freedom to respond however much they wanted. In addition, these types of interviews would allow me to ask further questions for clarification purposes because, as the researcher, I had some control over the line of questioning (Creswell & Creswell, 2018). For example, I could encourage the participants to ask additional questions or add more comments at the end of their interviews. This would give them another opportunity to elaborate on their thoughts and increase the possibility of adequately covering matters of interest.

Another more important reason was because I would be able to maintain some consistency in each interview, which was vital as I was focusing on comparing data sets for similarities

and differences among the stakeholder groups. Moreover, conducting semi-structured interviews seemed more appropriate considering the study context and its possible influences on the participants. Creswell and Creswell (2018) suggested that not all people are equally articulate and perceptive especially when being asked questions in a somewhat contrived setting and that the mere presence of the researcher could bias participant answers. Therefore, I thought it best to incorporate some structure and have a list of sub-questions to fall back on in order to smoothly conduct the interviews (Corbin & Strauss, 2015).

Responses from the questionnaire were used to inform these sub-questions. For instance, the second item on the questionnaire asked participants to share their opinion on what they believed to be the general purpose of the TEF (Appendix C, D). However, many of them did not (or could not) provide much detail in their responses, which varied widely with no discernable pattern. The same occurred for the third item in which respondents were asked to choose what they perceived to be the three most important items on the TEF. The disparity in NEST responses for this item was particularly noticeable as many of them could not read the TEF in its original Korean form. I therefore thought it prudent to create a list of follow-up or sub questions for each main interview question in order to better guide the participants and encourage them to share their knowledge and understanding of the TEF to the best of their ability (Appendix E-G).

The contextual nature of these interviews meant that I not only had to take precautions to ask the right questions in order to elicit responses from the participants that were both authentic and meaningful, but I also had to remain as neutral as possible in order to prevent contaminating the interviewing process (Cohen et al., 2007; Schwandt, 2015). For these reasons, I tried my best to maintain a high level of self-awareness of my positionality and keep my assumptions about the participants to a minimum during the interviews and throughout the study (Corbin & Strauss, 2015).

Face-to-face semi-structured interviews with the 12 participants (i.e. four students, four NEST, four GEP administrators) were conducted in a conference room located on campus as it afforded convenience, familiarity, and privacy for everyone. They were recorded with a device that was placed as unobtrusively as possible between the participant and myself. Each interview lasted approximately one hour depending on how much the participant wished to share or discuss and consisted of the following:

- Basic information such as the number of times they completed or reviewed the TEF, years of study or employment
- Four open-ended questions to help gain insight on the primary question
- Supplemental questions to aid with understanding and clarification (asked only when necessary)

The study's four research questions were used as the main interview questions. All participants were asked the same interview questions to increase the comparability of responses across stakeholder groups. These open-ended questions gave participants opportunities to further elaborate on their opinions of the TEF that they had mentioned when filling out the questionnaire. Sub-questions were only asked if the participant required assistance in understanding the main interview questions or if further clarification was required (Appendix E-G).

To ensure that I accurately captured the entirety of these responses, I took additional steps. First, I reminded the participants (students and GEP administrators in particular) of the option to communicate in whichever language they felt most comfortable and repeated the interview questions in Korean upon request. I also took down notes of participant responses or utterances that seemed significant at the time. This was done as discreetly as possible and only when absolutely necessary as participants might have found the action of note taking off-putting if it was done continuously throughout the interviews (Cohen et al., 2007).

3.2.3 Focus groups

Focus groups were used in the study as a means of validating the responses collected from the semi-structured interviews as well as “sites for consolidating collective identities” that “allow for proliferation of multiple meanings and perspectives” (Denzin & Lincoln, 2008, p. 398). This was achieved through a) gathering data on participant attitudes, values, and opinions b) empowering participants to speak out in their own words c) encouraging participants to voice their opinions as a group and d) covering additional issues and elaborating on them more than would be possible in a simple survey (Cohen et al., 2007) within a collaborative group setting.

Focus group discussion prompts were based on each stakeholder group’s interview data. That is, after conducting the interviews and collecting the data, I analyzed it to a point at which I was able to identify emergent categories by following a methodical process (Figure 3.1). Significant issues and concerns stemming from the emergent categories that required further elaboration from the study participants were used to inform the focus group prompts (Appendix H). By putting multiple perspectives on the table through focus group discussions, I felt that I was able to acquire a more comprehensive and collective, rather than just individual, view of the TEF from the interaction among the participants. This included gaining access to particular memories, positions, and ideologies (Denzin & Lincoln, 2008) related to the primary question’s focus on discerning similarities and differences among stakeholder group perceptions of the TEF.

The focus groups were comprised of the same interview participants from each stakeholder group, and each discussion lasted approximately an hour to an hour and a half. They took place on campus but in a private, closed-off lounge area partitioned into three sections. I chose this more informal and neutral setting in the hopes of lessening the amount of self-consciousness or discomfort participants might have felt when interacting with each other for the first time. As Cohen et al. (2007) pointed out, the contrived nature of focus groups is both their strength and weakness in that the setting itself is unnatural and might produce less data

than interviews. However, because focus groups concentrate on a particular issue, they can "yield insights that might not otherwise have been available in a straightforward interview" (p. 376). For these reasons, I had hoped that providing the participants a safe, private environment would give them an opportunity to more freely construct their individual and collective knowledge of the TEF, while voicing their true thoughts on how they believed the TEF should be used and how NESTs should be evaluated.

All three focus group discussions were guided by a set of open-ended questions/prompts based on issues and concerns that emerged from the participants' individual interviews. In order to allow participants to introduce new ideas that they might not have not previously considered, discussion guidelines were made flexible and kept at a minimum. I did, however, encourage participants to speak up whenever possible to prevent the possibility of "inarticulate members [being] denied a voice" (Cohen et al., 2007, p. 377). Although I was in the same lounge as the stakeholder groups during these sessions, I did not sit with the participants. Rather, after providing the participants with the discussion prompts on a laptop computer, and setting up a recording device in an unobtrusive location, I took care to wait in a separate room within earshot in an effort to "decenter the role of the researcher" (Denzin & Lincoln, 2008, p. 397). I endeavored to intervene only upon request or when I felt that it was necessary so that the discussions were open and fluid yet stayed on point (Cohen et al., 2007). During these discussions, I made notes on what prompts needed clarification, what the participants focused on the most, and what, if any, discrepancies in responses (e.g. interview vs. focus group) there were in the hopes of better understanding the data during the analysis stage.

3.2.4 Analytic memos

In addition to using the above three data collection methods, I also wrote analytic memos throughout the study, especially during the data collection and analysis stages. As researchers must fully interact with their data to yield useful and relevant results, analytic

memos are used as a means of recoding the researcher's ideas for reflection and analysis purposes (Corbin & Strauss, 2015; Schwandt, 2015). This allows researchers to critically think about what they are doing, challenge their own assumptions about the data, and reflect on how the process of inquiry is taking shape (Saldaña, 2009).

Analytic memos can include notes or diagrams about the codes or any other ideas about the data that came up at the time or comments that can help explain or describe developing patterns among categories. Corbin and Strauss (2015) described analytic memos as undeveloped portrayals of thought that grow in "complexity, density, clarity, and accuracy as research progresses" (p. 117) that can be used to compare categories for similarities and differences and to integrate or combine later. In other words, analytic memo writing can function as a code and category-generating method (Saldaña, 2009). It is therefore a highly recommended practice for qualitative studies as it is necessary for researchers to keep track of the cumulative thinking that occurs during the data analysis stage to construct a theory (Saldaña, 2009; Charmaz, 2014; Corbin & Strauss, 2015; Schwandt, 2015).

For these reasons, I tried my best to engage in methodological analytic memo writing throughout the study in order to uphold my accountability and ethical responsibility to the participants with the aim of "present[ing] them fairly" and showing "the complexity and range of issues that participants are dealing with" (Corbin & Strauss, 2015, p. 120). This required much reflection and asking myself a multitude of questions that allowed for the exploration, discovery, and development of ideas within an interactive space (Charmaz, 2014).

3.3 Data analysis methods and procedure

A methodological procedure influenced by Charmaz (2005, 2014), Cohen et al. (2007), Saldaña (2009), and Corbin and Strauss (2015) was followed for each study participant and stakeholder group. This was done in order to systematically and inductively analyze the data collected and to avoid relying solely on thematization or patterns of frequency, which can

obscure the contextual influences on the data. As taking a grounded approach involved developing increasingly abstract ideas from participants' responses and looking for specific data to form emerging conceptual categories (Charmaz, 2005), qualitative data collected from the semi-structured interviews and focus group discussions were meticulously and methodically coded and analyzed using the following process:

1. 1st cycle coding (Initial and Values Coding)
2. 2nd cycle coding (Pattern Coding)
3. Categorizing (Pattern Coding for emergent categories)
4. Re-coding and re-categorizing (Theoretical Coding for categories)
5. Final categorizing (Theoretical Coding for core categories)

These coding decisions were based on the methodological needs of this study and thus "mixed and matched" (Saldaña, 2009, p. 49) in order to enable an analysis that was theoretically sufficient and could best answer the primary question guiding the study. In this study, achieving theoretical sufficiency refers to establishing categories 'suggested' by the data rather than through exhaustive 'saturation' as the notion of saturation is subjective and imprecise (Dey, 1999, as cited in Charmaz, 2014, p. 215). This concept guided and encouraged me to develop a level of patience and tolerance in order to keep myself open to the emerging data and to work with it rather than resisting it (Cohen et al., 2007; Charmaz, 2014).

Data collected from each method/stage were coded and analyzed in order to inform the next stage as "points of departure" (Charmaz, 2014, p. 31). That is, I engaged in theoretical sampling where the gathering of data was based on analysis of previously collected data (Corbin & Strauss, 2015). For example, questionnaire responses were used to get an initial understanding of participants' perceptions of the current TEF and then to subsequently revise and finalize the interview questions. This data was also used to create sub-questions that could help elicit more accurate responses from the study participants. In turn, data

collected from the interviews were coded, analyzed, and then used to identify relevant concepts and to more fully explore their properties. This information helped me to create the focus group discussion prompts that participants could critically reflect and elaborate on in a collaborative manner.

Coding, writing analytic memos, and making constant comparisons helped to raise analytic questions about the data early on in the study (Cohen et al., 2007; Saldaña, 2009; Charmaz, 2014). Constant comparison refers to the iterative process of comparing new data with existing data and categories across a range of situations, groups of people, and a variety of methods so that the categories fit all the data (Cohen et al., 2007). This involves breaking down data into smaller, segmented pieces, and then putting it back together by comparing each piece for similarities and differences, grouping similar data that share the same concept, and grouping these concepts to form categories, all of which eventually integrate around a core category (Corbin & Strauss, 2015; Creswell & Creswell, 2018). Because of this, the coding and analyzing process was not as simple or linear as presented in the steps above; rather, it required going back and forth multiple times as “each piece of data… can inform earlier data” (Charmaz, 2005, p. 517). However, following this iterative process was important as it helped to limit the “intrusions of biases and assumptions” that could occur in grounded studies (Corbin & Strauss, 2015, p. 55).

Due to the study’s context and research approach, further steps were taken into consideration to help provide a better understanding and explanation of the similarities and differences among stakeholder group perceptions of the TEF and their influencing factors. For instance, semi-structured interviews and focus group discussions were transcribed verbatim and analyzed in the language in which they were told and not in translation. This was to ensure that the language used by participants when sharing their opinions and assumptions corresponded with the language in which the events in question took place and to decrease the level of discrepancy that could affect the amount of details and effect of utterances. Following this process was necessary for the study as some of the discussions

among students and GEP administrators involved constructing knowledge and negotiating meaning of certain Korean terms used with regard to the TEF.

In addition, while I acknowledged the value of calculating frequencies and creating numerical data through statistical analysis in order to strengthen the study, I chose to also focus on illuminating qualitative information regarding stakeholder perspectives on student-completed TEFs because frequency does not always equal importance (Cohen et al., 2007). As Saldaña (2009) observed in one of his studies, “mere numeric frequency of a code or category from data analysis… is not necessarily a reliable and valid indicator of a central/core category.” He instead urged researchers to also consider the “summative power” of a less-frequently mentioned code that could encompass all major and minor categories in a study (p. 166). To support this, Charmaz (2014) described how sometimes even though participant responses did not quite align with initial research interests, she still “pursued other topics” that respondents defined as “crucial” and that unexpectedly “emerged as a recurrent theme” (p. 32).

The following process outlines the coding and categorization methods used to analyze the interview and focus group discussion data. A worked example of this procedure can be seen in Figure 3.2. Throughout this process, I repeatedly referred to my analytic memos to help “digest” the data and build concepts based on it (Corbin & Strauss, 2015, p. 296).

1st cycle coding

For the first step in the analysis process, I chose to use both Initial and Values Coding for the interview and focus group discussion transcripts as methods to attune myself to the participants’ language and perspectives of the current TEF (Saldaña, 2009). It was important to use more than one coding method at this stage because doing so could help broaden my perspective on the data and to remain open to exploring all possible theoretical directions indicated by the data (Charmaz, 2014). As Initial Coding is intended as a starting point and open-ended approach to see the direction in which to take the study, I engaged in detailed,

line-by-line coding as advised by Saldaña (2009) and Charmaz (2014). This entailed identifying short yet salient chunks, phrases, and utterances for “analytic import” (Charmaz, 2014, p. 109). While taking this opportunity to reflect on the content as a whole, I also endeavored to break down the data into more discrete parts in order to get gain perspective on the participants’ values, attitudes, and beliefs regarding the TEF. I therefore reviewed the transcripts again and applied Values Coding to refine the Initial Coding results.

Saldaña (2009) defines a value as “the importance we attribute to oneself, another person, thing, or idea” that is related to the level of personal meaning attributed to personal value. One’s value is also “influenced by the social and cultural networks” to which an individual belongs. An attitude is a learned, affective reaction to “the way we think and feel about oneself, another person, thing, or idea,” while a belief is “part of a system that includes our values and attitudes, plus our personal knowledge, experiences, opinions, prejudices, morals, and other interpretive perceptions of the social world” (pp.89-90). I therefore found Values Coding to be an appropriate method to apply to the initial results because it helped me to focus on all of the logical and emotional constructs needed to better understand participant perceptions and what particular factors affected those perceptions the most within the study’s EFL context. The task of determining which construct corresponded to the coded units was difficult and complex. However, I made every effort to assess each unit and utterance so that they “capture[d] the participant’s worldview or personal ideology,” which is considered to be more in line with a grounded perspective (Saldaña, 2009, p. 93).

2nd cycle coding

After the first cycle of coding, I applied Pattern Coding to its findings as a way to pull together and group similar chunks and codes into more meaningful and concise units of analysis. While numeric frequency of a code was mostly used to achieve this, the summative power of a code was also taken into careful consideration because the quality of a code sometimes held just as much importance as its quantity (Saldaña, 2009).

Pattern Coding is considered appropriate for second cycle coding, developing major themes from the data, searching for rules, causes, and explanations in the data, examining social networks and patterns of human relationships, and forming theoretical constructs and processes (Miles & Huberman, as cited in Saldaña, 2009). For these reasons, Pattern Coding was applied to not only second cycle coding but also to the categorizing stage of this analysis process.

Categorizing

During this stage, I endeavored to code in a more "simple, direct, analytic, and emergent" manner (Charmaz, 2014). I therefore applied Pattern Coding again to my second cycle coding data as an "initial analytic strategy" (Saldaña, 2009, p. 48) to identify emergent categories. By constructing and examining these emergent categories, I began to see a shift in data in terms of real, specific, concrete examples to more abstract, general concepts and constructs. Codes formed during this stage of the analysis process were used as 'emerging categories' in the study.

Re-coding and re-categorizing

In order to further transcend the data and find the core categories, I re-coded and re-categorized the data from the previous step by using Theoretical Coding and engaging in constant comparison of the 'emerging categories' with the existing data. This helped me to develop a set of theoretical codes that revealed possible relationships among the emerging categories. Theoretical Codes are code units capable of explaining and covering all other codes and categories formulated thus far in the study (Saldaña, 2009) and "integrate and solidify the analysis in a theoretical structure" (Charmaz, 2014, p. 19). This coding process further moved the data analysis in a theoretical direction (Charmaz, 2005) and helped me to develop 'categories' for the study.

Final categorizing

To take the final, culminating step toward forming the core categories in the data, I applied

Theoretical Coding and constant comparison again to integrate and combine the categories that developed in the previous stage. These core categories had to “consist of all the products of analysis condensed into a few words that seem to explain what ‘this research is all about’” (Strauss & Corbin, as cited in Saldaña, 2009, p. 163). It should be noted, however, that at this point, I found myself examining and reflecting upon the data through a more critical lens because I felt that this was the direction the emerging categories were taking me, which is something that Corbin and Strauss (2015) indicated might happen in grounded theory studies.

This repeated process of rewording and transforming the categories helped me to progress toward forming abstract constructs with richer interpretive meanings (Saldaña, 2009). The categories consisted of abstract concepts that explained variation as well as the main points made by the data, thus forming a final set of codes to use as ‘core categories.’ The Theoretical Codes used to represent the core categories connected and aligned the study’s data with its research questions. These core constructs also fit the list of criteria recommended by Strauss (1987). This included not only a sufficiently abstract and overarching concept but also a concept that frequently appeared in the data, that was logical and consistent with the data, that could be used to do further research, and that could grow in depth and explanatory power regarding category relationships (as cited in Corbin & Strauss, 2015, p. 189).

Figure 3.1 Data collection and analysis process

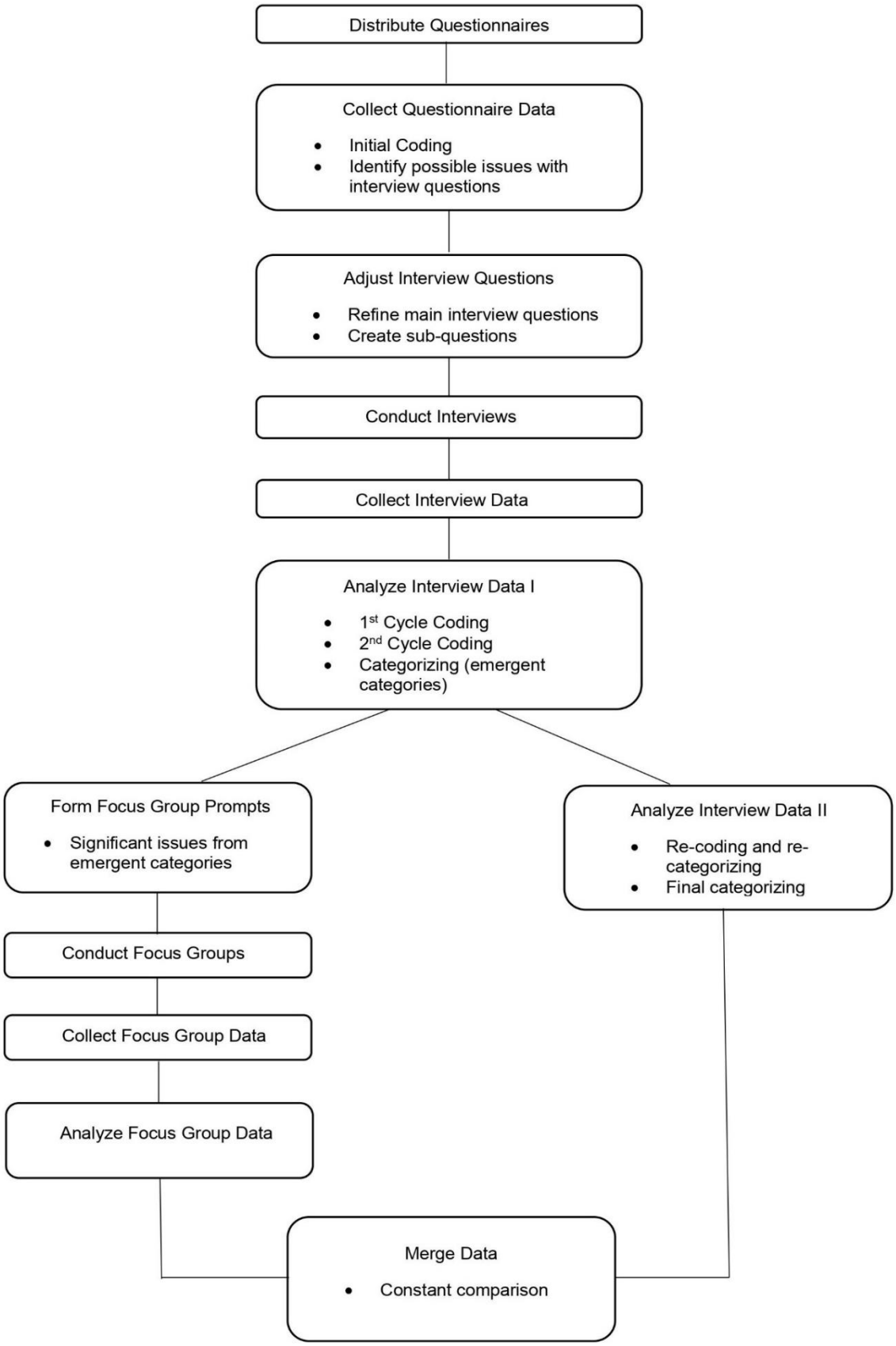
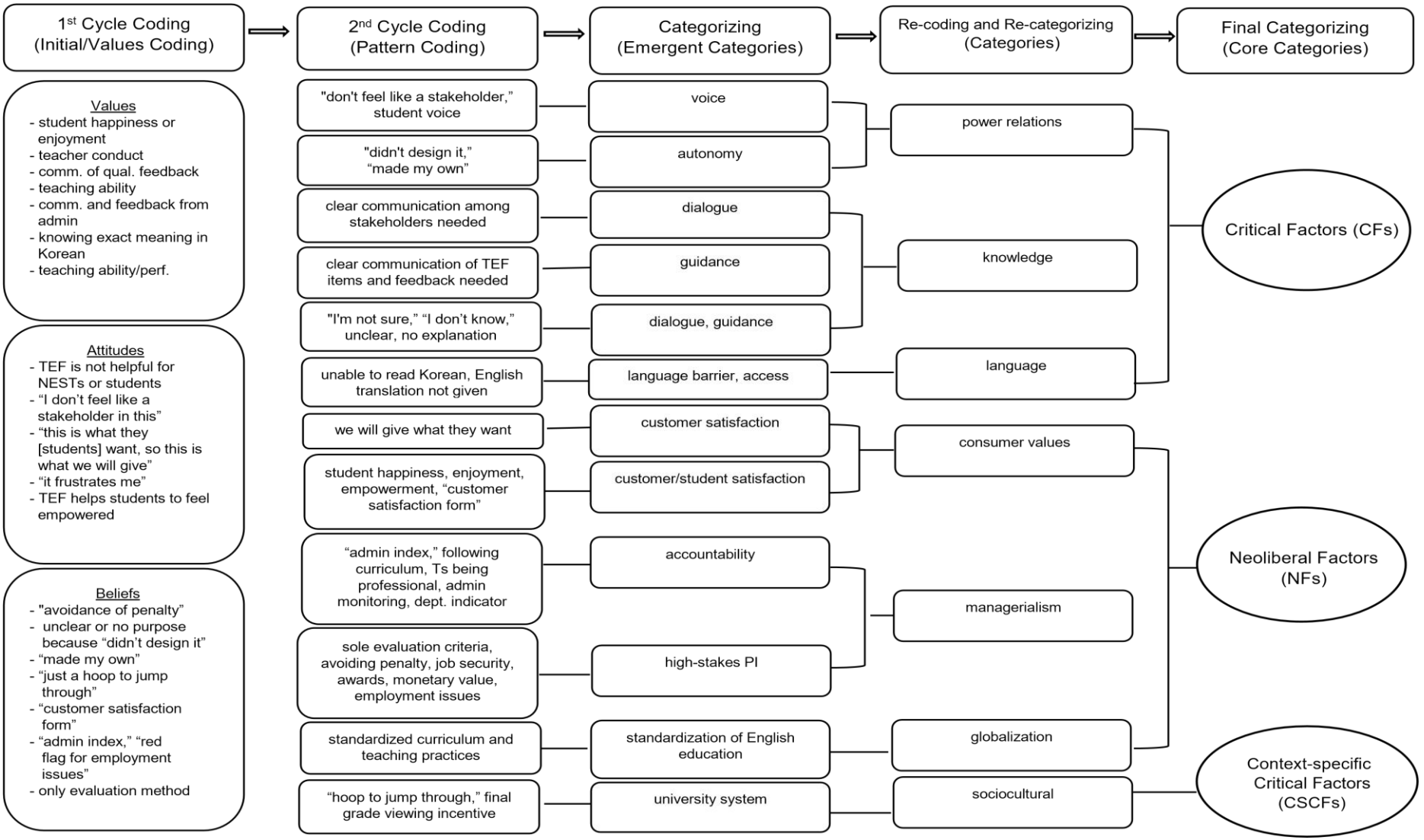


Figure 3.2 Worked example (NEST perceptions and influencing factors regarding TEF usefulness)



3.4 Ethical considerations

In addition to acquiring approval to conduct this study from my university's ethics committee (Appendix I), I endeavored to consider and strictly adhere to the following ethical aspects of this research study:

- Voluntary and informed nature of participation
- Anonymity and confidentiality of the participants and research data
- Assessment of possible harm to participants
- Anticipation of threats to validity of study results
- My own position as the researcher

Voluntary and informed nature of participation

All participants willingly took part in the study and full, written consent was obtained from them prior to the study (Appendix J). This was achieved by providing each individual ample information about the study that allowed them to understand the implications of their participation. For students and administrators, this was communicated in both Korean and English in order to avoid any misunderstanding and to help them make a fully informed decision. During this process, participants were also informed of a) their right to withdraw from the study at any stage without any pressure on my part and b) their right to gain access to copies of their interview and/or focus group discussion transcripts.

Anonymity and confidentiality

The privacy and anonymity of the participants was of utmost importance during this study. A great deal of effort was made to ensure that all individuals were provided anonymity and a good sense of security. This included assigning pseudonyms and repeatedly providing verbal and written assurances of privacy and confidentiality throughout the data collection stage (e.g. no voice recordings or personal identifiable data will be released or published for any reason).

In addition, I also informed participants of the research data protection and storage measures I would take during the study. For instance, all interview and focus group discussions were initially saved on a recording device and then deleted as soon as they were transcribed. Digital materials were saved on password-protected files, while hardcopies of transcripts, signed consent forms, data collection tools etc. pertaining to the study were stored in a locked cabinet in my personal office that could only be accessed with a specialized security card.

Assessment of possible harm

While the study did not focus on sensitive topics, minor or at-risk participants, or any other aspect of research that would require special arrangements or additional ethical approval, every effort was still made to anticipate and prevent any possible harm that participants might experience from the study. With regard to the student participants, I was aware of the high probability of them focusing on or worrying more about my position and identity as a GEP teacher rather than just a neutral researcher conducting a study. Because this perception could influence the way students reacted to me or what they would tell me while collecting data, I took extra care to gain their familiarity and trust by being as polite, understanding, and reassuring as possible. For instance, to prevent students from feeling fear of communicating their honest opinions about their professors or the TEF, I reiterated that they were merely sharing their own knowledge and thoughts, that there were no right or wrong answers, and that their anonymity would be protected at all costs.

For NESTs, I considered the possibility that they could become distressed or upset when sharing a negative experience regarding the TEF (e.g. receiving low TEF scores and its consequences). It was also possible that some might hesitate or fear to give critical opinions of the TEF as many already perceived it to be closely tied to their employment status at the university. In order to mitigate the potential stress or pressure they might have felt during the study, I continually reminded them that they could stop sharing or participating at any time, that they could always request to review their transcripts for approval, and that their privacy

was respected at all times.

Anticipation of threats to validity

In addition to making sure that participants were not subjected to any type of harm, I also considered different ways to prevent possible threats to the validity of their responses and study results. For example, although I distributed questionnaires to my own GEP students as well, I eventually chose students whom I had never taught but still fit the study participant criteria. This precaution was taken to avoid any conflict of interest or negative effect on the findings. Questionnaires given to my own students and the data collected from them were only used in connection with getting a general sense of how students viewed and understood the TEF.

Although all participants received the same questionnaire in order to increase the comparability of responses across stakeholder groups, some language adjustments were made for each target stakeholder group. For instance, I distributed questionnaires translated verbatim from English to Korean with appropriate vocabulary in consideration of student participants who perhaps did not have the necessary English skills to sufficiently articulate their thoughts about the TEF (Appendix D). This was done not only to improve the validity and reliability of their responses but also to demonstrate sensitivity towards the participants situated within this study's EFL context where they might have felt their English abilities lacking even if this was not actually true (Corbin & Strauss, 2015).

Another issue that I addressed regarding data validity involved the assumption that all participants would have opinions, be willing to share those opinions, or be able to express their opinions about the research topic (Cohen et al., 2007). As this could later become problematic, I gave participants the opportunity to indicate that they had no opinion or did not know the answer to some of the prompts in the questionnaire and interviews. I felt this was particularly important for the NEST questionnaire because it included the original, non-translated TEF as displayed online. By providing NEST participants the option of checking

off a box indicating that they could not read or understand the TEF, I had hoped to alleviate any pressure or anxiety they might have felt to write something rather than nothing. Had participants felt compelled to complete the questionnaire in any way possible, I thought it could have led to unreliable data.

In addition, I included NESTs perceived to be 'good' teachers by administrators and other faculty members; that is, teachers who had previously received teaching awards and monetary incentives based on their high TEF scores. This was to help ensure that the data collected from the study did not consist of too many dissatisfied responses or personally biased opinions towards the TEF from NESTs who maybe did not receive high enough scores or teaching awards. For both the students and NESTs, I also did my best to gain their trust and put them at ease so that they might feel more inclined to share their opinions in detail. Doing so was important because of the possibility of participants giving me responses or data that I had not anticipated collecting but that could help further my ideas within the study (Charmaz, 2014). I felt that doing this, in addition to maintaining the maximum level of objectivity in data collection and analyses throughout the study, could perhaps improve the quality of the participants' responses and validity in the study findings.

My positionality within the study

For any study, it is important for researchers to be aware of how their positionality can affect the data collection process as well as their comprehension and interpretation of the data. That is, it is crucial to consider how the researcher is positioned in relation to the researched (Lin, 2015) and to view positionality as a valuable research tool that can significantly impact the study. Positionality can be viewed as an element of reflexivity that includes numerous aspects of one's identity, such as age, race, class, gender etc. It is the understanding of the what, why, and how of research and being conscious of the role of power, privilege, and visibility in the process. By taking a reflexive approach to research and contextualizing one's multiple positions within a study, this can lead to more promising implications for data collection and analysis while, at the same time, prevent potential ethical issues (Jacobson &

Mustafa, 2019).

The influence of one's positionality can be observed in a number of critical qualitative studies (Tavakoli & Sadeghi, 2011; Perez-Milans & Soto, 2014; Lu & Hodge, 2019). For example, Tavaloki and Sadeghi (2011) conducted a study where the sources of data collection included the practitioner participant acting as critical ethnographer and on-site researcher. Through this study, they realized the necessity to take one's positionality in a serious manner as issues might arise from the participants' unwillingness to exhibit "vulnerability in sharing personal experiences" with the researcher in order to maintain their anonymity and avoid jeopardizing job opportunities (p. 370). An ethnographic study conducted by Perez-Milans and Soto (2014) included a researcher-practitioner who was positioned both as the researched and as an equal, collaborative partner with a colleague involved in the study. The recognition and acceptance of their positions within the study helped them to critically reflect on knowledge they had co-produced through both verbal and written dialogue from multiple perspectives in a deeper and more significant manner, which likely led them to richer data.

In addition, by engaging in meaningful reflexivity of one's positionality within the study, there is a higher possibility of attaining unexpected and enlightening data outcomes, as observed by Lu and Hodge (2019). In this narrative inquiry, the researcher noticed different relational dynamics between the two sets of student participants while conducting interviews. With the Swedish participants, she identified a more egalitarian student-teacher relationship whereas she observed more of a hierarchal, status-focused relationship with the Chinese participants. Coming from a Chinese background herself, the researcher continuously engaged in self-reflection through her journal and noted, for example, how she restrained herself from getting too personally involved with the Chinese students and tried to avoid imposing her feminist views that could have made them uncomfortable. As a result, the researcher was able to "initiate congenial and open collaboration" with the participants that led to a surprising and unintended change in a female Chinese student participant's attitude towards pursuing

a higher degree education, which was an act generally not encouraged in the student's culture.

Like this, by conceptualizing and embracing that my positionality consists of multiple facets and is an integral part of the research process, I felt that I was able to gain a more profound understanding of participant knowledge and perceptions regarding my study than if I had simply positioned myself as an outsider researcher. Because I saw my positionality as a advantageous research tool, this allowed me to further my capacity to critically and productively engage in more creative analysis and theorization of the data (Jacobson & Mustafa, 2019).

With regard to ethical considerations, as a practitioner-researcher, the need to protect the study participants and guard against impropriety while conducting my research was to be expected. This was especially so because I felt that I was already intimately acquainted with the study focus and context. From an ethical standpoint, it was therefore necessary to emulate the researcher in the study by Lu and Hodge (2019) and remind myself not to get emotionally involved in the study as this could lead to the loss of critical thinking ability (Corbin & Strauss, 2015). This was particularly important to consider in order to avoid siding with the participants when analyzing the interview and focus group data (Creswell & Creswell, 2018).

In order to address this issue, I made every effort to maintain a level of professional distance with the participants and engaged in constant critical self-reflection through analytic memo writing. In addition, I tried to remain attuned to how the participants perceived me (e.g. researcher, GEP teacher, colleague, employee) as "past and immediate identities" of both myself and the participants could "influence the content of interaction" (Charmaz, 2014, p. 71). In order to respect potential power imbalances of the participants during their interviews and focus groups, I withheld sharing my personal impressions, avoided leading questions,

and sought to involve the participants more as collaborators rather than just respondents (Creswell & Creswell, 2018).

CHAPTER 4: FINDINGS

The purpose of this study was to identify and examine underlying assumptions and understandings about student-completed teacher evaluations of native English speaking teacher (NEST) performance from three stakeholder perspectives: students, NESTs, and administration. This research aimed to provide insight regarding the similarities and differences in how stakeholders view the current, student-completed TEF's general purpose, usefulness, preferred use, and best way to evaluate NESTs as well as factors affecting these diverse perceptions.

Qualitative data for each stakeholder group was collected through questionnaires, semi-structured interviews, and focus groups[1]. Questionnaire responses were coded and used to modify and strengthen the four research questions that served as interview questions in order to elicit richer, more relevant answers. Interview sub-questions were also formed from the questionnaire data in order to help explain the wide discrepancy in responses. For instance, when students were asked to select the top three items from the TEF that they felt were most important and relevant to them, the overall questionnaire results showed no discernable pattern. As this was related to the third interview question (i.e. in what way do you consider the current TEF useful or not useful), interview sub-questions to this were created in order to prompt students to explain why certain items were important or useful to them. In addition, for the last open-ended question on the questionnaire (i.e. comment on what other areas of GEP teaching the TEF should evaluate), the majority of students did not provide any particular answer or simply left it blank. Because this was closely linked to the fourth interview question (i.e. what do you consider to be the best way to evaluate NESTs), sub-questions were added to give students an opportunity to voice their opinions and ensure that enough analytic details could be obtained.

[1] The TEF and all other data collection instruments (e.g. questionnaire, focus group prompts) used in this study are located in the appendices (Appendix A-H).

For each interview question, significant responses regarding the values, attitudes, and beliefs of each stakeholder group were coded, categorized, collated, summarized, and then later compared to examine similarities and differences. Significance was characterized by the “widespread agreement of the respondents on the issues” (Cohen et al., 2007, p. 465) which included unanimous or very considerable agreement. However, as numeric frequency is “not necessarily a reliable indicator of its importance or significance” (Saldaña, 2009, p. 189), the “summative power” of a response (p. 166) was also taken into consideration, as well as noteworthy utterances that shed light to context-specific responses that might have been difficult to understand otherwise. For instance, while neoliberal factors were not specifically mentioned with more frequency, upon closer inspection, it was an underlying category influencing many general critical factors. Significant interview responses from the emergent categories of the data were then used as a guide to identify each group’s perceptions and key concerns that were later discussed in the focus groups. Data from both the interviews and focus groups followed the same analysis procedure.

As a result of repeated coding, categorizing, and analytical reflection, three core categories that influenced stakeholder understanding of the TEF were formed: a) general critical factors b) context-specific critical factors and c) neoliberal factors. Within the core general critical factor category, there are four main categories: problematizing/accepting givens, language, power relations, and knowledge. Context-specific critical factors include sociocultural, sociopolitical, and historical categories, while neoliberal factor main categories are comprised of globalization, managerialism, consumer values, and university autonomy (Table 4.1). These categories were used to examine the similarities and differences in perceptions across the stakeholder groups regarding the study’s primary question. The way these three core categories and their main categories interrelate with each other and with existing literature will be reviewed in the discussion chapter.

Table 4.1

Factors and categories influencing stakeholder assumptions of TEF

Core Category	Main category	Emerging categories/factors
General critical factors (GCFs)	Problematizing/accepting givens	naming issues, reflection
	Language	access, barrier
	Power relations	autonomy, agency, voice, access, fear, anonymity
	Knowledge	dialogue, guidance, perceived truth
Context-specific critical factors (CSCFs)	Sociocultural	social hierarchy, cultural norm, age, university/education system
	Sociopolitical	political incidents
	Historical	previous experiences
Neoliberal factors (NFs)	Globalization	global influence, standardization of education
	Managerialism	business-oriented, accountability, high-stakes PI[a]
	Consumer values	customer satisfaction, student investment
	University autonomy	MoE[b]

[a] PI: performance indicator
[b] MoE: Ministry of Education

The findings in this chapter are organized and presented by interview question in order to "draw together all the relevant data for the exact issue of concern to the researcher, and preserve the coherence of the material" and to "provide a collective answer to a research question" in a clear and convenient manner (Cohen et al., 2007, p. 468). Italicized subheadings consist of keywords representing the stakeholder group participants' shared perceptions of that particular interview question. These perceptions were formed based on factors from the three core categories that emerged from the participants' interview and focus group discussion responses. In addition, the content under each subheading is illustrated with noteworthy utterances made by the participations in order to better understand and conceptualize the underlying factors that influenced their assumptions.

4.1 Q1: TEF general purpose

4.1.1 Student perceptions

Student responses on the TEF's perceived general purpose focused on three aspects:

- Curriculum improvement
- Communication improvement between students and NESTs
- "Punishment" for students and "summative purpose" for NESTs

Interview responses indicated a shared similar belief in that the TEF's intended general purpose was two-fold: overall curriculum and communication improvement. However, students also perceived that the TEF was actually used in a more negative way for both themselves as well as NESTs (i.e. punishment, summative purpose). Issues stemming from this interview data included the unclear connection between student learning goals and the TEF as well as the troubling effects of perceived anonymity regarding the TEF. These responses were used to form focus group prompts that students discussed to further elaborate on the TEF general purpose in a more critical way.

Curriculum improvement

Interview responses suggested that students assumed the TEF's general purpose to be that of curriculum improvement. This included evaluating teaching performance and for teachers to implement student feedback to help improve future classes and class quality in a general manner. Though students appeared to have a broad idea of what curriculum improvement meant to them, none of them specifically elaborated on this during their interviews. For instance, Yena repeatedly stated that the TEF pertained to future improvements to the curriculum rather than for current or formative use, but she did not provide any examples of this. Mia also agreed that the TEF was intended to help improve overall curriculum quality; however, she also thought that it was not clear what the TEF precisely evaluated and likened it to "licking the surface of a watermelon (수박 겉 핥기)." This seemed to refer to the unclear and vague TEF items that did not directly correspond with any particular English learning goals or NEST teaching practices, which was something that she repeated throughout her interview and during the focus group discussion.

Communication improvement

During their interviews, students understood the TEF to be a way to improve communication with their professors. For example, Jina saw the TEF as a "bridge" for Korean students to interact with NESTs while Yoon described it as a way to indirectly and tactfully communicate complaints to professors because it would be "uncomfortable" to do so otherwise (CSCF/sociocultural/social hierarchy).

All of the participants repeatedly spoke of anonymity and comfort levels regarding communication improvement with NESTs; for Yena, this was of particular importance:

> [I]t is, of course, not possible to directly approach someone who is in a higher position of authority and give them feedback, but with the TEF, you can deliver that feedback in written form and in a more anonymous way. (GCF/accepting givens, CSCF/sociocultural/social hierarchy, GCF/power

relations/anonymity)

In addition, interview responses indicated that students viewed the TEF's purpose was to promote honest communication that is not only straightforward and non-offensive (CSCF/sociocultural/social hierarchy) but also anonymous so that students would feel safe enough to provide more critical feedback (GCF/power relations/fear, anonymity). This sentiment of "safe communication" improvement through guaranteed anonymous TEFs was repeated and further emphasized in the focus group. All student participants agreed that if teachers and administrators were to set a clear precedent of guaranteed anonymity, then it would be possible to realize the TEF's intended purpose of improving communication with NESTs.

"Punishment" for students and "summative purpose" for NESTs

Interview responses showed that students assumed TEF results were used to "punish" both students and NESTs. All participants displayed a distinct lack of trust in the supposedly anonymous nature of the TEF. For instance, Yoon expressed a fear of retribution through "reverse evaluation (역평가/*yeok-pyeongga*)," while Mia mentioned the possibility of "reverse discrimination (역차별/*yeok-chabyeol*)" (GCF/power relations/fear) regarding their final grades depending on how NESTs respond to their TEF scores.

In addition to reflecting on the punishment aspect of the TEF for students, the participants also observed a summative purpose for NESTs. During their interviews, Mia drew on her experience working as a teaching assistant in main administration and referred to the TEF as a "teacher performance measuring stick (교수 잣대)" (NF/managerialism/high-stakes PI), while Jina commented on how the TEF appeared to be "a way for superiors to 'peck at' and pressure teachers (윗 사람 처럼 쪼아대는 느낌)" for NESTs. Yoon also assumed that Korean universities used low TEF scores as cause for termination as she recalled an instance when her current major professor claimed that he was dismissed from his previous university

teaching position for this reason (NF/managerialism/high-stakes PI).

Further discussion in the focus group revealed that Mia and Yena had experienced direct and threatening in-class confrontations with their Korean professors regarding their low TEF scores, while Jina and Yoon were made aware of these occurrences through their friends (CSCF/historical/previous experiences). These encounters included Korean professors teaching major classes angrily questioning "[t]hese are my [midterm] TEF results? What the hell is this? Who do you think you are?" Students' reactions and thoughts when confronted with these experiences and stories were those of panic and dismay (GCF/power relations/fear). All of the participants stated in the focus group that these experiences with their Korean professors negatively affected how they perceived the TEF. This attitude carried over to their relationships with NESTs and GEP classes (CSCF/historical/previous experiences) and further strengthened their belief that the TEF was not truly anonymous and how it was more than plausible for NESTs to use it to punish students (GCF/knowledge/perceived truth).

Table 4.2

Student perceptions of TEF general purpose

TEF general purpose	Main influencing categories and factors
Curriculum improvement	
Communication improvement	GCF/power relations/fear, anonymity
	GCF/accepting givens
	CSCF/sociocultural/social hierarchy
"Punishment," summative	GCF/power relations/fear
	GCF/knowledge/perceived truth
	CSCF/historical/previous experiences
	NF/managerialism/high-stakes PI

4.1.2 NEST perceptions

Overall, NEST interview and focus group responses showed a lack of consensus in details regarding the general purpose of the TEF among participants in that there was uncertainty and little overlap in their answers other than in a very broad sense. Their answers were influenced by a wide range of factors such as accountability, student satisfaction, fear, lack of knowledge, and language issues. The following NEST perceptions were chosen to help understand the train of thought and complexities behind the lack of consensus regarding the TEF's general purpose:

- Curriculum quality maintenance
- Student satisfaction and voice
- Summative, unclear, or no constructive purpose

Prior to presenting the findings for NESTs in this study, it should be mentioned that with the exception of John, who is fluent in Korean, the other participants had never fully understood the content of the TEF. None of them had direct access to an English translation of the TEF before taking part in this research (GCF/language/lack of access), and many of their answers began with "I think…," "I guess…," "I'm not quite sure…," "Maybe…," and "I don't know, but…" which indicated a vague understanding of the TEF. This was corroborated during the focus group when everyone agreed with Kelly's comment:

> We were all talking about the fact that when we arrive [at the university] we were not told about the TEF… we had no idea about it. At all. About the questions, about the purpose, or anything, and we all kind of learned as we went along.
> (GCF/knowledge/lack of guidance)

Considering their previous lack of knowledge of the TEF, it is perhaps not surprising that their responses seemed incongruent and lacked coherence at times. Because of this, the

overall idea of trying to examine NEST perceptions of the TEF that they were not familiar with seemed flawed. However, I endeavored to make note of these inconsistencies and relate them to the findings to the best of my ability.

Curriculum quality maintenance

In their interviews, all NESTs suggested that the overall purpose of the TEF was to provide a very basic and holistic overview of teaching performance in relation to maintaining a satisfactory level of curriculum quality and whether NESTs were conducting themselves in an overall professional manner (NF/managerialism/accountability). During his interview, John further surmised that the TEF's main purpose was to see whether the teacher was following the standardized curriculum determined by university administration (GCF/power relations/ lack of autonomy, NF/globalization/standardization of education). Kelly's response supported this when she commented on how administration was "just making sure you're doing the basic part of your job correctly" by reviewing final TEF results on a superficial level (NF/managerialism/accountability). However, other than these two responses, no other specific examples or details related to the formative use of the TEF for curriculum improvement or quality maintenance were mentioned by the remaining participants in either the interviews or focus group.

Student satisfaction and voice

NEST interview responses also suggested that the TEF did not adequately address English teaching goals, practices, or performance. Instead, participants such as John and Mark assumed that the TEF was an attempt to assess whether students were happy (i.e. to measure "student enjoyment level," "students' happiness level") (NF/consumer values/customer satisfaction). John further elaborated on this opinion in his interview:

> [T]his is going to sound really cynical, but… I think it's more of a customer satisfaction form… In terms of the teachers, once all the numbers for this is set up… this is going to sound very unprofessional, but it's saying that 'this is what they want, so this is what we'll give' kind of thing. I don't think

> that's how an academic institution should be run.
> (NF/consumer values/customer satisfaction, NF/managerialism/business-oriented)

In addition to John and Mark's thoughts, Kelly and Rob both mentioned the possibility of students using the TEF as a way to get their voice heard or to feel empowered to have a say (GCF/power relations/voice), but neither could offer any other specific or relevant reasons. However, during the focus group, all participants discussed and agreed that the TEF's primary concern was to measure student satisfaction levels with NESTs and GEP classes. Mark and Rob were particularly critical about this because they believed there were many external and possibly irrelevant factors affecting student happiness/satisfaction that were either not taken into consideration (Mark) or not explained fully by students through the TEF or any other means (Rob).

Summative, unclear, or no constructive purpose

In their interviews, all NEST participants expressed that the TEF's purpose was unclear to them but that it appeared to be mainly summative and negative in nature. Mark in particular thought that the TEF served no constructive purpose for NESTs. He also did not see any benefits for students because completing the TEF was "just a hoop for them to jump through to get their grade," referring to the university policy that makes it mandatory for students to fill out the TEF if they wish to receive their final grades a few days earlier than the actual grade release date (CSCF/sociocultural/university system). This was considered to be advantageous for students because it would allow them to have a few extra days to challenge their professors if they believed they deserved a higher grade in class. Matt's opinion about how the lack of TEF purpose for students was echoed by both Kelly and Rob in their interviews.

In addition, interview responses showed that in the absence of any other known evaluation technique or method, NESTs mainly viewed the TEF as a high-stakes PI and administrative

index used to assess teachers with numbers (John, Mark, Rob), present best teacher awards of monetary value (Kelly), and determine employment-related matters (Kelly, Mark, John, Rob) (NF/managerialism/high-stakes PI). Mark maintained:

> The main purpose is to avoid getting a bad TEF evaluation that puts them towards the bottom of their department's ranking because that could jeopardize their job. So I feel like the teacher's purpose is more focused on avoidance of penalty.
> (NF/managerialism/high-stakes PI, GCF/power relations/fear)

This negative view about the TEF and its purpose was further discussed in the focus group. Surprisingly, unlike the other NEST participants, Mark stated that he never reviewed the student-completed TEF or its numerical scores because of the lack of English translation of the items and qualitative feedback (GCF/language/lack of access). He never found it helpful, and therefore administered his own evaluation form tailored to his needs (GCF/power relations/agency). This disclosure during the focus group both startled and intrigued the other participants because of Mark's current position as the head curriculum coordinator and as a NEST who had received best teacher awards every year since the beginning of his employment. They assumed that a NEST who consistently achieved high TEF scores would be more supportive of and attuned with the form; instead, Mark demonstrated the opposite and firmly stated that the TEF held no purpose for NESTs because they did not design it (GCF/power relations/lack of autonomy, voice, agency).

The participants also discussed in length the discrepancy and temporal change in how the TEF was viewed by both administrators and NESTs. For instance, Rob commented on how he saw it "as a way to help teachers" and a "formative form of feedback" while "admin sees it as a way to judge my performance… and as a summative outcome." Kelly and John both mentioned "that point when it changed" where it was suddenly announced that the TEF was to be used as "an indicator of performance" (NF/managerialism/high-stakes PI). The

participants also attributed their varying answers regarding the purpose of the TEF to an absence of knowledge stemming from little guidance and communication from university administration.

Table 4.3

NEST perceptions of TEF general purpose

TEF general purpose	Main influencing categories and factors
Curriculum quality maintenance	GCF/power relations/lack of autonomy
	NF/globalization/standardization of education
	NF/managerialism/accountability
Student satisfaction and voice	NF/consumer values/customer satisfaction
	NF/managerialism/business-oriented
Summative or no purpose	GCF/knowledge/lack of guidance
	GCF/language/lack of access
	CSCF/sociocultural/university system
	NF/managerialism/high-stakes PI

4.1.3 Administrator perceptions

GEP administrator responses regarding the general purpose of the TEF were mostly influenced by neoliberal factors such as managerialism, globalization, and university autonomy. Their perceptions focused on the following:

- Curriculum improvement
- Student satisfaction assessment
- Summative or no constructive purpose

Although there was initially little consensus on how the TEF was used by administrators, some of this was resolved during the focus group discussions. Here, it is important to note that the focus group was conducted twice; the first focus group included all GEP administrators while the second focus group excluded one person (Jin) at the request of the remaining three (Hana, Hyuna, Siri). This was due to the perceived difference in authority and job status within the department that prevented these three participants from expressing their honest opinions in front of their superior (GCF/power relations/fear, CSCF/sociocultural/social hierarchy). The differences in responses between these two discussions were noted throughout the findings.

In addition, GEP administrators considered themselves different from administrative staff working for main administration because their responsibilities mostly consisted of handling NEST affairs, GEP elective classes, and acting as a go-between for NESTs and main administration. On the other hand, main administrators were seen as managing major classes and making all final decisions regarding GEP curriculum and NEST employment. Because of these reasons, a distinction was made between these two types of administrators for the purpose of this study: GEP administrators and main administrators.

Curriculum improvement

In their interviews, all GEP administrators briefly mentioned that one main purpose of the TEF was curriculum improvement by creating an opportunity for students and teachers to negotiate and communicate their needs through the TEF (GCF/knowledge/dialogue). However, this was mentioned on a superficial level, and most participants did not provide any specific examples or details. Only Siri suggested how NESTs could maybe use the TEF results to prepare course schedules and materials. She also expressed that it was main administration's responsibility to incorporate both students and NESTs' opinions regarding the curriculum by reflecting on the TEF results (GCF/knowledge/dialogue, NF/managerialism/accountabilty). In addition, some GEP administrators such as Jin stated

that the TEF's purpose was to make sure that student perspectives were acknowledged in order to improve class quality. However, later in her interview and during the first focus group, this emphasis on recognizing student views appeared to be more related to raising student satisfaction levels rather than giving them a voice in the curriculum (NF/consumer values/customer satisfaction).

Student satisfaction

Student satisfaction in GEP classes and NESTs was the most frequently mentioned factor throughout the study for all GEP administrators in their interviews and both focus group discussions. Interview responses revealed a business-oriented aspect of the TEF in that its perceived purpose was to "provide (제공/*jegong*) better quality classes to students" (Siri) by measuring and improving student satisfaction levels of the "services (서비스/*sseobiss*)" provided to them by NESTs (Hana, Jin). (NF/managerialism/business-oriented and accountability, NF/consumer values/customer satisfaction).

In her interview, Jin further stressed that "[a]ssessing student satisfaction levels is extremely important" because of its relation to "English education that will strengthen our students' global influence around the world" (NF/globalization/global influence) and implied a correlation between student satisfaction levels and English ability improvement. The importance placed on attaining and maintaining high student satisfaction was made even more apparent when Jin revealed in the interview that GEP administrators had been conducting a separate TEF for years without notifying NESTs or sharing any resulting data (GCF/knowledge/lack of dialogue). During the first focus group discussion, Jin explained how this special TEF was tailored for GEP classes and rigorously administered throughout the year. She also said that the specific purpose of this TEF was to improve student satisfaction levels by standardizing GEP curriculum (NF/globalization/standardization of education) and reducing the number of student complaints. This was considered important as all GEP administrators equated few student complaints to high levels of student

satisfaction regarding both the original and special TEF (NF/consumer values/customer satisfaction).

Summative or no purpose

Another collective assumption about the TEF's general purpose was its perceived summative aspect. In their interviews, Hana thought that the actual purpose of the TEF was "punishment for NESTs" and to determine best teacher awards, while Siri thought that main administrators used the TEF to make personnel decisions (NF/managerialism/high-stakes PI). Jin's answer supported these views:

> The most important thing is the evaluation results. That is really important for professors' renewal. We also use the results to rank NESTs and focus on 'low score ranking professors' to determine best teacher awards. Low scoring professors have private meetings with [main] administration to discuss their problems.
> (NF/managerialism/high-stakes PI)

Some GEP administrators also observed that the TEF did not serve any obvious purpose other than ranking NESTs. For example, during her interview, Hana shared her opinion that the TEF was mostly "for show" because "the government is telling us to do it" as a convenient way to leave written documentation or records to present to the Ministry of Education (MoE) upon request (NF/university autonomy/MoE, NF/managerialism/accountability). This opinion was discussed again in the second focus group where Hyuna and Siri also saw the TEF as a "satisfaction survey that is just a formality" that was recently enforced by main administration under the direction of the MoE (NF/university autonomy/MoE). All three GEP administrators surmised that the sudden scrutiny of student-completed teacher evaluations was likely because of new educational policy changes to address student complaints and protests stemming from a political scandal that involved the impeachment of former South Korean president Park Geun Hye

(CSCF/sociopolitical/political incidents). These changes included placing a heavier emphasis on attendance checks, absence policies, and class cancellations in the current TEF.

Table 4.4

GEP administrator perceptions of TEF general purpose

TEF general purpose	Main influencing categories and factors
Curriculum improvement	GCF/knowledge/dialogue
	NF/consumer values/customer satisfaction
	NF/managerialism/accountability
Student satisfaction	NF/consumer values/customer satisfaction
	NF/managerialism/business-oriented, accountability
	NF/globalization/global influence
Summative or no purpose	CSCF/sociopolitical/political incidents
	NF/managerialism/high-stakes PI, accountability
	NF/university autonomy/MoE

4.2 Q2: TEF usefulness

4.2.1 Student perceptions

Overall, in their interviews, students recognized a limited sense of usefulness regarding the TEF. Some displayed an ambivalent attitude in that it was not necessarily useful or useless, while others thought it was just okay or perceived the TEF's value as conditional. For example, Yena thought that the TEF could be useful but only if university administrators communicated the TEF feedback directly to NESTs and made sure that feedback was implemented into the curriculum. These perceptions were influenced by the following:

- Lack of clarity and relevance of TEF items
- Lack of interest and investment

- Fear

Responses from the focus group further supported and explained these influencing factors and assumptions. Student participants shared specific details and showed almost complete agreement when discussing the prompts. Findings from both the interviews and focus group suggested that students doubted the usefulness of the TEF and its validity as a result.

Lack of clarity and relevance

Interview responses indicated that the TEF items in particular were unhelpful to students for a number of reasons: their superficial nature, lack of clarity, and irrelevance to GEP classes. During her interview, Mia compared the usefulness of the TEF by repeating the expression "licking the surface of a watermelon (수박 겉 핥기)," referring to the seemingly meaningless purpose of the TEF and its items. She believed that the overly general and shallow nature of the TEF items did not help with core issues involving GEP class quality, curriculum improvement, or student learning because they did not elicit detailed or elaborate responses, nor did the items allow students to provide specific explanations for their answers.

The lack of clarity in the wording of TEF Likert scale items was also a great concern for the participants. All of them found the wording to be "vague, confusing, and extremely abstract" (Jina) in such a way that they were open to various interpretations that could also be affected by a multitude of factors unrelated to teaching performance or class curriculum. This confusion interfered with the students' ability to correctly comprehend the item and choose a number that best represented their thoughts and opinions (GCF/knowledge/lack of guidance). Three out of four students objected to the idea of Likert scale items as they believed that numbers were incapable of representing one's complex thoughts regarding the evaluation of teacher performance or curriculum quality.

However, unlike Yena, Mia, and Jina, Yoon seemed to be more understanding of the use of these items. In her interview, she shared her opinion that the TEF was deliberately made uniform and standardized in this manner in order to easily analyze the statistical results and thereby improve administrative efficiency (NF/managerialism/business-oriented). Nonetheless, later in the focus group, she agreed with Mia and the other participants in that the TEF items lacked clarity (i.e. they were too ambiguous, general, abstract, and vague) and did not elicit constructive student feedback about GEP classes. During the focus group discussion, students also showed difficulty understanding the relevance of these items when evaluating NEST performance as the TEF did not seem to reflect student English learning goals. Mia shared how "it feels like students are forced to rate teachers because the items don't reflect what students deem important," while Yena lamented on how "self-reflection items are just tossed out there" with no real meaning or connection between students and GEP classes (GCF/knowledge/lack of guidance). These responses, along with those from the interviews, indicated that students viewed the TEF to be largely unhelpful due to the lack of clarity and relevance in the items themselves but also because of the absence of guidance and knowledge about the form itself.

Lack of interest and investment

Student interview responses showed a general lack of interest in completing the TEF, which negatively affected the way they viewed the TEF's usefulness. All participants mentioned that they, as well as the majority of their friends and classmates, mostly completed the TEF to receive their final grades early but nothing more (CSCF/sociocultural/university system). Mia in particular, drawing from her experience as student body president, said that while she personally did her best to provide detailed feedback, she observed that most of her peers would fill out the TEF hastily and not give qualitative comments because it was "bothersome" and did not prove to be helpful for students currently taking the class (NF/consumer values/lack of investment). Jina thought that this behavior was likely because Korean students had a tendency of doing things "very quickly skimming through, paying scant attention (대충대충 /*daechoong-daechoong*)" and "[q]uick, quick, quick, quick!"

(CSCF/sociocultural/cultural norm).

During the focus group, students also frequently mentioned their past experiences from high school as well as other university major and elective classes in relation to how useful they viewed the TEF (CSCF/historical/previous experiences). Mia claimed that due to the rote learning and silent-method teaching style of Korean high schools (CSCF/sociocultural/education system), most students arrived at university accepting that that was just how things were: “[t]here’s nothing to give feedback on. It’s not like the professors will accept anything we say” (GCF/accepting givens). Yoon agreed and added that this shared general attitude was because there were no opportunities to ask teachers questions in the classroom (CSCF/sociocultural/education system).

Moreover, Mia and Yena both commented on how they did not believe that NESTs or university administrators really cared about student feedback because of the way TEF was “just tossed out there” without any visible follow up (GCF/knowledge/perceived truth, GCF/knowledge/lack of guidance). All of the students agreed with Mia when she said “student motivation starts with teachers” and that it was important to “emphasize necessity rather than compulsion or coercion” in order to encourage students to complete the TEF in a meaningful way. These responses, along with those in the interviews, suggested that students were greatly affected by previous experiences prior to attending university that led to a lack of trust towards teachers and administrators. They therefore did not perceive any benefits in carefully completing the TEF or providing detailed qualitative feedback for NESTs or GEP classes.

Fear

The most noticeable factor affecting how students perceived the TEF’s usefulness was fear related to anonymity. During her interview, Yena spoke about her fear and lack of trust towards the TEF and how she felt it was used by NESTs and administrators. For instance, she was convinced that NESTs who received low TEF scores had the technological means

to track down students and punish or treat them unfairly by giving them poor final grades (GCF/knowledge/perceived truth, GCF/power relations/fear, anonymity). In particular, she spoke of her high school and university experiences:

> There was an incident in high school where even though the evaluation form was supposedly anonymous, somebody saw that student ID numbers were partially shown on the results sheet… that was a problem… [a]nd our university's TEF is apparently connected to our final grades so… maybe my lack of trust and not wanting to speak up comes from the environment I grew up in
> (GCF/power relations/fear and lack of anonymity,
> CSCF/historical/previous experiences)

When this issue was further reflected upon in the focus group, all students openly shared this same fear towards both NESTs and Korean professors and how they believed the TEF could be used against them in different ways (GCF/power relations/fear, GCF/knowledge/perceived truth). Mia asserted at the beginning of the discussion "[t]his can be summarized in one sentence – if anonymity isn't guaranteed, we will be treated unfairly and punished accordingly." This statement was followed by students' doubts of the TEF's usefulness because they thought it highly likely that TEF responses were not sincere, thorough, or honest due to a lack of guaranteed anonymity when completing the form (GCF/power relations/fear and anonymity).

This led the students to believe that the subsequent TEF results could not be trusted. These misgivings appeared to stem from a number of reasons, such as personal experiences in the classroom where Korean professors used threatening language and directly confronted students about receiving low TEF scores (CSCF/historical/previous experiences). While this concern involving a lack of mistrust, fear, and anonymity was mentioned more in connection with Korean professors, students stated that these attitudes and experiences extended to

NESTs, thus adversely influencing the way they perceived the TEF and its usefulness regarding GEP classes (CSCF/historical/previous experiences, GCF/power relations/fear).

Table 4.5

Student perceptions of TEF usefulness

TEF usefulness	Main influencing categories and factors
** TEF not useful because of:*	
Lack of clarity and relevance	GCF/knowledge/lack of guidance NF/managerialism/business-oriented
Lack of student interest and investment	GCF/knowledge/lack of guidance NF/consumer values/lack of investment CSCF/sociocultural/education system, cultural norm CSCF/historical/previous experiences
Fear	GCF/power relations/fear, anonymity CSCF/historical/previous experiences

4.2.2 NEST perceptions

With the exception of Mark, NEST participants indicated a possibility for the TEF to be useful if used in a formative manner by providing an opportunity for teacher reflection, which in turn could affect professional development and curriculum improvement. In particular, the qualitative items were seen as having the most potential to be helpful for NESTs and their teaching practices. However, the vast majority of NEST responses in both the interviews and focus group were negative in nature regarding the pragmatic use of the TEF because of the following influences:

- Lack of clarity and relevance of TEF items

- Unhelpful or lack of feedback
- Lack of knowledge

Lack of clarity and relevance

NEST interview responses regarding the usefulness of the TEF included a lack of clarity in the wording of the TEF as well as the absence of relevant items regarding English language teaching, NESTs, and their teaching goals. Quantitative items were seen as superficial, too general, “really undefined,” (Mark) and consisting of “subjective expressions” (John) as well as “awkward word choice” (Kelly). Likert scale-based items in particular were considered to be unclear and unhelpful to students and NESTs alike (GCF/knowledge/lack of guidance). NESTs talked at length about how and why the TEF items were unrelated to their own teaching practices. They called attention to items that they believed only applied to administrative matters or knowledge-based courses as they did not seem to reflect English teaching or effective language teaching (GCF/knowledge/perceived truth, GCF/problematizing givens/naming issues. For instance, Kelly believed that Item 7 (satisfied with class content) did not apply to NESTs because anything related to class content was a decision made by main administrators as NESTs did not have any control over any curriculum-related changes (GCF/accepting givens, GCF/power relations/lack of autonomy and agency).

During his interview, Mark explained that the TEF was not helpful not only because NESTs did not design it (GCF/power relations/lack of agency and voice, GCF/problematizing givens/naming issues) but because the items were “just not really focused on our department.” He further elaborated on this by saying that the majority of the items were “just administrative, professional checkboxes” (NF/managerialism/accountability):

> Currently, my teaching goals are somewhat formed by the sort of top-down mandates that are given… I’m trying to balance any kind of pressure from administration about what they want me to teach and what

> I perceive students wanting to learn… we're asked to teach just really standard academic writing, so I'm trying to modify my class goals to a certain extent…
> (GCF/power relations/lack of autonomy, NF/globalization/standardization of education)

In the focus group, NESTs further discussed how the TEF was not useful because of the lack of clarity and definition in the wording, and they spent a good deal of time negotiating the meanings of terms such as syllabus vs. course schedule, and learning goals. They also noted that the TEF scores did not seem relevant to or aligned with the achieved goals. As a result, NESTs found the TEF to be generally unhelpful and questioned the accuracy and quality of the responses due to issues with clarity and relevance (GCF/problematizing givens/naming issues).

Unhelpful or lack of feedback

In addition to concerns regarding the lack of clarity and relevance of items, NESTs also expressed frustration with the feedback they received from both students and administrators. Although NEST responses indicated that qualitative items had the most potential to be useful, they found the actual qualitative feedback to be of little value. In their interviews, all of the participants voiced how they were not able to use qualitative student feedback to improve their teaching practices. This was because most students either did not provide any comments, gave poorly written English feedback irrelevant to teacher performance or class quality, or gave feedback in Korean that NESTs could not read (GCF/language/barrier).

For example, Kelly mentioned how students "might just comment on my personality… [s]o when I see that, it kind of frustrates me a little bit because I would like to have some suggestions for improvement," while Mark said "[t]he English responses tend to be 'I love your class!' or 'too much homework!' or really simple, not really valuable responses." Rob also stated that while he paid more attention to the qualitative comments rather than the

numerical scores, they were not helpful because he was not able to identify any teaching pattern issues in the written feedback that could help explain his low TEF scores. While all NESTs stated that student qualitative feedback was generally unhelpful, some also discussed the TEF feedback from main administrators and how it negatively affected their teaching or professional development. In his interview, John shared:

> [I]f I get a low score on that (the TEF), the next time, I'm not thinking about my… I'm not reflecting on my teaching. It's just how do I get that score back up? I mean, my job is… I'm on a contract. And so because of that, there's very little leeway in terms of how far or how long the probation or warning will go before they decide to say 'because of this, we will not renew your contract.' And that kind of fear is something that… is real. Having this kind of feedback… I mean, I almost want to say that it's more harmful than it is… helpful.
> (NF/managerialism/high-stakes PI, GCF/power relations/lack of agency, fear)

The focus group discussion echoed the participants' interview responses in that the vast majority of TEF qualitative feedback provided by students was considered to be inadequate and of no use to NESTs with regard to improving teaching practices. In addition, all of the participants agreed with John in terms of how the little feedback received from main administrators had the opposite effect of what was intended due to the high-stakes nature of the TEF. That is, instead of motivating NESTs to engage in professional development to improve practice or curriculum, it would instead make them only worry about and concentrate on improving their final TEF scores.

Lack of knowledge

Another reason why the TEF was viewed as not useful was because of the overall lack of knowledge NESTs had regarding the TEF. In their interviews, participants called particular

attention to the absence of communication and guidance from both GEP and main administrators regarding the TEF itself (GCF/knowledge/lack of dialogue and guidance, GCF/problematizing givens/naming issues). For example, none of the NESTs had any knowledge of what each individual item score represented, nor were they aware of how the TEF scores were calculated. As a result, NESTs did not find the TEF scores useful because of the lack of trust and knowledge in value placed on the numbers and what those numbers signified. They also added that not knowing how the TEF items applied to teaching goals or student learning goals or how to effectively interpret and use TEF results had a negative effect on both their professional development and curriculum improvement. During his interview, John reflected:

> I think one of the difficulties is that there isn't an explanation of the purpose, like from administration, for example, or whoever made the evaluation form… [t]here's no specific feedback… like, I don't feel like a stakeholder in this at all… the hardest part is not feeling like a person with the ability to choose or decide.
> (GCF/knowledge/lack of dialogue, GCF/power relations/lack of agency, autonomy)

Other participants also described the difficulties they faced regarding their positions as NESTs with little knowledge in relation to the TEF and teaching improvement. In her interview, Kelly stated the importance of all stakeholders being in possession of the same knowledge about the TEF. She emphasized the need for comprehensive guidance for both students and NESTs so that "students don't just click random numbers, as in somewhere in the middle. It encourages mediocrity. Like, don't try anything new, or the students may not like it, and then you're going to get a lower eval"
(GCF/knowledge/lack of guidance, NF/consumer values/customer satisfaction).

The lack of guidance and knowledge affecting the way NESTs viewed the TEF's lack of

usefulness was further discussed in the focus group. While talking about the issue of lack of communication among stakeholders, everyone concurred that English translations of the TEF were never provided to the NESTs despite the importance placed on it by university administrators (GCF/language/lack of access) and that not knowing what the TEF items meant or represented on a basic level prevented them from fully understanding and using the TEF to their advantage (GCF/problematizing givens/reflection). All of the participants agreed that inadequate knowledge and guidance of the TEF could render the results meaningless for NESTs and could likely have an overall harmful effect on their teaching and class quality, which could, in turn, negatively affect student learning.

Table 4.6

NEST perceptions of TEF usefulness

TEF usefulness	Main influencing categories and factors
** TEF not useful because of:*	
Lack of clarity and relevance	GCF/knowledge/lack of guidance
	GCF/power relations/lack of autonomy and agency
	NF/globalization/standardization of education
	NF/managerialism/accountability
Unhelpful or lack of feedback	GCF/language/barrier
	GCF/power relations/fear, lack of agency
	NF/managerialism/high-stakes PI
Lack of knowledge	GCF/knowledge/lack of guidance, dialogue
	GCF/language/lack of access
	GCF/power relations/lack of autonomy, agency
	NF/consumer values/customer satisfaction

4.2.3 Administrator perceptions

GEP administrators shared mixed reactions to the TEF and its perceived usefulness in their interviews. Their responses also changed and varied depending on which focus group discussion they participated in (i.e. all participants vs. Hana, Hyuna, Siri). While Hana, Hyuna, and Siri saw the TEF as being only somewhat useful on a superficial level for administrative purposes, Jin viewed the TEF in a more positive light; mainly, in regard to its role of assessing student satisfaction levels. However, all participants later agreed that the TEF was largely unhelpful in terms of improving NEST teaching performance or GEP curriculum quality because of the following:

- Lack of clarity and relevance of TEF items
- Unhelpful or lack of feedback
- Neoliberal factors

Lack of clarity and relevance

GEP administrator interview responses indicated that the TEF was too ambiguous and/or superficial to be of any specific or practical use to NESTs or students. During her interview, Hana used the Korean expression “[p]ut it on a nose, it becomes a nose ring; put it in your ear, and it becomes an earring” to describe the generic nature of the TEF. She further explained this was an apt phrase because the broad items on the TEF could be applied to any academic context: elementary, high school, or university. In addition, Hyuna commented that instead of providing clear, itemized questions, “[the TEF] subordinate factors are rolled into one,” making it problematic for students to give answers that accurately represent their opinions about their classes and professors.

The interview responses also suggested that the unclear and ambiguous wording of certain TEF items could confuse and mislead students into providing inaccurate responses. For example, Siri questioned Item 3 (syllabus and course schedule) because there could be times when professors would have to modify the course schedule due to unavoidable

circumstances; however, students would be obligated to give lower scores for this item because the professor technically did not follow the schedule as originally stated in the syllabus. Hyuna mentioned that many TEF items lacked the capacity to fully express students' complex thoughts. For example, regarding the student-teacher interaction items, she questioned whether "students can give a number score indicating that there was in-class interaction between professors and students, but does that technically mean that they improved their English abilities? This can be confusing for them" (GCF/knowledge/lack of guidance).

In addition to ambiguous wording, GEP administrators also commented on the seeming lack of relevance of the TEF to NESTs because of the absence of GEP-specific or English language learning/teaching items. This also made some of the participants question the value and usefulness of the TEF results and feedback during their interviews. For example, the only item that explicitly mentioned anything foreign language-related was Item 12 (foreign language %), but some understood this item to be mostly applicable to Korean professors giving lectures in English. Hana and Siri further expressed difficulty in understanding how this item could be applied to classes taught by NESTs and believed it to be "meaningless" because teaching English in English was to be expected by anyone. Concerns about the lack of clarity and relevance of TEF items were further discussed in the second focus group consisting of Hana, Hyuna, and Siri. They all lamented how the items were too general, vague, and irrelevant to NESTs. In particular, all three GEP administrators were critical of Item 6 (cultivating knowledge of the subject), as they did not see it at all relevant to English language teaching.

Unhelpful or lack of feedback

In their interviews, all GEP administrators said they mostly used the TEF to measure student satisfaction levels and reduce student complaint numbers (NF/consumer values/customer satisfaction). However, as TEF items did not seem to elicit specific, constructive responses from students, the participants considered this feedback to be unhelpful for NESTs with

regard to improving teaching performance and curriculum quality. For example, the TEF items were seen as very biased in favor of students (Siri, Jin), and the qualitative feedback was too “student-subjective” and “personal” to produce any constructive or practical feedback for NESTs or the curriculum (Hana, Hyuna). Common qualitative feedback from students included comments such as “Oh, I like this professor,” “This professor is really nice like an angel,” “He gives too much homework,” and “The professor has a cute dog,” all of which were seen as irrelevant or unhelpful to improving the curriculum or NEST teaching. Hana, Hyuna, and Siri also pointed out in each of their interviews how the TEF was only offered in Korean and that the majority of NESTs were unable to comprehend the meaning of any of the items, much less the qualitative feedback, without translation help from GEP or main administrators (GCF/language/barrier, lack of access). They perceived this to be a serious problem contributing to NESTs not being able to use the TEF or its data in any practical way.

During the first focus group, when asked to expand on the issue of lack of communication or exchange of feedback knowledge between NESTs and main administrators, Hana attributed this to “social, cultural, and language differences” affecting that relationship (GCF/problematizing givens/naming issues, CSCF/sociocultural). Jin further elaborated:

> I think the language barrier is the biggest yet most basic problem we have. I also think that main administration sees us [GEP] as somebody they slightly fear? Because they always have to rely on an interpreter… [i]t’s like they go through 영어울렁증 (‘English-nausea’)…
>
> (GCF/language/barrier)

In the second focus group discussion, Hana, Hyuna, and Siri further expressed their frustration with main administrators and their seeming lack of interest in providing translated feedback for NESTs. They believed that improved communication between these two stakeholders was highly unlikely because it was impossible for NESTs to have their opinions

heard by main administrators (Hana); therefore, the TEF was perceived as useless because NESTs did not have access to the meaning of its results or feedback (GCF/problematizing givens/naming issues, GCF/knowledge/lack of dialogue, GCF/power relations/lack of voice). Although Hana, Hyuna, and Siri had either personally translated qualitative TEF comments or had direct access to it, they did not see themselves as having enough authority to provide this to NESTs without express approval from main administration (GCF/power relations/lack of autonomy and agency). In addition to having difficulties with main adminstrators, Hana mentioned how "[i]t's rather pointless for us to give professors translated feedback because we're just assistants… [i]t would be more effective if somebody higher than us did that" (CSCF/sociocultural/social hierarchy, GCF/knowledge/perceived truth). This suggested that even if the quantity and quality of student feedback were sufficient and translated, NESTs would still not consider it useful unless a main administrator with more perceived power and authority were to provide it as this would make the feedback appear more trustworthy.

Neoliberal factors

During their interviews, every GEP administrator stated that student satisfaction was the most important factor to consider with regard to the TEF; therefore, the most useful items were those related to assessing student satisfaction levels (e.g. Items 5, 7, 8, student/teacher interaction items) (NF/consumer values/customer satisfaction). In particular, Jin repeated that Item 5 (sufficient feedback from professor) was the most useful as it reflected overall student satisfaction with the class and teacher performance. The GEP administrators also stated that items identifying student complaints (e.g. Items 3, 4, 5, 7, 8, 10) were important in assessing and improving student satisfaction levels (NF/consumer values/customer satisfaction).

In the second focus group, Hana, Hyuna, and Siri firmly stated that they did not use or pay any attention at all to the TEF and its numerical scores. Their only interest was in collecting negative qualitative feedback comments to a) measure student satisfaction levels ("[w]e only focus on negative qualitative comments because it has a direct connection with student

satisfaction") b) send a summary report to the director of the GEP and c) rank NESTs for teaching awards (NF/consumer values/customer satisfaction, NF/managerialism/high-stakes PI). This was an interesting response because it directly conflicted with how NESTs perceived their TEF scores and how they assumed main administrators used them (i.e. as a high-stakes PI directly affecting their job security).

These three GEP administrators also attributed the TEF's lack of usefulness to main administrators and the MoE. All three agreed that the TEF and its feedback was not helpful for anybody because the form itself was done "just for show" (Hana), again referring to recent political incidents that spurred the MoE to take action and issue changes to Korean universities (CSCF/sociopolitical/political incidents, NF/university autonomy/MoE, NF/managerialism/accountability). Hana added that it seemed like main administrators used the TEF as a written record or paper trail in an attempt show the MoE that the university was following government guidelines. To these participants, the TEF was not useful because it was mostly designed to appease students and higher authorities rather than to improve curriculum or class quality (NF/consumer values/customer satisfaction, NF/university autonomy/MoE).

Table 4.7

GEP administrator perceptions of TEF usefulness

TEF usefulness	Main influencing categories and factors
** TEF not useful because of:*	
Lack of clarity and relevance	GCF/knowledge/lack of guidance
Unhelpful or lack of feedback	GCF/language/barrier, lack of access GCF/knowledge/lack of dialogue GCF/power relations/lack of voice, autonomy, and agency CSCF/sociocultural/social hierarchy
Neoliberal factors	CSCF/sociopolitical/political incidents NF/consumer values/customer satisfaction NF/managerialism/high-stakes PI, accountability NF/university autonomy/MoE

4.3 Q3: Preferred TEF use

4.3.1 Student perceptions

Overall, students responded that the current TEF was not being used in the way they preferred. They identified certain underlying issues with the current TEF and considered a few advantageous ways to use it (GCF/problematizing givens/naming issues, reflection). All participants, with the exception of Yoon who initially thought that the current TEF "was okay," repeatedly mentioned that they would prefer to see the TEF used in a formative way to improve NEST teaching practices and GEP curriculum quality in terms of the following:

- Producing helpful feedback
- Increasing student investment and participation in TEF
- Giving students a voice

However, when asked to elaborate on these ideas and preferences on how to they would like to see stakeholder groups use the TEF, student responses in both the interviews and focus group tended to focus more on fear of retribution regarding their final grades.

Produce helpful feedback

Student interview responses emphasized the need for using the TEF for formative purposes. For instance, Jina expressed her hope that NESTs and administrators would place more focus on the midterm TEF because its results would yield more useful feedback that could be implemented throughout the semester for continuous improvement. In addition, Yena stated her preference of focusing on the TEF qualitative comment items that had the potential to provide detailed and valuable insight behind student responses. She believed that this type of information could help enhance class and curriculum quality.

Students also insisted that the TEF should be used as a safe, anonymous platform to engage in honest communication and provide helpful feedback that would not be used against them (GCF/power relations/fear, anonymity). When discussing the focus group prompt regarding the perceived importance of anonymity, Yena declared that “[t]he starting base score [that students give their professors] is a 4 or 5, and you don’t give any qualitative feedback” due to fear of unfair treatment or punishment. The rest of the participants all agreed with Yena, while they shared their own experiences with the TEF and professors (CSCF/historical/previous experiences). This fear seemed to prevent students from providing honest responses or any at all, much less constructive feedback.

In addition to general critical factors such as fear, context-specific critical factors that emerged from student responses also provided insight on their views. For instance, in the focus group, all students described how issues concerning authority and hierarchy in Korean culture made it impossible for them be frank with any of their professors in terms of giving them feedback. Yoon and Yena rapidly took turns describing how:

> [W]e have to observe authority and status everywhere, such as with our parents, but particularly towards professors, within the field of academics and within our own major departments… [p]rofessors make it so that students must submit to their authority… [a]s a student, you must sing praises of your professors' writing and research. Sing praises! Hallelujah! [T]o do otherwise would lead to getting 'marked or branded' by the professor
> (CSCF/sociocultural/social hierarchy and cultural norms, CSCF/historical/ previous experiences)

This attitude towards Korean professors carried over and was applied to NESTs as well. In the focus group, Jina returned to the notion of fear and observed how it was "easy to get scared in such situations" regarding their teachers. Therefore, by alleviating student fears and making it clear that the TEF was being used to anonymously collect honest, helpful feedback to be used in a practical and formative manner, she and the other participants believed this would lead to more positive outcomes for improving NEST teaching practices and GEP curriculum quality.

Increase student investment and participation in TEF

Another collective response to preferred TEF use included increasing student interest and participation in meaningfully completing the TEF through ample guidance. During her interview, Mia drew upon her experience as former student body president and reflected on the low participation and completion rates of the TEF. She believed that most students tended to "click on the same exact row of numbers, like, 4-4-4, and ignore the qualitative items entirely" just so that they can get access to their final grades earlier than others (CSCF/sociocultural/university system, NF/consumer values/lack of investment). Jina's interview response supported Mia's belief that not many student completed the TEF in a meaningful way. She considered it to be bothersome, not a priority, and that there was no particular reason to do it other than to view final grades quickly (NF/consumer values/lack of

investment). Other participants such as Yena attributed this low participation rate to fear of retribution (GCF/power relations/fear) but also that most students were not aware of the purpose and significance of the TEF (GCF/knowledge/lack of guidance).

During the focus group discussion, all students thought it was important for the TEF to be used as a way to foster interest in class and curriculum improvement. In order to increase student participation rates, Jina, Mia, and Yena asserted that NESTs and administrators should be required to provide sufficient guidance and encouragement to students to help them understand the value of their feedback and how it could affect NESTs and GEP classes (GCF/knowledge/lack of guidance and dialogue). Students also believed that clearly acknowledging that student feedback was being used to improve GEP classes was very important because “[i]f it doesn’t seem like they read our feedback, we don’t want to give any” (Yoon) (NF/consumer values/student investment).

Give students a voice

The participants believed that the TEF should be used as a platform to provide students an opportunity to have a voice in their university learning experience. However, during their interviews and focus group, none of the students showed any hope of freely voicing their opinions or making changes in their major classes (GCF/accepting givens), and this doubt carried over to the way they perceived the TEF in relation to GEP classes and NESTs (CSCF/historical/previous experiences). For example, in their interviews, Mia claimed that in order to use the TEF to its fullest, the items must reflect student wants and needs regarding their learning goals. For this to happen, students must be given the opportunity to provide opinions about this matter. Yena also indicated her preference for expanding and focusing on the qualitative comment items in order for students to communicate their personal thoughts and beliefs about NESTs and GEP classes.

During the focus group, however, students were quick to point out how they have no say or voice (GCF/power relations/lack of voice) and questioned “as students, are we even allowed

to voice [our opinions]... without professors getting offended?" (GCF/power relations/fear, CSCF/sociocultural/social hierarchy). Yena brought up how all Korean students go through compulsory rote learning in high school and university major classes, so they naturally did not have much to say or ask their teachers (CSCF/sociocultural/education system, CSCF/historical/previous experiences). At this, Yoon added that she barely noticed any communication between students and teachers in or out of the classroom. Mia presumed that this was due to the lack of opportunity to do so, while Jina simply said "[t]he answer is already given. There are very few professors who are open to hearing our opinions" (GCF/accepting givens).

Table 4.8

Student perceptions of preferred TEF use

Preferred TEF use	Main influencing categories and factors
Produce helpful feedback	GCF/power relations/fear, anonymity CSCF/sociocultural/social hierarchy, cultural norms CSCF/historical/previous experiences
Increase student investment in TEF	GCF/knowledge/lack of guidance CSCF/historical/previous experiences NF/consumer values/lack of investment
Give students a voice	GCF/power relations/fear, lack of voice CSCF/sociocultural/social hierarchy, education system CSCF/historical/previous experiences

4.3.2 NEST perceptions

During their interviews, NESTs shared their opinions on how the TEF should be used and then further reflected upon these issues in the focus group (GCF/problematizing givens/naming issues, reflection). They strongly expressed that they prefer to use the TEF in a formative manner in the following ways to improve their teaching practices as well as GEP curriculum quality:

- Improve NEST knowledge of the TEF
- Encourage NEST professional development

Improve knowledge

All NESTs indicated in both the interviews and focus group that they lacked even rudimentary knowledge about the TEF much less what other stakeholders' perceptions of it might be. They attributed this to the absence of knowledge influenced by a number of other factors such as language issues and power relations. The lack of access to an English translation of the TEF and guidance from main administrators meant that NESTs were not aware of how they were being evaluated by students, what their TEF scores meant, or what student needs and learning goals might be (GCF/language/lack of access, GCF/knowledge/lack of guidance).

In addition to language issues, all of the participants lamented the difficulty of trying to reach main administrators and the one-sided communication they usually engaged in (GCF/power relations/lack of access, GCF/knowledge/lack of dialogue). For instance, in his interview, John had hoped that main administrators would be more visible and provide an active, facilitative role so that NESTs could be more aligned with the university's goals in relation to their TEF scores (GCF/power relations/access, GCF/knowledge/guidance):

> [I]f I was told to use this [the TEF] as a way to improve my teaching, then I would like to see that is shared with me and my peers… [and] not just having somebody saying that you're weak in this area and push this area

up because I don't think that's how that really works.
(GCF/knowledge/lack of guidance and dialogue)

Kelly, as head teacher, and Mark, as curriculum coordinator, also expressed dismay in their interviews when discussing their failed attempts to approach main administrators to schedule meetings or the lack of dialogue in those meetings (GCF/power relations/lack of access). Unfortunately, both also believed that no changes would be made regarding this situation (GCF/accepting givens). This seeming lack of communication and trust among stakeholder groups was further discussed in the focus group. Kelly claimed that, based on her experience, "negotiations and ideas [with administration]… that doesn't exist right now… it appears they never listen to me" (GCF/knowledge/lack of dialogue), while Mark added that language barriers significantly separated the GEP and NESTs from the rest of the university (GCF/language/barrier).

NESTs also assumed that maintaining high TEF scores was a top priority and that there were no other teacher evaluation criteria (NF/managerialism/high-stakes PI). Therefore, they believed that the TEF should be used as a platform for meaningful dialogue with university administration especially as "admin doesn't know what we're teaching" (Mark) and to enrich their knowledge of each other in relation to the TEF and its desired results. John further added that if the TEF were used formatively as a channel for not only NESTs and administrators to engage in dialogue, but also for administrators and students to learn what each other's goals are, it could prove to be beneficial in improving overall GEP curriculum quality.

Encourage professional development

In both their interviews and focus group, NESTs thought that rather than only focusing on increasing final TEF scores, it was essential to use the TEF in a way that encouraged professional development in terms of teacher autonomy and agency, which could help improve GEP curriculum quality. During their interviews, Mark, John, and Kelly each

mentioned that the TEF should be used in a formative way to help NESTs engage in critical reflection of their own teaching practices in order to make improvements and to make sure that professional standards were being upheld. Mark believed that this was particularly important to do because of the change in autonomy regarding GEP curriculum. That is, he perceived a lack of teaching autonomy that affected the overall quality and relevance of the TEF in relation to GEP NESTs "now that we're all teaching the same standardized thing" (NF/globalization/standardization of education). This feeling caused him to conduct his own teacher evaluations independent of the current TEF so that he could critically reflect on his teaching and make necessary changes as he saw fit (GCF/power relations/agency).

In the focus group, John agreed with and appreciated Mark's efforts when he said "Mark saying that because he gets the high scores doesn't make him a good teacher… that he's not evaluating himself based on the numbers that are there… [it means] he's got his own standards." John also emphasized the importance of using the TEF on a "very human level," referring to the lack of autonomy and agency he felt due to his belief that administrators only saw him as a number (GCF/power relations/lack of autonomy and agency, NF/managerialism/business-oriented). Kelly's observation on how the TEF was mainly used as an indicator of performance and "I don't think we're encouraged to look at it and try and improve" showed how NESTs assumed the form was not used in a way to promote professional development (NF/managerialism/high-stakes PI, GCF/knowledge/lack of guidance).

Concerns related to lack of communication and trust among stakeholder groups were not limited to issues between NESTs and GEP/main administrators. During the focus group discussion, Rob commented on the difficulty in trying to improve one's teaching because NESTs rarely shared information with each other regarding classroom practices (GCF/problematizing givens/naming issues, GCF/knowledge/lack of dialogue). Drawing from her experience as the current head teacher, Kelly opined that this was due to NEST self-

interest and competitiveness stemming from the belief that evaluation scores were directly correlated to job security:

> [Q]uite a lot of people have said no. It's not my job to help other people, I'm here for my job… People are really competitive. The people at the bottom… you don't want to be at the bottom. So people are like, 'Yeah no, I'm not going to share my good idea of what I'm doing in class or how I'm getting high evaluations' because then everyone is getting high evaluations
> (GCF/knowledge/lack of dialogue, NF/managerialism/high-stakes PI)

Rob's heated comments following Kelly's explanation further illustrated how not using the TEF to support NEST professional development could have a negative effect on teacher autonomy, agency, and sense of legitimacy:

> Well, okay, but I'm willing to do anything… Right? Whatever it takes to do it. I don't care about my own pride… administration is going to start getting rid of teachers, and that's why I'm swallowing my pride and doing whatever it is I can to get my evaluations up! I have to increase my evaluation score for admin. How do I do that?
> (NF/managerialism/high-stakes PI, CSCF/sociocultural/education system, GCF/knowledge/lack of guidance)

Rob's outburst also revealed the frustration he felt regarding the perceived importance given to the summative use of the TEF over formative development as well as the lack of guidance from main administrators about this matter.

Table 4.9

NEST perceptions of preferred TEF use

Preferred TEF use	Main influencing categories and factors
Improve knowledge	GCF/knowledge/lack of dialogue and guidance GCF/language/barrier, lack of access GCF/power relations/lack of access NF/managerialism/high stakes PI
Encourage professional development	GCF/knowledge/lack of dialogue and guidance GCF/power relations/lack of autonomy and agency NF/managerialism/business-oriented, high stakes PI NF/globalization/standardization of education

4.3.3 Administrator perceptions

Interview and focus group responses regarding this question were relatively brief compared to others. However, the answers indicated that GEP administrators thought it best to use the TEF as a way to enhance class quality and teaching performance through the following:

- Improve knowledge through better communication and dialogue
- Improve student satisfaction levels

Improve knowledge

In their interviews, Hana, Hyuna, and Siri each asserted that they preferred the TEF be used as a source of communication and meaningful dialogue for both NESTs and main administrators in order to improve overall GEP curriculum and class quality. Hana thought that main administrators should analyze the TEF results themselves to discern the reasons for any low scores that some NESTs might have received and then use that information to

initiate dialogue with them. However, during her interview, Jin shared her opinion that “within our university system, main administration’s attitude is that general elective classes, such as the GEP, are not their responsibility.” This seemed to explain, to a certain degree, the general lack of communication and knowledge NESTs exhibited regarding the TEF and other administrative matters (GCF/knowledge/lack of dialogue, GCF/power relations/lack of access, CSCF/sociocultural/university system).

In addition to developing NEST knowledge through communication and dialogue, some GEP administrators also commented on using the TEF to help students overcome sociocultural difficulties and improve knowledge about their English learning and GEP classes. For example, during their interviews, both Siri and Hyuna observed how students struggle to directly communicate their concerns to NESTs due to language ability or their apprehension of approaching somebody of higher authority (GCF/language/barrier, CSCF/sociocultural/social hierarchy). Because of this, Siri thought the midterm TEF in particular should be used to as an opportunity for students ask questions and provide honest, written feedback to NESTs during the semester. Hyuna’s interview response suggested that students use the final TEF to gain insight on their English learning regarding what they have done and accomplished during the semester. Both Siri and Hyuna believed it necessary that students receive specific and sufficient guidance to help them fully understand the value and impact of the TEF on both learning and teaching practices (GCF/knowledge/guidance). Thus, by using the TEF in a more formative manner, these participants thought that all stakeholder groups could improve their knowledge of each other and thus ultimately improve class and curriculum quality.

During the first focus group discussion, when asked to reflect on the lack of communication between NESTs and main administrators, all GEP administrators agreed that language barriers were mostly to blame because “[l]anguage is the most basic, yet most difficult factor to overcome” (Jin). They added that if there were no language barrier, then better communication would be possible (GCF/language/barrier). However, in the second focus

group, Hana, Hyuna, and Siri expressed a deep lack of trust that communication between NESTs and main administrators could improve regardless of language-related matters. Hana scoffed “[e]ven a simple text can be helpful for NESTs but… [m]ain admin doesn’t give any information.” Hyuna then apologized for not sharing her honest opinion earlier because of Jin’s presence (CSCF/sociocultural/social hierarchy) and pointed out “it’s not possible for NESTs to give their opinions… so we think it’s extremely unlikely that communication with [main] admin can get better.” These comments suggested that while these three GEP administrators could envision using the TEF in a more productive way to improve NEST knowledge, they were highly doubtful this would happen because of main administration’s continued lack of involvement (GCF/accepting givens).

Improve student satisfaction

Although there were mixed responses, overall, GEP administrators focused a great deal more on students and student satisfaction regarding preferred TEF use compared to the other two stakeholder groups. For instance, in her interview, Hana stated that the TEF should be used in a way so that NESTs could adjust their teaching methods and practices to better reflect the students’ wants. She explained that this would raise student satisfaction levels and, in turn, likely please main administration (NF/consumer values/customer satisfaction). Hyuna’s interview response was similar in that she thought the TEF should be used in a contextual way to give precise and helpful feedback to NESTs that would be visibly implemented to satisfy the students.

Out of the four participants, Jin, the current head GEP administrative coordinator, placed the most emphasis on the need to improve student satisfaction levels through the TEF. She repeatedly mentioned in her interview that it should be mainly used to satisfy students’ needs and wants regarding their English education. She also claimed that closely monitoring this through the TEF would help students “learn English in a fun and easy way” and ultimately “have a global effect” when they enter society although no further explanation for this assumption was provided (NF/consumer values/customer satisfaction,

NF/globalization/global influence). Her response seemed to suggest a correlation between student satisfaction levels and their English capabilities.

These opinions were also reflected in the first focus group where Jin continued to focus on measuring and increasing student satisfaction of NESTs and GEP classes through the TEF. Other GEP administrators agreed with Jin regarding the need to reduce the number of student complaints because, as Hana reasoned, “[if] students don’t voice any complaints, then that means they’re satisfied with the class” (NF/consumer values/customer satisfaction). However, in the second focus group, Hana, Hyuna, and Siri offered a different perspective on preferred TEF use in relation to student satisfaction. For instance, Siri said that she wanted the TEF to be used as a way to produce valid student responses that could help NESTs. Although all three GEP administrators admitted that student satisfaction was at the core of their responses, they also said that this belief could not be helped due to their positions. That is, as administrative assistants, they were obliged to focus on improving student satisfaction levels as it was the main part of their responsibilities.

Table 4.10

GEP administrator perceptions of preferred TEF use

Preferred TEF use	Main influencing categories and factors
Improve knowledge	GCF/knowledge/lack of dialogue and guidance GCF/language/barrier GCF/power relations/lack of access CSCF/sociocultural/social hierarchy, age, cultural norms
Improve student satisfaction	NF/consumer values/customer satisfaction NF/globalization/global influence

4.4 Q4: Best NEST evaluation method(s)

4.4.1 Student perceptions

During the interviews and focus group, when asked to consider the best way to evaluate NESTs, students concentrated on modifying the current TEF but did not offer any suggestions on different methods or criteria to evaluate NESTs. Although Yoon initially stated in her interview that it did not require any particular changes, she later agreed with the rest of the students that NESTs should be evaluated with a tailored TEF that considered the following two aspects:

- Relevant, GEP-specific items
- Evaluating teaching practices as ongoing process

GEP-specific TEF

In their interviews, three out of the four student participants (Mia, Yena, Jina) repeatedly stressed the need for a separate TEF for NESTs that included more relevant, GEP-specific items. Mia and Yena in particular stressed the need to add more qualitative items that could elicit detailed responses from students regarding their opinions about the direction in which the class should go. This was due to their belief that there were inherent differences between certain types of classes offered the university. For example, they believed that language-based learning and content-based learning should not be treated equally (i.e. English language classes vs. Korean-taught content classes) because different learning and teaching methods were involved; therefore, it would not be appropriate to use the same TEF items to evaluate NESTs and GEP classes (GCF/problematizing givens/naming issues).

This reasoning also applied to the perceived disparity between major and elective GEP courses (CSCF/sociocultural/university system). With regard to compulsory major classes, Mia mused “[s]hould I say that major classes involve ‘injecting’ knowledge into students? … [but] GEP classes are elective, so there are different expectations in terms of how we learn.” Yena explained in her interview that content-based major classes, much like her high school

curriculum, relied on compulsory rote learning and required little to no interaction with professors (CSCF/historical/previous experiences, GCF/knowledge/lack of dialogue). This was in stark contrast to English learning where students were expected to actively engage in discussions with classmates and NESTs in their GEP classes. Mia and Yena therefore thought there should be a higher ratio of TEF items related to in-class speaking or communication when evaluating NEST teaching.

These opinions regarding context-specific factors were discussed and agreed upon by all participants in the focus group. All of the students, even Yoon, who initially stated in her interview that changes to the TEF were unnecessary, strongly suggested adding GEP-specific items to better evaluate NEST performance because GEP classes were just "too different compared to major [content-based] classes." In addition, although Yoon had originally thought that the current TEF was sufficient for evaluating NESTs, she expressed the need for NEST teaching consistency throughout the semester. This led her to suggest adding TEF items that check to see if all NESTs use the same teaching and student evaluation methods in order to ensure maintain class quality (NF/managerialism/accountability, NF/globalization/standardization).

Ongoing teacher evaluation process

Most student interview responses indicated that NESTs should be evaluated in a way that was much more formative and focused on improving teacher performance and class quality as an ongoing process. They were all opposed to using a final TEF at the end of the semester. For instance, Jina stated:

> I think the midterm and final evaluation should be the opposite. From administration's perspective, it seems like this [midterm evaluation] is too bothersome to do [analyze], so they only focus on the final results. But I think it's important to carefully build things up from the middle rather than the end.
> (NF/managerialism/business-oriented)

This opinion was echoed by the other students in their own interviews. Mia repeatedly insisted that mandatory NEST evaluations were needed throughout the semester rather than just as a single, summative assessment at the end of the course. This was due to the nature of GEP classes (i.e. language learning is a process, and focusing on the process is more important than the results) as opposed to content-based Korean taught classes where students were only required to be tested on their level of knowledge about a particular subject.

The participants also supported this view in the focus group. By shifting the focus from administering the TEF as a singular summative assessment to an ongoing formative process of NEST evaluation, Mia and Yena both predicted that this could lead to benefits such as more communication opportunities (GCF/knowledge/dialogue). They also thought that NESTs would be able to receive better feedback on a more frequent basis that could prove to be useful in improving the quality of current classes rather than just being provided a single set of numbers towards the end of the semester (NF/managerialism/business-oriented) because as Jina observed, “there’s no point in saying anything [through the TEF] after the whole course is over.”

Table 4.11

Student perceptions of best NEST evaluation method(s)

Ideal NEST evaluation	Main influencing categories and factors
GEP-specific TEF	GCF/knowledge/lack of dialogue
	CSCF/sociocultural/university system
	CSCF/historical/previous experiences
Ongoing teaching evaluation	GCF/knowledge/dialogue
(formative use > summative use)	NF/managerialism/business-oriented
	NF/managerialism/accountabilty

4.4.2 NEST perceptions

All NEST participants firmly stated that the use of the TEF as the sole evaluation criterion or performance indicator in a summative manner was inappropriate. Some participants such as John, Mark, and Rob agreed that a certain degree of standardization (i.e. using some general TEF items, following same professional standards) was needed to maintain consistency across departments to make the TEF more credible and to aid with administrative efficiency (NF/managerialism/business-oriented). However, everyone repeatedly expressed that the best and most comprehensive way to evaluate NESTs would be through the following:

- GEP-designed TEF
- Additional assessment criteria

GEP-designed TEF

During their interviews, NEST participants discussed the necessity of being evaluated with a GEP-designed TEF in order to improve teaching performance and class quality. In particular, John, Mark, and Kelly emphasized the need for GEP-specific items with a heavier formative focus and that were directly related to English language teaching. This was due to their belief that EFL classes and major content classes are not taught or learned in the same manner (GCF/problematizing givens/naming issues). For example, Kelly said “I think that Item 6 ‘This class helped to cultivate my knowledge of this class…’ That item is clearly not for us because this is not a language learning question or language learning statement.”

In addition, the participants believed that TEF items should be created by NESTs as this would allow for better quantitative and qualitative items that could elicit specific and relevant feedback to be used in a more formative way. Mark especially was adamant about this during his interview as a GEP-designed TEF would require NESTs and main administrators to engage in meaningful dialogue and for NESTs to have their voices heard:

> [I]f we're going to get any specific, helpful feedback for us, we need to design it because we know what we're teaching. Admin doesn't know what we're teaching. But the people who are actually teaching the courses need to design it; otherwise, it's not going to be useful. It just seems like common sense to me.
> (GCF/power relations/lack of voice, autonomy, and agency)

Creating a TEF tailored for GEP classes and NESTs was not further discussed in the focus group. However, other suggestions from interview responses included adding more 'how' and 'why' questions that would provide more valuable formative feedback for NESTs (John) and items that focused on different aspects of English learning and teaching (Kelly).

Additional assessment criteria

On top of the need for a GEP-designed TEF that could provide more autonomy and agency for NESTs, participants also discussed the need for additional assessment criteria from diverse perspectives that promotes both curriculum quality improvement and professional development. That is, all NESTs expressed desire for a fairer and more balanced formative NEST evaluation that considered all stakeholder perspectives and not just from students. One of the main critical comments repeatedly made by all NESTs was the perceived and sudden summative use of the TEF as a high-stakes PI and sole evaluation criterion for NESTs. During the focus group, Kelly recalled the moment when the TEF was first discussed in a faculty meeting well into her fifth year of employment at the university:

> [T]hat one point when it changed. That one meeting when [the GEP head director] announced that it was being used to determine performance and make decisions… Before that, I had never been told to check my evaluations, and I don't think I ever did before because I didn't know they were there.

(NF/managerialism/high-stakes PI, GCF/knowledge/lack of dialogue and guidance)

Rob also observed, “I look at it [the TEF] as a way that’s supposed to help teachers… as a formative form of feedback… [b]ut at the same time, I think admin just sees it as a way to judge my performance… they look at it as a summative outcome,” while Mark simply stated, “I think most of us think that this is the only way we’re being evaluated. There is no other evaluation” (NF/managerialism/high-stakes PI). Therefore, instead of using a single, generic form, all of the participants believed that an ideal NEST evaluation would require main administrators to review multiple assessment criteria influenced by all stakeholder groups’ perspectives (GCF/power relations/voice) before making a balanced and informed decision regarding curriculum changes, teaching practices, and personnel decisions (NF/managerialism/high-stakes PI). For instance, in the focus group, Rob asked to implement more neutral evaluation instruments (“can we have a performance index that has nothing to do with the students? To kind of balance everything?”).

In addition, NEST participants considered including a wide array of professional development related activities as part of one’s teaching performance to be essential in order to conduct an appropriate and more credible NEST evaluation. During her interview, Kelly believed it was important to recognize one’s contribution efforts to the program when evaluating NESTs such as acknowledging participation in GEP promotion events, department volunteer activities, and engaging in both official and unofficial discussions related to professional development. Taking such action could help NESTs “feel more like a person” rather than a number (John, interview) and help them establish a sense of autonomy and agency regarding their positions and professional development within the university, and perhaps have an overall positive influence on the curriculum as well. Unfortunately, in both the interviews and focus group, NESTs expressed doubt that this could happen because of their belief that main administrators would not be willing to develop a comprehensive NEST evaluation system (GCF/accepting givens).

Table 4.12

NEST perceptions of best NEST evaluation method(s)

Ideal NEST evaluation	Main influencing categories and factors
GEP-designed TEF	GCF/power relations/lack of voice, autonomy, agency
Additional assessment criteria	GCF/knowledge/lack of dialogue and guidance
	GCF/power relations/autonomy, agency
	NF/managerialism/high-stakes PI

4.4.3 Administrator perceptions

While all of the GEP administrators agreed that the current TEF was inadequate to fairly assess NESTs, they did not provide any suggestions related to additional evaluation criteria or other methods to improve this situation. Rather, they focused on the need to consider two points when assessing NEST teaching performance:

- Using GEP-specific TEF
- Considering external factors

GEP-specific TEF

Both interview and focus group responses indicated that all GEP administrators viewed the current TEF as unsuitable for NEST evaluation. This was because of the large number of items that were considered too general in nature or did not seem applicable to GEP classes or NESTs. For instance, in their interviews, Hana and Siri questioned the relevance of Item 6 (cultivate knowledge of the subject) in their interviews because GEP classes were language-based whereas all other types of classes were content-based. They also believed it would be better to remove Item 12 (foreign language %) as it only applied to Korean professors who were required to teach certain major classes in a foreign language (e.g. English or French literature classes). Instead, all of the participants were in favor of tailor the current

TEF with more comprehensive, language-specific items as GEP classes were considered to be quite different from content-based classes (“the current TEF items are too one-dimensional and lack depth”) (Jin, first focus group). They also supported increasing the number of qualitative questions that could elicit more detailed and useful feedback from students and thus help improve NEST teaching practices and class quality. These responses showed that GEP administrators thought it important to contextualize the current TEF in order to better align with GEP class teaching needs.

The participants also supported the idea of GEP-specific TEF item creation between NESTs and main administrators through clear communication and dialogue (GCF/knowledge/dialogue). During their interviews, both Siri and Jin said that, ideally, NESTs should be able to give their opinions about the TEF, have them acknowledged, and then implemented by main administration (GCF/power relations/voice and autonomy). This would help shape the TEF into a more appropriate and helpful source of knowledge for all stakeholder groups. However, while all GEP administrators supported this idea for improved NEST evaluation, most believed it highly unlikely that main administrators would consider changing the current teaching evaluation method (i.e. sole use of standardized TEF). In the second focus group discussion, Hyuna and Siri explained that a standardized TEF was easiest and most convenient for main administration in terms of efficiency, so to do otherwise would be too bothersome for them (NF/managerialism/business-oriented).

Consider external factors

Some GEP administrators also stressed the importance of considering external factors when using the TEF to evaluate NESTs. For them, it was necessary to take into account issues related to class types (e.g. major vs. elective) or language barriers in order to fairly assess NEST teaching performance. This was because they believed the effects of these factors on the quantity and quality of TEF qualitative feedback were too significant to ignore.

For example, Hana was particularly adamant that NESTs should be evaluated with a GEP-

specific TEF that purposely focused on EFL teaching. This was deemed necessary in order to offset student perceptions of class types and power relations that were reflected in the qualitative feedback. Hana explained in her interview:

> [S]tudents think that major classes are most important, then compulsory electives, and finally general electives. Even within compulsory electives, you know that GEP classes are different [because they're English language classes]… so it's extremely important to conduct a specific TEF and separate ourselves from major classes.
> (CSCF/sociocultural/university system, GCF/power relations/fear)

Interview responses also showed that Hana, Hyuna, and Siri perceived a significant difference in feedback content based on class type. As administrative assistants, all three had access to TEF results and feedback from multiple departments. They noticed that Korean major professors received brief yet relevant feedback regarding teaching methods or quality such as "[t]he professor used a lot of Powerpoint presentations in class" or "[t]he professor was good overall." However, in the second focus group, Hyuna and Hana both took turns describing how some NEST qualitative feedback written in Korean were quite different in nature. The content varied in degree depending on how students viewed their relationships with major professors as opposed to NESTs:

> NEST qualitative feedback is sometimes short, sometimes written like an essay and… includes all sorts of childish and nitpicky comments. It's probably because they [GEP classes] aren't major classes, so students only have to see NESTs once and never see them again. When you look at major class feedback comments, they [the students] are more careful with their word choice and don't use any strong language.
> (CSCF/sociocultural/university system, GCF/power relations/fear)

During this discussion, Hana, Hyuna, and Siri also mentioned that, while translating TEF

qualitative feedback, most student responses written in English were unrelated to NEST teaching. Instead, these comments seemed to focus more on the NEST's personality and character. They assumed that this could be partly due to students' low English vocabulary and writing abilities that prevented them from fully expressing their thoughts in detail about the NEST's teaching performance or class content (GCF/language/barrier). They also thought that considering such factors when assessing NESTs was necessary in order to avoid assuming that such qualitative feedback accurately represented student opinions about NESTs and GEP classes.

Table 4.13

GEP administrator perceptions of best NEST evaluation method(s)

Ideal NEST evaluation	Main influencing categories and factors
GEP-specific TEF	GCF/knowledge/lack of dialogue
	GCF/power relations/lack of voice
	NF/managerialism/business-oriented
	GCF/language/barrier
	GCF/power relations/fear
Consider external factors	CSCF/sociocultural/university system

CHAPTER 5: DISCUSSION

I conducted this study on the use of student-completed teacher evaluation forms (TEFs) for native English speaking teachers (NESTs) because of concerns stemming from my own personal and professional experiences. As both a NEST and the head of the Professional Development Committee (PDC) for the General English Program (GEP) at my university, I often encountered difficulty in finding opportunities for NESTs in the GEP to engage in reflexivity on either an individual level or within a group dynamic. I felt that this was mainly due to the overall lack of feedback on our teaching performance or practices, and because the only feedback we had access to was through a student-completed TEF that nobody seemed to fully understand. I also constantly felt a great deal of anxiety in the way university administration would use only a single, final numerical TEF score to assess my teaching performance that determined my yearly employment eligibility.

All of this led to a number of questions and concerns: why are NESTs in my program so unaware of how they are being evaluated even though administrators constantly remind them of the importance of raising their TEF scores? Why are such scores deemed sufficient in determining the effectiveness of NESTs whom students have had for only one semester? Can such student responses be wholly trusted and taken as fact? I therefore felt compelled to examine the similarities and differences in how the three main stakeholders at my university (i.e. students, NESTs, and administrators) view the use of a student-completed TEF to evaluate NESTs.

Through my study, I found that all stakeholder groups exhibit an overall general disapproval of the TEF and its use, which aligns with most existing student-based teacher evaluation literature. This includes the perceived lack of TEF formative use, heavy focus on its summative use, and main emphasis on student satisfaction improvement. For instance, findings show that general critical factor categories such as knowledge influence stakeholder opinions on how TEF data is being used (or not used) for formative purposes. Neoliberal

factor categories such as managerialism affect stakeholder perceptions of the TEF's summative use because of its high-stakes value, while consumer values have an impact on the focus on raising student satisfaction levels. In addition to these general critical and neoliberal factors, however, I also found a number of significant context-specific factors that influence stakeholder opinions regarding the study's primary question.

This chapter reviews the overall similarities and differences among stakeholder perceptions, core categories and factors that influence these perceptions, relationships between the findings and existing literature, and emergent issues and questions.

5.1 Similarities among stakeholder perceptions

Overall, stakeholders appear to share a negative perception regarding the use of the current student-completed TEF to evaluate NESTs. Students, NESTs, and GEP administrators view the TEF as inappropriate and/or insufficient to properly assess NEST teaching practices mainly because of issues regarding the following (Figure 5.1):

- Formative purpose of TEF not met
  - poor use of TEF and its data
  - focus on summative or punishment-like use
  - emphasis on accountability and student satisfaction
- Validity concerns
  - lack of clarity
  - relevance of items and feedback
  - lack of guidance
  - other factors

5.1.1 Formative purpose of TEF not met

Research indicates that teacher evaluations are conducted for two main reasons; namely, formative and summative assessments (Danielson & McGreal, 2000; Penny, 2003; Burden

2010; Spooren et. al, 2013). Formative assessments are generally associated with teacher development and improvement while summative evaluations are used to demonstrate accountability (Danielson & McGreal, 2000; Penny, 2003; Burden, 2008a). The findings in my study are consistent with current literature in this regard. All stakeholders agree that the general purpose of the TEF is two-fold: to aid with teaching performance and curriculum improvement (and/or quality maintenance) and for summative use regarding NESTs.

Stakeholders prefer that the TEF be used in a formative way for curriculum, class quality, and teaching performance improvement purposes that could also lead to better student learning. Interview and focus group responses show that, if utilized properly, TEF qualitative items could be beneficial for all stakeholders involved. Students see them as a way to improve communication and share specific opinions with NESTs, while NESTs think qualitative feedback has the potential to help improve their teaching practices and class quality that could better meet student needs. In addition, GEP administrators see qualitative feedback as a source of useful information for administrative purposes but also as a way for NESTs to reflect on their teaching.

However, overall responses also indicate that stakeholders assume the TEF data are currently not being used in the intended formative manner and is therefore not appropriate for evaluating NESTs. In my view, issues concerning a) the poor use of the TEF and its data b) the focus on TEF summative or “punishment” use (i.e. retribution, reverse evaluation) and c) the emphasis on accountability and student satisfaction indicate a need to more closely examine the rationale and dichotomy between the formative and summative aspects of the student-completed TEF at my university.

Poor use of TEF and its data

Literature has shown a number of issues resulting from the poor use of student-completed teacher evaluation data. Burden (2008a) suggested that even if these evaluations are intended for formative development, teachers do not gain any new knowledge from them.

This could be due to the perceived poor use and/or quality of TEF results and feedback by each stakeholder group. Studies have shown that students lack interest in evaluating their teachers because they are unaware of how the results are used. For instance, half of the student participants in a study conducted by Sojka et al. (2002) believed that professors did nothing with teacher evaluation results. Harvey (2001) mentioned in his report on student feedback in higher education about how students become “disenchanted” because they rarely receive feedback or see any changes made based on the opinions they provide (p. 14).

Similarly, I find that student participants in my study also exhibit a distinct lack of interest in the TEF for these reasons. That is, students do not seem to perceive a sufficient return on the time and effort they invest in providing detailed TEF feedback, nor are they aware of how their feedback is used. GEP administrator responses also align with these student perceptions. They believe that there could be improvement in student satisfaction with NESTs and GEP classes if students are able to see that their feedback has been acknowledged and applied in subsequent classes. GEP administrators also think that showing how student TEF feedback is being used in a more transparent way could lead to more students actively participating and completing the TEF. This, in turn, can provide NESTs with better feedback in both quantity and quality that can aid with formative teaching development. I believe that such student and GEP administrator perceptions indicate an overall need for improved guidance and knowledge regarding the TEF, without which can lead to both poor quality and poor use of its results.

I also find that this lack of knowledge regarding TEF data use affects not only students and GEP administrators but also NEST professional development, as observed by Chisman and Crandall (2007). In their study, they suggested that helping teachers analyze their teaching beliefs and the effects of their classroom practices can encourage them to “take risks” that can contribute to better teaching (p. 99). One way for teachers to actively reflect on their practices is for administrators to provide them with necessary information about the TEF and

its results. However, the majority of NEST responses in my study indicate a clear lack of TEF knowledge due to an absence of such dialogue and guidance. I feel that these general critical factors greatly affect NEST understanding of the TEF and interfere with NESTs attempting to engage in any type of formative development. This can be seen in Kelly's interview response where she explains that the lack of guidance for both students and NESTs can lead to mediocrity in teaching and class quality: "[l]ike, don't try anything new, or the students may not like it, and then you're going to get a lower eval."

Burden (2010) noted that many teachers are unaware of the purpose of the evaluation because it is not explained to them, much like the NESTs in my study. Although there is one bilingual NEST, all of the NEST participants appear to lack even the minimum information necessary to holistically understand the TEF itself, much less use its results and qualitative feedback for formative development. I find this absence of knowledge cause for serious concern because as Penny (2003) pointed out, not only do many student-completed evaluation users lack the appropriate skills and information to sufficiently handle or process these data, some may not even be aware of their own ignorance. Franklin and Theall (1989) also expressed concern for both teachers and administrators in that "administrators who are likely to use ratings for merit review should be well enough informed to avoid using ratings incorrectly and faculty who submit them should know enough to defend their own interests" (p. 18).

Kelly's response to Item 12 (foreign language use %) is an intriguing example that illustrates the issues observed by Franklin and Theall (1989) and Penny (2003): "I never look at the course… administration [questions]. What percentage was taught in a foreign language… well, I never use a foreign language, so there are some here that I never even look at." Her misunderstanding of this item is clarified by GEP administrators who explain that Item 12 is meant to be viewed from a Korean student perspective where English is a foreign language. GEP administrators Hana and Jin further note that this item is actually considered to be very important by main administrators. As demonstrated in Kelly's response, I believe it is

reasonable to assume that there could be a number of other TEF items that NESTs unwittingly misinterpret. Like this, if NESTs do not fully understand the purpose of the TEF or consider certain items not applicable to their own teaching practices, then it is likely that they cannot use the TEF in a formative and constructive manner, nor can they "defend their own interests" when necessary.

Stakeholder perceptions also indicate the need to consider the effect of language-related issues beyond what one would normally expect in research on student-completed TEFs within an EFL context. For example, both NESTs and GEP administrators scrutinize the lack of language access that NESTs experience regarding TEF qualitative feedback; that is, the assistance NESTs require from another party who is capable of providing the necessary translated materials in order to critically reflect on their teaching practices. GEP administrator responses suggest that main administrators are reluctant to provide NESTs access to said translated feedback because of extreme anxiety or 'English-nausea (영어울렁증)' when having to interact with NESTs.

In addition to language issues, I notice that both NESTs and GEP administrators find it challenging to make formative use of the TEF results/feedback because of power relations and managerialism-related factors. For example, three out of four GEP administrators discuss the extreme difficulty in providing translated TEF feedback to NESTs without express approval from main administrators due to their own perceived lack of autonomy and where they are positioned within the business-like hierarchy of the university. This is a concerning matter because without these translations, the majority of NESTs are not able to comprehend or use the TEF and its feedback to reflect on their teaching practices.

Focus on summative or "punishment" purpose

One leading criticism of teacher evaluation is the school administration's sole or heavy reliance on student-completed summative assessment as a means of determining teacher performance and competency rather than focusing on formative measures that can aid with

teacher development and quality control (Pennington, 1991; Wines & Lau, 2006; Bailey, 2009; Isoré, 2009; Troudi, 2009). My findings also support the literature in this matter. Rather than using its results and feedback to improve curriculum quality or teaching practices, stakeholders in my study generally view the use of the TEF in relation to summative purposes and assume that main administrators focus more on final TEF numbers and results rather than analyzing the progression of learning and teaching involved in GEP classes.

Theall and Franklin (2001) said that rather than using the TEF for formative purposes with regard to improving teaching quality and student learning outcomes, it is generally used as an efficient means to quickly assess and rank teaching performance based on a set of numerical scores. This perception of the TEF as being a quick, high-stakes, summative assessment method is shared and discussed by both students and NESTs in my study. For example, students mention in their interviews that the TEF's intended aim is to help improve communication with NESTs and curriculum quality throughout the course. However, Mia and Jina refer to the TEF as actually being a "teacher performance measuring stick" and "a way for superiors to 'peck at' and pressure teachers" with regard to NESTs, while Yoon claims that low TEF scores are cause for contract termination. Students express that the TEF should be conducted throughout the semester in order to better examine the student language learning process and language improvement as well as NEST performance.

NESTs also share these student perceptions. They express that, in absence of any other known evaluation method, the only apparent use of the TEF appears to be summative in nature. I find that their responses are indicative of how the TEF is mostly perceived as a high-stakes PI and administrative index used to assess and rank NESTs with numbers, present best teacher awards of monetary value, and determine employment issues. In my view, both student and NEST responses show that in order to get the best and most helpful results, the TEF should be used in a much more extended and thoroughly formative manner rather than just focusing on momentary, final results as they assume administrators do.

While literature indicates a heavy emphasis on the summative use of student-completed TEFs by administrators, some research studies also perceive a punishment-like aspect to it as well. Svinicki (2001) said students often believe that teachers who receive negative or critical feedback will retaliate or will somehow “come back and haunt them” (p. 18). Likewise, all students participants in my study describe this feeling as “fear of retribution” or “reverse evaluation” and believe that NESTs will retaliate by giving them poor final grades. Based on their interviews and focus group responses, I think this assumption is influenced by a sense of fear that can be explained through differences in perceived power among stakeholder groups. For example, students believe that teachers have more power over them; therefore, in their opinion, it is more than possible for NESTs to abuse that authority to track down students who provided negative written TEF feedback and use it against them. Because of this fear of punishment, I find that students, overall, do not think it is possible to bridge the gap between what they believe is the intended vs. actual purpose of the TEF.

Due to such fear of reverse evaluation and retribution (Svinicki, 2001), the notion of guaranteed anonymous communication and feeling safe when conveying feedback are highly valued by students (Gordon & Stuecher, 1992). Arranging secure platforms for students to voice their opinions and concerns could prove to be helpful in this respect (Barnes & Lock, 2013). I think that providing such ways that take into consideration these factors to assure students of their anonymity when completing the TEF can be a positive step towards improving constructive communication between students and NESTs and moving in the direction of TEF formative use. All stakeholder groups in my study agree with these suggestions. However, stakeholder perceptions also indicate a need to consider additional factors regarding this matter especially in relation to the study’s sociocultural context. For instance, students firmly state that it is nearly impossible to provide candid feedback to NESTs without going against cultural propriety. They feel that it is impolite and disrespectful to directly communicate their opinions to NESTs who are older and in higher positions of authority. I therefore find it necessary to critically examine stakeholder responses through a more cultural lens in order to gain a deeper understanding of their

perceptions.

Like students, NESTs also experience fear and anxiety regarding the TEF and the summative use of its feedback or results. Burden (2010) conducted a student-completed teacher evaluation study where NEST participants expressed a wide range of negative emotions linked to the lack of transparency regarding how TEF results are used and with contract renewal matters. These feelings mirror NEST views in my study. John and Rob both state a sense of fear when it comes to receiving notifications from GEP and/or main administrators about low TEF scores or unfavorable qualitative feedback because they end up focusing on job security-related anxiety instead of reflecting on their teaching practices. John, in particular, perceives such communication or feedback from main administrators about his TEF scores as being more harmful and stressful on both an emotional and professional level. I find this very concerning as research has shown that such summative practices lead to teachers concentrating on increasing evaluation scores rather than improving their performance or tailoring their curriculum to fit the evaluation criteria in an attempt to improve their scores (Simpson & Siguaw, 2000; Wines & Lau, 2006). For these reasons, it is difficult to see that the current student-completed TEF fulfills its formative purpose.

Emphasis on accountability and student satisfaction

Teachers value the formative use of student-completed TEFs for teaching and class quality improvement purposes. However, in an extensive examination of student-completed evaluation literature by Spooren et al. (2013), they observed how TEFs are frequently referred to as “personality contests” and “customer satisfaction” forms (p. 599), or “happy forms” as described by Harvey (2001). These alternative terms used to describe TEFs are indicative of how teachers perceive the usefulness of TEFs. The concern regarding this somewhat cynical view of TEFs was expressed by Barnes and Lock (2013) in their investigation of student perceptions of effective EFL teachers at a Korean university, in which they emphasized that teachers “should not merely pander to students’ opinions and

use this knowledge as the deciding factor in classroom decisions" (p. 19). NESTs in my study also refer to the current TEF as a 'customer satisfaction form' that serves little purpose other than to make students and administrators happy. Similarly, GEP administrators share a narrow view of the TEF and focus on its use to improve student satisfaction levels. I think these perceptions are highly problematic because, as Barnes and Lock (2013) suggested, to assess NESTs based on scores from a TEF that many see as merely a customer satisfaction form will likely compel them to concentrate on accommodating students in order to improve their numbers rather than reflecting on their teaching practices.

In addition to using student-completed TEFs to measure and improve student satisfaction, teachers and administrators also use them as a way to monitor accountability (Burden, 2008a; Danielson & McGreal, 2000). Burden (2008a) suggested that because of the demand for institutional accountability, universities encourage "the system of student-as-consumer evaluating teachers in the belief that popular teachers and courses offer student satisfaction…so that teaching is seen as an unreflective technical process and quality as synonymous with meeting pre-specified standards" (p. 490). NEST responses in my study support this opinion. For example, they discuss how the TEF aims to hold them accountable for maintaining basic teaching or professional standards. However, NESTs also assume they are only expected to make sure that students are sufficiently satisfied with their classes regardless of teaching quality or improvement in student English learning abilities.

GEP administrator responses share similarities with those of NESTs, and they provide specific examples of such. For example, GEP administrators emphasize the importance of Item 3 (syllabus) and Item 11 (make-up classes) because these items make it possible for main administrators to monitor NESTs in terms of whether they are performing their required duties. In addition, three out of four GEP administrators further perceive the TEF as a written method to prove accountability to the Ministry of Education (MoE) that the university is following national educational guidelines. However, I find that neither NEST nor GEP administrator assumptions regarding accountability lead to any positive contribution to the

formative use of TEF data in relation to assessing NEST teaching performance.

5.1.2 Validity concerns

TEF validity is regarded as the correctness of the findings based on how accurately the TEF items reflect what they purport to measure. Penny (2003) said that student ratings of teacher evaluation can be no more valid than the instrument used to collect the information. However, all stakeholder groups in my study are not convinced of the validity of the current TEF and its data because of issues relating to lack of clarity, relevance, guidance, and other factors. Therefore, they do not think the current TEF is a suitable method to evaluate NESTs, nor do they find it helpful for improvement or development purposes. Instead, I find that students, NESTs, and GEP administrators all stress the need for a GEP-specific TEF with a heavy formative focus in order to increase TEF validity and reliability.

Lack of clarity

Wines and Lau (2006) called attention to the "folly of using student evaluations of college teaching for faculty evaluation" because they found the majority of the items on such student-completed TEFs are unclear or ambiguous (pp. 177-178). This is particularly so for items that use rating scales (Block, 1998; Burden, 2008b; Spooren et. al, 2013). In my study, I find that all stakeholders question the value and validity of the current TEF because of a lack of clarity and guidance on the matter. Almost all of the stakeholders raise questions about the vague and confusing wording of TEF items that allow for a high degree of variance in interpretations and possible inaccurate responses, which echoes reactions made by NEST participants in a study conducted by Burden (2008a).

In addition, my findings indicate that three out of four student participants believe quantitative rating scale items are incapable of accurately representing one's complex thoughts when evaluating teacher performance or GEP class quality. NESTs also question the validity of the TEF's quantitative items as they are too general, undefined, and extremely subjective. GEP administrator Hana compares the imprecise and generic TEF items to the Korean

expression “put it on a nose, it becomes a nose ring; put it in your ear, it becomes an earring.” Like this, all stakeholders consider the Likert scale-based quantitative items to be unclear and limited and therefore object to using them to evaluate NESTs. Some might say that it is unreasonable to expect a group of individuals to understand and interpret TEF rating scale items in the same exact manner. However, in my opinion, if stakeholders engage in meaningful dialogue, it would be possible to revise items with clarity so that they align more with stakeholders’ values and thus minimize the level of discrepancy of TEF item interpretation. This could lead to improved TEF validity.

Relevance of items and feedback

For student ratings in TEFs to be valid, they should “reasonably relate to criteria of effective teaching” (Penny, 2003, p. 401). Some researchers question the validity of student-completed TEFs because although their purpose is to assess effective teacher performance and collect feedback efficiently, many of the items do not align with this aim (Block, 1998; Wines & Lau, 2006; Burden, 2008b). NESTs and GEP administrator perceptions in my study support the literature regarding the need for more relevance between TEF items and teaching practices in order to improve validity. For example, findings show that GEP administrators fail to see the connection between NEST/EFL teaching and items linked to content-based classes. In addition, NESTs perceive Item 3 (syllabus) and Item 11 (make-up classes) to be least important because it is unclear as to how they relate to improving teaching practices or class quality. This perplexity that NESTs exhibit regarding Item 3 reflects that of NEST participants in a similar study conducted by Burden (2008b) within a Japanese tertiary level EFL context.

For reasons like these, research suggests a need for more relevant TEF items, as observed in a student-completed evaluation study conducted by Sojka et al. (2002) where both students and teachers expressed their preference for TEF questions specifically relating to the teacher’s area of expertise. This view is also reflected in student, NEST, and GEP administrator responses in my findings. They regard the quantitative TEF items to be too

general and cross-curricular and are not particularly relevant nor useful to NESTs or English language teaching. Student participant Mia's opinion sums up this perception when she compares using the TEF to assess NESTs to "licking the surface of a watermelon" as it only very superficially measures teaching performance but does not get to the heart of the matter.

Overall stakeholder perceptions therefore appear to be in favor of designing a GEP-specific TEF that can improve communication and dialogue among stakeholder groups in order to increase TEF validity. Unfortunately, NESTs and GEP administrators believe it highly unlikely that main administrators would consider improving TEF item relevance as previous attempts to engage in such communication have failed. In my opinion, this resigned attitude is troubling because without the proper motivation, guidance, and meaningful dialogue needed to collaborate and construct more specific and relevant TEF items, it is difficult for any stakeholder to find any value or validity in them.

In addition to validity concerns about TEF items, another issue involves the lack of relevance between TEF data and teaching performance. Block (1998) found that students who assigned teachers the highest rating tended to give top scores and positive comments based on personality characteristics that were "indisputably about the teacher as an individual" and not relevant to teaching performance (p. 408). Some examples included equating the adjective "interesting" as something teachers do that "keep students awake" (p. 415). Findings in my study also reflect similar issues. NESTs and GEP administrators both view TEF qualitative items as largely invalid and unhelpful for formative use because there seems to be little to no correlation between the content in student feedback and NEST teaching performance. Rather, from their perspective, students tend to comment on a NEST's angelic personality or the amount of homework that is given, neither of which reflect classroom teaching practices. In this case, both stakeholders speculate that the low quality and relevance in student responses could be due to language issues where those with low English writing ability are only able to provide feedback in poorly written feedback.

Lack of guidance

Although the validity of student-completed teacher evaluation results relies mostly on students, it also depends on other users of the data such as faculty and administrators (Penny, 2003). A critical lack of knowledge of the TEF and the use of its data by any one of these stakeholders can pose a significant threat to its validity. For instance, not providing students with sufficient guidance on how to complete the TEF or why they should do it could lead to "mindless" evaluation behavior where students respond in a robotic manner (Dunegan & Hrivnak, 2003). Burden (2008a) also observed this worry in his study in which a NEST participant suggested that the absence of awareness in TEF knowledge could lead to students providing "non-committal responses further denuding the value of [TEF] feedback" (p. 484). In my study, this same concern is illustrated in Kelly's comment regarding the need for comprehensive guidance for both students and NESTs so that "students don't just click random numbers, as in somewhere in the middle" as this would have a significant impact on the validity of TEF results. Kelly's opinion is also repeated by GEP administrators Hana, Hyuna, and Siri.

The majority of stakeholder responses indicate the importance of critical factors such as meaningful dialogue and guidance in order to increase everyone's TEF knowledge level. For example, it seems plausible that the failure to inform students of the connection between their TEF feedback and its influence on NEST teaching or GEP class quality can further encourage their "paying scant attention (대충대충/*daechoong-daechoong*)" and "quick, quick, quick" attitude when completing the TEF. As both students and teachers want specific information from the TEF (Sojka et al., 2002; Burden, 2010), I believe that there is a strong need to inform students about the effects of their TEF feedback and provide them with detailed, systematic guidance on how to do evaluations.

Other factors

Research studies on teacher evaluation suggest that factors other than TEF clarity or relevance should also be taken under consideration when examining TEF validity, especially within EFL contexts. For instance, studies have shown that critical factors such as fear and anonymity can influence student feedback accuracy and quality (Gordon & Stuecher, 1992; Svinicki, 2001). Borg (2006) also emphasized the need to acknowledge the distinctive characteristics of language teachers and classes with regard to specific contexts as opposed to a general, all-encompassing milieu. This is because, unlike content/lecture-based courses, the dynamic nature of language teaching involves not only imparting information or facts to students, but also fluid knowledge related to culture and all aspects of life, as well as communication skills. In addition, since teaching practices can vary across different fields of study and disciplines, especially at a university level, and because EFL teachers are dependent on the social and cultural contexts they are situated in, some researchers question the overall validity of cross-curricular student-completed TEFs that are used in all courses (Neumann, 2001; Burden, 2008a; Lee, 2010).

My findings also support existing literature on this particular matter. All stakeholder groups agree that a GEP-specific TEF independent of other classes is needed to improve validity and reliability. In addition, students and GEP administrators stress the need to consider the influence of context-specific factors on student perceptions of the TEF. For instance, both stakeholder groups insist that differences between language vs. content/lecture-based classes as well as major vs. elective classes greatly affect the quality of student responses to the TEF with regard to NESTs and GEP classes. GEP administrators also perceive a noticeable difference in student feedback depending on class type (e.g. major classes vs. elective GEP classes). Hana, Hyuna, and Siri observe that qualitative feedback for major classes consist of short, polite comments about teaching methods or content, unlike feedback in GEP classes. They attribute this difference to the students' fear of offending their major professors whom they are required to see multiple times throughout their university career, and I find that student responses support this assumption.

In contrast to their major classes, students are only required to take two elective GEP classes and thus interact with NESTs just twice prior to graduation. Although there appears to be less fear involved compared to taking major classes, which allows some students to occasionally provide more candid feedback, GEP administrators state that the majority of such detailed feedback is unrelated to NEST classroom teaching performance. Rather, students tend to comment on the amount of homework or lack of flexibility in homework deadlines. I believe that this lack of relevance and quality in TEF feedback indicates a need for both GEP and main administrators to take into account a wide range of factors when evaluating NEST performance in order to avoid using a sole, flawed criterion to assess such a complex matter. I also think this further supports the need to closely examine context-specific critical factors when conducting similar research within EFL contexts.

5.1.3 Correlation of core and main categories regarding stakeholder perception similarities

The core categories (i.e. general critical, context-specific critical, neoliberal) and each of their main categories all interact with each other to affect stakeholder perceptions regarding the use of the current student-completed TEF to evaluate NESTs. My findings show that the general critical core category has the greatest overall influence on stakeholder perceptions regarding issues with the TEF's formative purpose and validity, and it also shares the most connections with the other two core categories.

General critical and neoliberal core categories

General critical and neoliberal core categories both intersect and interfere with the formative aspect of TEF use. In my view, there seems to be a strong correlation between power relations and managerialism. For example, GEP administrators discuss how their lack of autonomy (due to their perceived lower-ranking positions within the university's business-like hierarchy) prevent them from freely providing NESTs with translations of TEF feedback. I also find that NEST responses show a deep connection between the sense of fear they experience regarding the TEF and how they perceive its high-stakes, summative use by main administrators. The fear NESTs mention relates to receiving low TEF scores and the

loss of employment or job security, as seen in responses from John and Rob in terms of the anxiety they experience when receiving notifications from main administration about their TEF scores. I strongly feel that these examples show the significant influence of fear on the lack of TEF formative use, which leads to NESTs focusing on improving their final, end-of-semester scores instead of engaging in reflective teaching practices.

Findings in my study also show how general critical factors such as guidance and voice can have an effect on neoliberal TEF investment and student satisfaction. Students and GEP administrators suggest the need for providing better guidance to students and acknowledging their voice and opinions regarding the TEF. By doing so, students hope that their peers will recognize the value of actively engaging in and completing the TEF, which could lead to more helpful quantitative and qualitative feedback for NESTs. GEP administrators also suggest that acknowledging student voice and opinions can lead to increased participation in TEF completion and improved student satisfaction levels. In this regard, I speculate that issues with the lack of formative TEF use stemming from the correlation between lack of guidance and student investment can be resolved through improved communication and meaningful dialogue among stakeholders.

General critical and context-specific critical categories

I also perceive a correlation between general critical and context-specific critical core categories affecting TEF formative use. For instance, student responses reflect a serious concern regarding the fear of retribution, but I find there are sociocultural factors shaping this fear that further affects the way they complete the TEF. For instance, students are afraid that providing negative feedback could lead to NESTs punishing them by assigning poor final grades. However, they also feel that giving NESTs honest and critical feedback is impolite and disrespectful because NESTs are older, and thus automatically hold higher positions of authority. Here, I notice a sense of fear coming from both power relations as well as sociocultural factors, such as age and social hierarchy. As a result, TEF feedback and data provide little to no aid to NESTs in terms of formative measures that be used to improve

teaching practices and class quality.

I also find that the connection between these two core categories affects stakeholder views on TEF validity concerns. For example, both students and GEP administrators discuss how the differences between major vs. elective GEP classes, as well as content vs. language classes, can have an effect on the way students complete the TEF. In my view, these responses indicate that the level or type of fear students feel towards their major professors, as opposed to NESTs, influences the quantity and quality of the TEF results. Both students and GEP administrators therefore find it difficult to determine whether the TEF numerical scores or qualitative feedback indeed accurately represent students' honest opinions about NEST performance. These doubts then bring into question the overall validity of TEF results. For these reasons, I think a feasible way to approach the issue of validity improvement is to acknowledge and examine the wide range of factors influencing TEF data. This is particularly necessary for student-completed TEFs conducted within EFL contexts such as mine.

Within the general critical core category

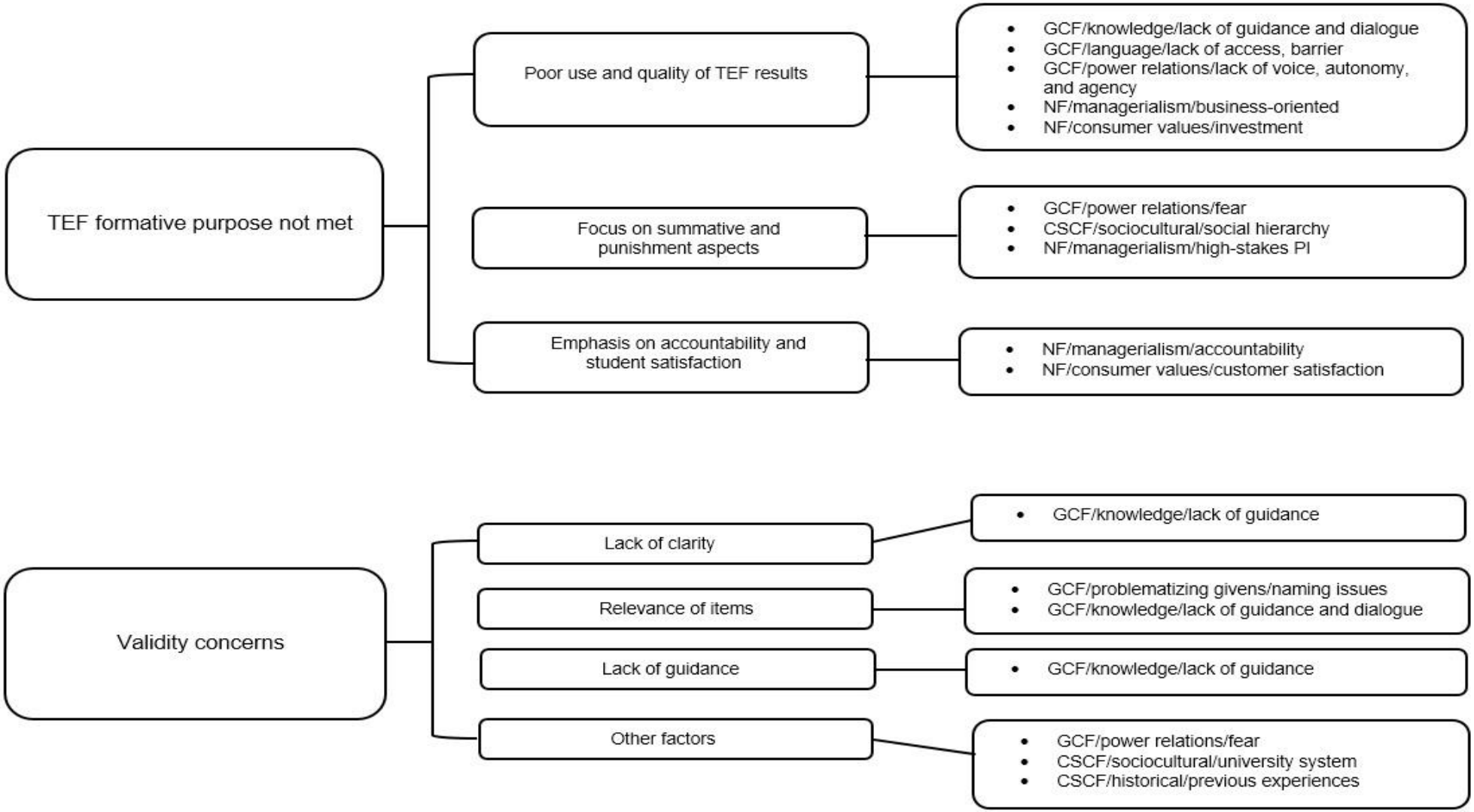
I find that main categories within the general critical core category also intersect and influence each other with regard to stakeholder perceptions of TEF use. In particular, the connection between knowledge and language appears to affect all stakeholders. For example, the way students are seemingly not provided with sufficient guidance on why or how to complete the TEF in a satisfactory manner is also sometimes further hampered by their low English abilities, which could lead to TEF feedback that is both low in quality and quantity. This not only harms TEF validity because it is difficult to discern whether student responses accurately represent what they wish to express, but it also prevents NESTs from reflecting on their teaching practices.

NESTs experience similar difficulties with the TEF and lack of guidance as they do not have access to an English-translated TEF and qualitative feedback. Based on GEP administrator

responses, it seems that NESTs are not able to receive this information or any other kind of assistance because of language barriers and anxiety that main administrators themselves experience (e.g. ‘English-nausea’). This combination of language issues and lack of guidance leads to a critical absence of TEF knowledge that again hinders NESTs from using its data in a formative manner.

I find the similarities across stakeholder group perceptions to be particularly significant for a number of reasons. In my opinion, these similarities indicate that problems stemming from the use of student-completed TEFs are not merely limited to NESTs who feel unjustly evaluated or inadequately equipped to understand and/or use the TEF. For instance, while it is not unexpected for NESTs to show concern about the lack of TEF formative use as it directly relates to their teaching practices, it is surprising to see how both students and GEP administrators also fail to see how the TEF is being used in its intended, formative manner. In addition, all stakeholder groups specifically lament the lack of clarity and relevance of the TEF items, as well as a conspicuous absence of guidance regarding its content and completion, which leads to further TEF validity concerns. Based on these findings, I believe it is important for researchers and practitioners within EFL contexts to examine all possible influencing factors and correlations among these factors in order to better understand stakeholder assumptions of student-completed TEFs particularly with regard to NESTs.

Figure 5.1 Similarities among stakeholder perceptions (main influencing categories and factors)



5.2. Differences among stakeholder perceptions

Overall, students, NESTs, and GEP administrators view the use of the current student-completed TEF to evaluate NESTs as inappropriate and insufficient. This is mainly because the TEF does not meet its intended formative purpose and because of concerns regarding the validity of TEF results. However, while all stakeholders arrive to this same conclusion, each group's perception is also marked by distinct, underlying factors (Figure 5.2):

- Student perceptions: fear and context-specific critical factors
- NEST perceptions: need for additional evaluation criteria and professional development
- GEP administrator perceptions: neoliberal and context-specific factors

5.2.1 Student perceptions affected by fear and context-specific critical factors

Studies have shown that students fear that teachers might punish them by giving poor final grades in retaliation for receiving low TEF scores or negative feedback (Simpson & Siguaw, 2000; Svinicki, 2001). Stakeholder perceptions from my study also support these views. For example, students indicate that fear of retribution or reverse evaluation by NESTs, which are based on previous experiences from both high school and university, negatively influence the way they complete the TEF. Because of this, students do not think the TEF is a good evaluation method for NESTs due to the resulting lack of validity in TEF data. Furthermore, Gordon and Stuecher (1992) suggested that anonymity has an important role in student-completed TEFs. In my study, I also find that a perceived lack of guaranteed anonymity, along with fear, negatively affects the quality and accuracy of responses on the TEF, thus rendering the TEF inadequate, ineffective, and invalid. Students indicate that it is urgent to assure anonymity, relieve fears, and build trust with NESTs in order to make effective use of the TEF, increase the validity of its results, and improve student learning.

Interview and focus group responses further show how fear and previous experiences in the

classroom can influence student attitudes towards the TEF. For instance, my findings suggest that students are not accustomed to candidly providing their opinions about their teachers or classes because they were never given the opportunity or encouragement to do so before. In the focus group, students shared anecdotes of those who had ventured to voice their thoughts about their classes only to be met with harsh words from their teachers. Student participants assume that this fear influences the way the majority of students assign numerical scores to quantitative items and how this leads to the tendency to provide short, non-critical qualitative feedback in order to avoid offending NESTs.

These assumptions support a study conducted by Svinicki (2001), who suggested that student beliefs about giving open-ended comments and their general lack of understanding and practice in giving it are two reasons why TEF feedback is low in both quantity and quality. This lack of understanding and practice can stem from sociocultural factors. For example, as observed by Svinicki (2001), Han (2005), and Burden (2010), for students who come from cultures with education systems where accuracy is valued in language learning (such as Korea), it could be difficult for them to provide appropriate written feedback. This is especially so if they have had little opportunity to practice using the target language or if they have not been given feedback on their evaluation feedback, such as the case with students in my study.

5.2.2 NEST perceptions affected by need for additional evaluation criteria and professional development

Like students and GEP administrators, NESTs indicate that the current student-completed TEF is not a suitable way to evaluate their classroom teaching performance. However, their main concerns differ from other stakeholder groups as they center on a) the need for additional evaluation criteria to improve validity and b) the importance of focusing on the TEF's formative purpose for professional development reasons. These concerns appear to be largely influenced by general critical factors such as lack of knowledge and power relations.

Need for additional evaluation criteria

Good teaching evaluation systems need to be linked to a set of clear standards and fair, reliable criteria (Isoré, 2009; Looney, 2011). Burden (2010) suggested that NESTs should be evaluated on the multidimensional aspects of EFL teaching to counter problems with content and validity that is frequently seen in student-completed evaluation questionnaires. This is to ensure that the TEF is not used as the sole summative evaluation criterion that might be based on a student's single affective reaction. My findings also support the literature in that, unlike students and GEP administrators, who do not mention any other teacher evaluation measures aside from a GEP-specific TEF, all NESTs repeatedly emphasize the necessity of additional assessment criteria especially with regard to main administrators and personnel decision-making.

In their extensive review of student-completed evaluation literature, Spooren et al. (2013) stated the importance of properly evaluating teacher performance and reducing the risk of high discrepancy among perceptions of what constitutes effective teaching. In order for this to occur, I believe stakeholders need to be well-informed about the TEF and engage in consistent and meaningful dialogue with the purpose of defining these constructs and designing multiple teacher evaluation instruments. NEST interview and focus group responses also reflect assumptions that improved access, dialogue, and guidance from both GEP administrators and main administrators are essential in creating a fair and comprehensive assessment system and making better use of the current TEF.

Focus on professional development

Alamoudi and Troudi (2017) stated that formative teacher evaluation can be used to make decisions regarding professional development, and Johnson (2000) described the need for teachers to be in a legitimate position to reflect on their own practices and improve on them. However, Burden (2010) observed how NEST participants in his study dismissed the TEF because they failed to see the formative value in it and its data regarding their teaching. Studies have also shown that NEST interests, needs, or input are not reflected in the use of

student-completed TEFs (Penny, 2003; Burden, 2010). All of these views are reflected in the NEST responses in my study. Compared to students and GEP administrators, NESTs emphasize the importance of not only including GEP-specific items to improve validity, but for those items to be designed by NESTs themselves in order to be of more practical and formative use.

For example, during the focus group discussion, Mark, who is the current curriculum coordinator and recipient of multiple teaching awards, surprised everyone by stating that he does not look at or use the current TEF at all because it does not help with improving teaching practices. He instead conducts a more specific and tailored TEF that helps him to better identify areas of improvement from his students' perspective. I find this demonstrates the importance Mark feels about gaining back a sense of autonomy and agency within his classroom. Failing to do so is problematic because, as Johnson (2000) suggested, teachers can experience feelings of confusion and diminished teaching abilities if evaluations results are not compatible with their own perceptions of good teaching. This, in turn, can cause anxiety and stress that is further intensified with the summative use of the TEF.

High-stakes TEF use can lead to a "performativity culture" where "professional development is more likely to become externally-driven" rather than self-directed (Tang & Choi, 2009, p. 15). This can be seen in comments made by Rob, who voices frustration in his inability to find any discernable patterns of poor teaching practices in the TEF's results and qualitative feedback. He also emphasizes that he would swallow his pride and do whatever it is he can to increase his TEF scores for main administrators, while John expresses his desire to be seen on a "human level" when it comes to teaching rather than just a number. I believe these examples show a need for both NESTs and administrators to take a more constructive and reflexive approach to assumptions about teacher evaluations.

Like this, NESTs appear to view the TEF more as a hindrance to their professional development due to a perceived lack of voice, autonomy, and agency. Kelly and Mark in

particular lament the feeling of not having a voice and not being part of the university because of the absence of communication and language problems through their experiences (or lack thereof) with the TEF. This feeling of isolation is echoed by NEST participants in a study conducted by Howard (2019) where they frequently struggled with language barriers and were constantly reminded of their outsider status as foreigners in Korea. I therefore think there is a need for more research and analysis examining NEST perspectives on teacher evaluation and professional development, especially within a Korean EFL context such as mine. This can perhaps lead to more awareness and administrative support, which can allay the damaging effects and difficulties NESTs experience when they endeavor to enhance themselves and their practice. Doing this could provide a more valid and reliable reason to use the TEF as a NEST evaluation method.

5.2.3 GEP administrator perceptions affected by neoliberal and context-specific critical factors

Teacher evaluation research has shown that university administrators tend to focus on neoliberal managerialism and consumer values, such as efficiently assessing teaching performance, increasing student satisfaction, and proving institutional accountability (Theall & Franklin, 2001; Chae & Hong, 2009). Similarly, in my study, GEP administrators are highly influenced by such neoliberal categories and factors. This is in contrast to student and NEST stakeholders who are more affected by general critical factors. However, my findings also indicate that university autonomy and sociopolitical factors (e.g. political incidents) also play a significant role in the way GEP administrators perceive the pragmatic value of the TEF.

Focus on student satisfaction

Although Spooren et al. (2013) conducted an extensive review of teacher evaluation literature, they were not aware of any recent studies that include administrators' attitudes towards student-completed TEFs. However, they did find that administrators preferred using TEFs to measure overall student satisfaction, much like the GEP administrators in my study.

My findings show that, unlike the other two stakeholder groups that discuss using the TEF in more effective ways to improve student learning or NEST teaching, GEP administrators appear to be primarily concerned with using the TEF to increase student satisfaction levels in GEP classes and with NESTs. Their responses are greatly affected by neoliberal factors, such as customer satisfaction, in that satisfying students' needs and reducing student complaints is seen as the equivalent of improving student satisfaction. I find that the issue here is the disparity between the deep concern that students and NESTs share regarding TEF validity and the way GEP and main administrators seem to mainly use TEF results to measure and improve student satisfaction levels without any consideration for validity. I believe it is necessary for meaningful dialogue to occur among stakeholders in order to bridge the gap in perceptions and together establish TEF knowledge constructs that could aid in valid TEF use.

Effects of sociopolitical factors on accountability and university autonomy

Student-completed teacher evaluation literature has revealed some common factors among stakeholder perceptions or attitudes, such as the emphasis placed on its summative value and student satisfaction. However, findings in my study indicate that context-specific critical factors should also be taken into consideration when examining TEFs and NEST evaluation, especially regarding GEP administrator perceptions. While the accountability aspect of TEFs has been referred to as "demonstrating the presence of adequate procedures for ensuing teaching quality" (Kember et al., 2002), GEP administrators provide more critical insight on their perceptions of TEF use. For example, with the exception of Jin, all other GEP administrators agree that TEF data is not helpful or well used by NESTs or main administrators because it is "just for show" and something that "the government is telling us to do." This perception seems to have been influenced by recent political incidents causing the MoE to take action and make policy changes in an effort to encourage educational institutions to hold themselves more accountable for the management of their students and teachers.

In order to better explain the reasoning behind this view, GEP administrators Hana, Hyuna, and Siri describe a political scandal involving the 2017 impeachment of former Korean president Park Geun Hye. Park's aide and confidante at the time, Choi Soon Sil, was also accused and convicted of crimes that involved not only matters of national interest but also on a personal level. Choi had solicited numerous academic favors for her daughter, Jung Yura, both during Jung's time in high school and at university. This included colluding with teachers and professors to help Jung illegally gain admission to a prestigious Korean university and to fabricate her grades as well as attendance record (Chung, 2016, 2017; Kim, 2016; "Choi Soon-Sil Jailed," 2017; Yang, 2017; Kim, 2018).

In addition to the already-growing neoliberal interest in controlling teachers' actions (Thomas & Yang, 2013; Raaper, 2017), these sociopolitical events seem to have influenced top-down reform that has been observed in other similar tertiary-level, Asian EFL contexts (Burden, 2008a, 2008b, 2010; Thomas & Yang, 2013). At the time of the scandal, the deputy prime minister of the Korean MoE stated "[w]e definitely feel responsible for [failing to properly] monitor the university" (Chung, 2016), while the Seoul Metropolitan Office of Education (SMOE) said it would oblige schools to review student attendance/absences through special grading committees (Kim, 2016). I think this shows why GEP administrators at my university seem even more focused on overall student satisfaction as well as attendance-related TEF items. I also speculate that such issues affecting university autonomy have likely caused most of the GEP administrators to see the TEF's general purpose as a mere formality given as an order from the MoE rather than an effective student-completed teacher evaluation method that could benefit all stakeholders involved.

5.2.4 Correlation of core categories regarding stakeholder perception differences

Findings show connections among the three core categories regarding the differences among stakeholder perceptions of the TEF. However, unlike the similarities in stakeholder perceptions where general critical factors can be seen as an overarching core category, I

find that each stakeholder group's difference in perception is marked by a distinctive core category correlation.

Student perceptions: general critical and context-specific critical categories

General critical and context-specific critical core categories greatly affect student assumptions of the TEF and its use to evaluate NESTs. In particular, fear and anonymity are the main underlying factors that can be seen in most student responses. In my view, this sense of fear and need for guaranteed anonymity is strongly correlated to their previous negative experiences with the TEF and teachers in both high school and at university. Because of this, students believe that the TEF is not truly anonymous and that NESTs have the ability to identify those who provide negative or critical feedback. As a result, students perceive TEF responses to be lacking in quality and accuracy, all of which have a negative effect on TEF validity. I believe that it is, therefore, important to both examine and take into account the effects of such context-specific critical factors and accordingly adjust the TEF as well as the way it is administered.

NEST perceptions: general critical and neoliberal categories

My findings suggest that general critical and neoliberal core categories together have an influence on how NESTs perceive the TEF and its use. NEST responses include the need for additional evaluation criteria in order to be fairly assessed and to offset problems with validity that is often seen in the sole use of generic, cross-curricular student-completed TEFs. This necessity is expressed in connection with concern regarding contract renewal and personnel decision-making. In my opinion, this shows how the high-stakes aspect of the TEF requires all stakeholders to engage in meaningful dialogue in order to find supplementary assessment methods that can be incorporated within a fair, comprehensive teacher evaluation system.

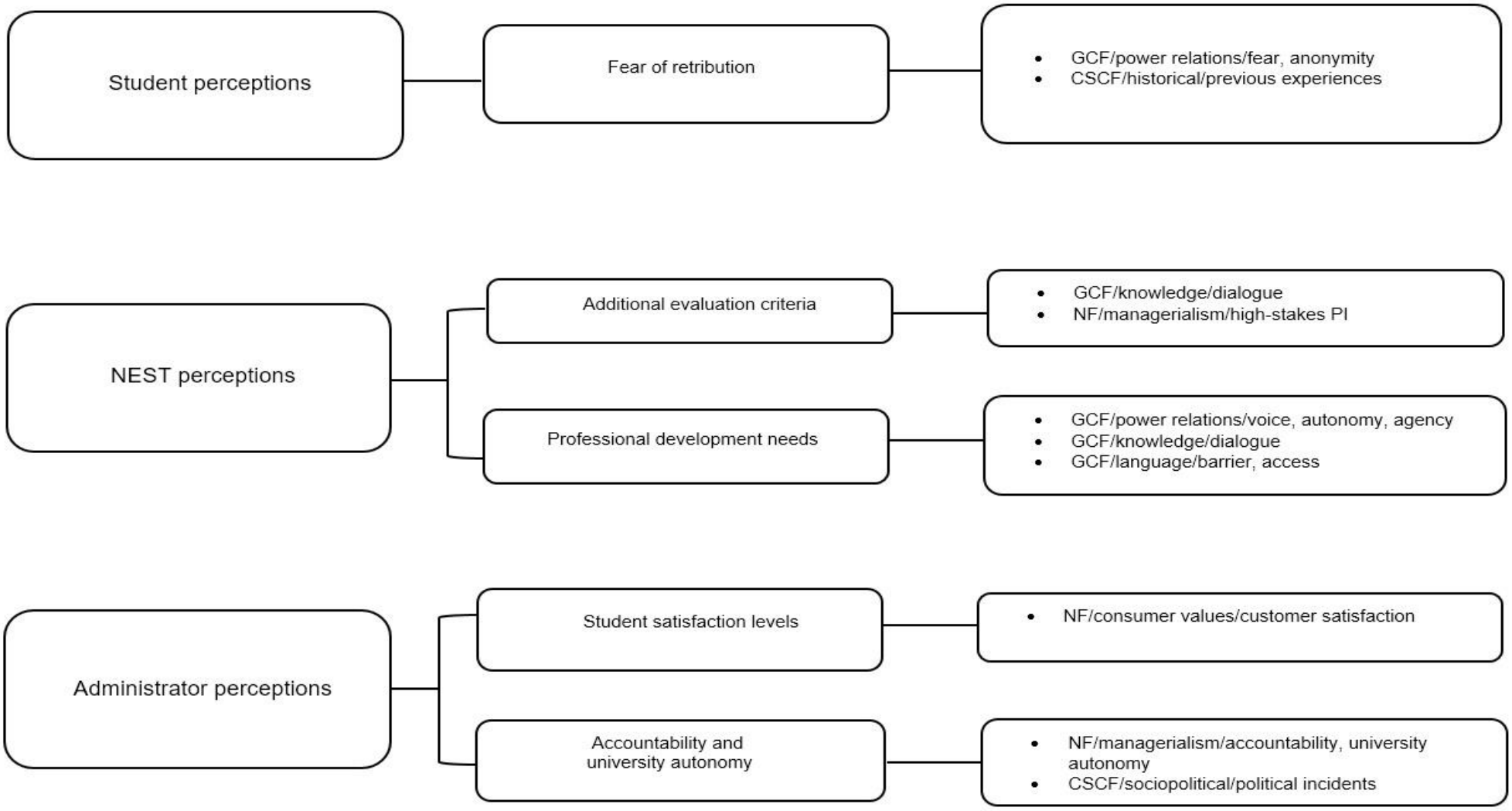
In addition to the correlation between these two core categories, the main categories within general critical factors also influence each other. For example, NESTs focus heavily on the

negative effects of the TEF on their professional development. They discuss concerns that stem from the lack of communication and language issues, which then lead to questions regarding their perceived weak positions within their professional context. This suggests an urgent need for NESTs and administrators to address their critical communication problems and actively negotiate ways to construct mutual knowledge of the TEF and its data use.

GEP administrator perceptions: neoliberal and context-specific critical categories

In general, I find that GEP administrator views of the TEF are mostly influenced by the neoliberal core category. Using the TEF to measure and improve student satisfaction levels is frequently mentioned in both interview and focus group responses for both NESTs and GEP administrators. However, some GEP administrator perceptions also reflect a particularly distinct correlation of core categories compared to those observed in the other two stakeholder groups: neoliberal and context-specific critical factors. For instance, Hana, Hyuna, and Siri discuss how the current TEF seems to fall short of being helpful to NESTs in a formative sense. Rather, they describe it as something that is done as a formality in order to comply with government guidelines that require universities to hold themselves more responsible for the management of their students and teachers. Findings in my study show that these GEP administrators connect this notion of accountability and diminished university autonomy to a political scandal that took place in 2017 which, they believe, led the MoE to implement more top-down education reform policies. In my view, their suggestion that this resulted in the university's increased focus on attendance-related matters and student satisfaction levels further demonstrates the importance of examining context-specific influences that researchers and practitioners might fail to notice at first.

Figure 5.2 Differences among stakeholder perceptions (main influencing categories and factors)



Overall, findings from my study suggest that stakeholders consider the current student-completed TEF to be generally inadequate and that there is a need to reassess the way in which NESTS are evaluated. A key factor influencing this opinion consists of stakeholder issues with language and fear that culminate in a lack of critical knowledge of the TEF, which then leads to concerns regarding TEF data validity. Another important factor involves neoliberal market forces that distort and impede the effective use of the TEF. This includes the high-stakes nature of the TEF and its summative use, as well as the concentrated effort on improving student satisfaction levels (i.e. emphasis on consumer values) rather than student learning or teaching practices.

This chapter evaluated the findings of my study in relation to existing and relevant literature. In the following final chapter, I will present the main implications of the findings, its contribution to knowledge within student-based teacher evaluation, the limitations of the study, and suggestions for future research and practice.

CHAPTER 6: CONCLUSION

The aim of this study was to examine the similarities and differences in how students, NESTs, and administrators view the use of a student-completed TEF to evaluate NEST performance at a Korean university. I sought to identify specific factors that influenced these perspectives in order to improve my understanding of how each stakeholder group perceives the TEF within their own situations. Using a grounded approach to analyze the qualitative data collected, findings show that all stakeholder groups similarly perceive a) a gap between the TEF's intended and actual purpose b) a lack of TEF usefulness and validity c) a strong preference for TEF formative use and d) a need for a more tailored TEF to evaluate NEST performance.

In addition, while all stakeholder groups arrive at the same conclusion that the current student-completed TEF used to evaluate NEST performance is insufficient and inadequate, each group's perception is also marked by distinct, underlying critical, context-specific factors. For example, student perceptions of TEF use on NESTs are significantly characterized by fear and context-specific critical factors, such as past historical experiences, whereas NESTs are primarily affected by their lack of TEF knowledge, the need for additional evaluation criteria, and lack of opportunity to professionally develop themselves because of a perceived lack of voice, autonomy, and agency. On the other hand, GEP administrator perceptions are profoundly influenced by neoliberal factors such as managerialism and consumer values (i.e. measuring and improving student satisfaction levels), as well as context-specific factors such as sociopolitical incidents.

Overall, study findings suggest that the singular, prevalent use of student-completed TEFs to assess NESTs in this context warrants a more thorough examination by all stakeholders involved. They also indicate the need to improve knowledge, practice, and student learning through the creation and implementation of a comprehensive NEST evaluation system rooted in the needs of both students and NESTs. This chapter presents the main

implications of the findings, its contribution to knowledge, limitations of the study, and suggestions for further research.

6.1 Implications of findings

The first implication of this study is that the current, student-completed TEF and its data lack validity; therefore, it is considered to be an insufficient and inadequate way to assess NESTs. In addition, students, NESTs, and even GEP administrators do not find the use of the TEF or its data to be wholly trustworthy. This is an important point to consider because if the data cannot be trusted, it cannot be considered credible either (Burden, 2008b). One way of improving the validity of the TEF could be to modify the items to make it more GEP or NEST-specific in order to yield more relevant and accurate data.

The second implication is that there is a serious need for a comprehensive NEST evaluation system rooted in the needs of both students and NESTs. Study findings and existing literature show that the singular and summative use of a generic, cross-curricular, student-completed TEF can have detrimental effects on all stakeholder groups. Therefore, I strongly feel that my university should implement an evaluation system consisting of multiple, contextual evaluation criteria that can examine NEST performance in a fair and balanced way. The system should reflect the complexity of EFL teaching and provide relevant, valid, and helpful data that can be used for formative purposes to bolster NEST teaching practices and professional development. This is because research has shown a positive causal connection between teacher knowledge construction within professional development and school development (Yoon et al., 2008; Tang & Choi, 2009). Furthermore, investing in continued professional development enables the growth of teacher knowledge and expertise that can lead to improved student learning and performance (Yoon et al., 2008; Murray, 2010; King, 2014).

The third implication of this study is that critical and context-specific factors can augment the

negative effects of student-completed TEFs on NESTs in a higher education EFL setting. Factors such as language, fear, and hierarchy sometimes intersect and lead to a critical lack of knowledge. For example, problems concerning NESTs being unable to read the Korean-written TEF items and qualitative student feedback are not surprising. However, language issues also appear to affect main administrators due to fear and anxiety they experience when required to communicate with NESTs (i.e. feeling 'English-nausea' and always needing an interpreter). Such overlooked, context-specific factors, in conjunction with perceived hierarchal differences, could have led to the lack of meaningful dialogue between NESTs and administrators. This, in turn, could have contributed to the NESTs' absence of knowledge regarding the TEF.

Lastly, stakeholder perceptions show that underlying neoliberal influences in EFL education have damaging effects on the way NESTs are evaluated. For instance, NEST, GEP administrator, and even some student responses indicate that business-oriented standards and practices have led university administration to use a single, cross-curricular student-completed TEF to assess NESTs for efficiency and accountability purposes. This leads to a significant lack of relevant, constructive feedback that NESTs need to develop themselves and their teaching practices. Using student-completed TEFs in such a summative manner and holding NESTs accountable for teaching practices that do not seem to apply to them can also diminish their sense of autonomy and agency.

6.2 Contribution to knowledge

Considering the extreme importance placed on English education in Korea, and the large number of NESTs teaching EFL in this country, there is surprisingly little research conducted on NEST teaching performance evaluation. Some studies examine student or teacher perceptions of what makes an effective EFL teacher in Korea; however, there seems to be a conspicuous absence of research that specifically focuses on NEST evaluation criteria or standards. This study can help fill the gap on student-completed TEFs and NEST

assessment in Korea from the perspectives of all stakeholders involved by providing more practical insight on how each stakeholder group views TEFs and what context-specific factors influence these perceptions.

For example, it is not difficult to anticipate language barrier related problems among stakeholder groups in studies within an EFL context. However, by delving deeper into this matter, my findings show that these language issues do not merely stop at NESTs being unable to communicate with students or administrators. Rather, for NESTs, they lead to a lack of guidance, and thus, the absence of critical knowledge of the TEF and failure to use TEF feedback in a formative manner. For some administrators, the thought of having to converse in English with NESTs about the TEF causes anxiety to the point where the term 'English nausea' is used to describe their feelings. I believe that studies such as mine could help identify unexpected, underlying issues that can help stakeholder groups understand each other better and encourage them to work together towards a common goal of improving teaching and learning.

In addition, although this study was conducted in Korea, there are implications in the findings that could inform wider international HE EFL contexts that are also under the influence of neoliberalism. This could be of particular use for NESTs within these contexts due to the paucity in this type of research. Studies and reviews of literature indicate that many perceive the use of student-completed TEFs as a main and valid way to establish reliable evaluation standards and policies with the purpose of improving teaching quality and performance. However, my study shows that the opposite of this occurs with NESTs and that no stakeholder group believes the singular use of a student-completed TEF can meet these purported aims. Rather, the findings speak to a wider presence and negative impact of metrics in HE. For instance, they suggest that stakeholder groups believe the use of and reliance on a student-completed TEF leads to a broad range of detrimental secondary effects (e.g. high levels of fear and anxiety, diminished sense of autonomy and agency, lack of professional development) that not only greatly impact NESTs but also students.

By examining the findings in this study, it is possible for stakeholders to increase their level of awareness regarding the need to improve TEF knowledge in relation to one's learning and teaching practices, as well as realize a need to establish a comprehensive teaching performance evaluation system. This can prove to be even more helpful for those in communities of practice that are significantly influenced by neoliberal administrative policies. The information gained from this study could also encourage similarly situated NESTs to dig a little deeper into concerns about how they are evaluated in a more holistic and collaborative manner regardless of the outcome. Doing so might embolden NESTs to problematize their givens and encourage them to find sound ways to conceptualize and uphold their own professional development values within their situated contexts.

6.3 Limitations of the study

One of the limitations of this study is the number of participants. Although I collected data from multiple sources to address this concern (Guba & Lincoln, 1994), I felt slightly apprehensive about the sample size and representativeness. My initial thoughts about this matter included recruiting more participants, which could lead to more theoretical sampling and saturation, and ultimately, increased validity of the findings. However, I decided to follow a suggestion made by Bowen (2008) and focus less on sample size and more on sample adequacy as the researcher does not always seek generalizability or representativeness. I therefore had participant samples drawn based on their prior knowledge of the TEF and their relevance to the primary research question.

Another limitation involves the validity of participant responses. For example, students would often answer a question or prompt in relation to their Korean professors and major classes before shifting their focus to GEP NESTs. It was sometimes necessary to keep them on track or ask them to clarify their thoughts to ensure the accuracy of their responses. However, this could also have been due to the heavy influence of their previous/historical experiences. GEP administrators such as Hana and Hyuna were former students at my

university and admitted that this sometimes influenced the way they answered the interview questions before self-correcting themselves. In addition, there is a concern with how participants often gave responses that covered or overlapped with multiple interview questions. It was therefore important to correctly tease out and organize the responses to their corresponding questions in order to properly analyze the data. Conducting a full pilot study prior to the main study could have perhaps helped to identify and address these issues regarding overlapping responses.

A final limitation involves my own position within this study. In addition to being the researcher, I am an active bilingual NEST, GEP colleague, and university employee. This unique position allowed me to gain insight on complex stakeholder perceptions of a student-completed TEF that is used to evaluate NESTs in my context. However, this same position also presented unwanted opportunities for me to get caught up in the participants' emotional responses during their interviews and perhaps even side with them. This would have led to validity and reliability issues in both the data collection and analysis stages, so I endeavored to keep my distance with the participants and engage in constant self-reflection in order to maintain a critical perspective throughout the study.

6.4 Suggestions for future research, practice, policy, and final thoughts

A number of suggestions for future research involve the aforementioned limitations of the study. This includes refining the semi-structured interview questions in order to elicit responses that more specifically and accurately align with the research questions. Increasing the number of participants to include main administrators who are more directly involved with the TEF could also help strengthen the findings of the study. These findings would provide NESTs with critical information needed to understand the TEF and to find practical uses for the data. In addition, to further reduce the risk of researcher bias in studies such as this, it would perhaps be helpful to have an additional, neutral researcher (e.g. non-GEP related) involved in data collection and analysis.

Encouraging and conducting more teacher evaluation research on NESTs at Korean universities could also resolve simple issues related to language barriers and knowledge among stakeholder groups. For instance, NESTs in my study frequently mentioned language-related difficulties with regard to accessing even basic information about the TEF (e.g. understanding the items, reading student qualitative feedback). However, NEST participants in studies conducted by Burden (2008a, 2008b, 2010) within a Japanese university EFL context did not mention any fundamental language-related difficulties, which is likely because one of the NEST hiring requirements stated by the Japanese MoE includes possessing a certain level of Japanese competency. This is perhaps a national educational policy that the Korean MoE should also consider implementing in order to, at the very least, reduce basic communication and knowledge issues that NESTs and main administrators frequently encounter. Addressing this need for distinct professional standards and robust assessments of NESTs in Korea on a national level could conceivably make a difference in NEST responses in future studies within this context.

Final thoughts

For the duration of this study, I immersed myself within my professional context and delved into student-completed teacher evaluation with regard to NESTs. I discovered the importance of maintaining an open and flexible approach to the emerging data. Doing this made me become aware of a number of unexpected (and sometimes troubling) opinions regarding the TEF and of how stakeholder groups perceived each other. While I was not surprised to see an overall negative view of the TEF by NESTs, I found it disturbing at how deeply it affected some of my colleagues. The absence of TEF knowledge and its neoliberal use (i.e. summative, high-stakes, focus on student satisfaction) has led to a multitude of problems, such as fear of job security, lack of autonomy and agency, and a general lack of professional development.

I was also taken aback at the degree of fear that students exhibited throughout the study.

Each student participant repeatedly asked for assurance of their anonymity during their interviews and focus group. Their assumption that NESTs would take offense to receiving low TEF scores and thus threaten or punish students through their final grades was startling. I think the lack of trust they show towards all professors and university administrators indicates a critical need for improved communication and sharing of knowledge. In addition, it was both surprising and disheartening to find that most participants seemed resigned to accepting their given situations. Although they were able to identify and problematize issues regarding the TEF, none of them showed any hope of change. As Penny (2003) commented, it is “unlikely that the use of student ratings will be abandoned” (p. 400), so even though the study participants believe that NESTs should be evaluated in a more fair, transparent, and valid way, it appears that they also doubt the possibility of this happening due to critical neoliberal constraints that HE institutions cannot forcibly remove.

On a personal level, I thought it sadly ironic that despite conducting this study, gaining a better understanding of the student-completed TEF from each stakeholder group perspective, and engaging in a very meaningful level of self-reflection, I still found myself anxiously focusing on my end-of-the-semester TEF scores like the rest of my colleagues. Rather than thinking about how I could use this feedback to better improve my teaching practices, I continued to worry about whether my final TEF score will allow me to stay at my job for another year. Unfortunately, these fears have been realized, and I am now currently engaged in a formal appeal process regarding the university administration’s decision to not renew my contract. Part of this process involves gathering TEF data from the past 10 years to present to a faculty personnel committee. This is because the administration perceives a direct correlation between my teaching performance and final TEF scores. Experiencing first-hand how TEF scores are utilized as a high-stakes PI has made me believe that NEST concerns regarding TEF data use are not entirely unfounded and that change is needed.

Through this study, I have come to feel strongly about encouraging all stakeholder groups to take a step back, look at the bigger picture regarding student-completed teacher evaluations,

consider all of the shifts that have occurred in higher education and in EFL, and work on creating a more valid NEST evaluation system that acknowledges these changes. This system should be comprehensive and satisfy the needs of all stakeholders in terms of providing them useful data for improvement purposes, encouraging them to engage in meaningful dialogue, and making full use of the evaluation data for NEST professional development. It should also take into account current, context-specific factors that can significantly affect the TEF itself or the way students complete TEFs. In order to achieve this at my university, it is necessary for students, NESTs, and administrators to work towards constructing knowledge on what constitutes good teaching practices. That is, to increase the possibility of evaluating NESTs in a more appropriate, useful, and valid way, stakeholder groups must continuously identify and balance mismatching expectations and perceptions of good or effective teaching in an EFL classroom (Alimorad & Tajgozari, 2016), while acknowledging the differences in language teaching practices depending on the context.

REFERENCES

Abelmann, N., Park, S. J., & Kim, H. (2009). College rank and neo-liberal subjectivity in South Korea: The burden of self-development. *Inter-Asia Cultural Studies*, *10*(2), 229-247.

Agger, B. (1991). Critical theory, poststructuralism, postmodernism: Their sociological relevance. *Annual Review of Sociology*, *17*(1), 105-131.

Alamoudi, K., & Troudi, S. (2017). EFL Teacher Evaluation: A Theoretical Perspective. In *Evaluation in Foreign Language Education in the Middle East and North Africa* (pp. 29-41). Springer International Publishing.

Alimorad, Z., & Tajgozari, M. (2016). A comparison of Iranian high school teachers' and students' perceptions of effective English teachers. *SAGE Open*, *6*(4), 2158244016679212.

Al-Mutawa, N.A. (1997). Evaluation of EFL primary school teachers' competencies in Kuwait. *Evaluation & Research in Education, 11*(1), 38-52.

Al-Mutawa, N., & Al-Dabbous, J. (1997). Evaluation of Student-Teachers' Performance in TEFL Practicum at Kuwait University. Retrieved from http://files.eric.ed.gov/fulltext/ED435195.pdf

Alptekin, C. & Alptekin, M. (1984). The question of culture: EFL teaching in non-English-speaking countries. *ELT Journal*, *38*(1), 14-20.

Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles*, *49*(9), 507-516.

Archer, L. (2008). The new neoliberal subjects? Young/er academics' constructions of professional identity. *Journal of Education Policy*, *23*(3), 265-285.

Bailey, K. M. (2009). 37 Issues in language teacher evaluation. *The Handbook of Language Teaching*, 706. Sussex: Wiley- Blackwell.

Barnes, B. D., & Lock, G. (2013). Student perceptions of effective foreign language teachers: A quantitative investigation from a Korean university. *Australian Journal of Teacher Education*, *38*(2), 2.

Barratt, L., & Kontra, E. H. (2000). Native-English-Speaking Teachers in Cultures Other Than Their Own. *TESOL journal*, *9*(3), 19-23.

Beck, J., & Young, M. F. (2005). The assault on the professions and the restructuring of academic and professional identities: a Bernsteinian analysis. *British Journal of Sociology of Education*, *26*(2), 183-197.

Bernstein, K. A., Hellmich, E. A., Katznelson, N., Shin, J., & Vinall, K. (2015). Introduction to special Issue: Critical perspectives on neoliberalism in second/foreign language education. *L2 Journal*, *7*(3).

Block, D. (1998). Exploring interpretations of questionnaire items. *System*, *26*(3), 403-425.

Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, *10*(1), 3-31.

Bowen, G. A. (2008). Naturalistic inquiry and the saturation concept: a research note. *Qualitative Research*, *8*(1), 137-152.

Burden, P. (2008a). ELT teacher views on the appropriateness for teacher development of end of semester student evaluation of teaching in a Japanese context. *System*, *36*(3), 478-491.

Burden, P. (2008b). Does the use of end of semester evaluation forms represent teachers' views of teaching in a tertiary education context in Japan? *Teaching and Teacher Education*, *24*(6), 1463-1475.

Card, J. (2005). Life and death exams in South Korea. *Asia Times*, *11*(30), 2005.

Çelik, S., Arikan, A., & Caner, M. (2013). In the Eyes of Turkish EFL Learners: What Makes an Effective Foreign Language Teacher? *Porta Linguarum, 20,* 287-297.

Chae, J. E., & Hong, H. K. (2009). The expansion of higher education led by private universities in Korea. *Asia Pacific Journal of Education*, *29*(3), 341-355.

Chang, D. & Park, C. (2011, May 10). Professor's presidency criticized. *The Tech*, 131(26). Retrieved from http://tech.mit.edu/V131/N26/kaist.suicides.html

Charmaz, K. (2005). Grounded theory in the 21st century. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research (3rd ed.) (pp. 507-531)*. Sage.

Charmaz, K. (2014). *Constructing grounded theory*. Sage.

Chávez, K., & Mitchell, K. M. (2020). Exploring bias in student evaluations: Gender, race, and ethnicity. *PS: Political Science & Politics*, *53*(2), 270-274.

Chisman, F. P., & Crandall, J. (2007). Passing the torch: Strategies for innovation in community college ESL. *Council for Advancement of Adult Literacy (NJ1)*.

Cho, J. (2012). Campus in English or campus in shock? *English Today*, *28*(02), 18-25.

Choi Soon-sil: South Korean ex-president's friend jailed (2017, June 23). *BBC News.* Retrieved from https://www.bbc.com/news/world-asia-40377564

Chun, S. Y. (2014). EFL learners' beliefs about native and non-native English-speaking teachers: perceived strengths, weaknesses, and preferences. *Journal of Multilingual and Multicultural Development*, *35*(6), 563-579.

Chung, A. (2016, November 21). University told to revoke place of Park's aide's daughter. University Word News. Retrieved from https://www.universityworldnews.com/post.php?story=20161121163433972

Chung, A. (2017, June 1). Push for jail terms over university admissions scandal. University World News. Retrieved from https://www.universityworldnews.com/post.php?story=20170601142810206

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education (6th ed.)*. Routledge.

Coldron, J., & Smith, R. (1999). Active location in teachers' construction of their professional identities. *Journal of Curriculum Studies*, *31*(6), 711-726.

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory (4th* ed.). Sage.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches (5th ed.)*. Sage.

Crookes, G., Davis, K. A., & Clair, N. (1995). Mainstream classroom teachers and ESL students. *TESOL quarterly*, *29*(1), 189-196.

Danielson, C., & McGreal, T. L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: ASCD.

Delandshere, G., & Petrosky, A. R. (1994). Capturing teachers' knowledge: Performance assessment a) and post-structuralist epistemology b) from a post-structuralist perspective c) and post-structuralism d) none of the above. *Educational Researcher*, *23*(5), 11-18.

Denzin, N. K., & Lincoln, Y. S. (2008). *Collecting and interpreting qualitative materials* (3rd ed.). Sage.

Dickey (2006). Assessing ethical standards for EFL teaching professionalism. *Information Technology, 50,* 60.

Dunegan, K. J., & Hrivnak, M. W. (2003). Characteristics of mindless teaching evaluations and the moderating effects of image compatibility. *Journal of Management Education*, *27*(3), 280-303.

Elyas, T., & Al-Sadi, A. (2013). Critiquing of Higher Education policy in Saudi Arabia: A Neoliberalism Approach. In *The Asian Conference on Language Learning*.

Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: A critique of student evaluations of teaching effectiveness. *Quality Assurance in Education*.

Ennser-Kananen, J., Fallas Escobar, C., & Bigelow, M. (2016). "It's Practically a Must": Neoliberal Reasons for Foreign Language Learning. *International Journal of Society, Culture & Language*, 1-14.

Evetts, J. (2012, May). Professionalism in turbulent times: Changes, challenges and opportunities. In *Propel International Conference Stirling* (Vol. 911).

Franklin, J., & Theall, M. (1989). *Who Reads Ratings: Knowledge, Attitude, and Practice of Users of Student Ratings of Instruction.* Northeastern Univ., Boston: Office of Instructional Development and Evaluation (ERIC Document Reproductive Service No. ED306241)

Ginns, P., Prosser, M., & Barrie, S. (2007). Students' perceptions of teaching quality in higher education: The perspective of currently enrolled students. *Studies in Higher Education, 32*(5), 603-615.

Giroux, H. (2002). Neoliberalism, corporate culture, and the promise of higher education: The university as a democratic public sphere. *Harvard Educational Review*, *72*(4), 425-464.

Gordon, R. A., & Stuecher, U. (1992). The effect of anonymity and increased accountability on the linguistic complexity of teaching evaluations. *The Journal of Psychology*, *126*(6), 639-649.

Guolla, M. (1999). Assessing the teaching quality to student satisfaction relationship: Applied customer satisfaction research in the classroom. *Journal of Marketing Theory and Practice, 7*(3), 87-97.

Han, S. A. (2005). Good teachers know where to scratch when learners feel itchy: Korean learners' views of native-speaking teachers of English. *Australian Journal of Education*, *49*(2), 197-213.

Hargreaves, A. (2000). Four ages of professionalism and professional learning. *Teachers and Teaching*, *6*(2), 151-182.

Harvey, L. (2001). *Student feedback: a report to the higher education funding council for England.* Univ. of Central England, Birmingham: Centre for Research into Quality.

Howard, A., & Donaghue, H. (2015). *Teacher evaluation in second language education*. New York: Bloomsbury Publishing.

Howard, N. J. (2019). Constructing professional identities: Native English-speaking teachers in South Korea. *The Qualitative Report*, *24*(7), 1478-1510.

Isoré, M. (2009). Teacher Evaluation: Current Practices in OECD Countries and a Literature Review. OECD Education Working Papers, No. 23. *OECD Publishing (NJ1)*.

Jacobson, D., & Mustafa, N. (2019). Social identity map: A reflexivity tool for practicing explicit positionality in critical qualitative research. *International Journal of Qualitative Methods, 18*, 1609406919870075.

Jeon, M. (2009). Globalization and native English speakers in English programme in Korea (EPIK). *Language, Culture and Curriculum, 22*(3), 231-243.

Jo, S. (2008). English education and teacher education in South Korea. *Journal of Education for Teaching*, *34*(4), 371-381.

Johnson, R. (2000). The authority of the student evaluation questionnaire. *Teaching in Higher Education*, *5*(4), 419-434.

Joye, S. W., & Wilson, J. H. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning*, *15*(4), 126-138.

Jung, H. (2008). The washback effects of English listening test in Korean CSAT on secondary school English learning and teaching. *The Linguistic Association of Korea Journal*, *16*(3), 143-162.

Kang, H.D. (2012). Primary school English education in Korea: From policy to practice. In Spolsky, B. & Moon, Y.I. (Eds.). *Primary school English-language education in Asia: From policy to practice* (pp.59.-82). New York: Routledge.

Kang, H.R. (2015, April 15). The all-work, no-play culture of South Korean education. *NPR.* Retrieved from http://www.npr.org/sections/parallels/2015/04/15/393939759/the-all-work-no-play-culture-of-south-korean-education

Kang, N. H., & Hong, M. (2008). Achieving excellence in teacher workforce and equity in learning opportunities in South Korea. *Educational Researcher*, *37*(4), 200-207.

Kasai, M., Lee, J. A., & Kim, S. (2011). Secondary EFL students' perceptions of native and nonnative English-speaking teachers in Japan and Korea. *Asian EFL Journal*, *13*(3), 272-300.

Kauppi, N., & Erkkilä, T. (2011). The struggle over global higher education: Actors, institutions, and practices. *International Political Sociology*, *5*(3), 314-326.

Kelchtermans, G. (2004). CPD for professional renewal: Moving beyond knowledge for practice. *International handbook on the continuing professional development of teachers*, 217-237.

Kember, D., Leung, D. Y., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment & Evaluation in Higher Education*, *27*(5), 411-425.

Kim, Y. S. (2001). A survey study of US EFL teachers in Korea. *KOTESOL Proceedings 2001*, *173*.

Kim, T. (2008). Higher Education Reforms in South Korea: Public—Private Problems in Internationalising and Incorporating Universities. *Policy futures in education*, *6*(5), 558-568.

Kim, S. (2012). Living as a welcomed outsider. *응용언어학*, *28*(2), 27-58.

Kim, T. Y. (2006). Motivation and attitudes toward foreign language learning as socio-politically mediated constructs: The case of Korean high school students. *The Journal of Asia TEFL*, *3*(2), 165-192.

Kim, R. (2016, December 5). Choi Soon-sil's daughter stripped of high school diploma. *The Korea Times*. Retrieved from https://www.koreatimes.co.kr/www/news/nation/2016/12/116_219515.html

Kim, J. (2018, February 13). Friend of ousted South Korean president gets 20 years in jail. *CNN News*. Retrieved from https://edition.cnn.com/2018/02/13/asia/south-korea-choi-soon-sil-sentence-intl/index.html

King, F. (2014). Evaluating the impact of teacher professional development: an evidence-based framework. *Professional Development in Education*, *40*(1), 89-111.

Koo, H. (2007). The changing faces of inequality in South Korea in the age of globalization. *Korean Studies*, *31*(1), 1-18.

Korea Institute for Curriculum and Evaluation (2001). *Curriculum, assessment and College Scholastic Ability Test in the Republic of Korea.* (PIM 2001-4). Retrieved from http://kice.re.kr/resrchBoard/view.do?seq=37615&s=kice&m=030106

Korea Ministry of Education and Statistics, Statistics Korea (2013). *Private education costs report.* Retrieved from http://kostat.go.kr/portal/korea/kor_nw/2/13/1/index.board?bmode=read&aSeq=311886

Kwon, O. (2000). Korea's English education policy changes in the 1990s: Innovations to gear the nation for the 21st century. *English Teaching (영어교육)*, *55*(1), 47-91.

Le Ha, P., & Barnawi, O.Z. (2015). Where English, neoliberalism, desire and internationalization are alive and kicking: Higher education in Saudi Arabia today. *Language and Education*, *29*(6), 545-565.

Lee, C. J. (2005). Korean education fever and private tutoring. *KEDI Journal of Educational Policy*, *2*(1).

Lee, C. J., Lee, H., & Jang, H. M. (2010). The history of policy responses to shadow education in South Korea: Implications for the next cycle of policy responses. *Asia Pacific Education Review*, *11*(1), 97-108.

Lee, J. J. (2010). The uniqueness of EFL teachers: Perceptions of Japanese learners. *TESOL Journal*, *1*(1), 23-48.

Lee, S., & Brinton, M. C. (1996). Elite education and social capital: The case of South Korea. *Sociology of education*, 177-192.

Lee, S.S. (2013, November 7). South Korea's dreaded college entrance exam is the stuff of high school nightmares, but is it producing "robots"? *CBC News*. Retrieved from http://www.cbsnews.com/news/south-koreas-dreaded-college-entrance-exam-is-the-stuff-of-high-school-nightmares-but-is-it-producing-robots/

Lin, A. M. (2015). Researcher positionality. *Research methods in language policy and planning: A practical guide*, 21-32.

Looney, J. (2011). Developing high-quality teachers: teacher evaluation for improvement. *European Journal of Education*, *46*(4), 440-455.

Lu, H., & Hodge, W. A. (2019). Toward multi-dimensional and developmental notion of researcher positionality. *Qualitative Research Journal, 19*(3)*.*

MacNell, L., Driscoll, A., & Hunt, A. N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education*, *40*(4), 291-303.

Master, P. (1983). The etiquette of observing. *TESOL Quarterly*, *17*(3), 497-501.

McDonald, M. (2011, May 22). Elite South Korean university rattled by suicides. *The New York Times.* Retrieved from http://www.nytimes.com/2011/05/23/world/asia/23southkorea.html

McLaren, P., Darder, A., Baltodano, M. P., & Torres, R. D. (2009). The critical pedagogy reader. Routlege Taylor and Francis Group.

Murray, A. (2010). Empowering teachers through professional development. *English Teaching Forum*, *48*(1), 2-11.

Nasser, F., & Fresko, B. (2002). Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, *27*(2), 187-198.

Neumann, R. (2001). Disciplinary differences and university teaching. *Studies in Higher Education*, *26*(2), 135-146.

Newton, E. H., & Braithwaite, W. E. (1988). Teacher perspectives on the evaluation of teachers. *Educational Studies*, *14*(3), 275-288.

Nicholls, G. (2000). Professional development, teaching, and lifelong learning: the implications for higher education. *International Journal of Lifelong Education*, *19*(4), 370-377.

Nixon, J. (1996). Professional identity and the restructuring of higher education. *Studies in Higher Education*, *21*(1), 5-16.

Nunan, D., & Lamb, C. (1996). *The self-directed teacher: Managing the learning process*. Cambridge: Cambridge University Press.

Park, G. P., & Lee, H. W. (2006). The characteristics of effective English teachers as perceived by high school teachers and students in Korea. *Asia Pacific Education Review*, *7*(2), 236-248.

Park, J. K. (2006). Professionalization of TEFL in Korea: The roads behind and ahead. *The Journal of Asia TEFL*, *3*(4), 113-134.

Park, J. K. (2009). 'English fever'in South Korea: its history and symptoms. *English Today*, *25*(1), 50.

Park, J.S.Y. (2010). Naturalization of competence and the neoliberal subject: Success stories of English language learning in the Korean conservative press. *Journal of Linguistic Anthropology*, *20*(1), 22-38.

Patton, M. Q. (2008). *Utilization-focused evaluation*. London: Sage Publications.

Pederson, R. (2012). Representation, globalization, and the native speaker: Dialectics of language, ideology & power. In Sung, K. & Pederson, R. (Eds.). *Critical ELT practices in Asia: Key issues, practices, and possibilities* (pp.1-22). Rotterdam: Sense Publishers.

Pennington, M. C. (1989). Directions for faculty evaluation in language education. *Language, Culture and Curriculum*, *2*(3), 167-193.

Pennington, M. C., & Young, A. L. (1989). Approaches to faculty evaluation for ESL. *TESOL Quarterly*, *23*(4), 619-646.

Pennington, M.C. (1991). Work satisfaction and the ESL profession. *Language, Culture, and Curriculum, 4*(1), 59-86.

Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, *8*(3), 399-411.

Pennycook, A. (1999). Introduction: Critical approaches to TESOL. *TESOL Quarterly*, *33*(3), 329-348.

Pennycook, A. (2001). *Critical applied linguistics: A critical introduction*. Routledge.

Perez-Milans, M., & Soto, C. (2014). Everyday Practices, Everyday Pedagogies. *Reflexivity in language and intercultural education: Rethinking multilinguism and interculturality*, 213-225. Routledge.

Pill, A. (2005). Models of professional development in the education and practice of new teachers in higher education. *Teaching in Higher Education*, *10*(2), 175-188.

Piller, I., & Cho, J. (2013). Neoliberalism as language policy. *Language in society*, *42*(1), 23-44.

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, *52*(2), 126.

Price, G. (2014). English for all? Neoliberalism, globalization, and language policy in Taiwan. *Language in Society*, *43*(05), 567-589.

Raaper, R. (2017). Tracing assessment policy discourses in neoliberalised higher education settings. *Journal of Education Policy, 32(3)*, 322-339.

Ramanathan, V., & Morgan, B. (2007). TESOL and policy enactments: Perspectives from practice. *TESOL Quarterly*, *41*(3), 447-463.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire. *Studies in Higher Education, 16*(2), 129-150.

Robinson, J. (1994). Social status and academic success in South Korea. *Comparative Education Review*, *38*(4), 506-530.

Ropers-Huilman, B. (1999). Witnessing: Critical inquiry in a poststructural world. *International Journal of Qualitative Studies in Education*, *12*(1), 21-35.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage.

Schwandt, T. A. (2015). *The Sage dictionary of qualitative inquiry*. Sage.

Seol, D. H. (2012). The citizenship of foreign workers in South Korea. *Citizenship studies*, *16*(1), 119-133.

Seth, M. J. (2002). *Education fever: Society, politics, and the pursuit of schooling in South Korea*. University of Hawaii Press.

Shin, H. (2007). English language teaching in Korea. *International handbook of English Language Teaching*, 75-86.

Sim, M. (2014). A qualitative case study of native-English speaking teachers in Korea. *Multicultural Education Review, 6*(2), 117-143.

Simpson, P. M., & Siguaw, J. A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education*, *22*(3), 199-213.

Śliwa, M., & Johansson, M. (2014). How non-native English-speaking staff are evaluated in linguistically diverse organizations: A sociolinguistic perspective. *Journal of International Business Studies*, *45*(9), 1133-1151.

Sojka, J., Gupta, A. K., & Deeter-Schmelz, D. R. (2002). Student and faculty perceptions of student evaluations of teaching: A study of similarities and differences. *College Teaching*, *50*(2), 44-49.

Sorensen, C.W. (1994). Success and education in South Korea. *Comparative education review*, *38*(1), 10-35.

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598-642.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques*. Thousand Oaks, CA: Sage.

Sung, K. (2012). Critical practices in Asia: A project of possibilities in the era of world Englishes. In Sung, K. & Pederson, R. (Eds.). *Critical ELT practices in Asia: Key issues, practices, and possibilities* (pp.23-54). Rotterdam: Sense Publishers.

Svinicki, M. D. (2001). Encouraging your students to give feedback. *New Directions for Teaching and Learning*, *2001*(87), 17-24.

Tang, S. Y. F., & Choi, P. L. (2009). Teachers' professional lives and continuing professional development in changing times. *Educational Review*, *61*(1), 1-18.

Tavakoli, M., & Sadeghi, S. (2011). An inquiry-based approach to teacher professionalism: Toward a trichotomous model of critical ethnography. *Reflective Practice, 12*(3), 361-373.

Theall, M., & Franklin, J. (1991). Using student ratings for teaching improvement. *New Directions for Teaching and Learning*, *1991*(48), 83-96.

Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions for Institutional Research*, *2001*(109), 45-56.

Thomas, M. K. & Yang, W. L. (2013). Neoliberalism, globalization, and creative educational destruction in Taiwan. *Educational Technology Research and Development*, *61*(1), 107-129.

Tran, T. T. T., & Do, T. X. (2020). Student evaluation of teaching: Do teacher age, gender, and qualification matter? *Educational Studies*, 1-28.

Troudi, S. (2009). Recognising and rewarding teachers' contributions. In M. Al-Hamly et al. (Eds.), *Proceedings of the 14th TESOL Arabic Conference: Finding Your Voice: Critical Issues in ELT* (pp.60-67). Dubai. TESOL Arabia Publications.

Wang, L. Y., & Lin, T. B. (2013). The Representation of Professionalism in Native English-Speaking Teachers Recruitment Policies: A Comparative Study of Hong Kong, Japan, Korea and Taiwan. *English Teaching: Practice and Critique*, *12*(3), 5-22.

Wennerstrom, A. K., & Heiser, P. (1992). ESL student bias in instructional evaluation. *TESOL Quarterly*, 271-288.

Wines, W. A., & Lau, T. J. (2006). Observations on the folly of using student evaluations of college teaching for faculty evaluation, pay, and retention decisions and its implications for academic freedom. *Wm. & Mary J. Women & L.*, *13*, 167.

Yang, H. (2017, May 31). Disgraced daughter of ousted South Korean president's confidante returns home. Reuters. Retrieved from https://br.reuters.com/article/us-southkorea-politics-chung-samsung-idUSKBN18R0OY

Yeom, M., & Ginsburg, M. (2007). Professionalism and the reform of teachers and teacher education in the Republic of Korea & the United States of America. *Asia Pacific Education Review*, *8*(2), 298-310.

Yoon, K. S., Duncan, T., Lee, S., & Shapley, K. (2008). The effects of teachers' professional development on student achievement: Findings from a systematic review of evidence. In *American Educational Research Association Annual Meeting*.

APPENDIX A – 2018 FINAL COURSE AND TEACHER EVALUATION FORM

구분	설문내용
학생자가평가문항 Student self-evaluation items	1. 나는 수업에 대비하여 충분한 준비를 하였다.(예습.복습) I was sufficiently prepared for class (i.e. previewed/reviewed class materials)
	2. 나는 이 수업에 적극 참여하였다 (출석, 질문, 과제, 시험, 상호작용 등) I actively participated in class (e.g. attendance, asked questions, completed assignments on time, quizzes/exams, class interaction etc.)
공통문항 -업적평가문항 -공통평점문항 Course/teaching evaluation items	3. 수업내용이 강의계획서에 맞게 충실히 진행 (실험/실습/실기 포함)되었다. The class adhered to the syllabus and course schedule (e.g. experiments, practical exercises, practicum).
	4. 수업내용에 적합한 강의방법과 교재 또는 자료가 사용되었다. The class content was taught through appropriate lecture methods and teaching materials.
	5. 교수는 학생에게 충분한 피드백 (시험, 과제, 질문 등)과 조언을 해주었다. The professor provided sufficient feedback and advice (i.e. regarding exams, assignments, and questions).
	6. 이 수업은 해당 교과목의 지식 함양에 도움을 주었다. This class helped to cultivate my knowledge of the subject.
	7. 전반적으로 이 수업 내용에 만족한다. (10 점척도) Overall, I am satisfied with the class content (on a scale of 1 to 10).
공통문항 (2017-2 추가) -공통평점 미적용 Course/teaching evaluation items (added Fall 2017) (excluded from evaluation score)	8. 교수방법 (언어표현의 정확도, 설명의 명확성, 사례제시의 적절성, 학습흥미의 유발 등)에 만족하고 있다. I am satisfied with the professor's lecture style and methods (e.g. accuracy of verbal expressions and explanations, appropriateness of examples used in class, contribution to learning motivation etc.)
	9. 강의의 진행속도는 학생의 능력과 이해수준을 고려하여 적절하였다. The lectures reflected the student's level of ability and understanding and progressed at an appropriate pace.
	10. 학생과 교수 간의 상호작용(질의·응답, 학생에 대한 관심과 이해 등)에 만족하고 있다. I am satisfied with the interaction between professor and student (e.g. Q&A, level of attention and understanding towards student etc.)
상호작용 (2017-2 추가) Interaction items (added Fall 2017)	상호작용 1. 강의 설명의 명확성 Interaction 1. Accuracy of lecture content and explanations
	상호작용 2. 강의 자료 및 사례제시의 적절성 Interaction 2. Appropriateness of lecture materials and examples used in class

	상호작용 3. 학습동기 및 흥미 유발 Interaction 3. Contribution to learning motivation and interest
	상호작용 4. 강의 시 언어표현의 정확성(목소리 크기, 발음 등) Interaction 4. Accurate use of vocabulary and verbal expressions (e.g. voice volume, pronunciation etc.)
	상호작용 5. 학생 질의에 대한 응답 만족도 Interaction 5. Satisfaction regarding responses to student questions
	상호작용 6. 학생에 대한 관심과 이해 Interaction 6. Level of attention and understanding towards student
수업개선용문항 Class improvement	11. 보강이 이루어지지 않은 휴강이 있었다. There were class cancellations that were not made up.
외국어강의문항 Foreign language item	12 이 수업의 약 몇 %가 외국어로 운영되었는가? Approximately what percent of the class was taught in a foreign language?
	13. 이 수업은 해당분야의 외국어능력 함양에 도움이 되었다. This class helped to develop my foreign language skills in this subject.
서술형문항 Qualitative feedback	20. 이 수업에서 특별히 좋았던 점은 무엇입니까? What did you like most about this class?
	21.수업내용이나 방법 등 개선할 점은 무엇입니까? What improvements should be made regarding class content, teaching methods etc.?

APPENDIX B – 2018 MIDTERM COURSE AND TEACHER EVALUATION FORM

구분	문항내용
공통문항 (5 단계 평점척도) 5-point Likert scale items	1.수업 운영 (출결 및 평가 기준, 강의방법 등)에 대한 공지가 명확하였다. The professor provided clear class guidelines (attendance & assessment criteria, lecture style etc.)
	2. 수업내용에 적합한 강의방법과 교재가 사용되었다. The class content was taught through appropriate lecture methods and teaching materials.
	3. 학생들과의 상호작용 (참여독려, 수업이해도 확인, 질의응답 등) 이 잘 이루어졌다. There was good interaction among students in class (e.g. the professor encouraged participation and checked student comprehension levels, Q&A)
서술형문항 Qualitative feedback item	4. 보다 나은 강의를 위한 기타 의견 또는 건의사항을 기술하십시오. Please provide any other opinions or suggestions for class improvement.

APPENDIX C – QUESTIONNAIRE EXAMPLE (NESTs)

The aim of this questionnaire is to examine the understanding and satisfaction that General English Program (GEP) teachers have regarding the current teacher evaluation form (TEF). Please take a few minutes to complete the following questions.

1. Have you reviewed the TEF items for classes that you have taught? If yes, then please indicate the number of times you have done so.

2. Please explain the general purpose of the TEF to the best of your knowledge.

3. The following is the current TEF. Please choose 3 items that are most relevant to you and your teaching goals. If you cannot read or understand the items, please check the box below.

□ I cannot read or understand the items

구분	설문 내용	√
학생자가평가문항	1. 나는 수업에 대비하여 충분한 준비를 하였다. (예습/복습)	
	2. 나는 이 수업에 적극 참여하였다. (출석, 질문, 과제, 시험, 상호작용 등)	
공통문항 (교원업적평가문항)	3. 수업내용이 강의계획서에 맞게 충실히 진행 (실험/실습/실기 포함) 되었다.	
	4. 수업내용에 적합한 강의방법과 교재 또는 자료가 사용되었다.	
	5.교수는 학생에게 충분한 피드백(시험, 과제, 질문 등)과 조언을 해주었다.	
	6. 이 수업은 해당 교과목의 지식 함양에 도움을 주었다.	
	7. 전반적으로 이 수업 내용에 만족한다. (10 점척도) 10 점 | 9 점 | 8 점 | 7 점 | 6 점 | 5 점 | 4 점 | 3 점 | 2 점 | 1 점	
수업개선용문항	8. 보강이 이루어지지 않은 휴강이 있었다. ① 없음 ② 1 회 ③ 2 회 ④ 3 회 ⑤ 4 회이상	
외국어강의문항	9. 이 수업의 약 몇 %가 외국어로 운영되었는가? ① 90% 이상 ② 80%~89% ③ 40%~79% ④ 20%~39% ⑤ 20% 미만	
	10. 이 수업은 해당분야의 외국어능력 함양에 도움이 되었다.	
서술형문항	11. 이 수업에서 특별히 좋았던 점은 무엇입니까?	
	12. 수업내용이나 방법 등 개선할 점은 무엇입니까?	

4. On the following scale, please indicate (√**)** how satisfied you are with the current TEF as a GEP teacher:

Very dissatisfied	Somewhat dissatisfied	Neither satisfied or dissatisfied	Somewhat satisfied	Very satisfied

4A. Please explain the reason for your answer.

5. Please comment on what other areas of GEP teaching you would like the TEF to evaluate, if any.

APPENDIX D – QUESTIONNAIRE EXAMPLE (KOREAN VERSION)

일반영어교육 수업평가 설문조사

이 설문지는 학생 여러분의 일반영어교육 (영어토론과발표, 영어쓰기읽기) 수업평가에 관한 전반적인 생각, 이해도, 그리고 만족도가 무엇인지를 알아보기 위한 것입니다. 이것은 시험이 아니므로 정답이 없으며 익명으로 제출하는 것입니다. 편안한 마음으로 주어진 질문에 답변 해주시면 감사하겠습니다.

1. 이전에 수업평가 설문에 응한적이 있습니까 (영토발, 영쓰읽 외 모든 강의 포함)? 만약에 '네'라고 대답하셨으면 대략 몇번 정도 하셨습니까?

2. 수업평가의 궁극적인 목적이 무엇이라고 생각하십니까?

3. 다음 내용은 현재 저희 대학교에서 실시하는 수업평가 설문 문항 목록입니다. 이 중에 일반영어 교육을 받는 학생으로서 가장 중요한 세가지 문항을 체크 해주십시오.

구분	설문내용	√
학생자가평가문항	1. 나는 수업에 대비하여 충분한 준비를 하였다.(예습. 복습)	
	2. 나는 이 수업에 적극 참여하였다.(출석, 질문, 과제, 시험, 상호작용 등)	
공통문항 -업적평가문항 -공통평점문항	3. 수업내용이 강의계획서에 맞게 충실히 진행(실험/실습/실기 포함)되었다.	
	4. 수업내용에 적합한 강의방법과 교재 또는 자료가 사용되었다.	
	5. 교수는 학생에게 충분한 피드백(시험, 과제, 질문 등)과 조언을 해주었다.	
	6. 이 수업은 해당 교과목의 지식 함양에 도움을 주었다.	
	7. 전반적으로 이 수업 내용에 만족한다. (10 점척도)	
공통문항 (2017-2 추가) _공통평점 미적용	8. 교수의 교수방법(언어표현의 정확도, 설명의 명확성, 사례제시의 적절성, 학습흥미의 유발 등)에 만족하고 있다.	
	9. 강의의 진행속도는 학생의 능력과 이해수준을 고려하여 적절하였다.	
	10. 학생과 교수 간의 상호작용(질의·응답, 학생에 대한 관심과 이해 등)에 만족하고 있다.	
상호작용 (2017-2 추가)	상호작용 1. 강의 설명의 명확성	
	상호작용 2. 강의 자료 및 사례제시의 적절성	
	상호작용 3. 학습동기 및 흥미 유발	
	상호작용 4. 강의 시 언어표현의 정확성(목소리 크기, 발음 등)	
	상호작용 5. 학생 질의에 대한 응답 만족도	
	상호작용 6. 학생에 대한 관심과 이해	
수업개선용문항	11. 보강이 이루어지지 않은 휴강이 있었다.	
외국어강의문항	12 이 수업의 약 몇 %가 외국어로 운영되었는가?	
	13. 이 수업은 해당분야의 외국어능력 함양에 도움이 되었다.	
서술형문항	20. 이 수업에서 특별히 좋았던 점은 무엇입니까?	
	21. 수업내용이나 방법 등 개선할 점은 무엇입니까?	

4. 일반영어 수업을 듣는 학생으로서 현재 쓰이고 있는 수업평가에 관한 자신의 만족도에 해당하는 칸을 체크 해주십시오.

매우 낮음	낮음	보통	높음	매우 높음

4A. 선택하신 만족도에 대한 추가설명을 해주십시오.

5. 일반영어교육 수업평가에 추가적인 문항을 만들 수 있다면 무엇을 반영하고 싶습니까?

APPENDIX E – INTERVIEW QUESTIONS (STUDENTS)

Main Question	Prompts (optional)
General Information	a) What year student are you? b) Which GEP classes have you taken? c) How often do you complete the TEF?
1. What do you think is the general purpose of the current TEF?	a) In your opinion, what is the purpose of using TEFs in general? TEF 를 쓰는 일반적인 목적이 뭐라고 생각하세요? b) In your opinion, what does the TEF evaluate? TEF 가 구체적으로 무엇을 평가 한다고 생각하세요?
2. How do you think the TEF is useful or not useful?	a) As a GEP student, what are your English learning goals? 일반영어 강의를 들은 학생으로서 영어를 배우는 목표/영어학습목표를 예로 들자면? b) In your opinion, do you think the TEF addresses your English learning goals? In what ways? TEF 와 학생의 영어학습목표가 잘 부합된다고 생각하세요? 어떤식으로? c) In your opinion, how well do you think the TEF addresses English teaching performance and effectiveness? In what ways? TEF 와 외국인 교수님 강의 및 강의방식에 대한 효과성이 잘 부합된다고 생각하세요? d) Which TEF items do you personally think are most useful/not useful to you and your English learning goals? Why? TEF 문항 중 학생들한테 가장 도움을 많이 주는 것들은 어떤거라고 생각하세요? 가장 도움 안 주는 문항은? 왜?
3. What is your most preferred way to use the TEF?	a) As student, in what way would you prefer the TEF to be used? TEF 를 어떻게 사용하면 학생들한테 가장 많은 도움이 될까요? 가장 이상적인 방법을 생각해주세요. b) In your opinion, do you think it is currently being used in this way? If not, why do you think that is?
4. What do you think is the best way to evaluate NESTs?	a) In your opinion, what are some other aspects of English teaching that could be evaluated in order to benefit students? 현재 TEF 문항 말고도 영어강의 평가하는데 있어서 추가적으로 평가를 했으면 좋겠다는 부분이 있으세요?

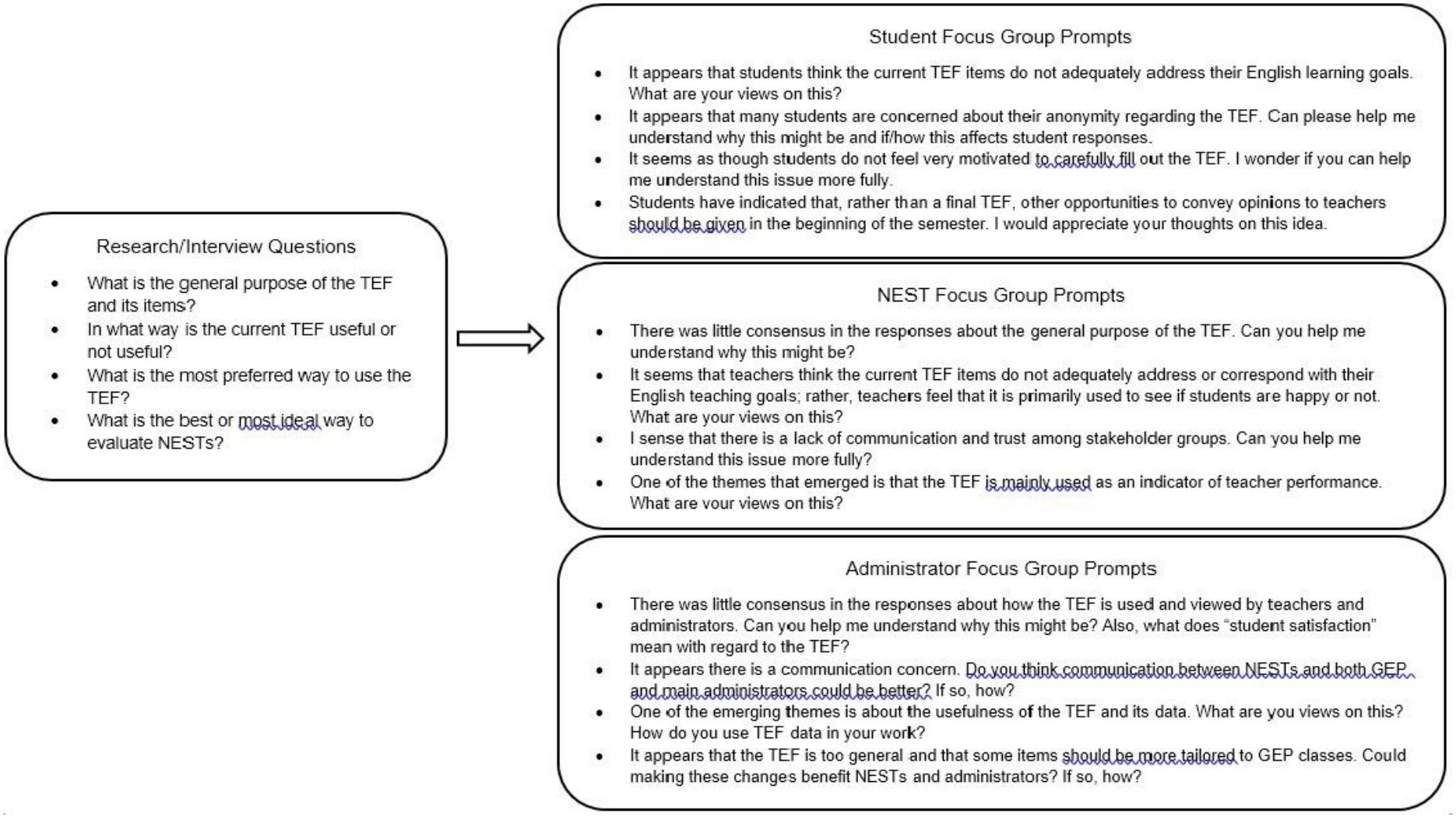
APPENDIX F – INTERVIEW QUESTIONS (NESTs)

Main Question	Prompts (optional)
General Information	a) How long have you been teaching at our university? b) What subjects do you teach? c) How often do you review the TEF?
1. What do you think is the general purpose of the current TEF?	a) In your opinion, what is the purpose of using TEFs in general? b) In your opinion, what does the TEF evaluate?
2. How do you think the TEF is useful or not useful?	a) As a GEP teacher, what are your English teaching goals? b) In your opinion, how well does the TEF address your English teaching goals/needs and effectiveness? In what ways? c) Which TEF items do you think personally are most/least useful to you with regard to achieving these teaching goals? Why?
3. What is your most preferred way to use the TEF?	a) As a GEP NEST, in what way would you prefer the TEF to be used? b) In your opinion, do you think it is currently being used in this way? If not, why do you think that is?
4. What do you think is the best way to evaluate NESTs?	a) In your opinion, what are some other aspects of English teaching that could be evaluated and that can benefit your teaching practice?

APPENDIX G – INTERVIEW QUESTIONS (ADMINISTRATORS)

Main Question	Prompts (optional)
General Information	a) How long have you been working as a GEP administrator? b) What administrative work do you do regarding the GEP?
1. What do you think is the general purpose of the current TEF?	a) In your opinion, what is the purpose of using TEFs in general? b) In your opinion, what does the TEF evaluate?
2. How do you think the TEF is useful or not useful?	a) What do you consider are the goals for the GEP from an administrative perspective? b) As a GEP administrator, how do you use the TEF and its data regarding administrative matters? c) How well does the TEF address your administrative needs regarding NESTs? In what ways? d) As a GEP administrator, which TEF items do you think are most useful regarding administrative matters? Why?
3. What is your most preferred way to use the TEF?	a) In your opinion, how can the TEF be used to benefit the GEP and administration? b) In your opinion, do you think it is currently being used in this way? If not, why do you think that is?
4. What do you think is the best way to evaluate NESTs?	a) In your opinion, what are some other aspects of English teaching and NESTs that could be evaluated in order to benefit the GEP or administration?

APPENDIX H – FOCUS GROUP PROMPTS



APPENDIX I – ETHICAL APPROVAL FORM

UNIVERSITY OF EXETER

GRADUATE SCHOOL OF EDUCATION

St Luke's Campus
Heavitree Road
Exeter UK EX1 2LU

http://socialsciences.exeter.ac.uk/education/

CERTIFICATE OF ETHICAL APPROVAL

Title of Project: Examining student-based teacher evaluation and stakeholder satisfaction in a Korean EFL university context

Researcher(s) name: Kyung Rhan Kim

Supervisor(s): Philip Durrant
Vivienne Baumfield

This project has been approved for the period

From: 20/10/2017
To: 13/07/2021

Ethics Committee approval reference:

D/17/18/03

Signature: Date: 29/08/2017
(Dr Christopher Boyle, Graduate School of Education Ethics Officer)

APPENDIX J – STUDY CONSENT FORM

GRADUATE SCHOOL OF EDUCATION

CONSENT FORM

I have been fully informed about the aims and purposes of the project.

I understand that:

there is no compulsion for me to participate in this research project and, if I do choose to participate, I may at any stage withdraw my participation

I have the right to refuse permission for the publication of any information about me

any information which I give will be used solely for the purposes of this research project, which may include publications

If applicable, the information, which I give, may be shared between any of the other researcher(s) participating in this project in an anonymised form

all information I give will be treated as confidential

the researcher(s) will make every effort to preserve my anonymity

…….. ……..
(Signature of participant) (Date)

……
(Printed name of participant)

One copy of this form will be kept by the participant; a second copy will be kept by the researcher(s)

Contact phone number of researcher(s):……..

If you have any concerns about the project that you would like to discuss, please contact:

……
OR

……

Data Protection Act: The University of Exeter is a data collector and is registered with the Office of the Data Protection Commissioner as required to do under the Data Protection Act 1998. The information you provide will be used for research purposes and will be processed in accordance with the University's registration and current data protection legislation. Data will be confidential to the researcher(s) and will not be disclosed to any unauthorised third parties without further agreement by the participant. Reports based on the data will be in anonymised form.