IN-SERVICE TEACHER EVALUATION AND ASSESSMENT AT ENGLISH PREPARATORY PROGRAMS OF THE UNIVERSITIES IN TURKEY: AN EXPLORATORY MIXED METHODS STUDY

Hasan SAVAŞ

MAY 2021

IN-SERVICE TEACHER EVALUATION AND ASSESSMENT AT ENGLISH PREPARATORY PROGRAMS OF THE UNIVERSITIES IN TURKEY: AN EXPLORATORY MIXED METHODS STUDY

A DISSERTATION SUBMITTED TO THE
GRADUATE SCHOOL
OF
BAHÇEŞEHİR UNIVERSITY

BY

HASAN SAVAŞ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE DEPARTMENT OF ENGLISH LANGUAGE EDUCATION

MAY 2021

REPUBLIC OF TURKEY
BAHÇEŞEHİR UNIVERSITY
GRADUATE SCHOOL

16/06/2021

We hereby recommend that the dissertation prepared under our supervision by **Hasan Savaş** entitled **"In-service teacher evaluation and assessment at English preparatory programs of the universities in Turkey: An exploratory mixed methods study"** be accepted in partial fulfillment of the requirements for the Doctor of Philosophy Program in English Language Education.

Prof. Dr. Kenan DİKİLİTAŞ
Thesis Supervisor

Recommendation concurred in:

Committe Members

Prof. Dr. Derin ATAY

Prof. Dr. Sumru AKCAN

Assoc. Prof. Dr. Nihal YURTSEVEN

Assist. Prof. Dr. Tuncer CAN

Approved:

Assoc. Prof. Dr. Burak KÜNTAY
Director

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Hasan SAVAŞ

Signature :

ABSTRACT

IN-SERVICE TEACHER EVALUATION AND ASSESSMENT AT ENGLISH PREPARATORY PROGRAMS OF THE UNIVERSITIES IN TURKEY: AN EXPLORATORY MIXED METHODS STUDY

Savaş, Hasan

Doctoral Dissertation, Doctor of Philosophy Program in English Language Education

Supervisor: Prof. Dr. Kenan Dikilitaş

May 2021, 241 Pages

This study aims to investigate the scope of teacher evaluation practices implemented at English preparatory programs of the universities in Turkey. Additionally, it is to suggest a context-specific teacher evaluation model through gaining the insights of EL teachers and administrators. The study bares an exploratory sequential mixed methods design in a sequence of phases. Through the qualitative phase, an online questionnaire were constructed. The aim here was to administer the questionnaire to the complete population of EL teachers and administrators. Following the piloting study, the questionnaire was delivered to the population through emails, 630 of whom responded back. This was the quantitative phase. The results revealed that the respondents had positive opinions towards teacher evaluation under certain circumstances and types of teacher evaluation practices were not more frequently used than each other. Teacher evaluation needed to pursue developmental aims while avoiding specific detrimental factors affecting teacher development. The study also revealed that there was no relationship between teacher evaluation and the years of teaching experience. Gender and the type of university that EL teachers work at did not have a significant effect on the opinions of EL teachers. However, educational background of EL teachers, their positions at the institutions and whether teacher

evaluation is systematic or not made a significant difference. Considering the overall findings of the study, a development-oriented teacher evaluation model (DTEM) was suggested to be used at the English preparatory programs of the universities in Turkey. Further implications, limitations, and suggestions for further research were also discussed.

Keywords: Teacher Evaluation, English Language Education, Professional Development of Teachers, Qualities of Teachers, Questionnaire Construction.

ÖZ

TÜRKİYE'DEKİ ÜNİVERSİTELERİN İNGİLİZCE HAZIRLIK PROGRAMLARINDA HİZMET İÇİ ÖĞRETMEN DEĞERLENDİRMESİ: KEŞFEDİCİ KARMA YÖNTEMLER ÇALIŞMASI

Savaş, Hasan
Doktora Tezi, İngiliz Dili Eğitimi Doktora Programı
Tez Yöneticisi: Prof. Dr. Kenan Dikilitaş

Mayıs 2021, 241 sayfa

Bu çalışmanın temel amacı, Türkiye'deki üniversitelerin İngilizce hazırlık programlarında uygulanan öğretmen değerlendirme uygulamalarının kapsamını araştırmak ve İngilizce öğretim görevlileri ve yöneticilerinin konu hakkındaki görüşlerini ölçmektir. Ayrıca bu yöndeki uygulamaların mesleki gelişim üzerindeki potansiyel etkilerini de dikkate alarak duruma özgü bir öğretmen değerlendirme modeli önermektir. Çalışma, nitel ve nicel veri toplama araç ve analizini bir dizi aşamada birleştiren keşfedici sıralı karma yöntem tasarımını ortaya koymaktadır. Nitel araştırma aşamasında, içerik ve tematik analizler uygulanmış ve bu sayede ana çalışma için anket maddeleri oluşturulmuştur. Buradaki amaç, İngilizce öğretim görevlilerinin ve idarecilerinin oluşturduğu araştırma evrenine uygulanacak bir anket oluşturmak olmuştur. Oluşturulan anketin güvenilirliğini ve geçerliliğini saptamak için uygulanan pilot çalışmasının ardından, anket e-postalar yoluyla tüm araştırma evrenine gönderilmiştir. Çalışmanın bu nicel aşamasıdır ve 630 katılımcıdan veri elde edilmiştir. Anketin sonuçları katılımcıların belirli koşullar altında öğretmen değerlendirme uygulamaları hakkında olumlu görüşlere sahip olduklarını ve uygulanan öğretmen değerlendirme türlerinin anlamlı ölçüde birbirlerinden daha sıklıkla kullanılmadığını ortaya koymuştur. Profesyonel gelişimi etkileyen zararlı

faktörlerden kaçınarak, öğretmen değerlendirmelerinin gelişimsel hedefleri amaçlaması gerektiği saptanmıştır. Çalışma ayrıca öğretmen değerlendirmesi ile öğretmenlik deneyimi arasında anlamlı bir ilişki olmadığını ortaya koymuştur. Cinsiyet ve görev yapılan üniversite türü katılımcıların görüşleri üzerinde anlamlı bir etkiye sahip değildir. Ancak, eğitim seviyeleri, kurumlardaki pozisyonları ve öğretmen değerlendirmesinin sistematik olup olmadığı, İngilizce öğretim görevlileri ve yöneticilerinin görüşlerinde anlamlı bir farklılık ortaya koymuştur. Araştırmanın genel bulguları dikkate alınarak Türkiye'deki üniversitelerin İngilizce hazırlık programlarında kullanılmak üzere gelişim odaklı öğretmen değerlendirme modeli önerilmiştir. Çalışmanın ek çıkarımları, sınırlamaları ve ileriki araştırmalar için öneriler sonuç bölümünde ayrıca tartışılmıştır.

Anahtar Kelimeler: Öğretmen Değerlendirmesi, İngiliz Dili Eğitimi, Öğretmenlerin Mesleki Gelişimi, Öğretmen Nitelikleri, Anket Geliştirme.

To my wife Elif and our baby to be born

ACKNOWLEDGMENTS

I would like to express my special appreciations and gratefulness to my esteemed teacher and thesis supervisor, Prof. Dr. Kenan Dikilitaş, who gave me a golden opportunity to embark on the journey that is my PhD studies. Having supported me from the beginning to the end, he always guided me put all this study well above the level of simplicity and into something concrete. But for his priceless support and supervision, I would not have come to know about the various new knowledge regarding this specific research and many other things. He was always there when I needed his support and guidance. I am thankful to him. He is my role model.

I would like to extend my sincere gratitude to Prof. Dr. Derin Atay, who opened a door for me to understand analytical thinking in depth and helped me focus on purposeful reading. Her inspirational comments and critical questions were always invaluable, both for this specific study and in the development of my academic career and skills. I feel honored to be a student of hers. I cannot express my gratitude enough to Prof. Dr. Sumru Akcan, who gave me the support while designing this dissertation and provided me encouragement to continue. I am grateful for her friendly and supportive apporach. My special thanks are for Assist. Prof. Dr. Tuncer Can for accepting my request to be a committee member and providing constructive feedback for the betterment of the dissertation. I would also like to thank Assoc. Prof. Dr Nihal Yurtseven for all her help and sincere productive comments.

Above all, I am extremely grateful to my parents, Şahin and Hediye Savaş, and to my sister, Handan Savaş, for their unconditioned love, prayers, caring and sacrifices they made throughout my life and academic career. They have always been by my side.

Last but by no means least, my deepest gratitude goes to my caring, loving and supporting wife, Elif Nur Savaş. Your endless love and inspiration during hard times encouraged me to walk this path. Thank you from the deepest of my heart. I dedicate this dissertation to you and our baby to be born.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

AR	Action Research
CELTA	Certificate in Teaching English to Speakers of Other Languages
CFA	Confirmatory Factor Analysis
CPD	Continuous Professional Development
DELTA	Diploma in Teaching English to Speakers of Other Languages
DTEM	Development-oriented Teacher Evaluation Model
EFA	Exploratory Factor Analysis
EFL	English as a Foreign Language
EL	English Language
ELT	English Language Teaching
EP	Exploratory Practice
In-Ser TE	In-service Teacher Evaluation
KMO	Kaiser-Meyer-Olkin Measure of Sampling Adequacy
L2	Second Language
LS	Lesson Study
MA	Master of Arts (degree)
MTEM	Marzano Teacher Evaluation Model
PHD	Doctor of Philosophy (degree)
PR	Practitioner Research
RP	Reflective Practice
SEM	Structural Equation Model
TEFL	Teaching English as a Foreign Language
TESOL	Teaching English to Speakers of Other Languages
TR	Teacher Research
VAM	Value-added Model

Chapter 1

Introduction

1.1 Overview of the Study

Teacher evaluation is considered a vital element in tool evidence from a range of sources that provide information regarding educators' performance in the classroom for the purposes of decision making and, as such, can lead to improvements in the teaching practice (Richards & Schmidt, 2013; Quirke, 2015). This means that through reflective monitoring schools should be active in developing and identifying opportunities for English language (EL) teachers at all levels and offer means for teacher development and learner progress -which constitute key elements for the successful integration of better practices and considerations into the teacher evaluation system (Donaghue & Howard, 2015; Thompson & Schademan, 2019; Tucker & Stronge, 2005). According to Mann (2004), gathering systematically triangulated evidence is essential for the development of an effective teacher evaluation program, which may include classroom observations (summative evaluation by managers or external evaluators; and formative assessment by peer teachers and professional development staff). Similarly, teacher evaluation models (i.e., value-added models) measuring student achievement, student evaluations of teachers, evaluation by the line manager or principal, teacher self-evaluation, and teacher portfolios of work can provide additional sources of evaluation to ensure triangulation (Bacher-Hicks et al., 2019; Donaldson, 2009; Quirke, 2015). There are a variety of factors that may affect how schools engage in teacher evaluation practices; the quality of teaching and learning is closely linked to that of student achievement, teacher quality and teacher development (Khan, Khan, Hussain, & Shaheen, 2017). In recent years, EL teachers have become more autonomous in decision making, analytical and critical of their practices, which has led to deeper and more critical instructional dialogue with mentors and supervisors (Kim & Danforth, 2012). Therefore, hierarchical and top-down relationship between administrators and teachers, observers and observees, supervisors and supervisees, or

teacher trainers and trainees might pose potential challenges in sustaining dialogic and a constructive relationship between parties (Kim & Danforth, 2012; Mann & Walsh, 2015).

For example, ongoing dialogue and constructive feedback are vital in sustaining the continuation of observation cycles (Iyer-O'Sullivan, 2015). However, as reported by King (2015), teachers have a stressful occupation, which leads to poor classroom performance - especially when accompanied by unsuitable managerial evaluation styles. Howard (2012), on the other hand, argues that a managerial stance in education may also result in stress, since this leads to a sense that professionalism is questioned, which may not address their actual developmental areas and pedagogical needs. Similarly, as stated by Riera (2011), the evaluative approach is "fraught with risks, which may damage rather than nurture the fragile enthusiasm and commitment to continuous improvement" (p. 54). In relation to this, "management controls have been gradually replaced by the practice of self-monitoring" (Kydd, 1997, p. 116).

Keeping all these complexities in mind in consideration of pre-sessional programs in Turkey, focus on collaboration between administrators and EL teachers in terms of the latter's professional development must to be sustained. There exists a lack of well-organized and focused teacher evaluation models that could hold the potential to open doors for educators to improve themselves professionally and to be evaluated for developmental purposes (Copland, 2015; Howard & Donaghue, 2015). Therefore, this thesis focuses on the vastly underexplored issue of the state of teacher assessment and evaluation in Turkish higher education - specifically in the context of English language teachers working in pre-sessional programs.

1.2 Purpose of the Study

This mixed-method study aimed to investigate the scope of teacher evaluation programs and practices implemented at English preparatory programs at state and foundation universities in Turkey. The study was conducted in two inter-related research stages (i.e., exploratory sequential mixed methods design). The first of these focused on data collection through semi-structured interviews and guided written reports obtained from administrators and EL teachers, in which the evaluative

practices procedures were identified and discussed to gain deeper insights into teachers' perspectives of their professional evaluation (See Appendix A).

At the second stage, on the basis of the analysis of this qualitative data set, an online questionnaire was constructed with the purpose of administering to the entire population of EL teachers and administrators working at English preparatory programs of the universities in Turkey (See Appendix B). Through the questionnaire, the complete respondents' perceptions of the evaluation and assessment procedures were elicited in order to draw up a context-sensitive teacher evaluation model for Turkey.

1.3 Significance of the Study

Teacher evaluation is essential for the design, implementation, and offering of an effective education to students - as well as for the continuous improvement of teaching practices. However, there is dearth of research exploring how EL teachers perceive their evaluation systems, despite this constituting one of the key pillars of professional development. To address this gap, we developed the questionnaire on the basis of the initial qualitative findings so the items would be based on realistic perceptions and practices of EL teachers. There is also a need to develop an EL teacher evaluation model for this specific context on the basis of the emerging needs, gaps, and expectations of multiple parties, which are built on the foundations of existing models. This being the case, the study is significant in two ways, since it will address an underexplored area of research in EL teacher evaluation and assessment and will reveal the state of EL teacher evaluation in Turkey with a widely administrated questionnaire. The thesis aims to generate an in-depth body of knowledge regarding EL teachers' perspectives of how they are evaluated, which offers program leaders and policy-makers numerous implications regarding the professional developments of EL teachers, students' achievements and institutional gains (Amrein-Beardsley & Holloway, 2019; Darling-Hammond, 2012; Moskal et al., 2016; Tucker & Stronge, 2005).

1.4 Research Questions

The study aims to explore the state of in-service teacher evaluation, EL teachers' perceptions of their evaluation practices, and the potential impact of such

evaluation on teacher development. To address this aim, the following research question guided the first qualitative stage:

1. What are the indicators of teacher evaluation from the perspectives of EL teachers and administrators working at English preparatory programs of the universities?

Based on the themes that emerged from the findings of the initial qualitative data, an online questionnaire was prepared and given to administrators and EL teachers working at English preparatory programs in Turkish universities - which formed the second stage of the study. We subsequently explored the same research question, drawing on a wider range of data collected from 630 respondents. Two general questions were raised at the second stage - both of which are divisible to specific questions (i.e., subscales). The four specific questions of the first set (question 2 below) were related to the four subscales of the questionnaire, as specified. The six specific questions in the second set (question 3 below) related to various continuous and categorical data obtained from the questionnaire. The research questions of the second stage included:

2. What are EL teachers' and administrators' views about the state of teacher evaluation in relation to the type of evaluation, impact on PD, and any intervening factors?

3. How do EL teachers and administrators view teacher evaluation in relation to their experience, educational background, working context, person's role, gender and specific evaluation practice?

1.5 Framework of the Study

Table 1 demonstrates the overall framework of the study. The first column represents the questions posed, while the second denotes the data collection procedure used to address each question. Meanwhile, the last column gives a brief account of the data analysis methods used to analyze the collected data.

Table 1

Overview of the Research Questions, Data Collection Instruments and Data Analysis

Research Questions	Data Collection Instruments	Data Analysis
1. What are the indicators of teacher evaluation from the perspectives of EL teachers and administrators working at English preparatory programs of the universities?	1. Guided Written Reports. 1. Semi-structured Interviews.	1.Two-staged Qualitative Analysis: a. Content & Thematic Analyses
2. What are the EL teachers' and administrators' views about the state of teacher evaluation in relation to the type of evaluation, impact on PD, and any intervening factors?	2. Online Questionnaire, Section 1 Section 2 Section 3 Section 4	2. Descriptive & Inferential Analysis a. Frequency Analyses b. Comparative Analyses
3. How do EL teachers and administrators view teacher evaluation in relation to their experience, educational background, working context, person's role, gender and specific evaluation practice?	3. Online Questionnaire Demographic Information Section 1 Section 2 Section 3 Section 4	3. Descriptive & Inferential Analysis a. Independent Samples T-Test b. One-way ANOVA c. Pearson Correlation Test

Chapter 2

Literature Review

2.1 Teacher Evaluation

Of comparable definitions of teacher evaluation, Borg (2019) refers to the term as "the process through which judgements about teacher quality are made and it is widely recognized to be a key element in effective educational systems" (p. 3). According to British Council's manual for teacher evaluation (2012), teacher evaluation is described as a vital element in improving student learning outcomes, gathering evidence from a range of sources that inform on teacher performance and using this to support teacher improvement in practice. To help schools improve, to identify opportunities for teachers of all levels to develop further, and to deliver accountability for self and learner progress are additional required dimensions to create a successful teacher evaluation system (Gordon & McGhee, 2019; Mielke & Frontier, 2012). In some cases, teacher evaluation serves to be used as a function of replacing, rewarding and/or dismissing teachers (Adnot, Dee, Katz, & Wyckoff, 2017; Dee & Wyckoff, 2015; Ridge & Lavigne, 2020).

In their earliest applications, most teacher evaluation practices lacked the sophistication that they have today - only exploring teachers' moral, ethical and personal traits (Ellet & Teddlie, 2003) or helping administrators make employment decisions (Carreiro, 2020). Danielson and McGreal (2000) state that most evaluation systems depended on the two dimensions of either *satisfactory* and *unsatisfactory*; while other systems incorporate rating scales of *low*, *medium*, or *high*, or *needs improvement*, *satisfactory* or *outstanding*. Although Danielson and McGreal appreciate the second type of evaluation as an improvement over the first type and offering greater objectivity, they believe that even the second type falls short of potential for precise evaluation because there is no agreement over what governs the ratings.

After the 1970s and 1980s, teacher evaluation began to consider rating scales, behavior checklists, categorizing systems and/or narrative records (Stronge, 1997).

According to Moss (2015), teacher evaluation during 1990s focused on standards-based performance indicators, as well as an expansion of the evaluation process, and was more consistent with creating standards for students and comprehensively descriptive of what teachers should know and be able to do; detailed behavioral rating scales with explicit standards covering multiple domains and including multiple levels of performance represented teacher evaluation.

Teacher evaluation today has become more sophisticated and its appropriate and contextual necessity cannot be denied except by those who are not familiar with the complexity of teaching profession (Tucker & Stronge, 2005). Teacher evaluation has a significant impact on the majority of teachers and practitioners deserving to develop professionally and improve teaching skills with high motivation (Derrington, 2011; Howard & Gillickson, 2010; Taylor & Tyler, 2012). The characteristics of and approaches to teacher evaluation are to foster dialogic mediation, collaboration and self-evaluation, and result in engagement and learning (Howard & Donaghue, 2015). However, Danielson and McGreal (2000) argue that teacher evaluation is for the most part characterized by the top-down communication between an evaluator and a teacher, which in the absence of clear criteria can be idiosyncratic. Instead, according to Tarhan, Karaman, Kemppinen, and Aerila (2019), teacher evaluation practices need to prioritize bottom-up processes which foster teacher empowerment and professional development based on expectations and what needs to be done to perform better. Rather than giving priority to high-stakes decisions by administrators on teachers' performance or renewal of contracts, the primary focus needs to be on low-stakes decisions such as betterment in teaching and student learning (Ridge & Lavigne, 2020). Tarhan et al. (2019) also suggest teacher evaluation should pursue a systematic process with a clear objective, preparation, and an expected product through different types of data collected and rigorous methods analyzed.

2.1.1 Purpose of teacher evaluation. Moskal, Stein and Golding (2016) state that the purpose of teacher evaluation practices in higher education is approached by teachers mostly in three ways; the first is staff simply perceives evaluations as mandated policies; the second is staff see formal evaluations as used for informing institutional purposes (e.g., promotion, judgment); and the third is that staff see formal evaluations as used for informing individual perceptions and developing institutional context. To Idapalapati (2019), the purpose of teacher evaluation should

be suggesting appropriate training to teachers to improve teaching in addition that it serves decision-makers to retain or terminate a teacher's contract; meanwhile, it is clearly stated in contemporary studies that teacher evaluation needs to consider other aspects such as teacher development, betterment in teaching and student achievement (Amrein-Beardsley & Holloway, 2019; Gordon & McGhee, 2019).

Many teachers at the beginning stages of their careers have difficulty expressing their competence in a new context and need assistance in order to achieve their goals (McIntyre, Hobson, & Mitchell, 2009). To attain these goals and help EL teachers become more effective, Akcan et al. (2017) draw attention to the need for the enhancement of teacher qualities and qualifications such as language proficiency, openness for professional development and self-reflection, character and pedagogical knowledge. There are a number of formal and informal ways to provide professional assistance to these educators, but feasibly two most common formal ways include mentoring and classroom observation (Hobson, Ashby, Malderez, & Tomlinson, 2009). Bailey (2006) asserts while mentoring is usually done by a more experienced teacher or by an administrator, observations can sometimes be termed as evaluative supervisions performed by peers or instructional experts and are sometimes conducted in the form of surprise visits and sometimes with prior notice to teachers; the first type is generally considered threatening to teachers (Bailey, 2006). In this sense, Moskal et al. (2016) suggest that evaluation practices and procedures seem to be more appealing to teachers as long as participation and engagement are positively supported. However, although teacher evaluation in almost all contexts is defended by administrators to serve the purpose of providing feedback and guidance for improving professional practice, in reality, the opposite occurs (Moskal et al., 2016). Goe, Biggers and Croft (2012) express in their study that teacher evaluation could become a tool to help teachers improve, yet school administrators often lack training in how to use evaluation results to guide teachers toward professional growth; however, outcomes of reliable and valid evidence need to serve the purpose of improving teacher performance and student learning, while taking accountability and improvement into consideration.

To give an example, accepting student achievement in teacher evaluation as the only criteria for teacher effectiveness should not be the ideal outcome (Little, Goe, & Bell, 2009). Al-Mutawa (1997) defines teacher evaluation as "research-based evaluation, seeing the evaluator as a collector of descriptive data on specified aspects

of the teachers' performance" (p. 42) and refers to the term as a method of evaluation of teachers by their own classroom work as a means of professional development and aiming to identify unexpected discrepancies of external evaluation. Al-Mutawa in his study aims to understand the language competencies of EL teachers in Kuwait. An external evaluation was carried out by the researcher through classroom observations, and - in addition - self-evaluation was requested by teachers using the same evaluation form to compare results. The criteria included three broad areas with sub-topics such as language level, lesson planning and implementation. The results showed that teachers mostly saw themselves more competent than the external evaluator in every criterion. Al-Mutawa (1997) suggests developing linguistic and teaching competencies of EL teachers as the criteria for recruitment, designing INSET training courses throughout the school-year to cater for the components in the study, and combining external and self-evaluation since their goals and procedures help to define the standards of evolving profession (Al-Mutawa, 1997).

Skinner and Hou (2014) aimed to find out differences and similarities between university supervisors' (academicians) and host teachers' (mentors) expectations and perceptions on teachers' teaching performances. The results revealed conflicting conclusions for two parties. First, the host teachers were realistic, but the supervisors were idealistic in practice. Second, host teachers showed a more soft-minded approach towards trainee teachers, while supervisors had a hard-minded approach. University supervisors, when compared to host teachers, also focused more on teaching abilities of the trainees than students' learning. In order to decide on shared purposes and expectations, host teachers or supervisors, if working together, need to agree on issues, which is an important factor to assess teaching performance more reliably (Skinner & Hou, 2014).

Abedi and Faltis (2015) state that teacher evaluations should be used for improvement, personnel decisions regarding payment, advancement and for better student performance. To Abedi and Faltis (2015), "teacher assessments play an integral role in instruction, placement, promotion, and efforts to ensure that students and teachers receive the support they need for success" (p. 7). By the provision of constructive feedback through diverse evaluative methods, institutions should provide continuous professional development opportunities to teachers for their own institutional goals, enhanced quality in teaching and student success (Idapalapati, 2019). The outcomes of teacher evaluations should also aim to contribute to teachers

in both cognitive (e.g., teacher's content knowledge of subject areas) and noncognitive domains (e.g., teacher's level of motivation and engagement and other psychological factors) in terms of valuable information in understanding teachers' instructional strategies and student performance (Abedi & Faltis, 2015; Borg, 2019).

2.1.2 Mode of teacher evaluation. Bi (2017) highlights a critical aspect considering effective teacher evaluation by stating that a pitfall is to rank teaching evaluation ratings from high to low because variations occur in the practices of teaching. For this reason, it is not constantly possible and feasible to draw conclusions from teacher evaluation, since many other factors play roles in addition to individual performances (Bi, 2017).

Brown (1994) underlines the use of checklist to avoid prejudicial judgments and move toward more objective ways of observation by parts of the checklist being devoted to teachers' strengths, as well as suggestions for the betterment of his or her work. However, EL teachers seem to pay significantly less attention to the feedback from supervisors after their classroom observations since the structure of the evaluative practice is judgmental rather than developmental (Bacher-Hicks et al., 2019; Moskal et al., 2015; Quirke, 2015). In Wei's study (2015), teachers reported that they further changed their teaching practices based on administrator feedback. The main reasons were that the sudden visit of a department head had made both them and their students nervous, causing them to make mistakes; the administrators observed their English classes two/three times a year without guidance or feedback; therefore, the conclusion from their classroom observations were incomprehensive; the administrators neither read their lesson plans nor understood the language proficiency level of the students. As a result, the administrators could not understand the modes of classroom activities and assessment tasks, and hence the suggestions were sometimes in contrast with the results and feedback received from student surveys (Wei, 2015). Even under the best circumstances, according to Tucker and Stronge (2005), unannounced observations (e.g., sudden visits to classrooms) as in the previous study can be artificial by nature, suggest an inspection approach to supervision, have limited validity based on the skill of the observer, are narrow in scope, and involve a small sample of teacher's actual work with students.

Kim and Danforth (2012) aimed to examine how EL teachers and supervisors perceived pedagogical practices. The data was collected by conducting interviews

and classroom observations. Classroom observations were enriched by a post-observation conference with the participants. Both parties as EL teachers and the supervisors believed that their experience of teaching, supervision, administration, EFL qualifications and trainings outweighed the other group. The findings of the study suggested that mutual dialogue was desirable to reconceptualize the roles, as well as an engagement in critical discussions with the other party. On this specific issue, the hierarchical relationship between administrators and teachers, observers and observees, supervisors and supervisees, or teacher trainers and trainees interrupt the potential dialogic and constructive supervision sessions; Kim and Danforth (2012) label this phenomenon a "top-down" relationship (p. 70). In support of such a case, according to Tucker and Stronge (2005), 99.8%of administrators conduct classroom observations for the direct source of teacher evaluation. Dependence on no more than observations or an abundance of other types of evaluation practices that pursue developmental purposes for teachers tend to produce significant problems. Tucker and Stronge recommend a balanced approach to teacher evaluation, which promotes student achievement and teachers' professional development. Similarly, Boysen, Kelly, Raesly and Casner (2014) give importance to the responsibilities of administrators during the process of teacher evaluation for the purpose of teacher development. Boysen et al. (2014) support the explicit training of faculty and administrators in order to avoid misinterpretations and transferring the greater burden of responsibility on administrators for leading the way in improving the use of teaching evaluations. Knop (as cited in Bailey, 2006) introduces three approaches to EL teacher supervision and observation; the first is *the scientific approach*, which involves competency-based education and the use of interaction analysis; the second is *the democratic approach*, which views supervision as therapy and ego counseling; and the third one is *clinical supervision*, in which the supervisor and teacher determine the goals of the observation (p. 6).

Howard (2015) concentrates on the voices of in-service teachers and investigates to what extent their input is valued and used during observation processes. The purpose is to understand if there is an "observer paradox" (Mclntyne, 1980) and what are undergone by evaluating teachers as the fundamental aims of evaluations are to meet the standards, to satisfy needs of students, to confirm teacher quality and effectiveness, and to reach individual learning outcomes both for students and teachers. Regarding the results of the study, Howard (2015) concludes that

appreciation of teachers' voices is limited, and that even when they have the opportunity to have pre-observation meetings, because of hierarchical and top-down communication, the effect on teaching is not helpful. Evaluations and observations, if not carried out on developmental purposes and systematically (e.g., unannounced observations) tend to reduce teachers' desire to communicate and receive feedback. On the students' side, they are not given the chance to provide comments on lessons observed; even when they are silent, it is assumed that everything has gone well or vice versa (Howard, 2015, pp. 207-208). Bellibaş and Gedik (2016) compared the implementations of teacher evaluation programs in a group of educational institutions both in U.S. and Turkish contexts. To collect data, they held semi-structured interviews with the teachers and the principals and compared documents they received. The results showed that an adequate supporting structure is needed for ineffective teachers, and there seems to be a need for specialists to provide continuous support to teachers. Bellibaş and Gedik (2016) conclude that an ill-defined and unsystematic implementation of evaluation systems may result in discouraging teachers and that a clearly-defined process should be preferable.

2.1.3 Process and product-oriented teacher evaluation. Wei (2015) states that as long as criteria and procedures are defined imprecisely and unclear, neither summative nor formative evaluations of teachers will result in better teaching. The reasons are basically related to three main domains; the first is inconsistent definitions of good English classroom teaching; the second is conflicting feedback from surveys and classroom observations; and the last one is unethical practices and strategies to cope with the formative teacher evaluation practices. To Wei (2015), "formative assessment does not seem to have an effect when its feedback is not helpful in improving the results from high-stakes summative assessment" (p. 619). The underlying reasons for this are that teachers have difficulty in identifying the gaps between their actual and expected performance, resulting in losing confidence and not trusting the validity and reliability of the teaching evaluation practices; they may stop the process of self-assessment (Wei, 2015). Stein, Spiller, Terry, Harris, Deaker, and Kennedy (2013) suggest that unless the provision of quality monitoring and well-structured continuous feedback are sustained in institutions, evaluation may not be willingly accepted by teachers. Support structures and rewards may also

encourage teachers to pay more attention to their individual developments as they hold institutional goals, individual beliefs, views and experiences (Stein et al., 2013).

Quirke (2015) suggests a comprehensive seven-step teacher evaluation system (i.e., ASPIRA), which is designed by the complete participation of the shareholders of a higher education institution (teachers, chairs, and managers). The system advocates an appraisal program that brings formative and summative principles into a powerful professional development resource by taking feedback and guidance into consideration. As opposed to classical teacher evaluation practices, the system supports individual and guided teacher appraisal practices. According to Quirke, teachers should be allowed to collect multiple-sourced data in the form of portfolios that are clear guidelines on the form and details on how they are to be reviewed. "A guided discussion framework between the teacher and supervisor with the pair summarizing the year's work is the approach that most teachers prefer" (Quirke, 2015, p. 103). The steps that make up the ASPIRA system are summarized as follows:

Stage 1 Goals and Reflection - Teachers specify their annual workload and responsibilities inside and outside the classroom. They submit their plan to their supervisors.

Stage 2 The first meeting - Supervisor and teacher hold a meeting on the detailed process and agree on detailed applications and outcomes. It is called 'goal review'. Then important highlights and competencies are reflected in 'teaching portfolio'. The format of the portfolios is left to teachers (hardcopy or softcopy).

Stage 3 Semester and data collection - Observations, peer observations, student evaluations, and supervisor feedback are sample sources. However, it is suggested that rather than observing teachers by supervisors which may alter the classroom dynamics, cannot avoid personal prejudices, cannot observe as a student and teaching styles, and cannot be linked to student achievement, a wide variety of self, peer, student, team and supervisor observations in formats such as audio, video, blitz, twenty-minute, full-lesson and unseen interpretations for the teacher should be used to evaluate teachers. Teachers need to have the chance to observe both teachers and supervisors. And the most crucial issue is to leave the evaluation choice to teachers.

Stage 4 The mid-cycle review - The teacher and supervisor usually meet and discuss what have been done, what data has emerged and what next step(s) can be taken during the second semester. The evaluations of self-, student, and peer-

observations/evaluations are discussed together. “Such a collaboration, including teachers, students and supervisors in reaching a consensus, has the potential to lead to an improvement in both teaching and learning” (Quirke, 2015, p. 109).

Stage 5 Semester two data collection - ‘The second cycle monitoring’ stage contains data collection through the second semester as agreed in the mid-year cycle review. Different types of materials and techniques will be used and reflected on portfolios.

Stage 6 Manager’s summative report - The summative write-up is the manager’s annual report that should be covered for every step and objective of teacher’s portfolio. This allows managers to see that supervisors know what have been done and teachers have done throughout the year. This also provides the institution a framework of the years, portfolios, and teachers’ professional development.

Stage 7 The end of cycle discussion - this meeting looks back and reviews the previous year, concludes the whole process and helps plan the next year’s program. The teacher’s own portfolio including self-reflections, student evaluations and feedback, and reference to supervisor’s report are documented to be used as a reference for further guide both to teachers and supervisors. This is referred to as a final discussion record form (Quirke, 2015, pp. 103-111).

2.2 Teacher Evaluation Models

In terms of teacher evaluation implementations, triangulated sources of systematic information gathering are vital to carry out teacher evaluations; a) academic models (i.e., value-added models, the framework model) that try to measure gains in student achievement or pinpoint factors in teacher effectiveness; b) classroom observations (by managers, other teachers or external evaluators); c) student evaluations of teachers; d) judgments made by the teacher’s line manager or principal; and e) teacher portfolios of work (e.g., self-evaluation) are assumed as the most common approaches (Bacher-Hicks et al, 2019; Darling-Hammond, 2015; Donaldson, 2009; Quirke, 2015; Stein et al., 2013).

There are basically few standards-based models of teacher evaluation that are commonly used in educational contexts. The following subtitles provide the basic definitions and descriptions of these theoretical models. And the other types of

teacher evaluation practices and strategies (e.g., classroom observations, self-evaluation) are also provided under the related subtitles.

2.2.1 The value-added model (VAM). In order to evaluate teacher effectiveness, the VAM, in basic terms, measures how a certain teacher contributes to the leaning achievement progress of their students (Jackson, Rockoff, & Staiger, 2014; Strøm & Falch, 2020). The VAMs measure the effects of teachers on student achievement and describes any analysis using longitudinal student test score data to study the effects of educational inputs on achievement (McCaffrey & Lockwood, 2008).

For applications, the background of learners' success levels and actual success levels of students are compared to each other after a specific time of teaching period (i.e., academic term), then through calculating the increased/decreased progressive test score averages of students, how effective a teacher on student achievement is decided (Koedel, Mihaly & Rockoff, 2015). The VAMs refer to "a family of statistical models that are employed to make inferences about the effectiveness of schools/teachers and attempt to extract from the data on score trajectories estimate of the contributions of schools or teachers to student learning" (Braun & Wainer, 2007, p. 867). The future of a teacher in terms of whether their contract will be renewed is decided by the estimates of the VAM scores (Hanushek, 2009, 2011). However, according to Koedel et al. (2015), "a caveat is that the long-term labor supply response is unknown" (p. 191).

Although in theory the VAMs is to allow the effectiveness of different teachers to be measured by presenting their results through student test scores, some have raised objections because the VAM does not provide a randomly-selected student-teacher match in the classroom, does not consider external factors such as school opportunities for facilities, socio-economic context of the educational setting, parental involvement, or factors out of teacher's control, and causes a ceiling effect, which suggests that a teacher who is assigned to students with already high scores or vice versa will not seem to have helped them, as student scores will stay the same (Braun & Wainer, 2007; Goldhaber & Hansen, 2013; Kane, McCaffrey, Miller, & Staiger, 2013; Todd & Wolpin, 2003).

In terms of the advantages, the VAMs may provide to measure teacher performance and school accountability and have the potential to keep track of student

success as a reflection to evaluating teachers while they lack common structural interpretation obtained from estimates (Sass, Semykina, & Harris, 2014). However, there are still problematic areas to consider while implementing the VAMs to evaluate teachers. For example, in their comprehensive study measuring the accountability and longitudinal validity of different uses of VAMs in educational settings, Braun and Wainer (2007) came to a conclusion that "given the complexity of educational settings, we may never be satisfied that VAM can be used to appropriately partition the causal effects of the teacher, the school and the student on measured changes in standardized test scores" (p. 889).

If to provide a clearer frame to the scope of VAM, according to Darling-Hammond (2015), evaluating teachers in terms of students' success or students' test scores would be problematic. A list of other influences may also intervene such as a) school factors such as class sizes, curriculum choices, instructional time, availability of specialists, tutors, books, computers, science labs, and other resources, b) prior teachers and schooling, as well as other current teachers—and the opportunities for professional learning and collaborative planning among them, c) peer culture and achievement, d) differential summer learning gains and losses, e) home factors, such as parents' ability to help with homework, food and housing security, and physical and mental support or abuse, f) individual student needs, health, and attendance.

Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012) elaborate on this specific model by stating that value-added models of teacher effectiveness are inconsistent; teachers' value-added performance is affected by the students assigned to them; and value-added ratings can't disentangle the many influences on student progress. Furthermore, Bacher-Hicks, Chin, Kane, and Staiger (2019), in a comparative study, draw attention to numerous criticisms regarding the predictive validity of VAMs to evaluate teacher performance while comparing the results of VAMs (i.e., test-based measures), classroom observations and student surveys (non-tested grades). First, the evaluation of teachers through non-tested measure types tend not to capture teachers' causal effectiveness whereas test-based measures at least provide more reliable and valid results; second, each of the evaluation types may not alone help decide teacher effectiveness unless they contain valid predictive information about teachers' performance following random assignment (Bacher-Hicks et al., 2019).

As can be understood, although the VAM has the potential to inform both administration, teachers, and even other parties about student achievement and the effectiveness of educational input, the model has still important limitations to evaluate teachers in a broad scope because the only dependence of evaluation criteria is on students' achievement test scores, no measurement can be obtained for higher-thinking skills or professional development of teachers (McCaffrey, 2010); based on student test scores, fear of negative consequences may damage teamwork and collegiality and criticizing and praising individual teachers fail to take into account the other work done by others on student achievement (Marshall, 2012).

2.2.2 Danielson's framework model. Developed by Charlotte Danielson, of the most commonly used standards-based teacher evaluation models applied in educational contexts, The Framework Model (See Appendix C) seeks to help teachers and administrators improve their specific skills and guide classroom observations to become more organized (Danielson, 1996, 2007). The Framework Model is grounded on four distinctive domains, which are divided into 26 components and 76 smaller elements, and covers the four essential responsibilities of teachers as 1) Planning and preparation, 2) Classroom environment, 3) Instruction, and 4) Professional responsibilities (Danielson, 2007, 2013).

The first domain of the Framework Model covers demonstrating content and pedagogic knowledge, knowledge of students, demonstrating knowledge of resources, setting educational outcomes, designing coherent instruction and student assessment; the second domain covers creating an environment of rapport and respect, organizing student behaviors and managing classroom procedures; the third domain includes engaging students in learning, use of discussion and questioning techniques, using assessment in instruction, and demonstrating responsiveness and flexibility; and lastly the forth domain is related to reflecting on teaching, developing professionally, maintaining accurate records, participating in a professional community, and showing professionalism (Alvarez & Anderson-Ketchmark, 2011; Danielson, 2011). According to Danielson (2014), the model follows three basic phases (See Appendix D) to conduct classrooms observations as a means of teacher evaluation:

1) Pre-conference;
    a. the observer schedules a pre-conference with the teacher,

b. the teacher provides a lesson plan,
c. the pre-conference data is recorded on the observation form,

2) Lesson plans;
   a. the teacher provides a lesson plan that includes four QSAC elements (objective, core curriculum standards, activities, assessments),

3) Post-conference;
   a. reflective conversation takes place (e.g., comments, suggestions, recommendations).

The advantages of such a standards-based teacher evaluation model include the provision of comprehensive discussions between teachers and administrators, the possibility on reflecting on specific practices, a more evidence-based feedback through use of checklist-based observations, a positive influence on student learning, and the empowerment, rather than judgment, of teachers (Donaldson, 2009**;** Heneman, Milanowski, Kimball, & Odden, 2006**;** Mielke & Frontier, 2012; Sartain, Stoelinga, & Brown, 2011). However, although this model is the most commonly used teacher evaluation model, which aims to support teacher improvement and student success (Hernon, 2019), the actual uses of the framework model do not always provide the expected consequences and may yield several complications because poor applications are implemented by untrained administrators, evaluation is more on teachers with almost no concrete beneficial effect on students' learning, and in addition, understanding of the evaluative outcomes mean differently to teachers and administrators (Donaldson, 2009; Moss, 2015).

2.2.3 The marzano focused teacher evaluation model (MTEM). Another model mostly used by educational institutions to evaluate teacher effectiveness and competency is the Marzano Teacher Evaluation Model (MTEM). Claiming teacher evaluation practices do not measure teacher quality accurately (with specific criteria) or efficiently (comprehensively), Marzano (2012) states that “an evaluation system that fosters teacher learning will differ from one whose aim is to measure teacher competence” (p. 14). MTEM (Marzano, 2007, 2014; Marzano, Frontier, & Livingston, 2011) is gathered under three basic constructs - a) routine strategies, b) content strategies, c) strategies enacted on the spot - comprising of 41 individual items representing each title in a comprehensive approach with specific aspects (See Appendix F). According to Marzano (2007), *comprehensiveness* refers to all the

elements that are associated with research-based student achievement; and *specificness* refers to the model identifying classroom strategies and behaviors at individual level. Marzano (2013) states the model is planned to correlate with student academic achievement and includes four domains as; Domain 1: Classroom Strategies and Behaviors, Domain 2: Preparing and Planning, Domain 3: Reflecting on Teaching, and Domain 4: Collegiality and Professionalism (p. 4). The four domains include 60 elements: forty-one in Domain 1, eight in Domain 2, five elements in Domain 3, and six elements in Domain 4 (Marzano, Frontier, & Livingston, 2011), which can briefly be summarized as follows:

Domain 1 focuses on *pedagogical strategies* that have a direct link with student achievement. Domain 1 addresses what teachers do in the classroom. The model is based on the premise that each lesson has distinct characteristics, routines, and processes. A model built to support teachers as they develop their skills must necessarily reflect the complexity of their work. But not all of the 41 elements need to be, nor should be, observed in a single lesson. Domain 1 breaks down teaching into "thin slices" for richer diagnostic and feedback purposes.

Domain 2 focuses on *planning and preparing* for units of instruction and lessons within units. Because these elements are directly related to Domain 1, the better a teacher prepares, the more effective are his or her instructional choices.

Domain 3 addresses deliberate practice. It encourages teacher self-reflection in the areas of evaluating personal performance and developing and implementing a professional growth plan. When teachers receive specific and focused feedback using a common language of instruction, they increase their expertise and, subsequently, student performance.

Domain 4 is the backdrop for the other domains and encourages a supportive culture. It addresses *collegiality* and *professionalism***,** emphasizing opportunities to observe and discuss strategies. This domain supports teacher participation in lesson study, instructional rounds, teacher-led professional development, and professional learning communities in which teachers collaboratively examine evidence of student learning and the impact that specific instructional strategies have on learning (Marzano, 2013).

According to Marzano (2013), the types of classroom observations (See Appendix E) can be implemented within the structure of various modes and implementational purposes; firstly, classroom observations have two modes as the

announced and the unannounced; in addition, the purposes are four as the formal or the informal ones, and the targeted observations or the walkthroughs. Formal observations are conducted only by pre-informing the teachers as announced ones. Informal observations are both announced or unannounced, which mostly last for 10 minutes long and are used for annual teacher evaluation. The targeted observations are announced and planned including feedback to be used for specific practices. The walkthroughs are unannounced observations, which usually last for 3-10 minutes long and are again used for annual teacher evaluation (Marzano, 2013). Concerning the fundamental purpose(s) of teacher evaluation, Marzano, Frontier, and Livingston (2011) took the opinions of more than 3,000 teachers in order to find as to whether teacher evaluation should serve either for teacher competency measurements or for teacher development or for the both. The results of the study conducted by Marzano et al. (2011) revealed that the vast majority of respondents (76%) believed teacher evaluation should be used for both measurement and development but that development should be the more important purpose. Although the study by Marzano et al. (2011) suggests that MTEM seems to consider teacher competence and professional development, as well as student achievement, this teacher evaluation model lacks a comprehensive scientific proof because there are few studies with large-scale samples that assess the correlations between teacher observation scores and student achievement using MTEM (Basileo & Toth, 2019).

2.2.4 The marshall principal evaluation rubrics. Another commonly used teacher evaluation model is The Marshall Principal Evaluation Rubrics (Marshall, 2011). This evaluation model (See Appendix G) is mainly based on classroom observations using scoring rubrics and grading scales (i.e., effective to ineffective; 1 to 5) and aims to get principals to make enough classroom visits to see daily reality, to ensure every principal has a good eye for instruction, to polish principals' skills at giving feedback to teachers, to decide how and when to use the rubric, and to keep student learning at the center of supervisory conversations (Marshall, 2013). Applied by school principals or teacher trainers in the institutions, "these practices will give teachers a stronger voice, use principals' time more effectively, and make teacher evaluation a real player in dramatically improving teaching and learning" (Marshall, 2012, p. 53).

Marshall (2017) supports the idea that observational practices should not follow the commonly-accepted procedures such as conducting a pre-observation conference with the teacher, observing a full class taking detailed notes, writing-up the evidence and conducting a post-observation conference because they are time-consuming, announced in advance which does not show real conditions in classroom, and the detailed feedback of them are overwhelming, unhelpful and dishonest to stakeholders. Instead, in order to evaluate teachers, Marshall (2013, 2017) suggests conducting 1) unannounced, frequent, short-timed observations to classrooms; 2) obtaining observation skills to give honest feedback; 3) detailed and supportive written feedback following observations; 4) the use of rubrics, items of which are agreed on between the teachers and the observer at the beginning of the year; and 5) check of student progress without waiting for the student achievement scores obtained at the end of the year (Marshall, 2013, pp. 4-5). However, Marshall (2017) himself declares that this model lacks research-based evidence in terms of its effectiveness and accuracy as it has not been applied in large-scale academic studies yet.

2.2.5 The 4Q approach model. Smith's (2008) 4Q approach model is another comprehensive teacher evaluation model which draws on four major sources of information under the theme 'evaluating and enhancing teaching' (See Appendix H). Smith's (2008) model is made up of four distinct categories to collect data for teacher evaluation; the first is self-reflection which focuses on responses to profession feedback, peer feedback, student feedback, self-evaluation of teaching, teaching portfolio and continuous professional learning; the second is peer review which focuses on learning outcomes, constructively aligned subject, teaching and learning strategies, course content and materials, teaching performance and teaching management; the third is student learning which focuses on student assessment results, student learning, progression and completion, student reflection upon knowledge, skills and dispositions, and quality of student portfolio work; and the last one is student experience which focuses on student focus groups and experience survey, student feedback on university resources and services, student engagement in learning communities and student complaints data.

Smith (2008) states the 4Q approach allows teachers to be more systematic about their analysis of their own teaching effectiveness, which is personally useful

for promotion and meaningful because it improves and enriches an understanding of teaching experiences. Moreover, Smith believes that the variety of different data sources and types it draws on gives a more comprehensive picture of the total teaching performance than any one of the methods, techniques or resources listed would have given alone. Smith (2008) provides a series of dimensions of the practice that must be a typical teacher evaluation instruments such as clear aims and objectives, appropriate amount of work, appropriate assessment and clear criteria and standards, clarity of explanations, promotion of deep learning/understanding, lecture/tutorial management, motivating of students through encouragement, sequence, integration and organization of content for coherence, teacher's knowledge of field, teacher-student interaction, and overall ratings item.

All these proportions are suggested to be implemented by *the Action Learning Cycle*, which is integrated into the 4Q model of Smith's teacher evaluation model (See Appendix I). Smith proposes a cycle for teacher evaluation which first starts from 1) teaching, 2) observing, 3) evaluating, 4) reflecting, 5) analyzing, 6) planning, 7) (re)designing, and lastly 8) acting; and afterwards, the cycle starts from the beginning (i.e., teaching). This evaluation cycle is carried out under the four main sources of information (i.e., evaluating and enhancing teaching) and the data obtained out of the 4Q model.

2.3 Teacher Evaluation Strategies

Having discussed the standards-based teacher evaluation models, there now needs to be a place for several strategies and techniques as well as research-based reflective practices that are implemented for the purpose of teacher evaluation. The following subtitles will discuss the evaluative strategies and their relation to teacher competence, effectiveness and development.

2.3.1 Classroom observation and teacher evaluation. Donaghue (2015) describes observation-based teacher evaluations as primarily observed lessons followed by feedback sessions, which should aim to discuss the lesson and design as well as how teaching standards can be improved. Classroom observations, according to Borg (2018), are one of the commonly used practices to measure teaching quality and monitor what is done in classrooms. According to Randall (2015), due to the fact

that among many observational practices and documents provided in academic articles teacher appraisal and evaluation are major uses of observation, there is a need for an observation and counseling framework in the form of a feedback-directed counselor and client relationship for EL teacher-training contexts. However, observation and feedback practices are mainly evaluative when implemented by managerial staff even though those should include critical friends, trainers and managers with interaction to each other, and the main purpose needs to yield helpful guidance to teachers (Marzano & Toth, 2013).

To evaluate teachers' performance and help observational procedures become objective, Randall (2015) recommends the use of checklists and instruments (codification, recording and counting) by managers and administrators and asserts the most complete accounts of observation are video and audio recordings as they open doors to more detailed feedback. Although such procedures are objective evidence, implementation is a problematic issue due to unqualified teacher trainers and managers as well as to crowded schedules of teachers and managers (Darling-Hammond, 2015; Randall, 2015). In addition, evaluation practices of observation can be devastating for teachers as Watson-Davis (2009) describe them as 'a poke in the eye with a sharp stick' and state that inappropriate observation schedules, extensive paperwork, marking ticks on a form or presence of observer may cause unnatural classroom environment. Observer reliability in the forms of inter-rater reliability - different observers grading the same lesson - and intra-rater reliability - the same observer grading several cases by the use of the same criteria - should be successfully maintained if to pursue effective outcomes for teacher evaluation (Ho & Kane, 2013; Pianta & Hamre, 2016).

Classroom observations are management driven set of practices, and if implemented with a purpose in which clearly described feedback is provided, they can play an important role in promoting EL teachers' professional development while helping to keep track of what is going on in classrooms in terms of teaching and learning (Howard & Donaghue, 2015). Kane and Staiger (2012) define ideal classroom observation as a uniqueness to define positive modes in teachers' instructional practices and name specific weaknesses that they may have. According to Wermke and Forsberg (2016), in teacher evaluation contexts, one may not realize how teachers or administrators deal with classroom observations without thinking how these practices relate to teacher autonomy, and such legislated educational

practices have the potential to limit teacher autonomy in critical consequences. To give an example, De Lima and Silva (2018) investigated how teachers and department heads perceived and experienced the implementation of a classroom observation system in a teacher evaluation context. To collect and analyze data, a survey for teachers and department heads was used, and in-dept semi-structured interviews with administrators who were responsible for classroom observations were held. The study surveyed 96 teachers and 15 department heads and the results showed that classroom observations were reported an inadequate teacher evaluation instrument; such a teacher evaluation system did not contribute to teachers' professional development; on the other hand; the department heads had dilemma between their management roles as teacher observers and professional roles as teacher peers; cultural norms of individual professional autonomy which design relations among colleagues prevented the department heads from using information, which were gathered through classroom observation, to nourish in-depth professional discussions about concrete educational practices; and lastly the teachers perceived the complete process negatively as the applications were seen as a threat to their professional autonomy (De Lima & Silva, 2018).

Lasagabaster and Sierra (2011) analyzed a wide range of EL teachers' attitudes toward classroom observations. 185 school teachers completed a questionnaire which focused on the role of observations in language classrooms. Although the majority of teachers expressed positive feelings about observations, they still felt unsecure about their experiences when observed. To the teachers, this insecurity stemmed from the observer's belief and they gained almost no benefit from the observations. Another problem was that observations did not provide much means for improvement. The observer and the observed did not work in coordination while preparing projects - and that was another problem. Lasagabaster and Sierra (2011) suggest that if shareholders in an institution work in cooperation, the assessment and evaluation process could be more ideal and other teachers may want to take part in activities related to observation. On the other hand, to Rantanen (2013), courses assigned to teachers should be properly suitable to teachers' personalities and teaching abilities if they are expected to teach better and evaluation is expected to generate positive results. Rantanen states assignments given by matching the most suitable course with the right teacher will give tools for administrators to improve the quality of effective teaching. Taking diverse teaching styles and the various personalities of teachers into

consideration are also regarded as significant dimensions that administrators need to keep in mind (Rantanen, 2013). In their study, Taylor and Tyler (2012) found that it is also possible to change teacher effectiveness with a long-lasting improvement of teacher skills. Taylor and Tyler applied a one-year-long observation-based teacher evaluation program and worked with EL teachers who were in their mid-careers. By using a quasi-experimental analysis method, the results of the study showed that a teacher who is evaluated throughout the complete academic year might become more productive and the productivity might increase as years go by; if a teacher learns new information about their professional growth during the evaluation period, this supports him/her develop new profession-related skills; long-lasting implementation increases teacher effectiveness via improving educational production; and in addition, such a well-structured evaluation system can serve as an identification mechanism for selective termination of the lowest performing teachers (Taylor & Tyler, 2012).

Devos (2014), in terms of more specific implementations, suggests six P-dimensions of observation to be employed in EFL classrooms - namely, 1) purpose, 2) predicate, 3) perceived parties, 4) participants, 5) profiteers, and 6) products; and - following these - observing for developing, observing for understanding, and observing for improvement as the fundamental outcomes of the previously mentioned dimensions happening between the observer and the observed (student teachers, novice teachers) or between peer observers. Devos (2014) emphasizes the ability to benefit from mutual understandings, stating that “it should be made clear to the observed person what the observer would like to observe, and the observed should also include items that they would like the observer to watch in order to improve” (p. 26). While delivering trainings over observations or guiding EL teachers to take action in self-reflection, Shoffner (2009) suggests triggering the affective domains on teachers such as establishing positive relationships, helping teach meaningfully, and improving personal impact on learners if to consider effective self-reflection.

In terms of the steps that ought to be taken in classroom observations, in order to provide a list of weak examples which may lead to misunderstandings and discouragement on the observees' side, Iyer-O’sullivan (2015) provides the following commonly asked questions for pre-observation sessions, which are considered discouraging:

- This is an interesting source. How are you going to use it?
- Do you think pupils might find the activity too easy/difficult?
- Why did you arrange the exercises in this order?
- Do you think pupils will have enough time to complete this activity?
- Why did you choose these activities?
- What steps will you take if this class is not able to do this activity?
- What strategies have you planned for the difficult pupils? (pp. 72-73).

The following list includes commonly asked, yet discouraging questions for post-observation sessions, which may also be seen as much problematic just as the pre-observation questions listed above:

- So... How do you think that went?
- What do you think went well?
- What would you do again?
- What did not work so well? Why do you think that?
- What would you do differently?
- Are you still confident and anxious about the same issues? Why? (Iyer-O'sullivan, 2015, pp. 72-73).

Iyer-O'sullivan (2015) emphasizes the importance of dialogic mediation and suggests observers/mentors not to aim to control the whole dialogue, but rather to encourage observees/mentees to initiate the conversation and talk about any part of the lesson. To give examples, as suggested by Iyer-O'sullivan (2015), the following questions and statements can be optional for pre-/post-observation sessions:

- What would you like to talk about?
- Which part of your lesson would you like to talk about?
- Please tell me about your lesson.
- Please feel free to talk about any aspect of your lesson (p. 80).

Overall, in order to sustain a more comprehensive teacher evaluation through classroom observations, Copland (2015) suggests the application of post-observation feedback sessions, which are offered in CELTA (Certificate in English Language Teaching to Adults) programs as follows:

- *Self-evaluation:* Trainee discusses their own lesson and highlights strong and weak aspects. Turns are generally short,
- *Trainer feedback:* Trainer critiques trainee's lesson with positive and negative evaluation, and provides advice and suggestions,

- *Peer feedback:* Trainees provide feedback to their peers on the strengths and weaknesses of the peer's lessons. Turns, if long, tend to provide a lot of descriptive comment,
- *Questioning:* Trainers ask trainees a series of questions about particular sections of their lessons,
- *Clarification:* Trainers and trainees talk about things not directly relevant to the lesson taught, for example, assignments. This is the only phase that is often trainee-initiated (p. 137).

2.3.2 Peer observation and peer evaluation. Hendry, Bell, and Thompson (2014) define peer observation as a form of observational learning in which an educator watches another colleague's teaching without necessarily judging their practices or being required to pursue summative evaluation. “Peer evaluation is the process through which a teacher is assessed by a colleague rather than by a school leader, line manager or external evaluator” (Borg, 2018, p. 32). Rather than that, according to Rayan (2013), peer observation will undoubtedly yield developmental results provided that the observer and the observed have a positive attitude and mutual trust towards each other and pre-decide on their lessons objectives together with the positive aspects of the teaching session. Teachers discuss the minutes of the observation together after all with critical eyes, and such a process is named peer observation of teaching (POT): (i) Pre-POT, (ii) POT, and (iii) Post-POT; it should follow a collaborative process and be formative in nature while the common goal should be achieving professional development and helping teachers to become more reflective (Rayan, 2013). According to Cosh (1999) “in a reflective context, peer observation is not carried out in order to judge the teaching of others, but to encourage self-reflection and self-awareness about our own teaching” (p. 25). For Motallebzadeh, Hosseinnia and Domskey (2017), peer observation is a technique by which colleagues agree to observe each other’s classes and, by doing so, help each other to develop teaching, solve problems, share their opinions, and exchange mutual support. In their study with Iranian EFL teachers, Motallebzadeh et al. (2017) aimed to investigate if peer observation as an evaluative and reflective tool could influence EL teachers’ professional development. The results showed that peer observation plays a significant role in improving teachers in circumstances in which desirable facilities such as financial aid are provided by the policy makers. And mutual

reflection may help teachers develop teaching strategies and understanding the perspectives of other teachers (Motallebzadeh et al., 2017). Furthermore, in terms of the effects of peer observations, just the experience of witnessing others first-hand is helpful in gaining and developing in teaching skills. Educators should observe each other and express their ideas in a way that help each side to appreciate differences (Murphey & Yaode, 2007).

Having been discussed, the notion of classroom observations often conjures up negative feelings of evaluation for many teachers and are thought to be in line with threatening or negative experiences (Devos, 2014; Richards & Farrell, 2005). Devos (2014) suggests the implementation of non-evaluative observations (e.g., peer-observations) more frequently, since they are the types that are better welcomed by teachers and foster personal, as well as professional growth. Richards and Farrell (2005) assert that observing another teacher has the potential of triggering reflections about one's own teaching and providing an "objective" point of view of the lesson; the observed teacher has the chance to collect information about the lesson that he/she might not otherwise be able to gather. To Head and Taylor (1997), peer observations are supportive in nature rather than evaluative, and so such practices should lead teacher to learn from and support each other. Trotman (2015) similarly states that "peer observation involves monitoring a lesson or part of a lesson given by a colleague in order to gain an understanding of a specific aspect of either teaching, learning or classroom interaction" (p. 181).

In a study conducted with six EL teachers, Yook and Lee (2016) investigated how teacher development methods could be effective on both teacher's self-evaluation and individual teacher identity. One of the findings showed that, based on teacher interviews, theory-grounded education was ineffective and had no benefit for the teachers. Instead, peer observations offered them a systematic way to provide more opportunities to observe other teachers in action without carrying any pressure for teacher evaluation or summative feedback (Yook & Lee, 2016). A reflective peer evaluation study, which Trotman (2015) conducted with 12 English teachers at a state university in Turkey, resulted in that EL teachers prefer peer observation sessions as a professional development tool when compared to being observed by a trainer with a checklist for judgmental evaluation purposes.

2.3.3 Self-observation and self-evaluation. According to Copland (2015), while peer-evaluation facilitates observational skills, distributes turns in the feedback conference, or supports self-development, it may be criticized for putting trainees in a critics' position. Copland draws attention to the concern of self-driven feedback in self-evaluation and states it is an important skill to learn how to reflect on actions and to develop as professionals where strengths are identified and retained while weaknesses are recognized and advanced. In this regard, Donaghue (2015) in a study points out the significant potential of self-feedback, reflection, and professional growth through the analysis of the audio recordings of post-observation meetings and the interviews with the supervisors. The aim here was to discover if the evaluative side of meetings had either prescriptive or collaborative style, and was to understand how the case of evaluation might end up both for the teachers and the supervisors. The teachers, in such a case, had the risk of failure or loss of job; on the other hand, the supervisors might have either the responsibility of advisors or evaluators. The result of Donaghue's study was that both supervisors thought they were more open to discussion and acted as they were supporting teachers. However, the analysis of data showed the opposite. Both supervisors always had the control of the meetings with the teachers and dominated the style of teaching, which teachers were supposed to do so. They believed they provided constructive feedback; however, they in reality evaluated teachers by the evaluation criteria that the institution imposed on them. Overall, Donaghue (2015) expresses the importance of constructive feedback to foster self-reflection because observation and feedback processes need to be maintained in order to identify if institutional demands are limiting or changing the way supervisors give feedback; in addition, reducing the number of classroom observations and introducing peer-observations or self-observations with specific teachers may provide an alternative; institutions need to review observation instruments and their use in feedback as these artifacts can influence the way feedback is given, at worst by dominating proceedings, leaving little space for trainees to talk or reflect; and teachers need to have the opportunity to develop and reflect on their expectations of the feedback meetings.

In terms of longitudinal reflection-driven practices, Murphey and Yaode (2007) conducted a portfolio-based self-evaluation project in China, pointing to such portfolios as one of the primary tools for teacher development as - through these - teachers could see their own professional lives reflected in them, offering a means of

exhibiting a teacher's strengths as well as weaknesses and general reflections for self-evaluation. On the administrators' side, it is highlighted in the study that *appreciative inquiry*, which can be defined as teacher trainers' or administrators' concentration on strengths rather than on areas for inspective evaluation, increases the motivation of teachers and leads to increased effectiveness. In a parallel manner for self-directed outlook, the 5R framework developed by Bain, Ballantyne, Mills, and Lester (2002) allows for critical reflection on experiences as implementing five fundamental domains to enhance better teacher evaluation in personal and professional development. The 5 domains include *reporting* the context of the experience (what happened? what was done?), *responding* to the experience (personal observations, feelings, thoughts), *relating* the experience to knowledge and skills one already has, *reasoning* about the significant factors/theories to explain the experience, and *reconstructing* one's practice by planning future actions for a similar experience (Bain et al., 2002). As it is clear that practices which foster self in the forms of observation, reflection and evaluation help teachers monitor their own weaknesses and strengths through self-observation and reflect upon their own effectiveness on teaching through development-based evaluation practices.

2.3.4 Mentoring for teacher evaluation. Mentoring refers to the practice of a more experienced teacher offering advice to a lesser experienced one to facilitate the growth and learning of another (Malderez, 2001); it helps teachers raise awareness about and adapt to their roles as teachers (Schwille et al., 2007). Mercado and Mann (2015) point to the importance of mentoring in teacher evaluation and draw attention to the roles of mentors as facilitators of reflecting and self-monitoring, negotiation of meaning, interaction of knowledge or experience, and internal establishment of personal development for teaching. Mentors facilitate self-evaluation through reflection, support action research/reflective practice, have a key role to develop insider view and reflective thought on novice teachers and sustain the balance between an enlightening approach and meeting standards (Mercado & Mann, 2015).

Mentors are expected to spend time with mentee teachers in formative, developmental, and supportive ways rather than following only quality assurance and performance evaluation, and "are the providers of documentation, tools, resources, information, and most importantly encouragement and support" (Mercado & Mann, 2015, p. 48). To this end, mentoring seems to create a shared platform in which

mentees and mentors work together as a result of dialogic and mutually beneficial relationship. Similarly, in a study aiming to find a framework for assessing mentoring quality, Arnold (2006) points to the strength of the relationship between the mentor and the mentee. The best mentors are those who share their mentees' sense of humor and give utmost importance to their feelings and experiences. Arnold points out that the language used by the mentor also affects the quality of the mentoring process.

In most educational contexts, experienced teachers are assigned as mentors to novice/beginning teachers in aim to provide guidance on being analytical, improving their teaching, teacher knowledge and competence as well as student success (Norman, 2011; Pylman, 2016); however, the opposite occurs and mentoring is limited to emotional support and advice-sharing rather than developing analytical rationales on decision-making for teaching (Stanulis & Bell, 2017). In order to avoid such an insufficient process, Stanulis and Bell (2017) suggest the Attentive and Targeted Mentoring (ATM), the domains of which are a) *valuing the voice of beginning teacher* through including concerns, questions and teaching goals; b) *targeted* by having a clear focus for improvement that is of instructional consequence; c) *mentoring* by modeling a co-learning stance, using data from student work, observing, videotaping as a basis for mentoring conversations and discussing a vision of effective teaching the pair is working toward. Lai (2010) summarizes the concept of a comprehensive mentoring program 'Guidelines on Mentoring' by stating experienced teachers need to take the role of mentors out of good will through providing formative feedback and frequent support (i.e., observations, pre and post lesson conferences), and mentors should be an adviser, a role model, a critical friend, a collaborator by which goals and outcomes become effective and self-improving for teachers with good level of knowledge in reflection. Similarly, for a well-designed mentoring program, Mercado and Mann (2015) exemplifies the sequence for collaborative mentoring process by maintaining a collaborative action plan (lesson plans and learning outcomes) designed within dialogic mediation for continuous development as follows: 1) Pre-Observation: Establishing needs and concerns, 2) Formative Class Observation: 'Off the record', 3) Action Plan: Setting goals and objectives, 4) Implementation: Classroom observation(s), 5) Assessment of final outcomes.

Concerning implementational procedures, Lai (2010, as cited in McIntyre and Hagger, 1993), similar to the program suggested by Mercado and Mann (2015), clarifies the baselines of a university-initiated mentoring program for ELT context under three domains; the first is *Relational Dimension* that creates a relationship between the mentor and mentee to achieve a personal transformation, the second is *Contextual Dimension* that brings together the specific context and learning needs to fit in the organization and culture of particular community, and the third is *Developmental Dimension* that develops a dynamic and reciprocal communication for personal and professional development of both.

Freeman (2004) expresses the importance of mentoring that does not limit the process to only identifying weaknesses but also to the strengths and such a perspective provides benefits to professional development. Under the guidance of Dewey's inspirational ideas, experiential learning and reflective thought are key elements needed to avoid routine teaching and learning (Freeman, 2004) while other dimensions, such as reflection, self-monitoring, and self-reflection are suggested (Burns, 2010; Roberts, 1998). Regarding the roles of mentors, Mann (2005) points to a collaborative mentoring process keeping the importance of letting teachers self-evaluate and reflect. Mentoring in teacher evaluation should mostly take place under structural processes such as orientation meetings, subject-related collaboration, developmental workshops, shared workloads, and team teaching (Ellis, Alonzo, & Nguyen, 2020; Norman, 2011).

Skinner and Hou (2014) searched for the university supervisors and host teachers' perspectives of novice teachers' teaching performances. The results showed that host teachers and supervisors, by sustaining a positively affective atmosphere with the novice teachers, played a key role in assessing teaching practice and performance in a more reliable partnership (Skinner & Hou, 2004). In Peru, according to Mercado and Mann (2015), the Instituto Cultural Peruano Norteamericano (ICPNA) has a system of mentoring aiming to enrich teacher evaluation and provide insights into to teachers' performances, levels of progress, teacher needs, degree of professionalism, and motivation to continue with development. Academic supervisors in the program act as mentors by providing opportunities to prospective teachers to teach real students under systematic guidance. 45 hours of seminars are offered, alongside 27 hours of classroom observations over a ten-day period. Following this, every month prospective teachers

observe mentors' classes each day, take notes, discuss issues and plan the lessons of their mentors. The observation objectives are also set beforehand. Teachers read theoretical articles and books to reflect on and learn, implement micro teaching lessons, and then design their own authentic classes and teach real students. Teachers Exchange ideas with each other (peer-co-learning, discussion circles) while they are expected to write reflection papers for the evaluation of the whole process and self-evaluation, as well (Mercado & Mann, 2015).

In terms of individual-sampled studies, Johnson (2003) conducted a case study which narrates her own mentoring experience to a non-native English-speaking teacher at a large U.S. university in the Midwest. Through her observations, it is stated that the influence of personal values on teacher identity are significant enough to be researched more. Accounts from other teachers may help a lot in improving teaching, and allow others to analyze and evaluate their own experiences in more depth (Johnson, 2003). Bullough (2005) conducted a case study in which a secondary school teacher and two prospective teachers were assessed by mentoring. The aim was to expand their identities as a teacher by mentoring. To collect data, the researcher used weekly e-mails, interviews, a mentoring log, and mentor seminar transcripts. The mentor took part in the study both as a teacher and an equipped mentor. And the conclusions were that as mentoring programs are educational and interpersonal, required support given by managers plays an important role in enhancing teacher identities. A mentor's sense of self will develop, and teachers will gain a lot if the identity is formed and people included in a mentoring program are closely connected to each other. Mentoring programs can also help develop specific teaching skills and help mentors to feel more assured in their work (Bullough, 2005). Similarly, in her study aiming to measure the validity and reliability of the mentor role inventory in Turkish context, Koç (2011) states the most important factor teachers expect their mentors to adopt is having relevant competence to make teacher candidates qualified enough at implementing effective teaching strategies. The participants who were in their final year at a teaching training program offered by a Turkish university drew attention to mentors' knowledge about the physical set-up of school rules, being well-equipped in providing support, having field experience, and acting in a less evaluative manner. These were considered important for an effective mentoring (Koç, 2011).

Yavuz (2011) contributes to the context of mentoring in Turkey with a study focused on its potential negative outcomes, using results gained from six ELT undergraduate students and a relatively inexperienced teacher as mentees. The mentees stated it was difficult for them to transfer what they had previously learnt into practice. At this point, the mentor played a vital role in addressing their concerns with feedback to ease the transferring process of the knowledge to teaching procedure. The mentees also highlighted that when a problem came up, they had some difficulty in coping with it. In response to this, the mentor had the role to teach certain strategies, such as effective classroom management and effective teaching (Yavuz, 2011). In another study, which was conducted with eleven voluntary in-service EFL teachers working as EL teachers in a foundation university in Turkey, Dikilitaş and Mumford (2016) mentored the teachers to conduct a collaborative teacher research. Dikilitaş and Mumford (2016) concluded the study by pointing to several suggestions as a) *Moral Support,* which refers to overcoming a lack of confidence and reassurance in the face of intellectual and methodological challenges through regular meetings, b) *Pedagogic Support,* which involves guidance on research article structure, c) *Language Improvement,* which is collaborative feedback between mentees, d) *Democratic Dialogue,* which is power of negotiation leading to new insights, e) *Sharing the Research Workload,* which relates to ease of time and collaboration on the workload, and f) *Drawbacks of Collaboration,* which refers to dissatisfaction with a partner's lack of commitment.

An experimental study conducted by Thompson and Schademan (2019) disclosed the in-class experiences of co-teaching pairs (mentee and mentor teachers), specifying both evaluative and developmental consequences of mentoring for EL teachers. The first is *Negotiating the difference* referring to collaboration depended on getting to know one another personally and continued with their mutual interest in engaging their different positionalities; the next is *Sharing authority* which was enacted in linguistic, physical, and interactional spaces of co-teaching practice and the language of "co" denoted a non-hierarchical, collaborative relationship; the next is *Co-mentorship* referring to shared curricula and activities that they typically used to teach content; the next is *Coaching in the moment* referring to instructional conversations carried out in front of students in a formative assessment manner, collaboratively produced and purposefully focused on understanding and improving student learning, and besides modeling for students how to collaborate productively

with others (Thompson & Schademan, 2019). Thompson and Schademan (2019) report that "two teachers engaged in resolving the many inherent contradictions around sharing a practice. Once they did so, they were better able to evolve their own practice and direct their attention toward student thinking and learning" (p. 7).

Wallace (1991) similarly calls for a collaborative approach involving bringing data back to the feedback conference for discussion and for the employment of counseling skills in exploring and facilitating learning. This represents more of a proactive than a theory-driven approach, which requires the use of instruments to focus on micro-teaching both for individual reflection and a common factor in all approaches (Randall, 2015). Randall (2015) embodies three wide-ranging approaches to counseling and feedback in mentoring for teacher evaluation; the first is *the behavioral approach* involving exploration, new understanding and action with training and control of behavior through positive feedback, which is common in training and inspection contexts and is highly prescriptive; the second is *the cognitive/behavioral approach* which focuses not only on how to perform better and what has happened but also on seeking guidelines for future practice and planning; and the third is *the humanistic/person centered approach* which focuses on feedback sessions to provide a platform for teachers to analyze themselves individually and arrive at new understandings in line with non-judgmental guidance, empathy, and active listening.

As can be understood from the analyses of the studies and literature which cover mentoring to assess and evaluate teachers, it is so crucial in order to create an environment where mentees can advance further and improve themselves in terms of both affective domains and academic capabilities (Hudson, 2013). Mentors, as facilitators of all steps and as role-models for implementational practices, ought to overtake responsibility and help ensure advancement is ongoing. The collaboration of both parties undoubtedly provides countless benefits. For mentoring programs, in addition to emotional support and effective communication, mentors are expected to represent role-models and support teachers in terms of pedagogical knowledge, content knowledge, and student characteristics (Achinstein & Davis, 2014; Ball, Thames, & Phelps, 2008; Hudson, 2009, 2010). The process should also encourage continuous professional development for both mentors and mentees (Masters, 2009); this ought to be a two-way relationship that fosters mutual development (Rush et al, 2008).

2.3.5 Research-based teacher evaluation. According to Johnson (2009), EL teachers today are required to tackle a number of issues, which include the professionalization in a complex socio-economic and technological world, gaining an understanding the relationship between student learning and teacher's professional training, resisting the politics of accountability that are shaping global educational policies, and being able to empower themselves to create learning opportunities that are sound, contextually appropriate, and equitable for students. In addition, terminologies such as *teacher development* and *professional development* are seen continuous developmental practices and applications for in-service teachers (Head & Taylor, 1997; Johnson, 2009; Thornbury, 2006).

Regarding teacher evaluation, Tarhan, Karaman, Lauri, and Aerila (2019) point to the importance of on-going personal and professional development of teachers as the base for evaluative practices while expressing that teacher evaluation needs to be purposeful with an expected product. However, to Darling-Hammond et al. (2012), many practitioners, researchers and policy makers come to an agreement that teacher evaluations fail to help teachers contribute to self-development or student learning. Therefore, it is clear that teacher evaluation is highly connected with teacher development, and providing descriptive explanations to this specific issue would be a need to provide insights to this connection.

First of all, Er, Ülgü and Sarı (2013) point to a distinction between teacher-related terminologies for teacher education and teacher training; "the former is concerned with theoretical issues whereas the latter is concerned with practical ones" (p. 43). Richards and Farrell (2005) state that training is related to development, and continual training of teachers is important even after their period of formal training is over. To Richards and Farrell, a distinction again should be made between teacher training and teacher development; teacher training generally focuses on understanding basic concepts and principles and applying them in the classroom. Some issues of importance include using effective strategies to open a lesson, adapting the textbook to match the class, implementing group activities effectively, questioning effectively, using classroom aids and resources and providing effective feedback; while development, on the other hand, refers to a teacher's longer-term growth in terms of understanding themselves as a teacher in addition to their profession as teaching and involves the examination of different aspects of teaching and reflective review of what happens in the classroom (Richards & Farrell, 2005).

For Yook and Lee (2016), providing practical in-service courses to enhance EL teachers' individual proficiency and exposing them to methods which are easily-applicable to the teaching of four skills is needed. Uzun (2016) discusses the effectiveness of a teacher training program in the ELT department of a Turkish university. Participants as prospective teachers in the study stated that their opinions, needs and expectations were not taken into consideration in the formulation of a teacher training program, thus rendering a number of classes both ineffective and useless. Courses which allowed participants to create and develop their own materials, which were applicable to their own classroom settings, were the most contributory to their professional life. According to Uzun (2016), meeting teachers' professional expectations and satisfying their academic needs were of paramount importance. For this reason, Johnson (2009) states that the provision of constructive feedback and guidance, the sharing of mutual knowledge, self-directed, collaborative and inquiry-based learning and additional combination of theory and practice in teaching under the inspiration of Socio-Cultural Theory (SCT) will undoubtedly provide countless benefits to teachers in terms of improvement. On this approach, for instance, Lantolf and Thorne (2006) put much emphasis on SCT, describing it as "the theory of mediated mental development and most compatible with theories of language with focus on communication, cognition and meaning rather than on formalist positions that privilege structure" (p. 4). Rather than short-termed (e.g., daily presentations) seminars and in-service workshops/training carried out by institutions, there is a need for continuous and longitudinal professional development practices in which teachers can gain the opportunities to enjoy the experience of hands-on practices and continuous reflection for self (Kumaravadivelu, 2006). Similarly, Engin (2015) states that the Vygotskian view of SCT emphasizes two dimensions; the first being the social level - the dialogic nature of understanding between two parties, and the second, the individual understanding of self. According to Engin (2015), such a notion is vital to teacher training contexts since most teacher training activities are highly social and interactive; furthermore, interactions involve scaffolding the development of teachers, which is the help provided to teachers in the context of tutorials by a more able teacher in educational settings. To conclude, Er et al. (2013) provides an overall frame of the necessary steps for EL teachers to follow; the first step is administrators responsible for decision-making need to be convinced of the importance of teacher education, the second is teacher education is an integral

component of the quality assurance in language teaching evaluation and should be through employing bottom-up strategies, the third is EL teachers need to learn how to cooperate for best practices in learning environments, and lastly professional development should be an essential evaluative criterion motivating the teachers to take part in continuous professional development programs willingly. Keeping such criteria in mind, it may be a necessity to examine the research-based models for teacher development while providing a variety of implications for evaluative practices.

In Regards to self-evaluation and construction of knowledge among EL teachers which are two significant shares of teacher evaluation (Kim & Danforth, 2012; Khan et al., 2017), a variety of developmental practices may be considered (Walsh, 2015). Such practices may first be gathered under the umbrella term Practitioner Research (PR) and Reflective Practice (RP) (Borg, 2007, 2010; Darling-Hammond, 2006). These may simply be defined as small-scaled and classroom-based research models conducted by teachers in the aim of developing themselves professionally, seeing their own weaknesses and strengths by self-evaluation, and improving student learning while they continuously monitor their own developmental progress. For Nakamura (2014), three commonly-used types of practitioner research are Action Research (AR), Exploratory Practice (EP) and Reflective Practice (RP) - each of which is context-specific research, where teachers have the opportunity to observe, collect data, analyze and reflect on themselves. The consequences in language education contexts are fruitful for both EL teachers and learners.

One type of PR is Action Research (AR). AR proposes a four-step model - planning, action, observation, and reflection - to help teachers actively find answers to their own contextual problems or questions in mind in actual classroom settings. A definitive understanding is given by Burns (2010) describing AR as ‘problematising’ by questioning and developing new ideas to a subject area while investigating and exploring personal teaching context as ‘the reflective teacher as researcher’. Self-realization and reflection are at the center of AR; practitioners involved in action research are expected to explore what they are doing, why they are doing it and what the impact has been after doing it (Mertler, 2019). As Cabaroğlu (2014) similarly states as a result of an AR study with prospective EL teachers, “participants pursued a wide range of topics for their action research projects thus indicating the diverse

individual interests, and the findings also pointed to a growth in self-efficacy as a result of participation in action research" (p. 85).

Another type of practitioner research is Teacher Research (TR), which may be described as self-directed classroom research conducted by reflective teachers in order to gain insights both for better student learning and teaching experience (Roberts, 1993). TR searches for the answers to teachers' own pre-identified pedagogical questions. By delivering relevant research questions, teachers carry out classroom-based research to collect data and analyze and infer meanings. The systematic collection and interpretation of data help teachers improve both the teaching and learning process. Stating the effectiveness of TR, Dikilitaş (2015) draws attention to the act of development for teachers in their own contexts as they take active participation in conducting research while constructing knowledge together with their students. Atay (2008), together with the participation of 18 EL teachers who conducted classroom-based research, investigated their attitudes toward classroom research and its effects on teaching practices. The findings of the study suggested five arguments for teachers' side - specifically, a) the development of research skills; b) an increased awareness of the teaching/learning process; c) a renewed enthusiasm about teaching; d) collaboration with colleagues; and e) the overcoming of general problems in research. Dikilitaş, Smith, and Trotman (2015) also point to the following issues that TR practices can help EL teachers in their own contexts by encouraging them to reflect critically on current teaching practices, raising their awareness of new practices, encouraging them to examine and review their beliefs, helping them gain further insight into their teaching, improving their motivation, and heightening their awareness of learners.

Mutual observation and jointly reflection on classroom teaching practices also constitute a means of assessing teachers' performance and improvement (Saito, 2012). In order to sustain a shared culture in an institution, Lesson Study (LS) provides the platform where teachers support inquiry, share discussions, experiences and a set of pedagogical practices (Cochran-Smith & Lytle, 2001; Grierson & Gallagher, 2009). According to Godfrey, Seleznyov, Anders, Wollaston and Barrera-Pedemonte (2019), "lesson study is an approach to teacher professional development (PD) involving collaborative planning of a lesson, live lesson observation and reflective discussion" (p. 325). A study conducted by Nami, Marandi and Sotoudehnama (2016) aimed to engage EL teachers in LS. The study yields a variety

of results in terms of observation and feedback. As it is the nature of lesson study to collaborate, to plan lessons, to implement the lesson plans and to observe/monitor the process, the teachers in the study provided their own individual comments. The focus was mostly on the feedback generally dealing with teachers' performance, the lesson plan content, and the lesson plan pedagogy. Peer-feedback and social-constructivist side of the process-oriented period led the subject to adaptation of more positive side of peer feedback and constructivist guidance (Nami et al., 2016).

The aim is not to always find answers, but rather to explore puzzles and questions EL teachers may have surrounding their own teaching contexts in order to increase quality of life in the classroom (Allwright, 2003; Gieve & Miller, 2006). Without taking action against problems in the classroom, teachers may start to explore and discover various pedagogical and/or curricular issues while they continue having regular lessons together with their students, and for their own improvement or satisfaction of their curiosity, they consult Exploratory Practices (EP) (Allwright, 2005; Allwright & Hanks, 2009). Allwright (2005), largely held as the originator of EP, defines it as the collaboration of teachers and students acting together in order to deepen their understandings of life in the classroom. Teachers and students search understandings for puzzles in a classroom-based research without abandoning pedagogical activities and practices. It seeks an understanding of what happens in classroom. That is, it is questioning what is inherently so? (Hanks, 2015). By conducting EP in the classroom, teachers as researchers in their own context can gain a sense of confidence to continue their journey, realize some shortcomings of their research along the way and modify their journey and/or embrace their own discovery (Hiratsuka, 2016).

2.3.6 Reflective practices for teacher evaluation. In order to avoid top-down evaluative practices (i.e., institutional appraisals), team-works, critical friends, peer processes, and other types of collaborative efforts open doors to EL teachers to evaluate themselves and develop professionally (Mann, 2005). Teachers should not be forced to engage predetermined practices of teacher evaluation (Bowen, 2004; Tomlinson, 2003); instead, reflective tools encourage the relation between different forms of knowledge and allow teachers to connect this received knowledge with more experiential, personal and contextual knowledge (Freeman & Hawkins, 2004). Mann and Walsh (2015) suggest a series of reflective tools to enhance reflection on

teacher evaluation; the first is *tool-reflective* which includes video and audio recordings to create opportunities for experiential learning and applications to promote self-evaluation; the second is *stimulated-recall* which uses video and audio data with a critical friend to provide particular procedures for the recall of specific changes or events; the third tool is *critical incident analysis* which focuses on problematic or critical incidents to help teachers recall and work with particular incidents; the next tool is *materials analysis* which encourages novice teachers to reflect on coursebooks, and other teaching materials and can be developmental if there is time for evaluation; the next tool is *questionnaires* which are for the ongoing process of self-evaluation; the next is *narratives* which prize reflective writing, storytelling, diaries or journals to develop individual self-evaluation; the next is *portfolios* which facilitate the importance of reflection through recording experiences of novice teachers and can provide a balance where individuals and peers can assess their own competencies and skills; the next is *team-teaching* which establishes collaborative partnership for self- and peer-evaluation in teaching and can be used to foster discussion about the roles and expectations of team teaching; the next is *peer observation* which uses ethnographic notes to improve and evaluate especially novice teachers while encouraging objectivity for all parties; the next is *interviewing* by which ethnographic interviewing may facilitate peer-understanding and self-awareness; the next is *focus groups* which promote opportunities to co-constructed learning communities for evaluative discussion of learners, materials, or methodology; the next is *critical friendship* which uses comments on how colleagues might develop sustainable critical friendship; the next is *collaborative peer conversations* which are non-evaluative in nature, and cooperative development directs the focus on self-development and evaluation; and the last tool is *individual writing* the process of which can prompt introspection, reflection and self-evaluation while auto biographical writing can be a sample for change (Mann & Walsh, 2015, pp. 22-23).

2.3.7 Conclusion. It is apparent that conducting the models of Practitioner Research (PR) either in the form of Action Research (AR), Teacher Research (TR), or Exploratory Practice (EP) and/or the types of Reflective Practices (RP) assists educators as practitioners to step forward in their professional development. As the nature of such practices are context-specific, inquiry-based and reflective-driven,

teachers benefit at maximum level in terms of theoretical, pedagogical and content knowledge. While a small number of researchers (Mercer, 2006; Riera, 2011) indicate that teacher appraisal systems in the forms of teacher evaluation have the potential to inform professional development (PD) and to promote more effective teachers and learning, opposing arguments dominate them, as it is often seen as a time-consuming and intensive task when the focus on quality assurance is lost, or the concept of professional development is totally absent (Hutchinson, 1995; Kandil, 2011; Miller & Young, 2007; Moss, 2015). It is clear, however, that teacher evaluation aimed at improving teachers' practice and professional development, as well as student success through a well-organized, focused, specifically defined criteria, can ensure many of the advantages hypothesized and teaching profession can hugely benefit from the outcomes (Hurley, 2013; Marzano, 2012).

Keeping all this in mind regarding teacher evaluation practices, Marzano and Toth (2014) clearly put emphasis on the requirement of active participative roles of teachers who need role-models and professional training to help students as skilled facilitators, to guide them to becoming autonomous learners, to equip them with tools to work collaboratively or individually apply and solve complex real-world problems. According to Basileo and Toth (2019), findings from teacher evaluations should be used to provide feedback to educators and guide their professional development, which - in turn - should positively impact student achievement.

In summary, as Darling-Hammond, Amrein-Beardsley, Haertel, and Rothstein (2012) assert teacher evaluation has the potential to become a valuable part of a productive teaching and learning system, which initiates accurate information about teachers, helpful feedback, and well-grounded personnel decisions. In their words, “by ensuring that evaluators are trained, evaluation and feedback are frequent, mentoring and professional development are available, and processes are in place to support due process, evaluation can become a more useful, supporting practice” (Darling-Hammond et al., 2012, p. 14). More specifically, grounded on a recent study conducted in Finland, Tarhan et al. (2019) present the fundamental concept of teacher evaluation practices and conclude that teachers should be evaluated; a) for teacher empowerment and increasing quality, b) based on needs for teacher education and professional development endeavors, c) during individual meetings with principals and group meetings with colleagues, and d) by school administration.

As can be understood, there is a need to elaborate more on several teacher-related themes in brief details as in the following sub-titles in order to understand more about the external and the internal dimensions of teacher evaluation.

2.4 Qualities of Teachers for Teacher Evaluation

Johnson (2009) states that teachers' normative ways of behavior, values, assumptions, and attitudes are embedded in the classrooms, their ways of acting and interacting might be affected by L2 teaching and learning processes. The purpose of teacher evaluation and training is expected to streamline these aspects and bring them in line with new findings and understandings of the teaching profession (Danielson & McGreal, 2000); for this reason, in language programs around the world important decisions are regularly made about the lives of language teachers based on hiring decisions, decisions about whether the teacher's work has been satisfactory, staff retention or cut-back decisions, and decisions about promotions and raises (Borg, 2019; Brown & Wolfe-Quintero, 1997). Most teacher evaluations depend on simple dichotomous criteria, such as ranging from *satisfactory*, to *needs improvement* or *unsatisfactory*. Others are based on rating scales, i.e., 1-2, representing items as *low*, *medium*, or *high* as well as *seldom*, *frequently* or *always* (Danielson & McGreal, 2000).

In terms of the purposes of teacher evaluation, it has earlier been discussed that the mode of evaluative practices should also be taken into consideration. According to Danielson and Mcgreal (2000), summative assessment and formative assessment of teachers differ in purpose; while the summative seeks consequential decisions, formative evaluation aims to enhance the professional skills of teachers. As suggested by many researchers, teacher evaluation needs to provide both identification of teacher skills and professional development (Marzano, 2012; Moss, 2015). According to Haefele (1993), screening out unqualified teachers, recognizing and helping outstanding practices, providing constructive feedback to teachers, offering directions for professional development, and helping teachers and administrators collaborate for student success are the necessary purposes in evaluating teachers.

Teacher evaluation is one of the key aspects that impacts the quality in teaching; defining teacher quality is fundamental in promoting the evaluation

procedures for this purpose (Mitchell, Robinson, Plake, & Knowles, 2001). Becoming a qualified teacher directly affects what students learn, how and how much they learn, and how they interact with others while learning, which makes teachers' influence key to the outcomes to be achieved (Stronge, 2007). Teachers need more comprehensive types of evaluative applications in order to search for professional development and enhancing quality of instruction (Sandilos, Sims, Norwalk, & Reddy, 2019). Dependence on single-measured teacher evaluations (i.e., value-added models) may not provide the scope of authentic results as well as teacher effectiveness because the intervening domains change from class to class or from year to year, which again relates to continuous nature of teacher development and evaluation (Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012).

Khany and Darabi (2014) draw attention to the evolutional process of EL teacher qualities and effectiveness from the 1960s and 1970s to the current era in the form of structuralism in linguistics, behaviorism in psychology, and focus on communicative language teaching (CLT) in foreign language education. In the 1980s, following these, important researchers in the field of language education (Richards & Rodgers, 1986; Nunan, 1987) began to criticize the structuralist view (i.e., grammar-based language teaching) and behavioristic approach (i.e., Audio-lingual method), while placing more emphasis on CLT. Functioning rather than mechanical teaching, focus on forms rather than focus on form, and/or communication, fluency, meaning rather than structure and accuracy gained more importance in EL teaching (Sanchez, 2004; Khany & Darabi, 2014). The dominance of method-influenced approaches to EL teaching later evolved to take on another perspective known as the post-method approach. Brown (2001), in this sense, explains that the post-method approach combines linguistic accuracy and fluency in EL teaching, opens doors to teachers to apply authentic and communicative practices for their learners, and provides opportunities for learners to engage in collaboration while fulfilling their potentials.

Savignon (2007) states that "EL teacher's engagement with language learners constitutes a fascinating arena for the study of social interaction" (p. 207) and discusses the interrelationship between CLT and post-method approach by elaborating on the responsibilities that teachers should possess and on the possible attainments that learners can receive. If the aim is to carry out the principles of CLT and post-method understandings, according to Savignon (2007), the following

qualities are to be tracked by EL teachers; a) sustaining learners' needs to effectively communicate in the target language, b) teacher-generated responses to social and technological updates, c) from theory to classroom application approach, which has taken place instead of a direct application of top-down methodologies, d) EL teaching as the collaborative and context-specific human activity, e) empowerment of EL teachers as practitioners and theory builders in addressing the language needs of learners. In order to understand more about the relationship between teacher evaluation and the qualities of effective teachers, the progressive evolution of teacher qualities and competences for teacher evaluation can be summarized in table 2 (Danielson & Mcgreal, 2000):

Table 2

Evaluation Criteria of Teacher Qualities From Past to Today

Decade	Research on Teaching
1950s	Teacher traits (e.g., voice, appearance, trustworthiness, enthusiasm)
1960s & 1970s	Teacher effectiveness (e.g., improving basic skills, designing classroom observation techniques), Clinical supervision (e.g., research on enhancing teacher effectiveness and basic skills acquisition), Learning styles
1980s	Expectancy studies (e.g., behavioristic approach), Discipline models (e.g., instructionally focused teacher development), Cooperative learning, Brain research (e.g., motivation, retention, transfer)
1990s	Critical thinking, Content pedagogy and knowledge, Alternative assessment, Multiple intelligence, Collaborative learning, Cognitive learning theory, Constructivist classrooms, Authentic pedagogy, Engaged teaching and learning, Teaching for understanding
2000s	Constructivist classrooms, Authentic pedagogy, Engaged teaching and learning, Teaching for understanding.

Taking into account a more comprehensive context and globally recognized contemporary understanding, Kumaravadivelu (2003) adds three elements to the list

of roles an educator must embody - namely that of; a) passive technicians, b) reflective practitioners, and c) transformative intellectuals. It may consistently be assumed by many that effective English teaching relies on being dedicated to applications of prescribed principles and techniques in the classroom (Richards & Rodgers, 2001); however, it has been recognized that theorized methods may not successfully fit in the scope of EL teachers' own contexts, but rather "artificially transplanted" amid what can otherwise be deemed a post-method pedagogical era (Kumaravadivelu, 2006).

Kumaravadivelu (2006) conceptualizes the post-method pedagogical era with three all-encompassing schemes; a) particularity, b) practicality, and c) possibility and states that "post method pedagogy must be sensitive to a particular group of teachers teaching a particular group of learners pursuing a particular set of goals within a particular institutional context embedded in a particular socio-cultural milieu" (p. 171). *Particularity* refers to contextual and localized teaching and learning environments in which cultural, social and local dimensions are facilitated; *Practicality* refers to methods which enable teachers to practice from the theories or theorize what they practice because a theoretical method becomes impractical if it is not practiced; and lastly, *Possibility* refers to consciousness that triggers identity formation and social transformation both for teachers and learners (Kumaravadivelu, 2006). Considering this, "post-method teachers are autonomous, analysts, strategic researchers and decision-makers. Such teachers are also reflective as they observe their teaching, evaluate results, identify problems, find solutions, and try new techniques" (Can, 2009, p. 3). The post-modernist view builds a connection between students and teachers, meaning and context, space and time, the knower and the known (Richards, 2001; Slattery, 2006); furthermore, schooling in the post-modern view encourages the search for wisdom through theological experiences, the creation of cooperative learning environments, and the commitment to reverent, democratic models of schooling (Fahim & Nazari, 2014).

The applicable side of the post-method approach in language education contexts is commonly discussed among researchers (Pennycook, 1989; Prabhu, 1990; Richards, 1990) as pursuing more theoretical procedures rather than actual practice (Kamali, 2014). For Kumaravadivelu (2001), a post-method pedagogy must facilitate the advancement of a context-sensitive language education based on a true understanding of local linguistic, sociocultural, and political particularities, rupture

the reified role relationship between theorists and practitioners by enabling teachers to construct their own theory of practice and must tap into the sociopolitical consciousness that participants bring with them in order to aid their quest for identity formation and social transformation. Searching for the best method should perhaps not be the ultimate purpose for EL teachers and researchers, there is move from a post-method to "beyond methods" era (Kumaravadivelu, 2003); hence, one must search for the best method to apply in their own classrooms (Can, 2009). In an attempt to avoid counter-arguments against this approach and to promote more practical learning and teaching platforms for language classrooms, Kamali (2014) provides three comprehensive suggestions, summarizing the literature as follows:

- *Teacher metamorphosis;* changing a teacher to *a jack of all trades* - package of material developer, syllabus designer, examiner, reformer and the like - can be an asset. However, it requires certain expectations to be met - namely a good salary, responsibility, trust, and respect. Making teachers responsible for what they do, trusting them for the decisions they make, and respecting their opinions are the panacea by which the post-method approach can survive,
- *Teacher education;* Kumaravadivelu (as cited in Kamali, 2014) offers a five-step modular model for teacher education which is called KARDS (i.e., Knowing, Analyzing, Recognizing, Doing, and Seeing). Knowing refers to personal, procedural and professional knowledge; Analyzing covers learners' needs, motivation and autonomy; Recognizing covers teachers' identities, beliefs and values; Doing covers theorizing, dialogizing and teaching; and lastly Seeing covers the perspectives of learners, teachers and observers (Kumaravadivelu, 2012, p.125). Of this model, Kamali (2014) asserts "the authority can assure that a teacher does not make an alibi for his lack of knowledge, because there is no lack of knowledge. The practice comes from pure theory and vice versa" (p. 828),
- *Teacher freedom;* letting EL teachers exploit their professional, procedural, and personal knowledge with the post-method approach in sight, "any learner, and any teacher have specific needs and giving freedom to the 'post-method teacher' to choose the way of teaching is a key to success" (Kamali, 2014, p. 828).

Arikan, Taser and Sarac-Suzer (2008), pointing to the approach in the Turkish context, suggest a list of the qualities that EL teachers need to possess, which are having personal strategies to teach, maintaining positive teacher-student interaction and reinforcement, being an example, being knowledgeable on target cultures, having correct pronunciation of English sounds, and teaching with effective classroom materials and using technology.

A range of qualities and qualifications seem to be required for EL teachers to provide quality education, sustain their effectiveness and move forward in terms of skill development; such qualities are highly related to various dimensions, including the content and pedagogical knowledge of teachers, their sense of identity, autonomy and characteristics, as well as contemporary adaptation to advances in the field of language education. As stated by Kumaravadivelu (2001), "language pedagogy must be sensitive to a particular group of teachers teaching a particular group of learners pursuing a particular set of goals within a particular institutional context embedded in a particular sociocultural milieu" (p. 538).

Teacher evaluation should not merely be subject to high-stake or top-down decisions about teachers, which in turn have the potential to drive competent and committed teachers away from the schools and even causing them to leave the profession (Johnson, 2015). Hereafter, teacher evaluation should not solely focus on examining a single component of the qualities of teachers, but rather than those it should use a variety of data sources, multiple types of classroom observations (e.g., peer observations, self-observation), is timely and conducted by expert evaluators and provide constructive and meaningful feedback to teachers in several aspects (Darling- Hammond et al, 2012). Such aspects can be discussed under the following subtitles in aim to provide more comprehensible view to teacher evaluation.

2.4.1 Teacher effectiveness. According to Stronge (2007), teacher effectiveness can be defined in a variety of ways; one of these ways is through students' achievements, which is mostly conducted by evaluating teachers through value-added model, while another alludes to performance ratings obtained from evaluators or supervisors. Still, other measures of teacher effectiveness (e.g., mentoring, classroom observations, VAMs) may rely on comments from stakeholders, such as students and administrators (Strøm & Falch, 2020). Teachers have profound influences on their students and their own professional development

(Lupascu, Pânisoară & Pânisoară, 2014; Rockoff, 2004); and therefore, it is not so easy to measure how these influences take place and what outcomes should be taken into consideration.

In addition, it is not the teacher that is solely responsible for the outcomes of teaching. There are many variables outside the teacher control that might affect measures of effectiveness. Yet teachers' characteristics and behavior, in addition to their expertise, remain important factors. These characteristics and behaviors can be summarized as preparation, classroom management, planning, temperament, methodology, and monitored progress (Duta, Tomoaica, & Pânisoară, 2015; Hammond, Bransford, & LePage, 2005). High teacher effectiveness is positively related to a teacher's performance in classroom and student success (Moradi & Sabeti, 2014; Rahimi, 2014; Saberi Duta, Tomoaica, & Pânisoară, 2015), with criteria such as enthusiasm, creativity, flexibility and adaptability and success of students (Darling-Hammond, 2000), as well as teacher persistence, enthusiasm, commitment and instructional behavior, which can also be reflected in student outcomes in terms of achievement, motivation, and self-efficacy (Tschannen-Moran & Hoy, 2001). Furthermore, calmness, tolerance, sense of humor, friendliness and well-preparedness are also several other suggested indicators of teacher effectiveness (Lupascu et al., 2014).

To enhance teacher effectiveness, pedagogical practicum accepted as the theoretical knowledge provided to pre/in-service teachers seems not to fulfill the required expectations in teacher development (Caires & Almeida, 2005). The gap concerning theory and practice is one phenomenon commonly discussed among researchers in the field, and this points to the difference between what teachers learn through lectured knowledge and how they apply it in real-life (Nilsson, 2008; White, 2009). According to Hobjilă (2012), "given the importance of pedagogical practicum and the generalization of the practice, specialized literature is generous regarding the presentation of multiple aspects on this matter" (p. 319).

As is clear, in addition to the provision of pedagogy and content knowledge to teachers, practical applications matter in teacher effectiveness. At this point, Zhang (2009) asserts that effective teachers must possess a number of abilities, ranging from cognitive skills to personal characteristics, as well as the ability to adapt pedagogical and content knowledge to classroom operation. Zhang suggests a list of six dimensions to clarify effective teachers - namely; 1) academic qualification and

publication; 2) preparedness and subject knowledge; 3) personality trait and personal style; 4) connectedness with students; 5) motivation and enthusiasm; and 6) classroom operation.

Park and Lee (2006) point to English proficiency, pedagogical knowledge and socio-affective skills of EL teachers in terms of the characteristics of effective EL teachers. Er, Ülgü and Sarı (2012) posit the importance of developmental requirements and personal development among educators, which are vital in enhancing their effectiveness, while undoubtedly increasing the quality of education. Meanwhile, Teng (2019) points to three significant issues in terms of improving teacher effectiveness and autonomy; 1) various benefits of conducting action research for professional development, curricular improvement, comprehension of reflective practices (as cited in Hopkins, 1993; Burns, 1999); 2) reflecting on their teaching practices in order to reevaluate benefits of certain practices; and 3) teacher collaboration for creativity, exchange of lesson ideas, and professional development rather than following THE traditional curriculum (e.g., top-down approach).

Within a broader scope, Tucker and Stronge (2005) define EL teacher effectiveness with a comprehensive list of indicators. They draw attention to a variety of domains, ranging from content knowledge to pedagogical knowledge, from professional development to achieving student success. They posit that effective EL teachers; a) have formal teacher preparation training, b) hold certification and are certified within their fields, c) have taught for at least three years, d) are caring, fair, and respectful, e) hold high expectations for themselves and their students, f) dedicate extra time to instructional preparation and reflection, g) enhance instruction by varying instructional strategies, activities, and assignments, h) present content to students in a meaningful way that fosters understanding, i) monitor students' learning by utilizing pre- and post-assessments, providing timely and informative feedback, and re-teaching material to students who did not achieve mastery (Tucker & Stronge, 2005).

2.4.2 Teacher autonomy. Gupta and Baveja (2014) describe teacher autonomy as “the ability to make their own decisions about what to do rather than being influenced by someone else or told what to do” (p. 162), and refer to independence, self-sufficiency, self-government, self-rule as synonymous to the term; the concept of autonomous teacher, with their ability to grow professionally, focuses on teachers’

capacity for self-directed professional development by continuously updating their knowledge. Erss and Kalmus (2018) describe teacher autonomy as involving the "liberty to act" and "freedom from constraints" and make a distinction between professionalism and professionalization - the former of which refers to autonomy itself, while the latter refers to developmental practices to improve teacher autonomy. EL teachers need to gain their freedom control by others and, at the same time, in order to be free from control must be self-directed (Benson, 2011). According to Teng (2019), teacher autonomy is a multifaceted concept that combines the self-directed professionalism, roles and development of EL teachers, while it is recognized as a means of supporting EL learners' autonomy.

According to Hermansen (2017), the perception of professionalism has been in constant flux and teachers need to keep up with new developments in their field of expertise and remain informed of key sources for professional autonomy; teachers' organizational roles and routines in school settings direct them to consequences for instructional development in a collective autonomy platform. Following this perspective, collaboration between teachers and administrators is sustained, and shared knowledge turns into "promise" rather than "threat" (Hermansen, 2017, p. 8). Erten (2015) suggests first equipping teachers with pedagogic knowledge in order to help them become more autonomous and effective, while teachers additionally need research skills training as a prerequisite for new identity formation and the ability to take on the role of teacher-researchers (Coşkun-Ögeyik, 2009; Dikilitaş, 2015). On this, Erss and Kalmus (2018) suggest providing support and educational resources, as well as collaborative platforms between teachers and local administration where shared responsibility takes place. A comprehensive approach to developing teacher autonomy according to Gupta and Baveja (2014) is summarized as follows;

- *Characteristics of teacher autonomy:* Negotiation and reflection skills, knowledge of the institution, and continuous development through observation, reflection, thoughtful consideration, understanding, experience, evaluation of alternatives,
- *Domains of teacher autonomy:* Teaching and assessment, curriculum development, and school functioning, and professional development,
- *Factors affecting teacher autonomy*: Professional competence and received support, teacher's personal belief system, teacher's intrinsic and extrinsic motivation, and psychological constraints,

- *Ways to develop teacher autonomy*: Developing abilities and willingness to implement changes, self-awareness, responsibility, participation and collaboration, changing roles in the classroom, and lastly creating a social network for professional development such as action research, self/peer-observation, reading literature, or attending conferences (pp. 167-176).

Zonoubi, Rasekh and Tavakoli's (2017) investigation into EL teachers' development of self-efficacy and autonomy through professional learning communities (PLC) reveals that EL teachers obtained the benefit of pedagogical self-efficacy, perceived language proficiency and collective autonomous perception by following a six-month developmental program (i.e., PLC). After undergoing the process, teachers mostly reported having enjoyed experiencing innovative teaching strategies and improved skills, while novice teachers reported having obtained better classroom management skills and decision-making capabilities, which contributed to becoming more autonomous (Zonoubi et al., 2017).

The provision of autonomy support is appreciated through the social context (Ryan & Deci, 2000; Reeve, 2009); when autonomy is supported, teachers tend to take the perspective of their students, introduce activities that vitalize and support the psychological needs, provide explanatory rationales for their requests, communicate using informational language, acknowledge and accept expressions of negative affect, and display patience (Reeve & Cheon, 2016). Reeve and Cheon (2016) point to self-determination theory, which suggests fulfilling inherent psychological needs in terms of engagement, positive functioning, and well-being through the social context and theory-based instructional guidance by professionals. As long as such developmental programs are interconnected with reflective practices, EL teachers continue developing a sense of who they are (i.e., self-image) and what they do (i.e., professional identity); indicators of which are teacher belief, values and (re)construction of their roles as language teachers in relation to their peers and varying contexts (Burns & Richards, 2009; Farrell, 2011). An understanding of teacher identity is related to how educators come to "figure out" who they are, through the realms in which they participate and how they relate to others within and outside of these worlds (Urrieta, 2007). For Farrell (2011), it is not a case of forcing teachers to change, but to encourage them to reflect on their practice in the forms of group discussions and reflections. In addition to sustaining positive environment at work place between EL teachers and the principals, supporting them professionally

in their own context enhances teacher autonomy and teaching abilities (Javadi, 2014).

When it comes to the Turkish context, teacher autonomy seems to have several indicators from classroom teaching to professional proficiency. Ulas and Aksu (2015) investigated areas in which Turkish teachers feel autonomous, dividing these into the dimensions of a) autonomy in instructional planning and implementation; b) autonomy in professional development; and c) autonomy in determining the framework of the curriculum. However, the actual implementations and the final decisions of these pre-mentioned issues are not left to teachers, because the curriculum and teaching practices are dictated to teachers by institutions (Öztürk, 2012). Dincer (2019) states that the autonomy of EL teachers in Turkish context are affected by external factors, such as the curriculum, classroom environment, and salary, as well as internal factors, such as self-evaluation, the joy of teaching, and student motivation; helping teachers reveal their own potentials by collaboration with decision-makers and engage in autonomy-supportive practices to take individual actions in form of self-directed way will result in better education.

2.4.3 Teacher knowledge and competency. Teacher knowledge is considered an important considered in teacher competency during the process of recruitment. Pedagogical knowledge and teaching experience matter greatly, since a teacher's content knowledge and intellectual ability is not sufficient in itself and does not directly translate to effective teaching. For example, it is observed that while there is not necessarily a strong relationship between teacher knowledge itself and student learning, there is such a relationship between teachers' various other abilities - such as verbal/professional abilities, as well as the quality/quantity of their students (Strauss & Sawyer, 1986). More to the point is that certification is not a guarantee that effective teaching will follow. According to Stronge (2007), in some cases, variation among teachers with the same licensure and certification is more than teachers with varying certifications. This point highlights the importance of teacher preparation, performance monitoring, and evaluation. In such a case, integrated capabilities and the roles of teachers may generate us a more specific understanding.

The pedagogical and content knowledge (PCK) of EL teachers are both accepted as significant competences in language education. Shulman (1986) first drew special attention to this concept by stating that "PCK identifies the distinctive

bodies of knowledge for teaching and goes beyond the knowledge of subject matter per se to the dimensions of subject matter knowledge for teaching" (p. 8). Shulman (1987) extended his definition by adding more categories to teacher knowledge as content knowledge (CK), PCK, curriculum knowledge, general pedagogical knowledge (PK), knowledge of learners and their characteristics, as well as knowledge of educational contexts, purposes or values. More concisely, according to Shulman (1987), a) content knowledge is the understanding of the facts and structures of content domain; b) general pedagogical knowledge is broad principles and strategies of classroom management and organization that appear to generalize across different subject matter domains, c) curriculum knowledge is the understanding of the materials and the programs for teaching; d) knowledge of learners is the understanding of their characteristics and learning styles that they bring with themselves; and e) knowledge of educational contexts and values constitute understandings of classrooms, governance of school settings and character of communities and cultures in terms of education-related concerns.

In teacher education programs, Evens, Elen, Larmuseau, and Depaepe (2018) prioritize the necessity of explicit provision of all three components as content knowledge (CK), pedagogical knowledge (PK), and pedagogical content knowledge (PCK) while claiming that each specific domain complements the other. In their study, Evens et al. (2018) compared the application results of segregated and integrated provision of CK, PK and PCK to EL teachers, with the results showing that the integrated but explicit teaching of PCK yielded better understandings and developments for teachers.

The results of another study conducted with Turkish EL teachers by Erten (2015) showed that 73% of participants (56 EL teachers) indicated the lack of real-life teaching experience, insufficient contextual awareness, and an experienced mismatch between the ideal and the actual, while over 30% of participants cited insufficient language proficiency and teaching methodology, as well as a lack of good roles models, who could guide them in the form of in-service teacher development. On the other hand, 85% of participants reported a need for improvement in terms of more real-life teaching experience, striking a balance between theory and real-life teaching, and training opportunities to become more autonomous researcher-teachers.

Regarding the development of PCK in the Turkish EL teaching context, Atay, Kaslioglu and Kurt (2010) investigated the outcomes of its applications in the form of micro-teaching, over the course of a teaching skills training program. Following the application, it was found that 1) opportunities to apply the theoretical knowledge they gain in the methodology course, contributed to the development of their PCK; 2) teacher educators established environments to guide PTs into active thinking and designing materials for content and pedagogical competency; and 3) detailed feedback should be an integral part of the whole experience. Atay et al. (2010) conclude that prospective/novice teachers "need to understand the content they want to teach but they also need to understand how to unpack and present the content so that students can learn with understanding" (p. 1425).

Given the contemporary advancements in the educational use of ICT, according to Angeli and Valanides (2009), "citizens of information-age societies are required to be able to think critically, problem solve, collaborate with others, communicate, use various technologies, take initiatives, and bring diverse perspectives in the learning situation" (p. 154). Teachers need technological capabilities, in addition to content and pedagogical knowledge (Koehler, et al., 2007). The technological knowledge (TK) of EL teachers has undeniably gained much importance. Mishra and Koehler (2006) claim that the CK, PK, TK of teachers can be developed and employed independently, and that at the end, TPACK emerges itself for pedagogical purposes. TPACK, or Technological Pedagogical Content Knowledge, can also be formed by adding up all dimensions in order to make it clear that various characteristics and competences of teachers come together to allow for effectiveness, indications of which, in addition, can further be implemented for teacher development practices (Baran, Canbazoglu Bilici, Albayrak, & Tondeur, 2019; Voogt & McKenney, 2017).

Koh and Chai (2016) provide a seven-factor teacher knowledge framework for the integration of TK into the CK and PK of contemporary teachers, whilst they conclude that the framework is still open to evolve and applications for teaching are based on teachers' contextual conditions. The TPACK framework suggested by Koh and Chai (2016) is summarized as follows:

- TK: Technological knowledge about the use of ICT-related tools,
- PK: Knowledge about learning, instruction, assessment and students,
- CK: Knowledge about the subject matter,

- PCK: Knowledge about content representation for teaching and learning,
- TPK: Knowledge of pedagogically sound ways to use specific ICT tools,
- TCK: Knowledge about how ICT tools can be used to create subject matter,
- TPACK: Synthesized knowledge that reflects elements of technological, pedagogical and content knowledge (p. 246).

2.4.4 Teachers' roles. In order to facilitate the continuation and betterment of learning, EL teachers need to hold a variety of proficiencies. As language learning is not a lecture-based system in which knowledge benefits from being shared rather than transmitted and constructed together with the active participation of learners and teachers, EL teachers must plan, set, apply, examine, and reflect on every aspect of classroom practice (Chan, Sprat, & Humphreys, 2002; Yıldırım, 2012). They are expected to be active and engage EL learners in the complete process (Koçak, 2003). To Spratt, Pulverness and Williams (2005), EL teachers in the classroom must inhabit the role of planners, informers, managers, observers, involvers, parents/friends, diagnosticians, and resources. Teachers are the providers of knowledge and instruction, directors and managers of teaching and learning environment, judges, leaders, evaluators, controllers, facilitators and mediators (Saban, Koçbeker, & Saban, 2007). According to Simonian and Robertson (2002), the fundamental dimensions for EL teachers' roles create a space for reflective discourse, providing the occasion for reflection, guiding the process, and mediating between the classroom and the outside world.

Harmer (2001) defines EL teachers and their roles in his well-known work "The Practice of English Language Teaching" as actors on the stage, orchestral conductors directing conversation by setting pace and tone, or as gardeners planting seeds and watching them grow. According to Harmer (2001), "a teacher's role may change from one activity to another, or from one stage of an activity to another. As long as teachers are fluent at making these changes our effectiveness as teachers is greatly enhanced" (p. 57). All roles aim to facilitate students' progress in some way or other. The roles teachers inhabit can briefly be listed as follows:

- *Controllers* take the roll, tell students things, organize drills, or read aloud. They, in various other ways, exemplify the qualities of a teacher-fronted classroom and act as transmitter of knowledge to learners,

- *Organizers* get students to perform various activities, involving giving them information, telling them how they are going to do the activity, putting them into pairs or groups, and finally closing things down when it is time to stop,
- *Assessors* help students whether or not they are getting their English right. This is where teachers offer feedback or correction and grade students in various ways.
- *Prompters* help students get involved in activities where they are not sure what do to next or how to get started,
- *Participants* from time to time prefer taking part in exercises or activities together with students to acting as an organizer or controller,
- *Resource,* as a role, comes into play when students may still have need of their teacher as a source of knowledge while avoiding spoon-feeding,
- *Tutors* point students in the right direction when they are working on longer projects; in such cases, teachers are combining the roles of prompter and resource, acting as a tutor.
- *Observers* observe for feedback, keep track of what students do, and also observe for evaluation of the materials and exercises,
- *Teachers as a teaching aid,* apart from the roles, make use of their mimics, gestures and expressions to convey meaning, be a language model for their students, and provide comprehensible input (Harmer, 2001, pp. 58-66).

2.4.5 Conclusion. Considering these, definitions of quality teacher - in addition to various teacher knowledge types and roles - require the ability to work with a diverse community of learners, more intellectually rigorous and meaningful instruction, teacher ability to engage students in activities, and the ability to demonstrate positive dispositions for teaching (Mitchel, et. al., 2001). It would not be wrong to support the idea that EL teachers are not only effective when they have more teaching experience than others or when their students get high grades; teacher effectiveness depends on several domains, and each domain is interrelated and complimentary of the other, which is supported by Barry, Pendergast, and Main (2020) as “teacher effectiveness has a powerful impact on student performance and a teacher evaluation process that supports professional growth can be a key lever for improving teaching quality” (p. 1). To Barry et al. (2020), a standardized mechanism

is to be brought to a more goal-oriented and better-organized teacher evaluation by considering the relationship between the teacher, the evaluator and the organization for whom they work, the proficiencies of the evaluator including how they deliver feedback, and the supplementary teacher development plan.

Chapter 3

Methodology

3.1 Introduction

While chapter one provided an overview of the problems the study seeks to focus on and its significance, the second provided the grounds for the study by offering an extensive review of the related issues - elaborating on some of the works directly related to the current study. Chapter three will move forward to deal with steps taken as part of this study. In this chapter, research context and research design, participants, data collection instruments, data collection procedures, and the data analyses of both the qualitative and the quantitative studies, as well as the construction process behind the formulation of the questionnaire will be provided.

3.2 Research Context

In Turkey, English preparatory programs at universities serve as pre-sessional units that take place in the year before undergraduates enter faculties or vocational schools. At English preparatory programs, students aiming to study at various departments and faculties first receive English language education over the course of an entire academic year. Students take a placement test at the beginning of the academic year, and based on the results of this level-scored English test, are appointed to English classes in the forms of beginner/elementary level (A1/A2 level), intermediate/upper-intermediate level (B1/B2 level), and advanced/proficient level (C1/C2 level). At most of these programs, a modular system is followed and each module lasts around eight weeks. Achievement tests, progress tests, educational short-term projects, and similar evaluative applications are the indicators whether a student will move to a next module (e.g., language level) or repeat the same module. On the other hand, students usually attend 20 to 28 hours of English classes per week, and weekly class hours of classrooms differ as English levels of students change. At the end of every academic year, those who successfully manage to

complete upper-intermediate and/or advanced level are considered to have completed the English preparatory program successfully and can start studying at their faculties.

For the teaching side of the system, EL teachers holding Bachelor's or Master's degrees in English language education, language and literature, linguistics, and/or translation and interpreting are put in charge of instruction. Teacher evaluation, in the form of diverse formative and summative practices, is mostly conducted by administrators, professional development unit (PDU) members or teacher trainers. It can be claimed that the evaluative practices applied towards EL teachers lack academic and theoretical background in addition to limited developmental purposes. The absence of a national and a specific research study investigating teacher evaluation in Turkey, which bears developmental purposes for EL teachers, constitutes the basics of this particular research study.

3.3 Research Design

This research essentially explores the perceptions of EL teachers and administrators working at English preparatory programs in Turkish universities as the indicators of how teacher evaluation practices are implemented and serves the development of a scale based on these perceptions for teacher evaluation using a measurement instrument, which was developed as an online questionnaire. This kind of research requires the use of qualitative and quantitative research, so the design of the study can be categorized as a mixed-methods research design. According to Clark and Creswell (2014), in a mixed-methods design, based on the research questions, both qualitative and quantitative data are collected and analyzed in a convincing and rigorous way. The collation of two types of data at the same time by placing one on another or by building one on top of the other is called triangulation and adds to the validity of research. Mixed-method studies can be defined as research designs incorporating both qualitative and quantitative approaches, by paraphrasing the data and making inferences along with their quantitative analysis (Christensen, Johnson & Turner, 2014; Teddlie & Tashakkori, 2010). Additionally, Dörnyei (2003) states that implementation of either merely quantitative or qualitative scales alone cannot provide enriched data; the former is a highly structured data collection tool, which either asks about specific information or offers various responses to respondents for

statistical analysis purposes, while the latter is made up of truly open-ended items that provide data for exploratory purposes (p. 14).

In this specific study, an exploratory sequential mixed methods design as suggested by Creswell and Creswell (2017), which is one of the approaches in mixed method research, was employed. This design takes place in two-way interactive steps model. Based on this model, the first stage of this study consisted of two qualitative data collection tools in the form of semi-structured interviews and guided written reports. Then collection and analysis of the quantitative data was the second stage. The second stage, in which quantitative research took place via an online questionnaire, was accomplished by using and applying the results of the first stage (i.e., qualitative stage). The aim here was to reach the entire population of EL teachers and administrators working at the English preparatory programs of the universities in Turkey: According to Creswell (2002), "survey research designs are procedures in quantitative research in which investigators administer a survey to a sample or to the entire population of people to describe the attitudes, opinions, behaviors, or characteristics of the population" (p. 376). Table 3 briefly clarifies the procedures of the complete study as the following:

Table 3

Summary of the Complete Study

Steps and Dates	Brief Explanations
Step 1 (February, 2018)	All participants were given the informed consent forms to fill out, and the form was explained in detail for the first qualitative stage of the study.
Step 2 (February - April, 2018)	Implementing guided written reports and semi-structured interviews: 1. How would you define teacher evaluation in your own words? What is your understanding of the term? 2. What do you believe is the rationale behind teacher evaluation? 3. What teacher evaluation practices and procedures are followed in your institution? 4. What positive results and/or outcomes do you believe are achieved through teacher evaluation? 5. What drawbacks and/or negative outcomes do you believe are observed through teacher evaluation? 6. In what specific aspects do you believe those practices and procedures have effects on teachers and their professional development? Could you provide detailed examples, please? 7. Is there anything you would like to mention more regarding the topic?
Step 3(a) (February - April, 2018)	Guided written reports; participants were given some appropriate time to answer the questions, and they sent them back to the researcher.
Step 3(b) (February - April, 2018)	Semi-structured interviews; one to one, voice recorded interviews were held with the participants.

Tabel 3 (cont.d)

Steps and Dates	Brief Explanations
Step 4 (April - May, 2018)	The researcher collected and kept record of all written/recorded data for further analysis.
Step 5 (May - June, 2018)	The researcher analyzed the content of the data and specified the themes that emerged from the written reports and the interviews.
Step 6 (June - September, 2018)	Finalizing the exploratory stage (i.e., qualitative study) and designing the questionnaire based on the emerging findings.
Step 7 (September, 2018 - February, 2019)	Delivery of the designed questionnaire for piloting: Analyzing the questionnaire for reliability and validity.
Step 8 (February, 2019 - May, 2019)	Finalizing and construction of the online questionnaire.
Step 9 (May, 2019 - September, 2019)	Delivering the online questionnaire to the entire population of EL teachers and administrators working at the universities in Turkey through Google Forms; beginning to keep and store data for further analysis (i.e., the main study, the quantitative stage).
Step 10 (September, 2019 - May, 2020)	Analysis of the findings of the questionnaire; findings and discussion.
Step 11 (June, 2020 - January, 2021)	Designing the suggested development-oriented teacher evaluation model (DTEM) and the related practices for English preparatory programs of the universities in Turkey.

3.4 The Qualitative Study

3.4.1 Participants of the qualitative study. The first stage of the study involved the participation of 50 EL teachers working at English preparatory programs in both foundation and state universities around Turkey. The number of the participants from the foundation universities was 32, while 18 were drawn from state universities. They were contacted for the study through purposeful sampling at the beginning and through snowball sampling as the study continued. Twelve of the participants were administrators, including vice-principals, level coordinators, and unit supervisors; while twelve were the heads of continuing professional development (CPD) units or heads of curriculum development units (CDU), who were in charge of evaluating teachers in their institutions. Twenty-six participants were EL teachers who taught English on a daily basis in their institutions and had no authority to evaluate teachers but in a state of being evaluated. A brief summary of

the participants' demographic information at the first stage is demonstrated in the following table:

Table 4

The First Stage - Demographic Information of the Participants and Universities

Type of the university	Number
State	14
Foundation	17
Total	31
Number of EL teachers and administrators	
State	18
Foundation	32
Total	50
Position at the university	
EL Teacher	26
Administrator	24
Total	50
Years of teaching experience	
0-5 years	12
5-10 years	16
10+ years	22
Total	50
Gender	
Male	17
Female	33
Total	50

3.4.2 Data collection instruments of the qualitative study. For the qualitaitve study, two data collection instruments, guided written reports and semi-structured interviews (See Appendix A), were used to construct the questionaire while gathering data for answering the first research question of the study.

3.4.2.1 Guided written reports. Guided written reports included seven open-ended questions and were delivered to the participants so as to provide answers to the first research question of the study, which aimed at finding out what specific teacher evaluation means to them, what practices are carried out and/or what procedures are followed in the institutions where the participants work. The questions of the reports were determined in line with the research questions. This instrument, as a supplementary data collection instrument to the semi-structured interviews, enabled the researcher to collect more in-depth data while the interviews were being held.

Before the implementation of guided reports started, the researcher contacted the participants either by phone calls or emails to make the process clear for all and to answer any questions they had about the questions. Each participant was required to provide information regarding teacher evaluation experiences that were taking place in their institution. The reports assisted the participants reflect upon their perspectives, understandings and experiences.

3.4.2.2 Semi-structured interviews. Semi-structured interviews were carried out concurrently with the rest of the participants, with whom the researcher had chance to meet in person. The objectives and the questions of the semi-structured interviews were the same with guided written reports. The interviews took place after working hours when all parties were available. The interviewer was the researcher of the study. The interviews were face-to-face meetings with the participants and voice-recorded. Each interview lasted from 10 to 20 minutes. Before beginning voice recordings, the participants were given appropriate time to have a look at the questions and to think about their answers, which also helped reduce their anxiety. When they were ready to start, the voice recordings started and some opening questions such as questions about demographic information or the institution they worked at were asked. While the interviews were being conducted, the researcher also filled the contact summary sheet (Miles & Huberman, 1994). As the nature of semi-structured studies let the researchers ask additional questions to clarify any ambiguities and to go deeper into any superficial answers, additional questions were asked to the participants when necessary. Finally, the participants were asked if they would like to talk more about the topic.

3.4.2.3 Trustworthiness of the qualitative study. Credibility in qualitative studies play an important role as the interpretation of data can seem to be subjective and therefore the results may not be considered as generalizable (Creswell & Creswell, 2017). In order to overcome such bias and to contribute them to improve, the researcher followed the member-checking procedure with the respondents to enhance credibility and received support from an outsider researcher (i.e., the advisor of the study) for peer-reviewing to examine the process and then to provide feedback to the researcher (Patten & Newhart, 2017). In regards to further trustworthiness, using two different data collection tools sustained triangulation of data sources (Pilot

& Beck, 2014). In order to increase authenticity of sources and trustworthiness of data, the data collection design was created in line with the research objectives and the questions (Cohen, Manion, & Morrison, 2007). Moreover, the data - guided written report entries and semi-structured interview transcriptions - were kept under record completely and a consistent connection between data and findings was sustained.

The initial data used to construct the questionnaire were obtained from a group of teachers (N=50) working at different universities in Turkey, who were asked to provide answers to the item-based seven open-ended questions on the guided written reports and semi-structured interviews concurrently. This step followed a two-stage design; a qualitative stage to gather baseline both for quantitative data regarding teacher evaluation and qualitative responses to the items suggesting further opinions used to address the topic. From the responses of EL teachers, 115 items were listed (open coding stage). This initial data was then grouped around salient themes (axial coding stage), which were identified by the first author. The researcher and the advisor of the study moved 54 out of 115 items from the questionnaire, sustaining an inter-rater agreement with 61 items at total. The emerging items were written into a questionnaire using a 5-point Likert-type scale (5=strongly agree to 1=strongly disagree (See Appendix B).

3.4.3 Data collection procedures of the qualitative study. The process of the qualitative study started with the examination of the ways by which EL teachers' performances were evaluated. Next, the researcher formulated seven open-ended questions based on the related literature. The participants answered the questions freely and without any constraint being imposed on them. The data collection procedure was initiated by purposeful sampling of the participants working at the same university with the researcher, and then the researcher used his own connections with other universities and asked his colleagues to help reach other universities by snowball sampling. After the participants were selected, informed consent forms were delivered to each participant. The form had information about the procedures to be followed during the study - what the participants were expected to be doing for the study and the confidentiality and voluntary nature of the study - the participants were clearly informed that pseudonyms would be used for their names, their voice recordings and written reports would be confidential, and they had

the right to refuse or stop participating in the study. Following this step, semi-structured interviews and guided written reports were implemented concurrently.

3.4.4 Data analysis of the qualitative study. Data analysis was conducted using two different sources, which were guided written reports and semi-structured interviews. Both were carried out thoroughly by the researcher. To begin with the guided written reports, following the collection of the data, the researcher reviewed the entries of the reports several times to identify concepts. The pre-coding and coding stages were implemented and examined to maintain a consistent coding system. In order to enhance credibility of data analysis process, member-checking post-meetings were held with the participants. After forming an agreed coding system, the researcher generated exclusive categories to categorize the coding. In other words, as it is the nature of qualitative research, inductive data analysis was established; patterns/codes and then more abstract units (e.g., categories) were created to organize the data. In addition to qualitative theme-based analysis, the number of each participant's answers was counted to calculate the percentages of answers in each category. Those represent the specifications of content and thematic analysis methods as suggested by Creswell and Creswell (2017).

The analysis of semi-structured interviews was quite similar to the analysis of the guided written reports except the transcription step. After all participants were interviewed and their voices were recorded, the researcher transcribed the recorded interviews. The transcriptions of the interview recordings were typed on word documents. Following that, the researcher read the transcriptions and contact summary sheets iteratively. The pre-codes were identified, cross-checked with another researcher who was informed with the study and the categories were formed. The predominating perceptions of the participants and factors related to teacher evaluation were grouped under categories. Finally, categories were grouped into comprehensive themes.

3.5 Construction of the Questionnaire

During the construction process of the questionnaire, all the comprehensive and progressive steps suggested by Dörnyei (2003) were followed from the very beginning stage, creating as many items as possible to allow for the gathering of qualitative exploratory data (e.g., through unstructured/semi-structured interviews, written reports, discussion groups etc.). In addition, the following specific tips were noted as suggested by Dörnyei (2003):

- Use short and simple items,
- Avoid ambiguous sentences,
- Avoid negative constructions (e.g., use of no, not),
- Avoid double-barreled questions (e.g., an item asking two questions),
- Include both positively and negatively worded items (avoid making respondents marking only one side of a rating scale).

There are various critical aspects to include while constructing questionnaires, including fundamental domains such the layout of phases, which were specifically considered throughout the complete process. Those criteria which were followed by the researcher of the study can be summarized as follows:

- Three types of data about the respondents can be obtained by; a) factual questions (e.g., demographic information on our questionnaire), b) behavioral questions (e.g., Section 2 on our questionnaire), and c) attitudinal questions (e.g., Section 1, Section 3, Section 4 on our questionnaire).
- Advantages are; a) ease of researcher time, b) ease of researcher effort, c) fewer financial resources. Huge amount of data in a short and affordable time will be collected; processing data will be fast and relatively straightforward by a computer program (Dörnyei, 2003, pp. 8-10).

  And during the construction of the questionnaire, several critical steps were also followed, which are detailed as the following;
- Decide on the general features such as the length, the format, and the main parts,
- Write effective items, draw up an item pool, select and sequence items, and write appropriate instructions,

- Pilot the questionnaire and conduct item analysis (Dörnyei, 2003, pp. 16-17).

3.5.1 Construction of the content of the questionnaire. An item tool consisting all the emerging constructs related to the research focus was created by analyzing the content of all answers to the guided written reports and the semi-structured interviews. In relation to this, Dörnyei (2003) suggests conducting a small-scale qualitative study beforehand to provide information on the relevant issues and states it is not an easy task to decide on what to cover and eliminate from the content of the questionnaire items, it would be better to create an item pool, and explicit content specifications will help decide what to limit or put on the range of questions. "The initial design of the questionnaire design should focus on clarifying the research problem and identifying what critical concepts need to be addressed by the questionnaire" (Dörnyei, 2003, p. 31). The researcher and the advisor of this study combined several content-related constructs and, as a result, the number of the items decreased to 81 and finally to 61 - which formed the full range of questions (i.e. items) within the questionnaire. The first template of the questionnaire was formed out of the participants' perceptions of teacher evaluation. The detailed examinations of this analysis were presented as the findings of the first research question of the study under chapter four.

3.5.2 Piloting the questionnaire. In order to implement an academic questionnaire to the target population, several complementary steps need to be taken. A piloting delivery to a sample respondents' group (Dörnyei, 2003) or getting opinions of experts in the field (Baykal, 2015) can help maintain a more valid and reliable scale.

An integral part of questionnaire design is piloting (field testing), which is conducted with a smaller target set compared to the original population of the research as Creswell (2002) states, "a pilot test of a questionnaire or interview survey is a procedure in which a researcher makes changes in an instrument based on feedback from a small number of individuals who complete and evaluate the instrument" (p. 390). Collecting feedback, deciding on problematic items to exclude, deciding items in terms of difficulty in understanding, the time needed and clarity of the instructions are several issues to consider. According to Dörnyei (2003), piloting

is a time-taking process, but "this is usually much more than was originally intended for this phase of the research and continuously discussing every aspect with a colleague should be considered" (p. 65). On this specific study, the researcher himself and the advisor of the study made discussions on every step of the construction process and also made required changes when needed by following the whole process suggested by Dörnyei (2003), explicitly by discussions with colleagues and experts (interviews, comments), asking them to answer questions one by one and having analyzed all items, asking for further general comments.

In order to pilot the questionnaire, it was converted to a Google forms format and sent to a number of 274 EL teachers and administrators from various universities around Turkey, 108 of whom responded voluntarily. For the piloting groups, the same guidelines were followed and the respondents were asked to provide comprehensive feedback. Following the piloting, necessary changes and corrections were made. As suggested by Dörnyei (2003), the following three steps were also followed; missing responses by several respondents may indicate a problem with that specific item, the range of the respondents by each item (avoid items endorsed by everyone/no one or add additional items relatively), and internal consistency of multi-item scales is measured. The next subtitles provide detailed explanations of how the questionnaire was constructed in an aim to obtain a well-organized, reliable, and valid questionnaire.

3.5.2.1 Reliability of the questionnaire. The alpha reliability coefficient from this pilot was .908. Therefore, it was considered sufficiently reliable for use with the main study. In the light of feedback from the piloting, minor adjustments were made to wording and formatting, and along with a participant consent form (which gave information about the study, the researchers, and participant rights), the questionnaire was sent to every single EL Teacher and administrator of the English preparatory programs of the universities in Turkey via emails including the link of the online questionnaire for their responses.

This questionnaire was divided into four sub-scales, each of which was designed to address each of the research questions brought up at this stage. Appendix J is the case summary which shows the number of valid cases or people answering questions in the questionnaire. The Cronbach internal consistency reliability estimate of the questionnaire proved to be high r=.908 as represented in Table 5 below.

Table 5

Cronbach Alpha Reliability Estimate

Cronbach's Alpha	N of Items
.908	61

3.5.2.2 Validity of the questionnaire. In terms of assessing the validity of a questionnaire, there are two general methods: first, examining questions in it for their relevance to the subject under investigation and/or asking for experts' views on their appropriateness (Baykal, 2015); second, running a factor or principal component analysis to see if the questions cluster around the general factors or sub-scales of the questionnaire as hypothesized by the researcher (Beavers et al., 2013; Child, 2006). On this, Child (2006) states that in factor analysis, the exploratory methods have the longest and best-established record and have been used to explore the possible underlying structure in a set of interrelated variables (p. 6). In the case of this questionnaire, both of previously mentioned approaches were adopted. That is, the questions were first examined by reviewing the literature to make sure that they were appropriate for the purposes of the study. Besides, the researcher and the advisor of the study studied the relevant items and themes to decide on the finalized version of the questionnaire. A principal component analysis was also run to see if the questions clustered around identifiable major variables. The results of the analysis as detailed below show that four extracted factors together accounted for more than 50% of the variability in the data, which is enough for accepting the instrument's validity.

In this analysis as the first output, there is a large correlation matrix; the coefficient determination of which, that comes at the bottom of the table, tells us whether there is multicollinearity (high correlations between variables or questions of the questionnaire) or not. We need a lack of multicollinearity to believe that our questions address different aspects of a construct. If this value is smaller than .00001, we can ascertain that we actually do not have multicollinearity. Otherwise, we should eliminate the problematic questions. This co-efficiency of determination value for our questionnaire is .897 E-20, which is much smaller than .00001. The output also generates two other values, namely KMO and Bartlett's test of sphericity indices. These two values are given in the following table:

Table 6

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.934
Bartlett's Test of Sphericity	Approx. Chi-Square	26913.17
	df	1830
	Sig.	.00

KMO shows sampling adequacy for the entire questionnaire and Bartlett's test of sphericity is related to correlations between adjacent variables. For KMO to be acceptable, it needs to be larger than .5 and if this is the case, it is indicative of enough people answering questions in the questionnaire. It is clear that more than enough people have answered questions in our questionnaire with KMO equal to .934. Also, the fact that Bartlett's test of sphericity is significant (.00 < .05), means that almost no two adjacent variables have correlated with each other highly.

There is also an anti-image covariances matrix with correlation values on its diagonal indicative of sampling adequacy for every single variable. If this value is above .5 for a question, it means that enough people have answered that particular question. This was the case for almost all questions in this questionnaire.

Another piece of information is presented in the Total Variances Explained table. This table can be generated using two methods: first, asking for the extraction of variables explaining variances above one eigenvalue; and second, requesting the program to extract factors equal to the number of questionnaire's sub-scales to see if they explain more than 50% of the variance. Both tables were generated in this analysis. Table 11 shows that extracting four factors accounts for slightly more than 50% of the variance which is just beyond what is needed for considering the questionnaire valid. Appendix K too shows that extracting more factors does not add to the model fit substantially since the additional six factors extracted explain only 14% of the variance. The scree plot (See Appendix L) also shows that four factors before the elbow explain the largest portion of the variance, and in the Appendix M, the rotated component matrix likewise represents factors and questions (items) loading on them. Therefore, we can conclude that the questionnaire has been valid with four factors extracted.

Table 7

Total Variance Explained With Four Factors

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	12.68	20.79	20.79	12.68	20.79	20.79	11.04	18.11	18.11
2	11.16	18.30	39.10	11.16	18.30	39.10	10.27	16.85	34.96
3	4.29	7.04	46.14	4.29	7.04	46.14	4.79	7.86	42.82
4	2.66	4.37	50.51	2.66	4.37	50.51	4.68	7.68	50.51
5	1.84	3.02	53.53						
6	1.67	2.75	56.29						
7	1.41	2.32	58.61						
8	1.27	2.08	60.69						
9	1.16	1.90	62.60						
10	1.07	1.76	64.36						
11	.99	1.63	66.00						

Extraction Method: Principal Component Analysis.

Having provided the results for the Exploratory Factor Analysis (EFA) of the questionnaire, the next necessary step is to measure the confirmation of the factor analysis by applying Confirmatory Factor Analysis (CFA). CFA is a type of structural equation model (SEM) that can measure the relationship between observed variables and latent variables (Brown, 2006). In the study, goodness of fit indices, which are the most frequent in literature, were used. The diagram for Confirmatory Factor Analysis (CFA) is provided as the following:

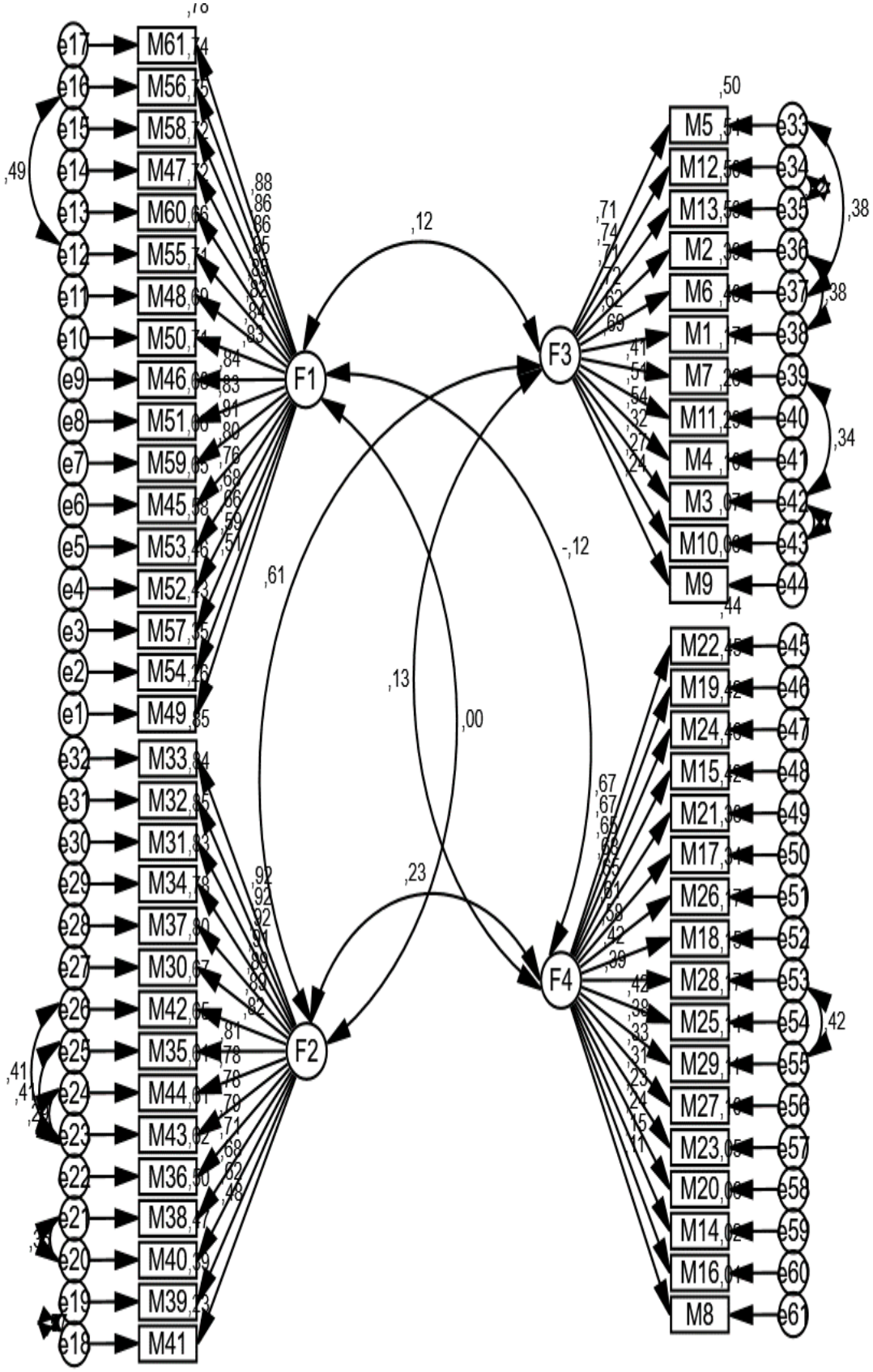

Figure 1. Diagram for confirmatory factor analysis

Criteria regarding the goodness of fit for confirmatory factor analysis are presented in Appendix N, and it was found that the analysis results and the fit statistics calculated with the confirmatory factor analysis were at an acceptable level with the previously determined factor structure of the scale. Standardized factor loads, t values and explanatory (R^2) values formed by the items are provided in Appendix O. Having examined the standardized coefficients, it was found that the factor loads were high, the standard error values were low, and the t values were significant. These results confirm the construct validity of the predetermined factor structure.

3.6 The Main Study

3.6.1 Participants of the main study. The participants of the main study (i.e., the second stage) included 630 EL teachers and administrators working at English preparatory programs in both foundation and state universities all around Turkey. In Turkey, 202 universities functioned over the course of the academic year of 2019/2020, 129 of which were state universities and 73 foundation universities, according to the official website of Council of Higher Education of Turkey. Two hundred seventy-two (272) of the participants were from the foundation universities and 358 were from the state universities. They were reached for the questionnaire study through purposeful sampling at the beginning, and the online questionnaire as the study continued in order to reach a wider demographic (i.e., the complete population) around Turkey. One hundred thirty-one of the participants included administrators, such as vice-managers, level coordinators, and/or unit supervisors, who were in charge of evaluating teachers in their institutions. Four hundred ninety-nine of the participants were regular EL teachers, who taught English on a daily basis in their institutions and had no authority to evaluate teachers. Eighty-two of the participants had a teaching experience between 0-5 years, while 213 participants ranged between 6-10 years, and 335 had more than 10 years of teaching experience. The study included 87 participants with a Bachelor's degree, 191 participants with a Master's degree, 117 participants continuing their Master's degree studies, 94 participants with a Doctorate Degree, and 141 participants continuing their doctoral studies. In terms of gender, 203 were male, and 427 were female. A summary of the demographic information at the second stage is shown in the following table:

Table 8

The Second Stage - Demographic Information of the participants and the Universities: Online Questionnaire

The type of the universities in Turkey	Number
State	129
Foundation	73
Total	202
The number of EL teachers and administrators	
State	358
Foundation	272
Total	630
Position at the university	
EL Teacher	499
Administrator	131
Total	630
Years of teaching experience	
0-5 years	82
6-10 years	213
10+ years	335
Total	630
Educational Background	
Bachelor's Degree	87
Master's Degree	191
Master's Degree (Continuing)	117
PhD	94
PhD (Continuing)	141
Total	630
Gender	
Male	203
Female	427
Total	630

3.6.2 Data collection instruments of the main study. For this stage of the study, the online questionnaire was delivered to all EL teachers and the administrators of the English preparatory programs of the universities in Turkey to gather data for answering the second and the third research questions of the study.

3.6.2.1 Online questionnaire. Based on the content and thematic analysis of the findings of semi-structured interviews and guided written reports, an online questionnaire was designed in order to reach every EL teacher and administrator working at English preparatory programs of the universities in Turkey. The

questionnaire included 61 items at total, and all the items were gathered under four sections (i.e., factors) to search answers for different aspects regarding teacher evaluation. Section 1 was for the opinions of EL teachers and administrators about teacher evaluation; section 2 was for the types of teacher evaluation practices offered to EL teachers; section 3 was for the impacts of teacher evaluation on teacher development; and section 4 was for the teacher evaluation-related factors affecting teacher development.

3.6.3 Data collection procedures of the main study. After gathering opinions about the items, the researcher created the sample questionnaire and conducted a pilot study to make a valid and reliable data collection instrument. Lastly, the researcher gathered data in terms of variables and made a comparative analysis. This part of the study constitutes the quantitative side of the study (the second stage). It has the features of a questionnaire descriptive design as suggested by Fraenkel, Wallen and Hyun (2011).

The data collection procedure of the main study followed a purposeful sampling design to reach the complete population. The researcher downloaded the list of the official websites of all the state and the foundation universities in Turkey from the official website of Council of Higher Education in Turkey (www.yok.gov.tr). Through the official websites of English preparatory programs of the universities, the e-mail addresses of all academic staff were reached. And then the link of the online questionnaire, which was constructed on Google Forms, was sent to every EL teacher and administrator working at English preparatory programs through e-mails. The informed consent form of the online questionnaire was provided to the participants at the beginning of the online questionnaire before they started participating the study (See Appendix B). The form consisted of almost the same information and explanations as those defined in the first stage - what the participants were expected to do and the confidentiality and voluntary nature of the study - they had the right to refuse or stop participating in the study at any time they wished.

3.6.4 Data analysis of the main study. This stage involved the analysis of the data collected by distributing the online questionnaire. As the first step the reliability and validity of the instrument were checked using Cronbach Alpha internal

consistency estimation procedure and factor analyses (i.e., EFA, CFA). Dörnyei (2003) states that multi-item scales are "key components of scientific questionnaire design" (p. 32) and single-items measuring an attitude, belief, opinions or values may not result in expected findings because they could mean different things to different people. "No individual item carries an excessive load; an inconsistent response and one item would cause limited damage" (Skehan, 1989, p. 11). For this reason, the solution is using multi-item scales, which refer to several differently-worded items that focus on the same target (i.e., one attitude, one factor). The questionnaire in this study was divided into four sub-scales as four factors, and each was designed to address the related research questions brought up.

For the research question 2 with its first two sub-scales as the respondents' opinions about teacher evaluation (i.e., Section 1) and the type of teacher evaluation practices applied (i.e., Section 2), after the scores of the similar items and their values were added up, frequency analysis was required, the results of which were calculated and averaged as totals ascendingly. This was the data analysis method used to answer the first two sub-scales of the research question 2. To address the other two sub-scales - the EL teachers' and administrators' views about the state of teacher evaluation in relation to impact of teacher evaluation on professional development (Section 3) and the intervening factors affecting teacher evaluation (Section 4), the respondents' answers needed to be compared with each other after computing the totals of the values assigned. We needed to calculate the mean of the values assigned to each question using the descriptive statistics menu in SPSS and multiply it by number of the respondents. This addition was conducted using the transform menu and SUM command in the Functions and Special Variables area of SPSS.

In order to answer the research question 3 with its sub-scales, firstly the examination of the significant differences of categorical data such as educational background, the type of the university the participants work at, their positions at the universities as EL teachers and administrators, their gender and the applications of systematic or non-systematic teacher evaluation was made by running One-way ANOVAs because these independent variables had more than two levels. On the other hand, the Independent-samples t-tests were used for investigating the possible differences attributable to gender and type of university EL teachers work at since these two independent variables had only two levels. Regarding the only one

continuous data (i.e., years of teaching experience) of this research question was measured by exploring its relationship with the opinions of EL teachers working at Turkish universities by running the Pearson correlation test since both the dependent and the independent variables were continuous.

To sum up, in order to reveal the results for the research questions, depending on the nature of levels of measurement, appropriate descriptive and inferential statistics were used to come up with the deductions and assumptions. The detailed explanations of these measurements were also provided in the results of the study while answering the research questions.

Chapter 4

Results

4.1 Findings of the Research Question 1

The results for the first research question 'What are the indicators of teacher evaluation from the perspectives of EL teachers and administrators working at English preparatory programs of the universities?' aimed to generate baseline for this specific research and eventually create a reliable and valid questionnaire for the target population of the main study. The complementary analyses for the questionnaire such as the reliability and the validity were not discussed in this chapter because they were presented in the methodology section of the study. Under this title, the results to the researh question as the indicators of teacher evaluation from the perpectives of the respondents were presented. Every section of the questionnaire was discussed by the relevant sub-titles just as they were administered to the entire population of the main study. Sections in the questionnaire as the following refer to the purposes for which the research questions were formulated:

- Section 1 = Respondents' Opinions about Teacher Evaluation
- Section 2 = Types of Teacher Evaluation
- Section 3 = Impact of Teacher Evaluation on Teacher Development
- Section 4 = Teacher Evaluation-related Factors in Teacher Development

The following four sub-titles reveal the overall analysis of the qualitative study and the creation of the items and the sections (i.e., constructs, factors) of the questionnaire. Here, complete 61 items were presented with the aim of providing a comprehensive explanation of the process by which the questionnaire was developed. In order to make the process more understandable and concrete, several sample quotes from the participants were also included right after the tables. The items were categorized under four constructs (i.e., section 1, section 2, section 3, section 4) just as they were presented on the original online questionnaire, which may provide a more comprehensible outlook to the process.

4.1.1 Section 1 of the questionnaire. The following table demonstrates the constructs and the items under the theme of *Your Opinions about Teacher Evaluation*, which ask EL teachers to provide their opinions about the characteristics of in-service teacher evaluation in general. The aim is to understand individual opinions and perceptions on in-service teacher evaluation.

Table 9

Numbers and Percentages of the Items Emerging from the Guided Written Reports and Semi-structured Interviews - Section 1 of the Questionnaire

Creation of Titles, Sub-titles and Constructs	Number of Participants	Percentages of Participants
Section 1: Your Opinions about Teacher Evaluation In the questions, *In-Ser TE* refers to In-Service Teacher Evaluation:	(out of 50)	(%)
1. In-Ser TE needs to be continuous and programmed	4	8
2. In-Ser TE needs to be conducted for teacher development	23	46
3. In-Ser TE needs to be conducted for administrative decisions	10	20
4. In-Ser TE needs to be compulsory	2	4
5. In-Ser TE needs to provide quality of instruction	19	38
6. In-Ser TE needs to provide quality of students' learning	6	12
7. In-Ser TE needs to provide quality of administration	17	34
8. In-Ser TE needs to be self-conducted	4	8
9. In-Ser TE needs to be conducted by peers	8	16
10. In-Ser TE needs to be conducted by administrators	12	24
11. In-Ser TE needs to be conducted by trained professionals	4	8
12. In-Ser TE needs to give teachers a scope for future development	11	22
13. In-Ser TE needs to detect teachers' areas of improvement	4	8

Thirteen main items emerged under the theme of perceptions of EL teachers on teacher evaluation, including the ideas that evaluation must be continuous, compulsory, conducted by peers, and/or focus on what teachers need to improve. The following listings provide sample quotes and excerpts from the participants:

Item 2: In-Ser TE needs to be conducted for teacher development. This was mentioned 23 times by the participants, which accounts for 46% of the results. Below are three sample quotes from the participants:

Participant 14: *"It is used for teachers' professional development."*

Participant 32: *"to help teachers in terms of professional development."*

Participant 41: *"Improvement of the teacher and, of course, the institution."*

Item 5: In-Ser TE needs to provide quality of instruction. This was mentioned 19 times by participants, accounting for 38% of results. Below are sample quotes from the participants:

Participant 1: *"The assessment process to decide the quality of the education that a teacher gives."*

Participant 11: *"Evaluating the efficiency of the teacher with regard to their ability to present a subject to engage the students".*

Participant 27: *"In my opinion, evaluation of teachers should be conducted for assessing professional competence."*

Item 12: In-Ser TE needs to give teachers a scope for future development. This was mentioned 11 times by participants, which accounts for 22%. Below are sample quotes from the participants:

Participant 22: *"...So basically, it should not be a judgment process. It should be a positive process, where the teacher who is evaluated actually receives advice on certain issues that the observer believes they need help with."*

Participant 28: *"In teacher evaluation, professional teachers who are experts in this area can help and guide the teacher. They can offer some alternative methods to the teacher to opt for going forward."*

Participant 48: *"In part, teacher evaluation is a process that allows a teacher to reflect on what they have been doing as per a particular situation whether in or outside a classroom in an attempt to provide data that can help guide or shape the conduct of future academic, professional, or personal endeavors."*

4.1.2 Section 2 of the questionnaire. The following table demonstrates the construct and the items under the theme of *Types of Teacher Evaluation*, which aims to reveal the teacher evaluation practices applied in EL teachers' current working context. They are supposed indicate what extent these practices are implemented and seen as part of teacher evaluation appraisal process in their institutions.

Table 10

Numbers and Percentages of the Items Emerging from the Guided Written Reports and Semi-structured Interviews - Section 2 of the Questionnaire

Creation of Titles, Sub-titles and Constructs	Number of Participants	Percentages of Participants
Section 2: Types of Teacher Evaluation	(out of 50)	(%)
14. Peer observations without feedback sessions	2	4
15. Peer observations with feedback sessions	13	26
16. Administrative observations without feedback sessions	18	36
17. Administrative observations with feedback sessions	6	12
18. Observations by external teacher trainers	1	2
19. Observations by internal teacher trainers	5	10
20. Mentoring without observations and feedback	3	6
21. Mentoring with observations and feedback	3	6
22. Regular individual appraisal meetings	2	4
23. Irregular individual appraisal meetings	12	24
24. Teacher appraisal forms	10	20
25. Students' reports on classroom performance	13	26
26. Self-evaluation practices	7	14
27. Administrative follow-chart reports (early leave, late coming, and/or health report)	9	18
28. Attending CELTA/DELTA/MA/PhD programs	4	8
29. Participation in conferences or in-service training courses	9	18

Sixteen main items, ranging from *Item* 14 to *Item* 29, emerged under the theme of types of evaluation offered to instructors. Several items came about from students' evaluations/online surveys, peer observations with/without feedback sessions, or observations from internal/external teacher trainers. The following listings provide sample quotes and excerpts from the participants:

Item 16: Administrative observations without feedback sessions. This was mentioned 18 times by the participants, which accounts for 36% of responses. Below are two sample quotes from the participants:

Participant 13: *"We have at least one class observation each term.*

Participant 35: *"Class visits are conducted by administrative coordinator."*

Item 25: Students' reports on classroom performance. This was mentioned 13 times by the participants, which accounts for 26% of responses. Below are two sample quotes from the participants:

Participant 1: *"End of year students' evaluation; an online platform where students need to evaluate their instructors with multiple choice questions."*

Participant 30: *"Students' observations based on surveys distributed at the end of each term."*

Item 24: Teacher appraisal forms. This was mentioned 10 times by the participants, which accounts for 20% of responses. Below are two sample quotes from the participants:

Participant 2: *"Management reports."*

Participant 50: *"Managers conduct performance reviews."*

4.1.3 Section 3 of the questionnaire. The following table demonstrates the items and the constructs under the theme of *Impact on Teacher Development*, which aims to find out to what extent teacher evaluation practices in EL teachers' context impact teacher development experiences.

Table 11

Numbers and Percentages of Items Emerging from the Guided Written Reports and Semi-structured Interviews - Section 3 of the Questionnaire

Creation of Titles, Sub-titles and Constructs	Number of Participants	Percentages of Participants
Section 3: Impact on Teacher Development In the questions, *In-Ser TE* refers to In-Service Teacher Evaluation:	(out of 50)	(%)
30. In-ser TE helps me develop my teaching skills	8	16
31. In-ser TE helps me evaluate my teaching performance	4	8
32. In-ser TE helps me understand my own teaching	9	18
33. In-ser TE helps me develop continuously	3	6
34. In-ser TE helps me see my weaknesses and strengths	23	46
35. In-ser TE helps me improve student success	6	12
36. In-ser TE helps me meet the pedagogical needs of the institution	5	10
37. In-ser TE increases my self-awareness	6	12
38. In-ser TE supports collaboration with other teachers	7	14
39. In-ser TE supports collaboration with administrative teams	1	2
40. In-ser TE helps me understand the pedagogical perspectives of other teachers	10	20
41. In-ser TE supports understanding the perspectives of administrators	8	16
42. In-ser TE helps preparatory programs improve	4	8
43.In-ser TE helps preparatory programs improve student success	5	10
44. In-ser TE helps preparatory programs achieve pedagogical objectives	6	12

Fifteen main items, ranging from *Item* 30 to *Item* 44, emerged under the theme of impact on teacher development, some of which include for example, teacher evaluation helps me develop my teaching skills, helps me see my weaknesses and strengths, or increases my self-awareness. The following listings provide sample quotes and excerpts from the participants:

Item 32: In-ser TE helps me understand my own teaching. This was mentioned nine times by the participants, which accounts for 18% of responses. Below are two sample quotes from the participants:

Participant 6: *"Teacher evaluation can help detect problematic areas and contribute to improvement."*

Participant 22: *"You can reflect upon the positive criticism… that when you do have a slight problem identified, you are also given suggestions on how to overcome it."*

Item 37: In-ser TE Increases my self-awareness. This was mentioned six times by the participants, which accounts for 12%. Below are two sample quotes from the participants:

Participant 16: *"If the feedback is given in an effective way, it can increase teachers' self-awareness."*

Participant 1: *"it can relate to a self-awareness in instructors."*

4.1.4 Section 4 of the questionnaire. The following table demonstrates the items and the constructs under the theme of *Teacher Evaluation-related Factors in Teacher Development, which* aims to find out the teacher evaluation-related factors that might influence teacher development.

Table 12

Numbers and Percentages of the Items Emerging from the Guided Written Reports and Semi-structured Interviews - Section 4 of the Questionnaire

Creation of Titles, Sub-titles and Constructs	Number of Participants	Percentages of Participants
Section 4: Teacher Evaluation-related Factors in Teacher Development	(out of 50)	(%)
45. Demotivating language used by the evaluator	19	38
46. Lack of professional guidance that teachers need	7	14
47. Lack of trust in evaluators	3	6
48. Lack of guidance on what teachers are expected to do	7	14
49. Stress of being evaluated	18	36
50. Lack of clear criteria of evaluation	8	16
51. Under-qualified professional development staff	9	18
52. Unannounced evaluative practices (e.g., observations)	3	6
53. Non-authentic evaluative practices		
54. Receiving negative feedback	5	10
55. Threatening procedures applied	5	10
56. Judgmental attitudes of evaluators	13	26
57. Fear of contract not renewed	12	24
58. Low quality of feedback given	4	8
59. Sense of favoritism among other teachers	8	16
60. Sense of non-satisfactory evaluation	12	24
61. Unconstructive feedback	10	20

Seventeen main items, ranging from Item 45 to Item 61, emerged under the theme of factors involved in teacher evaluation process; some of which include, for example, demotivating language used by the evaluator, the stress of being evaluated, the lack of clear criteria of evaluation, as well as under-qualified professional development staff. The following listings provide sample quotes and excerpts from the participants:

Item 49: Stress of being evaluated. This was mentioned 18 times by the participants, which accounts for 36% of responses. Below are two sample quotes from the participants:

Participant 10: *"It is a major cause of stress."*

Participant 32: *"If it is not done in a positive way, teachers may feel nervous, judged or under pressure."*

Item 51: Under-qualified professional development staff. This was mentioned nine times by the participants, which accounts for 18% of responses. Below are two sample quotes from the participants:

Participant 5: *"Teacher evaluation, if carried out by the so-called supervisors in a compulsory form, is likely to generate stress and professional dissatisfaction, and, consequently, resistance."*

Participant 16: *"I think in order to give really good constructive feedback, one must be trained and very skilled. For example, I once noticed what my colleague did wrong, but I didn't say as I didn't know how she would feel about it. I didn't feel brave enough to tell her that she made a mistake."*

Item 57: Fear of contract not renewed. This was mentioned 12 times by the participants, which accounts for 24%. Below are two sample quotes from the participants:

Participant 18: *"Summative outcomes such as job dismissal could cause stress on some teachers."*

Participant 50: *"If evaluation is used against teachers such as losing their jobs, then it helps nothing."*

4.2 Findings of the Research Question 2

The research question for the second stage 'What are the EL teachers' and administrators' views about the state of teacher evaluation in relation to the type of evaluation, impact on PD, and the intervening factors?' was answered by administering the questionnaire to the entire population as a part of the main study (i.e., the quantitative study).

To address the first two subscales of this research question, frequency analyses were needed after adding up values assigned to the related questions by the respondents regarding the EL teachers' and administrators' views about the state of teacher evaluation (Section 1) and its relation to the type of evaluation (Section 2).

To address the other two subscales - the EL teachers' and administrators' views about the state of teacher evaluation in relation to impact on PD (Section 3) and the intervening factors (Section 4), the respondents' answers were compared with each other. The addition was conducted using the transform menu and SUM command in

the Functions and Special Variables area of SPSS. Table 13 represents the number of the participants and whether there were any missing cases.

Table 13

Valid and Missing Cases in Each Sub-scale of the Questionnaire

		Section 1	Section 2	Section 3	Section 4
N	Valid	630	630	630	630
	Missing	0	0	0	0

4.2.1 Respondents' opinions about teacher evaluation. To come up with answers for the first subscale (Section 1) of the research question 2, 'Respondents' Opinions about Teacher Evaluation', we need to look at the participants' frequencies of answers. Appendix P is the frequency table and represents the overall values assigned to the questions by the respondents. The values are organized ascendingly in the first column. The second column shows the number of people assigning those overall values to the questions. Therefore, the overall values assigned by two people have reached 27 for example. If we add up numbers in the second column, we will come up with the overall number of the respondents. Columns three and four show these frequencies in percent and the last column shows the cumulative percentage with 100 at the bottom indicating that all respondents have answered all of the questions. The average of the overall scores the respondents have given to this set of questions is given in Appendix Q, which is equal to 50.89. If this value is divided by the number of questions, it will bring up 3.91. This number is indicative of the respondents' strong though not absolute positive opinion towards evaluation.

The histogram and the pie chart for the frequency of overall values assigned to questions about the importance of evaluation give us an impression of the way the values have been distributed (See Appendices R & S). Evidently, the values are clustered almost in the middle slightly tilted towards the right end or high values on the histogram. It also gives us a cross section of the assigned values in terms of their frequencies. In the pie chart, most respondents have assigned above-average values to this question, from which may be deduced as participants had somewhat positive opinions toward teacher evaluation.

In addition to overall frequencies presented above, the following bar chart is the indicator of the frequency counts assigned to every value on Section 1 of the questionnaire. As can be observed, the opinions of the participants are titled toward

the right side. This could again tell us about positive attitudes of the participants toward teacher evaluation under conditions it is conducted for particular purposes such as continuous and programmed for professional development, quality of instruction, student success, and providing both EL teachers and administrators a scope for future improvement. Each bar in this chart represents a value starting from 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree), to 5 (strongly agree).

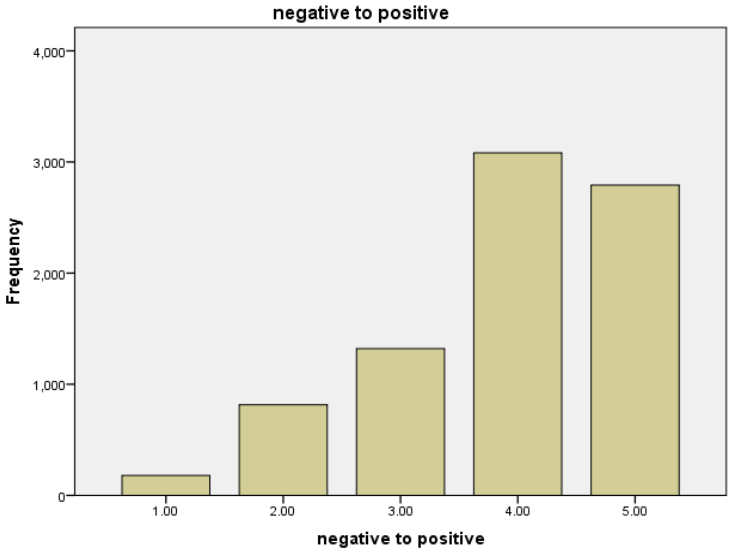

Figure 2. Bar chart of the overall frequency of the values assigned to questions showing the respondents' opinions about teacher evaluation

Lastly, regarding the findings of Section 1, Appendix T specifically provides more detailed picture of the findings for each item and demonstrates the added-up frequencies and percentages of the related items. Positive opinions stand for the items *'strongly agree'* and *'agree'*; neutral opinion for *'neutral'*; and negative opinions for *'strongly disagree'* and *'agree*' on the scale. As can be seen on the table, all of the items except item 10 'In-Ser TE needs to be conducted by administrators' with 282 participants (44.8%) on this section were regarded as positive elements to be considered for teacher evaluation and the participants were someway optimistic with fluctuating frequencies above the average.

If it was to be more specific, a) item 1 'In-Ser TE needs to be continuous and programmed' with 543 participants (86.2%), b) item 2 'In-Ser TE needs to be conducted for teacher development' with 596 participants (91.4%), c) item 5 'In-Ser TE needs to provide quality of instruction' with 686 participants (93.1%), d) item 6

‘In-Ser TE needs to provide quality of students’ learning’ with 556 participants (88.3%), e) item 11 ‘In-Ser TE needs to be conducted by trained professionals’ with 571 participants (90.6%), f) item 12 ‘In-Ser TE needs to give teachers a scope for future development’ with 609 participants (96.7%), g) item 13 ‘In-Ser TE needs to detect teachers' areas of improvement’ with 594 participants (94.2%) were regarded as positive domains for teacher evaluation.

4.2.2 Respondents’ opinions on types of teacher evaluation. To continue with the second subscale of research question 2 regarding “the types of teacher evaluation practices”, we needed to follow the same line of analysis.

In this part, the expression of Section 2 is equivalent to the types of evaluation, while the purpose is to figure out the kinds of evaluation that are implemented more. By types of evaluation, as questions 14 to 29 in the questionnaire reveal, it is meant, for example, whether the evaluation is done by a peer, an administrator, an external agent or a teacher trainer and also whether feedback is provided to teachers or not. Like frequency table for the previous research question, ascending numbers in the first column of Appendix X show the overall values assigned to the questions with the numbers in the second column pointing to their frequencies. Other columns show the percentages of the assigned values. To find out what answers the respondents have given to this section in terms of evaluation types, we need to look at the overall values they have accorded to questions and their frequencies. From Appendix X, we can see that the overall values of 25 to 45 have the highest frequencies. Each item in the table represents a value starting from 1 (never), 2 (seldom), 3 (sometimes), 4 (often), to 5 (always). The implication of this finding is that the frequency counts measurement of the type of evaluation practices carried out at the preparatory programs is not considerably different from each other; or in other words, none is more frequently-used from the perspectives of the respondents even though there are cases that have attributed slightly high values to this variable.

The histogram of the distribution of the respondents’ answers is given in Appendix Y. Unlike the former histogram, this histogram is slightly tilted to the left, meaning that the majority of the respondents have attributed less-than-average values to the types of teacher evaluation.

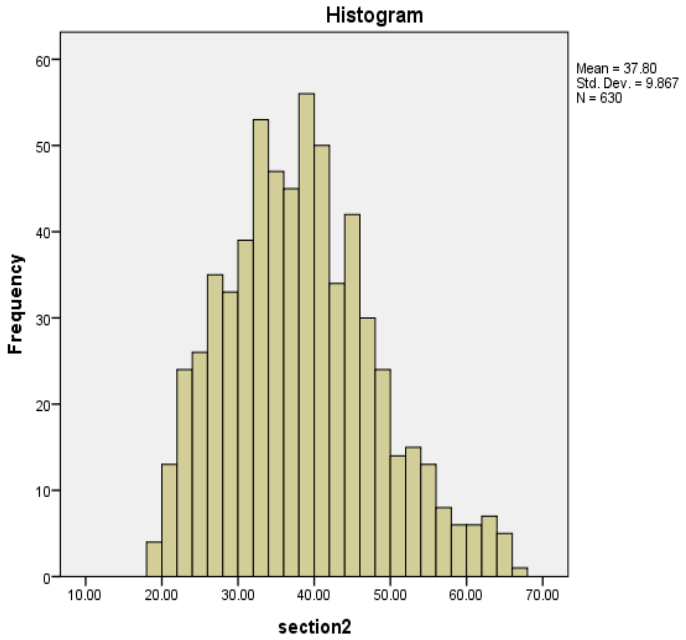

Figure 3. Histogram of the frequency of overall values assigned to questions about the type of evaluation

Similarly, as in the case of the previous subscale, we can calculate the average of the overall values assigned to the type of evaluation and divide it by the number of questions to come up with a representative measure of the types of evaluation (See Appendix Z). In this case, if we divide 37.80 by 16, which is the number of questions addressing this issue, the representative value will be 2.36, which is equal to a moderate average for the applications of types of evaluation (i.e., sometimes, seldom).

In addition to overall understandings presented above, the following tables 14 and 15 are the indicators of the frequency counts assigned to each value for Section 2. As can be observed, the opinions of the participants are mostly gathered around the values three, two, and one (sometimes, seldom, and never respectively) by being related to the types of teacher evaluation applied in the institutions. (a) Peer and administrative observations without feedback sessions, (b) mentoring without observations and feedback and (c) irregular individual appraisal meetings are the least frequently applied practices in the institutions, the results could again mostly draw our attention to negative-directed answers of the respondents toward these teacher evaluation types; that is, the mean scores of these items were below the average.

When it comes to the individual items with mean scores above or equal to the overall mean; (a) teacher appraisal forms, (b) students' reports on classroom performance, (c) classroom observations by administrators, internal trainers, mentors or peers with feedback sessions, (d) self-evaluation practices, (e) attending CELTA / DELTA / MA / PhD programs, and lastly (f) participation in conferences or in-service training courses may be seen as the practices that are more frequently applied ones in the institutions to evaluate teachers. However, all these results have close values to the mean score, which again draws our attention to the point that there is no specific type of teacher evaluation that is more frequently used than the others in the institutions.

Table 14

Mean Scores of each Item for the Types of Teacher Evaluation

	N	Mean
Questions	10080	35.75
Types of teacher evaluation	10080	2.23
M 14. Peer observations without feedback sessions	630	1.87
M 15. Peer observations with feedback sessions	630	2.34
M 16. Administrative observations without feedback sessions	630	1.58
M 17. Administrative observations with feedback sessions	630	2.17
M 18. Observations by external teacher trainers	630	1.69
M 19. Observations by internal teacher trainers	630	2.11
M 20. Mentoring without observations and feedback	630	1.83
M 21. Mentoring with observations and feedback	630	2.24
M 22. Regular individual appraisal meetings	630	2.20
M 23. Irregular individual appraisal meetings	630	1.89
M 24. Teacher appraisal forms	630	2.40
M 25. Students' reports on classroom performance	630	2.95
M 26. Self-evaluation practices	630	2.48
M 27. Administrative follow-chart reports (early leave, late coming, and/or health report)	630	2.33
M 28. Attending CELTA/DELTA/MA/PhD programs	630	2.76
M 29. Participation in conferences or in-service training courses	630	2.91
Valid N (listwise)	630	

Table 15

Frequency Counts for Individual Values for the Types of Teacher Evaluation

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1. Never	3426	33.9	33.9	33.9
	2. Seldom	1683	16.7	16.7	50.6
	3. Sometimes	2299	22.8	22.8	73.4
	4. Often	1508	14.9	14.9	88.3
	5. Always	1164	11.7	11.7	100.0
Missing	System	0.0	0.0		
Total		10080	100.0	100.0	

Apppendix AA is also to provide the overall frequencies and percentages of the items for Section 2 on the questionnaire. The result here provides answers of the participants for each type of teacher evaluation practices, which are applied in the institutions. As can be observed from the individual item-based results on this table, no specific type of evaluation can be thought to be used more frequently than the others. Other related details can also be observed on the bar chart, the histogram and the item-based frequency table (See Appendices BB, CC and DD respectively).

In terms of types of observations, when we have a look at items individually, mostly the ones with feedback sessions are preferred to the others without feedback sessions although their frequencies are again too low. Observations by (a) administrators (22.8%), (b) internal teacher trainers (24.8%), (c) mentors (26.2%) and (d) peers with feedback sessions (27.1%) are the most preferred types of observation at English preparatory programs. The following items seem almost never to be used frequently in the institutions: (a) peer observations without feedback sessions (7.4%), (b) administrative observations without feedback sessions (5.2%), (c) observations by external teacher trainers (11.9%), (d) mentoring without observations and feedback (12%), (e) irregular individual appraisal meetings (9%).

In terms of the other types of teacher evaluation, the results were as the following; (a) participation in conferences or in-service training courses (54.4%), (b) students' reports on classroom performance (50.6%), (c) attending CELTA / DELTA / MA / PhD programs (46.4%), (d) self-evaluation practices (35.9%), (e) teacher appraisal forms (33.2%), (f) administrative follow-chart reports (early leave, late coming, and/or health report) (30.8%), and (g) regular individual appraisal meetings (26.2%).

4.2.3 The impact of teacher evaluation on teacher development. The third subscale (Section 3) of the research question 2 in this category was: In which areas of development do the teacher evaluation practices help EL teachers develop professionally? This is where frequency counts cannot help us too much and instead of looking into which value was assigned more, we need to compute totals of the values assigned to the questions and compare them with each other. To do so we first should know which questions address this issue, that is the professional development of EL teachers. The questions in our questionnaire addressing this issue were items 31 to 45. So, as the first step we have to add up the scores that all of the 630 respondents have given to each question. To this end, we only need to calculate the mean of the values assigned to each question using the descriptive statistics menu in SPSS and multiply it by number of the respondents. The means and totals therefore are presented on Appendix EE.

To know which variables have impacted more on the EL teachers' professional development from their own points of view, we can simply examine the same table and sort the means either descendingly or ascendingly and match them up with their corresponding questions. Table 16 shows the mean values sorted descendingly, as well as the questions for which the mean values are calculated.

Table 16

Questions About EL Teachers' Professional Development and Their Mean Values Sorted Descendingly

Q37	4.09	Helps me meet the pedagogical needs of the institution
Q34	4.06	Helps me develop continuously
Q32	4.03	Helps me evaluate my teaching performance
Q31	4.02	Helps me develop my teaching skills
Q42	3.94	Helps me understand the perspectives of administrators
Q33	3.93	Helps me understand my own teaching
Q45	3.92	Helps preparatory programs achieve their pedagogical objectives
Q40	3.81	Supports partnership with administrative teams
Q38	3.79	Increases my self-awareness
Q44	3.76	Helps preparatory programs improve student success
Q36	3.73	Helps me improve student success
Q43	3.72	Helps preparatory programs improve
Q35	3.70	Helps me see my weaknesses and strengths
Q41	3.45	Helps me understand the perspectives of other teachers
Q39	3.43	Supports collaboration with other teachers

Since all of the mean values are above 2.5, it is clear that all variables in the questions are helpful to the professional development of EL teachers. However, if we review the content of the questions, we will find out more about the impact of the variables in the question and the way the participants reacted to them. Furthermore, a scree plot can represent the participants' reactions to the questions more expressively (See Appendix FF). The scree plot indicates that other than the items 37, 34, 32, and 31 respectively presented in Table 16, the impacts of others are almost the same. The most outstanding variable but is (a) item 37 'TE helps me meet the pedagogical needs of the institution'; and the other most effective items are (b) item 34 'TE helps me develop continuously', (c) item 32 'TE helps me evaluate my teaching performance', and (d) item 31 'TE helps me develop my teaching skills'.

4.2.4 Teacher evaluation-related factors affecting teacher development. Research question 2 at this stage was formulated to explore the intervening factors that affect professional development of EL teachers by teacher evaluation practices (i.e., Section 4). Of course, all of the questions in the questionnaire addressing this issue allude to elements that naturally affect professional development through teacher evaluation in a negative direction. Some of these elements are demotivating language by the evaluator, lack of professional guidance, lack of trust in evaluators, and stress of being evaluated. It is clear that all of these factors would undermine the efficacy of teacher evaluation. But we can follow a procedure very much similar to our procedure in answering the third section to see which factors are more detrimental to teacher evaluation from the perspective of the respondents. Thus, as in the previous analysis, we should add up the values assigned to these questions to get the total values and then calculate the means of the assigned values. Appendix GG shows the questions, the assigned values' totals, and their means.

In Table 17, the questions and the means of the values assigned to them are organized descendingly. This kind of organization makes it easier for us to figure out which elements are more detrimental to teacher evaluation practices from the perspective of the respondents. Understanding the impact of these factors and the degree to which they affect evaluation practices negatively by using a scree plot would be helpful, as in the case of the third question. The scree plot is a visualization of the effects of the factors but at the same it shows how different these impacts are from each other, that is whether there are any factor(s) whose effects are

substantially different from others (See Appendix HH). What the scree plot reveals is that the single most detrimental factor is a) item 49 'lack of guidance on what teachers are expected to do'. In addition, b) item 46 'demotivating language used by the evaluator, c) item 51 'lack of clear criteria of evaluation, d) item 47 'lack of professional guidance that teachers need', and e) item 50 'stress of being evaluated' could be seen as having significant effects on teachers negatively. Other factors though detrimental are not significantly different from each other in the negative influence that they exert. However, a) item 54 ‘non-authentic evaluative practices’, b) item 57 ‘judgmental attitudes of evaluators’, and c) item 52 ‘under-qualified professional development staff’ could be the three least detrimental.

Table 17

Questions Addressing Teacher Evaluation-related Factors Affecting Teacher Development, their Mean Values Sorted Descendingly

Q49	4.10	Lack of guidance on what teachers are expected to do
Q46	4.09	Demotivating language used by the evaluator
Q51	4.09	Lack of clear criteria of evaluation
Q47	4.07	Lack of professional guidance that teachers need
Q50	4.07	Stress of being evaluated
Q61	4.03	Sense of non-satisfactory evaluation
Q48	4.02	Lack of trust in evaluators
Q56	4.01	Threatening procedures applied
Q58	3.97	Fear of contract not renewed
Q59	3.97	Low quality of feedback given
Q60	3.93	Sense of favoritism among teachers
Q53	3.90	Unannounced evaluative practices (exp. observations)
Q55	3.90	Receiving negative feedback
Q52	3.76	Under-qualified professional development staff
Q57	3.70	Judgmental attitudes of evaluators
Q54	3.58	Non-authentic evaluative practices

4.3 Findings of the Research Question 3

The third stage of the analysis intended to uncover variations. These differences may stem from many sources. Notable among them include years of teaching experience, educational background, type of university the EL teachers work at, their positions at universities, their gender, and finally the application of systematic or non-systematic teacher evaluation. The question therefore can be formulated as follows:

3. How do EL teachers and administrators view teacher evaluation in relation to their experience, educational background, context of work, person's role, gender, and specific evaluation practice?
   3a. years of teaching experience
   3b. educational background
   3c. the type of university they work at
   3d. their positions at the universities
   3e. their gender
   3f. the applications of systematic or non-systematic teacher evaluation

This question is multifaceted. In fact, it involves six independent variables and one dependent variable. Five of the independent variables are categorical while one of them, namely years of teaching experience, is continuous. Regarding the continuous variable, we have no way but to explore its relationship with the opinions of EL teachers working at Turkish universities rather than seeking differences between their opinions. Differences among the participants of the study in terms of other variables can be answered by running One-way ANOVAs and Independent-samples t-tests depending on their levels. The Independent-samples t-tests will be used for investigating possible differences attributable to gender and type of the university the EL teachers work at since these two variables have only two levels.

4.3.1 Opinions in relation to years of teaching experience. The following table shows results of the Pearson correlation run between the years of teaching experience of the participants and totals of the values they have assigned to the questionnaire.

Table 18

Relationship Between Years of Teaching Experience and Opinion About Teacher Evaluation

		Years of teaching experience in your profession	totals
Years of teaching experience in your profession	Pearson Correlation	1	.00
	Sig. (2-tailed)		.83
	N	630	630
Totals	Pearson Correlation	.00	1
	Sig. (2-tailed)	.83	
	N	630	630

Table 18 is indicative of a very small Pearson correlation value (r=.00, p=.83 >.05) that can be obtained by chance only. The finding thus convinces us to categorically reject the existence of any relationship between the participants' years of teaching experience and their opinions about teacher evaluation.

To answer the second part of the question, it is necessary to check for the normality of distribution of the scores in the dependent variable, i.e., totals, and homogeneity of variances of the groups. The second assumption is checked by default as part of the analyses (t-test and ANOVA) but for the normality we can run either the Shapiro-Wilk or 1-sample K-S test. Both of these tests are available through the Explore menu in SPSS, the results of which are presented below in Table 20. And Table 19 provides case processing summary of the dependent variable.

Table 19

Case Processing Summary of the Dependent Variable

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Totals	630	100.0%	0	0.0%	630	100.0%

Table 20

Normality Tests

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Totals	.03	630	.055	.99	630	.31

a. Lilliefors Significance Correction

The results of both normality tests are non-significant (statistic for K-S=.03 & statistic for Shapiro-Wilk=.99, df=630, p1=.055 & p2=.31), pointing to the typical nature of the distribution. The following bell-shaped curve fitted on the totals' distribution too is indicative of its normality.

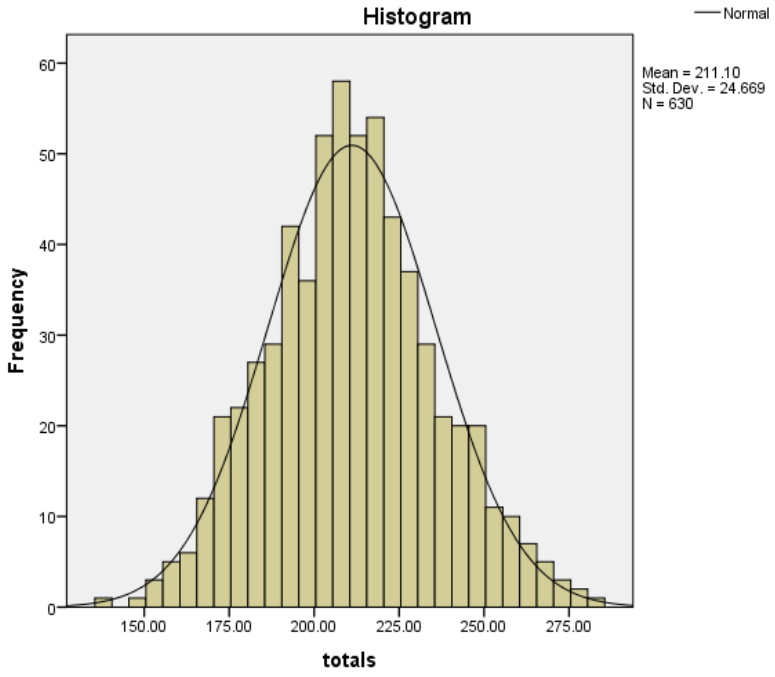

Figure 4. Normal curve representing the dependent variable's (totals') normal distribution

4.3.2 Opinions in relation to gender. As mentioned earlier, five of the independent variables are categorical. Of these categorical variables, gender and type of the university, have two levels whereas the other three variables have more than two levels. To compare the levels of these independent variables with each other and to see if they are different from each other in terms of the participants' opinions about teacher evaluation, we need to run Independent-samples T-tests for gender and type of university and three One-way ANOVAs for the others. Of course, we can run a single three-way ANOVA to investigate these three last variables' levels but since we are not interested in the interaction between the variables, and interpreting multiple-way ANOVAs is difficult, it would be better to run three separate One-way ANOVAs.

First, we can see if gender makes any difference in the opinions of the EL teachers by running an Independent-samples t-test. Table 21 presents group statistics.

In other words, it tells us how many males and females have answered the questions and what the means and standard deviations of their responses are.

Table 21

Group Statistics

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Totals	Male	203	208.48	25.30	1.77
	Female	427	212.34	24.29	1.17

Table 22 is the Independent-samples t-test table. In this table in the lower part, we have to first look in the first row at the first significant value which is the result of Levene's test of homogeneity of variances. If this value is above .05, we can assume that the groups have been homogeneous and proceed with our reading by looking at the second significance value, which is the result of the t-test. If this sig value is smaller than .05, we can conclude that there has indeed been a difference. If not, we should conclude the two groups have not been different in terms of their opinions about teacher evaluation. It is easy to see that the sig or p-value for the t-test is larger than .05 (df=628, p=.06) and therefore, we can conclude that there is not a statistically significant difference in the participants' opinions based on their gender.

Table 22

Independent-samples T-test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Totals									
Equal variances assumed	.087	.76	-1.84	628	.06	-3.86	2.09	-7.98	.26
Equal variances not assumed			-1.81	382	.07	-3.86	2.12	-8.04	.32

4.3.3 Opinions in relation to type of university. Regarding the subscale of research question, in this case, we are concerned with the type of university the EL teachers work at. Therefore, to figure out the possible difference, we have to run an Independent-samples t-test again. Table 23 represents the group statistics.

Table 23

Group Statistics

	Type of University	N	Mean	Std. Deviation	Std. Error Mean
Totals	Foundation	272	212.71	25.555	1.549
	State	358	209.87	23.936	1.265

In the first row of the second half of Table 24, we can see that the assumption of homogeneity of variances is met because Levene's statistic is non-significant. But, like gender, the difference caused by this variable too proves non-significant (df=628, p=.15), meaning that the kind of university the EL teachers responding to the questionnaire work at has not been influential in their attitude toward teacher evaluation.

Table 24

Independent-samples T-test

Totals	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Equal variances assumed	1.60	.20	1.43	628	.15	2.84	1.98	-1.04	6.73
Equal variances not assumed			1.42	562.85	.15	2.84	2.00	-1.08	6.77

4.3.4 Opinions in relation to educational background. The fourth subscale in this series concerned the participants' educational background. There were five levels to this variable and therefore a One-way ANOVA was run to see if the levels made any difference to the participants' opinions about teacher evaluation. Table 25 shows descriptive statistics of the levels.

Table 25

Descriptive Statistics of the Participants' Educational Backgrounds Totals

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Bachelor's Degree	87	203.35	29.44	3.15	197.08	209.63	138.00	281.00
Master's Degree (Cont)	117	210.66	22.39	2.07	206.56	214.76	162.00	272.00
Master's Degree	191	211.02	24.38	1.76	207.54	214.50	154.00	276.00
PhD (Cont)	141	215.00	24.37	2.05	210.94	219.05	146.00	279.00
PhD	94	213.10	22.28	2.29	208.54	217.67	171.00	271.00
Total	630	211.10	24.66	.98	209.17	213.03	138.00	281.00

Table 26 shows the results of Levene's test of homogeneity of variances. In this case the significance level is smaller than .05 meaning that the variances of the groups have not been very similar and the assumption is violated. But ANOVA is robust against violations of this assumption and the only problem that may occur in situations like this is that the test loses its power. That is, the finding may not be significant while in fact there is a genuine significant difference.

Table 26

Levene's Test of Homogeneity of Variances Totals

Levene Statistic	df1	df2	Sig.
2.59	4	625	.03

The main table in a One-way ANOVA test is the ANOVA table. The sig value in this table indicates if the levels of the independent variable are different from each other or not. This value in our test for the levels of educational background points to the existence of a difference because it is smaller than .05 (F=3.23, df=4, 625, p=.01). However, ANOVA is an omnibus test and does not reveal which levels have been different from each other. To find out about this, we have to run a post-hoc test and look for significant values for the differences between groups.

Table 27

Result of ANOVA Test for Educational Background Effect on Teachers' Opinions About Teacher Evaluation

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	7762.94	4	1940.73	3.23	.01
Within Groups	375021.75	625	600.03		
Total	382784.70	629			

The post-hoc test conducted for locating the difference between levels of the independent variable was Tukey HSD (Table 28). Reading the lines of the table indicates that the differences in the opinions of the EL teachers about teacher evaluation have been between teachers possessing bachelors and PhD (Cont) degrees and teachers possessing bachelors and PhD degrees. In other words, the significance in the ANOVA table is the result of difference in the opinions of these three groups of EL teachers. The means plot figure also indicates schematically that the differences, which have reached a significance level (p=.00; p=.05), have been between the EL teachers having bachelor's degree and those with a PhD and PhD (Cont.) degrees (See Appendix II).

Table 28

Post-hoc Test Run to Locate Which Educational Levels Had Different Ideas About Teacher Evaluation-Dependent Variable: Totals Tukey HSD

(I) Educational background	(J) Educational background	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Bachelor's Degree	Master's Degree (Cont)	-7.31	3.46	.21	-16.79	2.17
	Master's Degree	-7.66	3.16	.11	-16.33	.99
	PhD (Cont)	-11.64*	3.33	.00	-20.77	-2.50
	PhD	-9.75	3.64	.05	-19.71	.21
Master's Degree (Cont)	Bachelor's Degree	7.31	3.46	.21	-2.17	16.79
	Master's Degree	-.35	2.87	1.00	-8.22	7.50
	PhD (Cont)	-4.33	3.06	.61	-12.71	4.04
	PhD	-2.43	3.39	.95	-11.72	6.84
Master's Degree	Bachelor's Degree	7.66	3.16	.11	-.99	16.33
	Master's Degree (Cont)	.35	2.87	1.00	-7.50	8.22
	PhD (Cont)	-3.97	2.71	.58	-11.41	3.46
	PhD	-2.08	3.08	.96	-10.52	6.36

Table 28 (cont.d)

(I) Educational background	(J) Educational background	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
PhD (Cont)	Bachelor's Degree	11.643*	3.339	.00	2.50	20.77
	Master's Degree (Cont)	4.333	3.063	.61	-4.04	12.71
	Master's Degree	3.973	2.719	.58	-3.46	11.41
	PhD	1.893	3.261	.97	-7.02	10.81
PhD	Bachelor's Degree	9.750	3.644	.05	-.21	19.71
	Master's Degree (Cont)	2.439	3.392	.95	-6.84	11.72
	Master's Degree	2.080	3.086	.96	-6.36	10.52
	PhD (Cont)	-1.893	3.261	.97	-10.81	7.02

*. The mean difference is significant at the 0.05 level.

4.3.5 Opinions in relation to teachers' positions at the universities. The next variable to be investigated was EL teachers' roles and positions at their institutions and if they made any difference in their opinions. In this case the independent variable had three levels, the descriptive statistics of which are given in Table 29.

Table 29

Descriptive Statistics of EL teachers' Roles Totals

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
EL Teacher	501	209.76	24.99	1.11	207.57	211.95	138.00	279.00
Administrator	17	215.82	23.29	5.64	203.84	227.80	185.00	255.00
Both of them	112	216.35	22.75	2.14	212.09	220.61	154.00	281.00
Total	630	211.10	24.66	.98	209.17	213.03	138.00	281.00

The Levene's test of homogeneity of group variances (Table 30) shows that the homogeneity assumption was satisfied because it is larger than .05 (statistic=1.28, df=2. 627, p=.27).

Table 30

Levene's Test of Homogeneity of Variances for EL teachers' Roles Totals

Levene Statistic	df1	df2	Sig.
1.28	2	627	.27

The ANOVA table reveals that there has been a difference but as in the case of educational degree, we were in need of running a post-hoc test to know exactly where the difference lied.

Table 31

Result of ANOVA Test for EL Teachers' Roles Effect on Their Opinions About Teacher Evaluation Totals

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	4368.30	2	2184.15	3.619	.02
Within Groups	378416.39	627	603.53		
Total	382784.70	629			

The post-hoc test reveals that the difference has been between the practitioners performing only as EL teachers and practitioners having both the roles of EL teachers and administrators (p=.02). The figure, which is the means plot of the roles the EL teachers played, is an additional visualization of the difference that we spotted (See Appendix JJ).

Table 32

Post-hoc Test Run to Locate Which Roles Caused Differences in Ideas About Teacher Evaluation

Dependent Variable: Totals

Tukey HSD

(I) role	(J) role	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
EL Teacher	Administrator	-6.05	6.05	.57	-20.29	8.17
	Both of them	-6.59*	2.56	.02	-12.62	-.56
Administrator	EL Teacher	6.05	6.05	.57	-8.17	20.29
	Both of them	-.53	6.39	.99	-15.55	14.48
Both of them	EL Teacher	6.59*	2.56	.02	.56	12.62
	Administrator	.53	6.39	.99	-14.48	15.55

*. The mean difference is significant at the 0.05 level.

4.3.6 Opinions in relation to the application type of teacher evaluation. The final question in this series asked about the effect of the application of systematic or non-systematic teacher evaluation practices. This variable involved three levels: systematic evaluation, random evaluation, and no evaluation. The results of the One-

way ANOVA run on this categorical variable and its concomitant dependent continuous variable are presented in the last four tables and the means plot (See Appendix KK). Total mean scores for the descriptive statistics of the kind of evaluation is also presented on Appendix LL. Appendix MM provides the test of homogeneity of variances for this subscale with a significance level of p=.0.03 and so does the Appendix NN for the result of ANOVA test to investigate the effect of the kind of teacher evaluation practices on EL teachers' opinions between groups with a significance level of p=.01.

The last interpretation of this subscale together with the other tables and the means plot in appendix are exactly the same as interpreting other tables and figures of ANOVA that we dealt with earlier. In one sentence, as can be observed on table 33, the kind of evaluation makes a difference and the difference is between systematic and non-systematic evaluation from the perspective of the EL teachers queried in this study (p=.00).

Table 33

Post-hoc Test Run to Locate What Kinds of Evaluation Made Differences in the Opinions of EL Teachers About Teacher Evaluation

Dependent Variable: Totals

Tukey HSD

(I) Systematic Evaluation	(J) Systematic Evaluation	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Systematic evaluation	No systematic evaluation	6.92*	2.31	.00	1.48	12.37
	Random evaluation	2.61	2.42	.52	-3.06	8.30
No systematic evaluation	Systematic evaluation	-6.92*	2.31	.00	-12.37	-1.48
	Random evaluation	-4.31	2.64	.23	-10.52	1.90
Random evaluation	Systematic evaluation	-2.61	2.42	.52	-8.30	3.06
	No systematic evaluation	4.31	2.64	.23	-1.90	10.52

*. The mean difference is significant at the 0.05 level.

To sum up this section, a) gender and b) type of the university the EL teachers work at have no significant impact on EL teachers' opinions about the importance or effect of teacher evaluation. However, a) educational level, b) the positions that EL teachers have at the universities, and c) whether the evaluation practices are systematic or not do have a significant effect.

Chapter 5

Discussion and Conclusion

5.1 Introduction

As found through the findings of this specific study, no specific type of evaluation is used more frequently than the others. For teacher evaluations to both inform on EL teachers' strengths and weaknesses and boost professional development, they need to possess a number of interrelated features. These features are necessary in order to make the process more reliable. More reliable evaluation models could help us make better-informed decisions for EL teachers via more constructive feedback. One of these features is consistency just as Darling-Hammond (2015) states there is a need for consistent standards of good practice in teacher evaluation. By consistency, we mean a model that can bring about the same or similar results wherever applied, with results showing less fluctuations between evaluations, which again draws our attention to another finding suggesting implementations of systematic teacher evaluation practices. The model should also be guarded against bias such as lack of guidance on what EL teachers are expected to do and demotivating language used by evaluators. Prejudicial evaluations undeservedly inflate or deflate the ratings that EL teachers receive and this may damage trust in the process (Howard, 2012; Riera, 2011). Trust, therefore is the natural next feature required because another two findings of the study suggest that teacher evaluation needs to avoid lack of clear criteria of teacher evaluation while it is considering the pedagogical needs of the institution and continuous development of EL teachers. If teachers are not satisfied with the results of an evaluation, their voices should be heard and their complaints taken into consideration (Kamali, 2014; King, 2015).

There are various evaluation models each with their own specific characteristics. These include, for example, value-added model (VAM) which heavily relies on students' performance in exams; teacher observation models (e.g., the Marshall Rubrics) which rely on observing teachers' performance in the

classroom; as well as the Framework Model which judges EL teachers' performance based on their preparation and planning, the classroom environment that they create, their instruction methodology and their commitment to their professional responsibilities. All have their own specific gains and pitfalls and are mostly designed to evaluate teachers only by particular aspects without considering either their opinions and needs/interests or their needs in terms of professional development and future ambitions. If to provide a clearer frame to these limited scopes of teacher evaluation models, a sample study on the effectiveness of VAM was previously discussed; according to Darling-Hammond (2015), evaluating teachers by only student success on tests would be problematic because school factors, prior schooling and teachers, home factors or other individual factors - such as student needs, health or attendance - may intervene.

Having considered all these variables, final decisions on teacher effectiveness had better not be given by specific teacher evaluation models because teachers account for only a limited share of variation in student achievement. They need to gain more experience on their teaching competency or professional development. As can be concluded, evaluating teachers by teacher evaluation models, which work within a limited scope, may result in reduced quality.

Instead of focusing on theoretical assumptions, adaptation of teachers to innovations, or descriptive attitude studies by which teacher development had been searched, it has been understood that there is a need for platforms where teachers have the chance to implement studies, experience collaboration, make research, receive academic guidance (e.g., mentoring), get feedback, or develop their professional practices (Guskey, 2000). As a result of such an approach, the effectiveness of educators in terms of their professional career will increase, and student success will be well-affected. Quality of education will matter more than the quantity of evaluative practices. This is much desired, as stated by Goldhaber (2010), "the effect of increases in teacher quality swamps the impact of any other educational investment" (p. 1), while from another perspective; "more can be done to improve education by improving the effectiveness of teachers than by any other single factor" (Sanders, Wright, & Horn, 1997, p. 63); and "systems that help teachers improve and that support timely and efficient personnel decisions have more than good instruments" (Darling-Hammond et al., 2012, p. 12).

The Marzano as well as the Danielson models focus on teacher evaluation practices that are research-based and apply standards to evaluation by taking numerous competencies of EL teachers into consideration (Danielson & McGreal, 2000; Marzano, 2012; Marzano et al., 2011). All of these models have advantages and disadvantages. For example, an evaluation which is based on observation is dependent on the observer's skill and is as good as the observer himself or herself. Some of the evaluation procedures are also time-consuming and expensive (Kandil, 2011; Miller & Young, 2007; Moss, 2015). However, in line with the summary of the findings that emerged from this study (i.e., table 34), it is so apparent that teacher evaluation is part-and-parcel of any effective teaching program and guarantees the quality of the education that students receive. For these reasons and others that might have been slipped of the attention of the current researcher, it is suggested that all of the advantages of the mentioned models and the findings of this study be incorporated into a single model and an effective, low-cost, less labor-intensive, and more fast-paced development-oriented teacher evaluation model (DTEM) be created. By taking all these into account, the development of teachers maintains a powerful place if it is expected from teachers to take steps and actively participate in teacher evaluation practices (Hurley, 2013; Marzano & Todd, 2014).

5.2 Discussion of Findings of the Research Questions

Having analyzed all the data that was obtained from the complete study, eight overall findings emerged. Each finding was actually related to the research questions with their sub-categories as presented on Table 34. At this point, we must combine our findings and create a to-do list moving forward. A summary of the findings so far is:

Table 34

Summary of the Findings That Emerged From the Study

No.	Overall Findings
1	EL teachers have a positive view of teacher evaluation under specific conditions.
2	The type of teacher evaluation practices is not highly important from the perspectives of EL teachers.
3	The most important purposes of teacher evaluation from the perspectives of EL teachers include fulfilling the pedagogical needs of the institution, being conducted by professionals, supporting student achievement, enabling the continuous development of teachers, facilitating the self-evaluation of teachers and developing teaching skills - although other purposes are considered of a certain importance.
4	All detrimental factors, and in particular lack of guidance on what EL teachers are expected to do, demotivating language used by evaluators, stress of being evaluated, and lack of clear criteria of teacher evaluation should be avoided.
5	The degree of teaching experience, gender and the type of university have no impact on the opinions of EL teachers towards teacher evaluation.
6	The educational background of EL teachers makes a difference in their opinions on teacher evaluation.
7	EL teachers' positions/roles in their institutions as instructors, coordinators, trainers and/or administrators make a difference in their opinions towards teacher evaluation.
8	Whether teacher evaluation is implemented systematically or non-systematically makes a difference.

5.2.1 EL teachers' positive views on teacher evaluation. The first question of the study regarded the opinions of EL teachers and administrators regarding evaluations. By reading the literature and compiling guided written reports and semi-structured interviews with the participants in the first qualitative stage of the study, indicators were formed with the aim of creating an online questionnaire to be distributed to respondents to fit a Turkish academic context. Having read the extended literature, open-ended questions correspondingly helped the researcher create the structure of the questionnaire.

Research question 2 aimed to investigate the EL teachers' and administrators' views about the state of teacher evaluation in relation to the type of evaluation, impact on PD, and the intervening factors. Although the results did not show that EL teachers have a strong positive approach towards teacher evaluation, their opinions seem to turn to the positive side depending on the implementation of specific evaluative practices. The overall understanding of the findings of Section 1 can be summarized if we aim to draw a clear picture of the case. Focusing on the positive and the negative opinions of EL teachers toward teacher evaluation as follows will give us a scope of what can be done for further phases. This may indicate to us that

as long as evaluative practices a) yield developmental objectives; b) are programmed and continuous; c) provide quality of instruction and student success; and d) offer EL teachers scope for future improvement, there is always a place for teacher evaluation (i.e., the results of Section 1).

Randall (2015) specifically emphasizes the importance of evaluating EL teachers and training them in the light of exploration, awareness-raising, self-evaluation and solution. That may direct us to the point that creation of active listening and mutual empathy for personal, professional, and institutional ambitions should eventually result in more objective evaluative practices and constructive feedback for development. Otherwise, as long as teacher evaluation is conducted for purely evaluative purposes, it will be not accepted positively by EL teachers without even considering the type of evaluative practice, which can also be observed from the results of Section 1 (e.g., Item 3) and Section 2.

5.2.2 Types of teacher evaluation from the perspectives of EL teachers. As the findings of this specific study highlighted, the importance of evaluative practices - which need to foster teacher development and avoid any type of judgmental or negative practices - suggest a preference for platforms whereby educators feel at peace while being evaluated (Richards & Farrell, 2005). At the same time, those platforms would be preferable as long as they promote hands-on practice, monitoring, reflection as well as improvement in the common ground of teachers, students and institutions.

A variety of components - from mentoring to peer coaching or from in-class observations to self/peer-observations - could very well be integrated into teacher evaluation. Furthermore, in such cases, formative assessment will inevitably become involved in the process and connect itself to evaluative practices, which can be termed summative assessment (Randall, 2015). Therefore, role-modeling and longitudinal guidance - either by mentoring or in-class observations - alongside mutual development (e.g., team-teaching, peer-observation) and constructive feedback by peer-coaching or classroom observations perhaps proves most beneficial, which are again consistent with the findings of this specific study indicating that applications of different teacher evaluation practices are not significantly more frequent than the others (i.e., Section 2).

The practices of teacher trainers may also be required to undergo research in order to enable them to be a better role model for EL teachers as this study suggests teacher evaluation needs to be conducted by trained professionals (i.e., Section 1, item 11). This may allow teachers' commitment to conducting research and participating in collaborative teacher studies and/or other types of professional practices to strengthen, allowing for increased learning (Nunan, 1986; Mann & Walsh, 2015). This study confirms that teacher evaluation should provide both EL teachers and administrators a scope for future improvement (i.e., Section 1, item 12), Johnson (2009) likewise draws attention to the sociocultural and co-constructed nature of teacher evaluation. Taking this into consideration, guidance and collaboration between EL teachers and mentors, as well as professional support from academicians, are also suggested (Atay, 2008). Cheon (2016) asserts that self-determination should also be fulfill inherent psychological needs, such as engagement, positive functioning, and well-being through social context and theory-based instructional guidance by professionals.

In addition, enhancement of self-evaluation may be sustained through a variety of reflection tools (e.g., stimulated recall, critical incident analysis, narratives, portfolios, collaborative peer observations) when appropriate (Mann & Walsh, 2015). Such accounts would support teachers in gaining access to a variety of teacher development tools and additionally provide them with contexts to experience-reflective practicing tools. The overall understanding of the findings of Section 2 regarding the type of teacher evaluation practices here can be summarized and suggested as follows;

- No specific type of evaluation can be thought to be used more frequently than the others,
- Teacher evaluation practices with feedback sessions can be preferred to the ones without feedback sessions,
- Teacher evaluation practices may help support professional development and focus on specific pedagogical gains (e.g., teacher autonomy, pedagogical/content knowledge of teachers, pedagogical needs of institutions),
- Participation in conferences or continuous in-service training courses, students' reports on classroom performance, attending CELTA / DELTA / MA / PhD programs are also the ones with high percentages that are

mostly preferred practices which can be considered for the development of EL teachers while being used for evaluative purposes.

5.2.3 Purposes of teacher evaluation. The most important purposes of evaluation from the perspectives of EL teachers are continuous development of EL teachers, self-evaluation of teachers, developing EL teachers' teaching skills and pedagogical needs of the institution although other purposes are considered to be important to a certain degree. From this perspective, Quirke's (2015) seven-stage teacher appraisal program, ASPIRA, can be a good example as it suggests 1) the annual plan, 2) the conference, 3) first cycle monitoring, 4) mid-cycle review, 5) second cycle monitoring, 6) manager report, and 7) end of cycle discussion. In such a system, both parties - on behalf of the teachers and the institution - have the opportunity to set-up annual plans together, to see where and how to begin discussions, and lastly to observe progress before final decisions in the form of evaluation take place.

The other two findings in this vein included the fact that teacher evaluation needs to promote *the reflective self-evaluation of teachers* and *the development of teaching skills.* Teachers will undoubtedly show little willingness to improve unless a stress-free atmosphere can be sustained at their institutions just as one of the findings of this specific study suggests (i.e., item 50). This is unambiguously related to motivation both intrinsic and extrinsic. Intrinsically motivated teachers will most probably search for ways to go further in their teaching and professional development because another finding of this study revealed that EL teachers perceive teacher evaluation to help them evaluate their teaching performance while developing teaching skills (i.e., item 31, item 32, item 34). Likewise, extrinsically motivated teachers will expect their administrators, as well as teacher trainers/mentors, to build communicative-friendly relations with them. For instance, Dikilitaş, Wyatt, Hanks, and Bullock (2016) draw attention to the positive impacts of classroom research on EL teachers' careers in several aspects such as reflective, collaborative, theoretical, and practical unities.

Because continuous and professional development are found to be of utmost importance by the respondents of this study (i.e., item 1, item 2), self-evaluation through reflective practices (e.g., narratives, portfolios, team-teaching) can relatedly be among the alternative ways to improve EL teachers since *reflective practitioners*

and *transformative intellectuals* are two fundamental assets that Kumaravadivelu (2003) suggests for teacher qualities. Reflecting on teaching practices together with action research for professional development are also highly suggested (Burns, 1999; Teng, 2019). Howard (2015) claims that today teacher evaluation is mostly carried out with the purpose of inspecting teachers in order to establish whether or not effective teaching is taking place whereas the purpose needs to help teachers reveal their own effectiveness of pedagogic practices (p. 194). Such negatively-driven evaluative practices have the potential to put a heavy burden on teachers just as the findings of our study suggest so (i.e., section 4).

Another finding of our study is connected to the purpose that teacher evaluation should yield institutional benefits (i.e., item 37). As long as the aim is to increase teacher effectiveness and its optimistic reactions on the atmosphere of institutions, teachers are suggested to have alternatives for evaluation and not forced to engage in practices dictated by administrators (Bowen, 2004; Tomlinson, 2003). What is understood here is that teacher evaluation does not necessarily have to inspect teachers but let them explore, find ways, implement various practices which can help self-improvement together with their institutions. On the concrete side of this, in terms of provision of constructive feedback to teachers, Darling-Hammond et al. (2012) recommend institutional applications of evaluative practices by experts (e.g., mentors, teacher trainers, academicians) looking at multiple sources of data and provision of timely and meaningful feedback to the teachers, which is again in line with one of the findings of this study that teacher evaluation needs to be conducted by trained professionals (i.e., item 11). In connection with this, effective strategies for a reflective conference with teachers, according to Darling-Hammond et al. (2012), should include the followings:

- Beginning by talking about the components of teaching that were strong, such as respect and rapport building,
- Being prepared to ask a question such as: “Since I was just doing a brief walkthrough, I think I missed the stated objective. Can you tell me about that?”,
- Allowing the teacher to share the learning outcomes and where the lesson went after the administrator left the room,
- Asking any lingering questions that the administrator has about the direction of the lesson and the learning expected from the students,

- Evaluating this method of conferencing versus telling a teacher “That lesson was a failure.”,
- Relationship building to allow the teacher to feel “part of” the process versus feeling “outside” the process (p. 37).

Arneson (2014), to sum up, points out to several inhibited assumptions that observers need to consider before they get into teachers' classrooms such as focusing on teaching not on the teacher, leaving bias about the teacher at the door and the reign of objectivity (p. 42). Keeping such perspectives in mind, the strategies of Arneson seem to be positively correlated with the findings of our study, specific samples of which are a) teachers appreciate the importance of institutional goals and expectations while they wish to develop professionally and personally, b) mutual development, support in partnership, increase in self-awareness, understanding the perspectives of other teachers, and student success are also found to be important but not with high values (i.e., items 37, 34, 32, 31, & 42 in the questionnaire respectively). Having provided a detailed demonstration of these findings in Table 16, the items could be re-presented as follows; a) item 37 ‘TE helps me meet the pedagogical needs of the institution’, b) item 34 ‘TE helps me develop continuously’, c) item 32 ‘TE helps me evaluate my teaching performance’, d) item 31 ‘TE helps me develop my teaching skills’, and e) item 42 ‘TE helps me understand the perspectives of administrators.’

5.2.4 Drawbacks of teacher evaluation. All detrimental factors and, in particular, the lack of guidance on what EL teachers are expected to do, demotivating language used by evaluators, and a lack of clear criteria of teacher evaluation should be avoided (items 46, 49, & 51). The most negative moments in terms of evaluation mostly take place between administrators and teachers either during post-observation meetings or at the end-of-ear appraisal meetings (Devos, 2014; De Lima & Silva, 2018). Post-observation meetings mostly take the form of feedback given to educators about teaching and learning, whereas end-of-term meetings mostly involve the overall evaluation of performances of teachers throughout the academic year.

Regarding post-observation meetings, in order to avoid and/or eliminate the negativity of an evaluation, immediate feedback may not be the right choice, as teachers can feel anxious right after the lesson and tend to focus on what has or has not gone well. Thus, delayed feedback sessions may be a good alternative to

immediate feedback. Teachers and their observers should take time to think over specific areas to discuss and bring cases together through constructive dialogue (Iyer-O'Sullivan, 2015, p. 81). Three findings of our study a) 'stress of being evaluated, b) 'sense of a non-satisfactory evaluation', c) 'lack of trust in evaluators' also support such suggestions to ease negativity in teacher evaluation. Critical incidents, as Harmer (2008) and Thiel (1999) suggest in the forms of *magic moments* or *teaching highs/teaching lows*, have the capacity to enhance peer feedback and exchange ideas in the means of constructive post-observation sessions. Schön's (1983) perception of 'reflection-on-action' is highly linked to these post-observation feedback applications and that can even stand the base for critical incident dialogues among teachers.

Kilbourn et al. (2005) draw attention to the lack of quality feedback which needs to be given to teachers in a sensitive and productive manner and state that most of the time mentors or teacher trainers neglect learning how to give genuinely constructive feedback (p. 299). This conclusion of Kilbourn et al. is directly related to one of our findings, which pointed out the 'lack of guidance on what teachers are expected to do' and/or the 'lack of professional guidance that teachers need' (i.e., items 47 & 49). According to King (2015), whose work supports the previous conclusion, although organizations first aim to create professional development and maintain quality assured institutions by administrating teacher appraisal systems, evaluation is considered to be an unwanted and irrelevant task - especially in environments characterized by an authoritarian management style. "When considering the value of teacher evaluation for experienced teachers, the literature suggests a partial negativity" (King, 2015, p. 171). Building trust reciprocally between teachers and administrative staff therefore plays a very crucial role. Directly addressing this matter, Vodicka (2006) identifies the four elements of trust as a) compassion, b) consistency, c) communication, and d) competency. Compassion is the caring for other individuals that is central to a trusting relationship. Consistency is prevalent in most of the definitions of trust but feeling consistency itself is not enough to generate trust. Vodicka positions communication as the key, since leaders, who teachers think should be open, consider communication a strategy that breeds trust. Lastly, competence implies reputation and affiliation, but producing positive results is likely the best determinant of competence.

Any practices proposing negative consequences toward teachers need to be avoided. In a qualitative study by Burns (2000), during research interviews with the

participants it was found that there were several key areas where the inspection was seen as a threat and even the stress of the upcoming teacher evaluation period affected teachers' preparation for school practices. Such findings may suggest to us that negatively-directed teacher evaluation practices not only have damaging effects on the actual practices of teachers but also on future applications, which is also consistent with another result of our study 'demotivating language used by evaluators', ‘stress of being evaluated’ or ‘judgmental attitudes of evaluators’ (i.e., items 45, 49, & 56).

In a study conducted by Arneson (2014), the following quote from a teacher was noted; “my principal rated me lower on professionalism because of a parent complaint. But what about my side of the story?”. Arneson (2014) states, "the relationship between principal and teacher is delicate in nature as the principal typically serves as the primary evaluator for the teacher’s performance" (p. 5). Provision of fairness in a mutual relationship is clearly important, otherwise it may result in unequal power distribution between two parties. Allowance for effective communication needs to take place. Teachers must believe the principal is fair and equitable in his or her evaluation of teachers’ capabilities. Trust is a major factor in relationship building, just as mistrust is a critical factor in the breakdown of relationships. To Arneson (2014), for instance, negative barriers to good communication include the following; a) letting employees be the last to know, b) not responding to emails or phone calls, c) lack of open or effective communication, d) closed doors, and e) not listening and other poor communication skills (p. 28). These are also related to our findings, ‘lack of professional guidance that teachers need’, ‘lack of trust in evaluators’, or demotivating language used by the evaluator’ (i.e., items 45, 46, & 47). In order to avoid such problematic consequences in teacher evaluation, another finding of this study indicates that both EL teachers and administrators are to be provided a scope for future development (i.e., item 12), which in the end has the potential to increase the quality of instruction as well as well-trained administrators (i.e., items 5 & 11).

5.2.5 Educational background of EL teachers and their positions in their relevant institutions. Having found in this study that the amount of teaching experience, gender and university type of teachers had no impact on their opinions toward teacher evaluation (i.e., results of research question 3), it pays to consider the

nature of what individuals consider achievements and quality in their profession. Rather than focusing merely on the years that EL teachers have spent on their profession or simply on their gender, the characteristics of what they are achieving and have achieved so far need to be considered. In this sense, Master (2014) supports the view that "evaluation systems that account for the potential complexity and diversity of teachers' contributions by including flexible measures vis-à-vis their performance may be more responsive and ultimately more effective" (p. 224). Experienced teachers may enjoy the possibilities of empowering their own weaknesses, which have perhaps not been self-detected before, while inexperienced/novice teachers can explore areas to improve. Zonoubi et al. (2017) conclude in their study that experienced teachers had the chance to observe innovations in the field and teach better while, on the other hand, novice teachers improved themselves in several themes such as classroom management.

However, in a certain sense, the educational background of EL teachers and their positions/roles as instructors, coordinators, trainers, mentors and/or administrators make a difference in their opinions toward teacher evaluation. Having observed out of our study that there is not a difference between the more experienced and the less experienced unlike the difference between a teacher holding bachelor's degree and a teacher with doctorate's degree. The attitudes, expectations and interpretations of both parties may change from another, so a balanced process of implementations and the various types of developmental teacher evaluation practices need to take place (Tucker & Stronge, 2005). Following the beneficial outcomes of developmental evaluation practices, both parties, namely, teachers and administrators have much to gain. Kim and Danforth (2012), for instance, reject the top-down approach towards teacher evaluation due to the fact that it ignores the development of teachers but focuses purely on the detection of weaknesses. It would not be wrong to claim that treating both parties differently and separating the evaluative implementations from each other during teacher evaluation practices will most probably result in better consequences. Howard (2015) states unless every shareholder (e.g., teachers/teacher trainers, observers/observees, students, and administrators) is listened, it does not seem to be possible to develop professionalism and provide a worthwhile process under triangulation (p. 208). The reason for this can be that a teacher holding a BA degree may perhaps not have academic expectations for the future but professional development for personal career while a

teacher continuing or holding his/her PhD degree may wish to develop in academic practices for future accomplishment. As long as practical and pragmatic developmental practices are presented to the preferences of EL teachers and their educational levels and positions at the institutions are considered, drawbacks of teacher evaluation may be avoided. All those shareholders will high probably not resist to be evaluated but volunteer in being a part of the evaluation program.

Apart from the development of EL teachers, administrators or other staff who are in charge of evaluating teachers are preferred to be experts and well-trained professionals as the findings of the study suggest (i.e., item 11). In case of the nonexistence of professional teacher trainers or experts, the ones who have the authority to evaluate teachers should be offered the required qualifications, which would undoubtedly yield more improved results for a more promising teacher evaluation system (Boysen et al., 2006; Knop, 1980).

5.2.6 Systematic or non-systematic teacher evaluation. It is clear that a well-designed teacher evaluation system with clearly-defined objectives and ongoing procedures for the outcomes expected will result in more effective teacher development just as the findings of our study show; whether the process is systematic or non-systematic is of key importance. For instance, the literature suggests that unannounced classroom observations and evaluation practices that do not serve for teacher development are not preferred (Howard, 2015; Tucker & Stronge, 2005). Similarly, dependence on single-measurement criteria, such as student success does not fulfill teacher requirements due to the fact that fear of negative consequences can be observed and the other work engaged in by teachers can be overlooked (Darling-Hammond, 2015; Marshall, 2012; McCaffrey, 2010). To give a sample, with the aim to create a systematic teacher evaluation model of mentoring which serves for social constructivist approach in teacher development, the findings of a study by Dikilitaş and Mumford (2016) may be followed. To Dikilitaş and Mumford, the roles of both mentors and the collaborators can be clarified in several aspects; the mentors can support EL teachers on choosing a suitable research focus, a formal style in language, or provide moral support as reassurance of value of writing and play a key role in the process from finding initial focus to final written product while at the same time the collaborators can mutually support in drafting, selecting material for inclusion, grammar correction as ‘second pair of eyes' and shared workload and

responsibility in all stages. As can be observed on the above-provided mentoring system, based on the suggested literature, a teacher evaluation model adapted by an institution may better aim to;

- start with clear aims and objectives for practices,
- draw the pathway step by step for the upcoming practices,
- consider both professional domains (e.g., pedagogy, professional development) and affective domains (e.g., motivation, fear of judgmental evaluation),
- give teachers the opportunity to choose their preferred type of evaluative and developmental practice,
- open doors for teachers to observe what has been done and what can be done for future (Bain et al., 2002; Burns, 2000; Darling-Hammond, 2012; Smith, 2008).

Additionally, King (2015) regarding teacher evaluation in EL contexts states that when determining the value of a teacher, a teacher evaluation system leading to professional development as part of a reflective cycle, providing constructive feedback, and the need for proper training in appraisal are suggested. Otherwise, in absence of those, teachers had better be given the opportunity to reflect on themselves to identify ways to improve themselves since "the main reasons for disillusionment from teachers seem to be that appraisers are unskilled and biased, appraisal systems are useless, and observations are unrepresentative of reality" (King, 2015, p. 176). Such suggestions again draw our attention to the need for systematically implemented and purposefully designed nature of teacher evaluation practices. EL Teachers apparently deserve to be listened for the ways of their preferences for teacher evaluation practices and professional development tools in order to sustain better teaching.

5.3 Implications

A final concern regarding this study may be considered in the question: "What is the ideal model for teacher evaluation practices in terms of the expectations of EL teachers and administrators?" This is somehow related to our sense of responsibility toward our students and commitment to the ideals of offering quality education. It is,

at the same time, highly associated with the personal and professional growth of EL teachers (Borg, 2019; Burns, 2000; Darling-Hammond, 2012; King, 2015).

Inspired by the findings, we can claim that teacher evaluation is a necessary step in educational programs and its importance is not compromised in terms of amount of teaching experience, gender or university type. In other words, evaluation is necessary for all EL teachers regardless of these variations.

We can also observe that the educational background of EL teachers, their roles or positions and the way evaluation is conducted make a difference. Therefore, the model that can be suggested here would contain the following elements:

Table 35

Overall Suggestions for Teacher Evaluation Practices.

No.	Teacher evaluation practices and procedures should:
1	Be continuous and ongoing in aiding development.
2	Be clearly defined and implemented systematically.
3	Treat EL teachers with and without administrative responsibilities differently.
4	Treat EL teachers with different university degrees and academic backgrounds differently.
5	Be guarded against any type of negative factors or applications.

5.3.1 Development-oriented teacher evaluation model (DTEM). A teacher evaluation model which paves the way for the differentiated professional development of EL teachers would be suggested in alignment with the findings of the current study and the deductions obtained out of related literature. The model for teacher evaluation cooperates with the overall suggestions for teacher evaluation practices presented in table 35. The suggested *Development-oriented Teacher Evaluation Model* has been constructed upon four procedural steps to follow; *Planning Period, Preparation and Practicing Period, Evaluation Period,* and *Post-evaluation Period*. It aims to be implemented upon four distinctive principles; *Teaching, Professional Development, Student Achievement* and *Institutional Gains*. The model is presented and described in specific details as briefly as possible in table 36:

Table 36

Development-oriented Teacher Evaluation Model (DTEM)

PLANNING PERIOD	
Steps to Follow: *In order to decide on a teacher's specific preference of the four suggested principles offered in the model, one of the following scheduled and ongoing practices ought to be rolled out via:* *a. Systematic interviews with teachers (e.g., weekly, monthly),* *b. Systematic self-reports, minutes of diaries (e.g., weekly, monthly),* *c. Systematic volunteer classroom/peer/self-observations for diagnostic purposes (e.g., weekly, monthly).* *d.*	
Teacher's Name & Surname:	
Teacher's Signature:	
Position at the Institution:	
Level of Education:	
Academic Year:	
Preferred Area to be Developed and Evaluated:	1. TEACHING 2. PROFESSIONAL DEVELOPMENT 3. STUDENT ACHIEVEMENT 4. INSTITUTIONAL GAINS
Administrator's Name & Surname:	
Administrator's Signature:	
PREPARATION AND PRACTICING PERIOD	
Steps to Follow: *Teachers/mentors/teacher trainers/administrators should pick a specific area to be developed and evaluated on the pre-decided period (e.g., end of semester, end of academic year).* *a. Starting the process for personal or professional development (e.g., weekly, monthly).* *b. Participation for the scheduled meetings (e.g., weekly, monthly).* *c. Handing in the required reflective tools on time (e.g., weekly, monthly).* *d.*	
1. TEACHING: Enhancing the qualities of teachers in terms of teacher knowledge (TPACK), teacher competency, teacher effectiveness or teacher autonomy.	<u>Practices for Development:</u> *a. Mentoring (See Appendix OO, PP, & QQ)* *b1. Peer-Reflective practices and tools (See Appendix RR)* *b2. Self-Reflective practices and tools (See Appendix SS)* <u>Practices for Evaluation:</u> *a. Systematic meetings with the mentor via the use of reflective agendas for developmental purposes (Appendix TT & UU)* *b1. Classroom/Peer/Self-observations for developmental purposes in the form of development-oriented teacher evaluation tools via the use of specific checklists suggested (Appendix VV)* *b2. Reflective Self-evaluation tool suggested (See Appendix WW)* *c. Continuous formative assessment as well as final evaluation by administrators / coordinators / teacher trainers / mentors (See Appendix YY)*

Table 36 (connt.d)

2. PROFESSIONAL DEVELOPMENT: Continuous teacher development practices to be evaluated on a regular basis by professional teacher trainer(s) who have advanced academic and pedagogic background.	Practices for Development: *a. Research-based practices (i.e., AR, TR, EP, LS)* *b. Doing an MA & PhD* *c. Teaching Certificate programs (e.g., CELTA, DELTA, TESOL)* Practices for Evaluation: *a. Publishing articles (Appendix WW)* *b. Reflective Self-evaluation tool suggested (See Appendix WW)* *c. In-house/conference presentations (See Appendix WW)* *d. Continuous formative assessment as well as final evaluation by administrators / coordinators / teacher trainers / mentors (See Appendix YY)*
3. STUDENT ACHIEVEMENT Implementation of reflective tools to increase student achievement.	Practices for Development: *a. Teachers' reflective tools to keep track of and improve student success (See Appendix RR & SS).* Practices for Evaluation: *a. Student surveys for development-oriented teacher evaluation (See Appendix XX)* *b. Results of student test scores* *c. Reflective Self-evaluation tool suggested (See Appendix WW)* *d. Continuous formative assessment as well as final evaluation conducted by administrators / coordinators / teacher trainers / mentors (See Appendix YY)*
4. INSTITUTIONAL GAINS: Contribution to the testing system, material design, curriculum and syllabi of the institution.	Practices for Development: *a. Being a member of the institutional departments (e.g., material development unit, continuous development unit, testing center).* *b. Systematic constructive support by administrators / coordinators / teacher trainers / mentors* Practices for Evaluation: *a / b. Reflective Self-evaluation tool suggested (See Appendix WW)* *a / b. Continuous formative assessment as well as final evaluation by administrators / coordinators / teacher trainers / mentors (See Appendix YY)*
EVALUATION PERIOD	
Steps to Follow: The administrator(s) and the teacher should come together to discuss and provide detailed answers for the following questions to evaluate the complete process: *a. What has changed?* *b. What needs have emerged?* *c. What are the weaknesses and the strengths of the practices for the teacher?* *d. What are the weaknesses and the strengths of the practices for the students?* *e. What are the weaknesses and the strengths of the practices for the institution?* *f.*	

Table 36 (connt.d)

POST-EVALUATION PERIOD
Steps to Follow: The administrator(s) and the teacher should come together to discuss and provide detailed answers for the following questions to redesign and replan further developmental and evaluative practices: *a. What further steps need to be taken?* *b. What follow-up/complement areas need to be adapted for teacher development and evaluation?* *c. What follow-up/complement areas need to be adapted for student achievement?* *d. What follow-up/complement areas need to be adapted for institutional gains?* *e.*

5.4 Limitations

Similar to many studies, the present research has missed several aspects that could not be predicted or implemented by the launch of data collection procedures. Although this exposes the limits of the validity of the results, it should not serve to spurn other researchers to expand on knowledge in the field.

The first limitation pertains to the longitudinal implementation of the suggested model of teacher evaluation. The data collection process of the current study was an obstructive factor for the researcher, which should encourage more detailed implementations in future studies. The construction of the questionnaire, in particular, took more than one year. In addition to the construction process, the delivery step also took more than four months due to all EL instructors and administrators of the English preparatory programs of the universities in Turkey having been reached by e-mails sent step by step on a daily basis. Predictably, due to the nature of this means of correspondence, as well as the hectic nature of educators' working schedules, responses took more time than anticipated.

Another limitation involves the number of the participants for the qualitative stage. Although the variety of university types stood at an acceptable level, no more than 50 respondents volunteered to provide data for the construction period of the questionnaire.

Another limitation was the scope of the study that it is limited to higher education contexts. As the current study focuses on the EL teachers working in English preparatory programs at the universities, the findings may not be as informatively applied to pre-school, primary, secondary or high school levels of study.

Another limitation relates to the affective domains of teacher evaluation, which can be seen as a part of the process. Constructs such as motivation, attitudes, burn-out, and/or dedication to an institution were beyond the scope of the carrying out of this specific study. However, the integration of several related items into the questionnaire (e.g., section 3, section 4) somehow helped the researcher compensate for such lacking dimensions.

5.5 Recommendations for Further Studies

As the current study presents a comprehensive teacher evaluation system together with concrete practices, follow-up studies may adapt the relevant parts of the study for their own specific context and needs. Such an implementation will undoubtedly fill the gaps that this study has already opened.

Having mentioned the small-scaled number of respondents for the qualitative stage of the study, which pursues qualitative data collection procedures, more investigations can be obtained by such a method, and in-depth insightful outcomes will undoubtedly emerge together with the participation of a greater number of EL teachers.

Observations, interviews, focus groups or recordings in authentic settings in the form of non-numerical data will help understand the concept of teacher evaluation and opinions of EL teachers more in detail; that is, the related findings will have the potential to open doors to generate new ideas for further research.

The last suggestion may be related to the EL learners' side. As their achievements are crucial to assist educators for further educational implementations, effects and reflections of teacher evaluation on student achievement may further be investigated by the help of the findings of this specific study.

5.6 Conclusion

The study has been a success in terms of having reached a number of clear conclusions on the impact of teacher evaluations on teacher development by examining qualitative and quantitative findings sourced from educators and administrators themselves.

Having constructed a questionnaire based on the results of a number of preliminary investigations, further questioning prompted us to explore EL teachers'

views on their preferred manner of evaluation to find out what factors lay behind their opinions that could be quantified in terms of experience, background, position, gender, or other aspects.

Our results show that consistent commitment to implementing teacher evaluations for the sole cause of professional development, practiced at a level among peers, and considering individual EL teachers' particular concerns, without forgetting the main aim of improving the learning environment for students, is the ideal way forward in this regard.

The results of our analysis have the potential to help advance the way educational institutions map out their evaluative practices and make a further case for increasing the amount of dependence and support educators enjoy within their work, dissuading from traditional, top-down, authoritarian management styles, which are proven to have a negative impact on both the reliability of evaluations - by way of increasing anxiety among teachers - and, ultimately, quality of education, by way of dissuading potentially effective educators from developing in their profession.

REFERENCES

Abedi, J., & Faltis, C. (2015). Teacher assessment and the assessment of students with diverse learning needs, *Review of Research in Education, 39*(1), 7-14.

Achinstein, B., & Davis, E. (2014). The subject of mentoring: Towards a knowledge and practice base for content-focused mentoring of new teachers. *Mentoring & Tutoring: Partnership in Learning, 22*(2), 104-126.

Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement. *Educational Evaluation and Policy Analysis, 39*, 54-76.

Akcan, S., Aydin, B., Karaman, A. C., Seferoğlu, G., Korkmazgil, S., Özbilgin, A., & Selvi, A. F. (2017). Qualities and qualifications of EFL professionals: What do intensive English program administrators think?. *TESOL Journal, 8*(3), 675-699.

Allwright, D. (2003). Exploratory practice: Rethinking practitioner research in language teaching. *Language Teaching Research, 7*(2), 113-141.

Allwright, D. (2005). From teaching points to learning opportunities and beyond. *Tesol Quarterly, 39*(1), 9-31.

Allwright, D. (2005). Developing principles for practitioner research: The case of exploratory practice. *The Modern Language Journal, 89*(3), 353-366.

Allwright, D., & Hanks, J. (2009). *The developing language learner: An introduction to exploratory practice.* Basingstoke, England: Palgrave, Macmillan.

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

Alvarez, M. E., & Anderson-Ketchmark, C. (2011). Danielson's framework for teaching. *Children & Schools, 33*(1), 61-63.

Amrein-Beardsley, A., & Holloway, J. (2019). Value-added models for teacher evaluation and accountability: Commonsense Assumptions. *Educational Policy, 33*(3), 516-542.

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, *52*(1), 154-168.

Arikan, A., Taser, D., & Sarac-Suzer, H. S. (2008). The effective English language teacher from the perspectives of Turkish preparatory school students. *Online Submission*, *33*(150), 42-51.

Armstrong, H. (2012). Coaching as dialogue: Creating spaces for (mis) understandings. *International Journal of Evidence Based Coaching and Mentoring*, *10*(1), 33-47.

Arneson, S. M. (2014). *Building trust in teacher evaluations: It's not what you say; it's how you say it*. Corwin Press.

Arnold, E. (2006). Assessing the quality of mentoring: Sinking or learning to swim?. *ELT Journal*, *60*(2), 117-124.

Atay, D. (2006). Teachers' professional development: Partnerships in research. *Tesl-Ej*, *10*(2), 1-15.

Atay, D. (2008). Teacher research for professional development. *ELT Journal*, *62*(2), 139-147.

Atay, D., Kaslioglu, O., & Kurt, G. (2010). The pedagogical content knowledge development of prospective teachers through an experiential task. *Procedia-Social and Behavioral Sciences*, *2*(2), 1421-1425.

Bacher-Hicks, A., Chin, M. J., Kane, T. J., & Staiger, D. O. (2019). An experimental evaluation of three teacher quality measures: Value-added, classroom observations, and student surveys. *Economics of Education Review*, *73*, 1-15.

Bailey, K. M. (2006). *Language teacher supervision: A case-based approach.* New York: Cambridge University Press.

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

Bain, J.D., Ballantyne, R., Mills, C. & Lester, N.C. (2002). *Reflecting on practice: Student teachers' perspectives*, Post Pressed: Flaxton, Qld.

Ball, D. L., Thames, M. H., & Phelps, G. C. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59, 389-407.

Baran, E., Canbazoglu Bilici, S., Albayrak Sari, A., & Tondeur, J. (2019). Investigating the impact of teacher education strategies on preservice teachers' TPACK. *British Journal of Educational Technology*, *50*(1), 357-370.

Barnawi, O. Z. (2016). Dialogic investigations of teacher-supervisor relations in the TESOL landscape. *Cogent Education*, *3*(1), 1217818.

Basileo, L. D., & Toth, M. (2019). A state level analysis of the marzano teacher evaluation model: Predicting teacher value-added measures with observation scores. *Practical Assessment, Research, and Evaluation*, *24*(1), 1-6.

Baykal, A. (2015). A preoperative index for construct validity. *The 41st International Association for Educational Assessment (IAEA) Conference. October 11-16, 2015, Center for Educational Testing and Evaluation (CETE),* The University of Kansas, Lawrence, Kansas, *USA.*

Beavers, A. S., Lounsbury, J. W., Richards, J. K., & Huck, S. W. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, *18*(1), 6.

Bellibaş, M. Ş., & Gedik, Ş. (2016). A Comprehensive ınitiative for evaluating teacher performance: The case of Michigan (USA). *Journal of the Faculty of Education*, *17*(1), 113-130.

Benson, P. (2011). Teaching and researching autonomy (2nd ed.). London: Pearson.

Bernstein, D. J., Jonson, J., & Smith, K. (2000). An examination of the implementation of peer review of teaching. *New directions for teaching and learning*, *83*, 73-86.

Bi, H. H. (2018). A robust interpretation of teaching evaluation ratings. *Assessment & Evaluation in Higher Education*, *43*(1), 79-93.

Borg, S. (2007). Research engagement in English language teaching. *Teaching and Teacher Education*, *23*(5), 731-747.

Borg, S. (2010). Language teacher research engagement. *Language Teaching*, *43*(4), 391.

Borg, S. (2010). Doing good quality research. *JACET Journal (*50), 9-13.

Borg, S. (2013). *Teacher research in Language teaching*. Cambridge: Cambridge University Press.

Borg, S. (2018). *Teacher evaluation: Global perspectives and their implications for English language teaching. A literature review.* Delhi: British Council.

Borg, S. (2019). Contemporary perspectives on teacher appraisal. *United Nations Educational, Scientific and Cultural Organization*, 1-24.

Bourdieu, P. (1986). The forms of capital. Handbook of theory and research for the sociology of education. JG Richardson. *Greenwood*, *241*(258), 19.

Boysen, G. A., Kelly, T. J., Raesly, H. N., & Casner, R. W. (2014). The (mis) interpretation of teaching evaluations by college faculty and administrators. *Assessment & Evaluation in Higher Education*, *39*(6), 641-656.

Bowen, T. (2004). Continuous professional development. *The one stop magazine.*

Braun, H., & Wainer, H. (2007). 27 Value-Added Modeling. *Handbook of Statistics*, *26*, 867-892.

British Council (2012). *A guide to continuing professional development-Peer observations*. British Council.

Brockbank, A., & McGill, I. (2007). *Facilitating reflective learning in higher education*. McGraw-Hill Education (UK).

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy*. New York.

Brown, H.D. (2007). *Principles of language learning and teaching* (5th edit). New York: Longman.

Brown, J. D. (1994). *The elements of language curriculum*. Florence, USA: Cengage Learning, Inc.

Brown, J. D., & Wolfe-Quintero, K. (1997). Teacher portfolios for evaluation: A great idea or a waste of time. *The Language Teacher*, *21*(1), 28-30.

Bullough Jr, R. V. (2005). Being and becoming a mentor: School-based teacher educators and teacher educator identity. *Teaching and Teacher Education*, *21*(2), 143-155.

Burns, A. (2009). *Doing action research in English language teaching: A guide for practitioners*. Routledge.

Burns, J. (2000). *Improvement through ınspection?: An investigation of teachers' perceptions of OFSTED as a vehicle for improvement*. Trentham.

Cabaroglu, N. (2014). Professional development through action research: Impact on self-efficacy. *System*, *44*, 79-88.

Cakir, A., & Balcikanli, C. (2012). The use of the european portfolio for student teachers of languages (EPOSTL) to foster teacher autonomy: English language teaching (ELT) student teachers' and teacher trainers' views. *Australian Journal of Teacher Education, 37*(3), 1-16.

Can, N. (2009). Post-method pedagogy: Teacher growth behind walls. In *Proceedings of the 10th METU ELT Convention.*

Caires, S., & Almeida, L. (2005). Teaching practice in initial teacher education: Its impact on student teachers' professional skills and development. *Journal of Education for Teaching, 31*(2), 111-120.

Carreiro, D. M. (2020). Making a shift in educator evaluation. *Journal of Educational Supervision, 3*(2), 97-104.

Chan, V., Spratt, M. & Humphreys, G. (2002). Autonomous language learning: Hong Kong tertiary students' attitudes and behaviours. *Journal of Evaluation and Research in Education, 16*(1), 1-18.

Child, D. (2006). *The essentials of factor analysis*. Continuum International Publishing Group: London & New York.

Chism, N.V.N. (2007). *Peer review of teaching. A Sourcebook.* Jossey Bass: San Francisco.

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2014). *Research methods*. Pearson Education Limited.

Clark, V. L. P., & Creswell, J. W. (2014). *Understanding research: A consumer's guide*. Pearson Higher Education.

Cochran-Smith, M., & Lytle, S. L. (2001). Beyond certainty: Taking an inquiry stance on Practice. In. A. Lieberman & L. Miller (Eds.) *Teachers caught in the action: Professional development that matters* (pp. 45-58).

Coenders, F., & Verhoef, N. (2019). Lesson study: professional development (PD) for beginning and experienced teachers. *Professional Development in Education, 45*(2), 217-230.

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

Copland, F. (2015). Evaluative talk in feedback conferences. *Teacher Evaluation in Second Language Education*, 135-149.

Cosgun-Ögeyik, M. (2009). Evaluation of English language teaching education curriculum by student teachers. *Insan ve Toplum, 9*(1), 42-57.

Cosh, J. (1999). Peer observation: a reflective model. *ELT Journal, 53*(1), 22-27.

Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Upper Saddle River: Merrill Prenice Hall.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Danielson, C. (1996) *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching*. Alexandria,VA: Association for Supervision and Curriculum Development.

Danielson, C. (2011). Evaluations that help teachers learn. *Educational leadership, 68*(4), 35-39.

Danielson, C., & McGreal, T. (2000). *Teacher evaluation to enhance professional practice.* Princeton New Jersey, USA: Educational Testing Service.

Darling-Hammond, L. (2000*). Teacher quality and student achievement: A review of state policy evidence*. Center for the Study of Teaching and Policy, University of Washington.

Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, *57*(3), 300-314.

Darling-Hammond, L. (2015). Can Value added add value to teacher evaluation? *Educational Researcher, 44*(2), 132-137. doi:10.3102/0013189x15575346.

Darling-Hammond, L. (2015). *Getting teacher evaluation right: What really matters for effectiveness and improvement*. Teachers College Press.

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93(6), 8-15.*

De Lima, J. Á., & Silva, M. J. T. (2018). Resistance to classroom observation in the context of teacher evaluation: teachers' and department heads' experiences and perspectives. *Educational Assessment, Evaluation and Accountability*, *30*(1), 7-26.

Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management, 34*, 267-297.

Derrington, M. L. (2011). Changes in teacher evaluation: Implications for the principal's work. *Delta Kappa Gamma Bulletin*, *77*(3), 51.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199.

Devos, N. (2014). A framework for classroom observations in English as a Foreign Language (EFL) teacher education. *Dil ve Dilbilimi Çalışmaları Dergisi*, *10*(2).

Dikilitaş, K., ed. (2014). *Professional development through teacher research*. İzmir: Gediz university Press

Dikiltaş, K. (2015) Professional development through teacher-research In Dikilitaş, K., Smith, R., & Trotman, W. (eds.) *Teacher-researchers in action*. Kent: IATEFL (pp. 47-55).

Dikilitaş, K. (2015). "Teacher Research for EL teachers." In *Professional development for English language teachers: Perspectives From higher education in Turkey*, edited by S. Borg, 27-33. Ankara: British Council, Turkey.

Dikilitaş, K., Smith, R. C., and W. Trotman, eds. (2015). *Teacher researchers in action*. Canterbury, Kent: IATEFL.

Dikilitaş, K., & Mumford, S. E. (2016). Supporting the writing up of teacher research: peer and mentor roles. *Elt Journal*, *70*(4), 371-381.

Dincer, A. (2019). Language teacher autonomy in Turkish EFL context: Relations with teachers' autonomy and job satisfaction levels. *i-manager's Journal on English Language Teaching, (9)*2, 11-25.

Donaghue, D. (2015). Differences between supervisors' espoused feedback styles and their discourse in post-observation meetings. *Teacher Evaluation in Second Language Education*, 117-133.

Donaldson, M. L. (2009). So long, Lake Wobegon? Using teacher evaluation to raise teacher quality. *Center for American Progress*, 1-32.

Dörnyei, Z. (2003). *Questionnaires in second language research: Construction, administration, and processing*. Lawrence Erlbaum Associates: Mahwah, New Jersey.

Duta, N., Tomoaica, E., & Panisoara, G. (2015). Desirable characteristics defining to describe an effective teacher. *Procedia-Social and Behavioral Sciences*, *197*, 1223-1229.

Edwards, E., & Burns, A. (2016). Action research to support teachers' classroom materials development. *Innovation in Language Learning and Teaching*, *10*(2), 106-120.

Ellet, C. D. & Teddlie, C. (2003). Teacher evaluation, teacher effectiveness, and school effectiveness: Perspectives from the USA. *Journal of Personnel Evaluation in Education*, *17*(1), 101-128.

Ellis, N. J., Alonzo, D., & Nguyen, H. T. M. (2020). Elements of a quality pre-service teacher mentor: A literature review. *Teaching and Teacher Education, 92*, 1-13.

Elton, L. (1984). Evaluating teaching and assessing teachers in universities. *Assessment and Evaluation in Higher Education*, *9*(2), 97-115.

Engin, M. (2015). Artefacts in scaffolding the construction of teaching knowledge. *Teacher Evaluation in Second Language Education,* 85-99.

Er, M., Ülgü, S., & Sarı, İ. (2013). The journey of ELT teachers from apprenticeship to mastery. *Procedia-Social and Behavioral Sciences*, *70*, 41-51.

Erss, M., & Kalmus, V. (2018). Discourses of teacher autonomy and the role of teachers in Estonian, Finnish and Bavarian teachers' newspapers in 1991-2010. *Teaching and Teacher Education, 76*, 95-105.

Erten, I. H. (2015). Listening to practising teachers: Implications for teacher training program, *Procedia-Social and Behavioral Sciences*, *199*, 581-588.

Evens, M., Elen, J., Larmuseau, C., & Depaepe, F. (2018). Promoting the development of teacher professional knowledge: Integrating content and pedagogy in teacher education. *Teaching and Teacher Education, 75*, 244-258.

Fahim, M., & Nazari, M. (2014). Critical look at situation analysis of Jack. C. Richards: A post-modernist view. *Procedia-Social and Behavioral Sciences, 98*, 458-462.

Farrell, T. S. (2011). Exploring the professional role identities of experienced ESL teachers through reflective practice. *System*, *39*(1), 54-62.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education*. New York: McGraw-Hill Humanities/Social Sciences/Languages.

Freeman, D. (2004). 'Knowledge architecture: Some orienting references'. *TESOL Italy Plenary Address*. October 2004.

Freeman, D., & Hawkins, M. (2004). Collaborative education as critical practice in teacher education. In Hawkins & Irujo (eds.), *Collaborative Conversations among Language Teacher Educators. Alexandria,* VA: TESOL, pp. 1-13.

Fuller, F. & Bown, O. (1975). Becoming a teacher. In K. Ryan (ed.), Teacher Education (74th) *Yearbook of the national society for the study of education. Part II,* (pp. 25-52). Chicago: University of Chicago Press.

Gardner, H. (1983). Frames of mind: The theory of multiple intelligences. *NY: Basics*, 1.

Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences*. Hachette Uk.

Gieve, S., & Miller, I. K. (Eds.). (2006). *Understanding the language classroom*. Hampshire: Palgrave, Macmillan.

Godfrey, D., Seleznyov, S., Anders, J., Wollaston, N., & Barrera-Pedemonte, F. (2019). A developmental evaluation approach to lesson study: exploring the impact of lesson study in London schools. *Professional Development in Education, 45*(2), 325-340.

Goe, L., Biggers, K., & Croft, A. (2012). Linking teacher evaluation to professional development: Focusing on improving teaching and learning. research & policy brief. *National Comprehensive Center for Teacher Quality*.

Goldhaber, D. (2010). Teacher pay reforms: The political implications of recent research. CEDR Working Paper No. 2010-4.0. *Center for Education Data & Research*, 1-54.

Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the stability of measured teacher performance. *Economica, 80*(319), 589-612.

Gordon, S. P., & Mcghee, M. W. (2019). The power of formative evaluation of teaching. In M. Derrington & J. Brandon (Eds.). *Differentiated teacher Evaluation and Professional Learning.* Cham, Switzerland.

Grierson, A. L., & Gallagher, T. L. (2009). Seeing is believing: Creating a catalyst for teacher change through a demonstration classroom professional development initiative. *Professional Development in Education, 35*(4), 567-584.

Griffiths, C. (2007). Languge Learning Strategies: Students' and teachers' perceptions. *ELT Journal, 61*(2), 91-99.

Gupta, N. K., & Baveja, B. (2014). *Basics in education: Textbook for B. course.* New Delhi: NCERT.

Haefele, D. L. (1993). Evaluating teachers: A call for change. *Journal of Personnel Evaluation in Education*, *7*(1), 21-31.

Hanks, J. (2015). Language teachers making sense of Exploratory Practice. *Language Teaching Research*, 1-22.

Hanushek, E. A. (2009). Teacher deselection. In D. Goldhaber, & J. Hannaway (Eds.), *Creating a new teaching profession*. Washington, DC: Urban Institute.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 266-479.

Harmer, J. (2001). *The practice of English language teaching*. Longman.

Harmer, J. (2008). How to teach English. *ELT Journal*, *62*(3), 313-316.

Head, K., & Taylor, P. (1997). *Readings in teacher development*. London, UK: Macmillan.

Heneman III, H. G., Milanowski, A., Kimball, S., & Odden, A. (2006). *Standards-based teacher evaluation as a foundation for knowledge-and skill-based pay.* Philadelphia, PA: Consortium for Policy Research in Education.

Hendry, G. D., Bell, A., & Thomson, K. (2014). Learning by observing a peer's teaching situation. *International Journal for Academic Development, 19*(4), 318–329. doi:10.1080/ 1360144X.2013.848806.

Hermansen, H. (2017). Knowledge relations and epistemic infrastructures as mediators of teachers' collective autonomy. *Teaching and Teacher Education*, *65*, 1-9.

Hernon, S. (2019). *An exploration of principal perceptions of the Danielson framework for teaching in providing feedback and İmproving ınstructional leadership*. Seton Hall University Dissertations and Theses (ETDs).

Hiratsuka, T. (2016). Actualizing Exploratory Practice (EP) principles with team teachers in Japan. *System*, *57*, 109-119.

Hobjilă, A. (2012). Effective/ineffective in pedagogical practicum-mentors' considerations on the initial training of pre-school teachers. *Procedia-Social and Behavioral Sciences*, *33*, 318-322.

Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education*, *25*(1), 207-216.

Hong, H. Y., Lin, P. Y., & Lee, Y. H. (2019). Developing effective knowledge-building environments through constructivist teaching beliefs and technology-integration knowledge: A survey of middle-school teachers in northern Taiwan. *Learning and Individual Differences*, *76*, 1-9.

Howard, A. (2012). Teacher appraisal observations: The possibilities and impossibilities. In *Proceedings of the 2011 TESOL Arabia Conference*.

Howard, A. (2015). Giving voice to participants in second language education evaluation. *Teacher Evaluation in Second Language Education*, 193-210.

Howard, A., & Donaghue, D. (2015). *Teacher evaluation in second language education*. Bloomsbury Academic Publishing.

Howard, B. & Gullickson, A. (2010). Setting standards for teacher evaluation. In M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 337-353). San Francisco, CA: John Wiley & Sons.

Hudson, P. (2010). Mentors report on their own mentoring practices. *Australian Journal of Teacher Education, 35*(7), 30-42.

Hudson, P. (2013). Mentoring as professional development: growth for both mentor and mentee. *Professional Development in Education*, *39*(5), 771-783.

Hutchinson, B. (1995). Appraising appraisal: some tensions and some possibilities. *Higher Education, 29*(1), 19-35.

Hyland, F., & Lo, M. M. (2006). Examining interaction in the teaching practicum: Issues of language, power and control. *Mentoring & Tutoring, 14*(2), 163-186.

Idapalapati, S. R. (2019). Towards the Fairest Teacher Evaluation Model. *World Journal of Educational Research*, *6*(2), 168-187.

Iyer-O'Sullivan, R., Howard, A., & Donaghue, D. (2015). From bit to whole: Reframing feedback dialogue through critical incidents. *Teacher Evaluation in Second Language Education*, 69-84.

Jackson, C. K., Rockoff, J. E., & Staiger, D. O. (2014). Teacher effects and teacher-related policies. *Annu. Rev. Econ.*, *6*(1), 801-825.

Javadi, F. (2014). On the relationship between teacher autonomy and feeling of burnout among Iranian EFL teachers. *Procedia-Social and Behavioral Sciences*, *98*, 770-774.

Johnson, K. E. (2003). "Every experience is a moving force": Identity and growth through mentoring. *Teaching and Teacher Education*, *19*(8), 787-800.

Johnson, K. E. (2009). *Second language teacher education: A sociocultural perspective*. New York, NY: Routledge.

Johnson, K. E. (2009). Trends in second language teacher education. In A. Burns & J. C. Richards (Eds.), *A Cambridge guide to second language teacher education* (pp. 20-30). New York: Cambridge University Press.

Johnson, S. M. (2015). Will VAMs reinforce the walls of the egg-crate school? *Educational Researcher*, *44*(2), 117-126.

Kamali, J. (2014). Post method survival. *Procedia-Social and Behavioral Sciences*, *98*, 824-829.

Kandil, A. (2011). 'Lesson observation: Professional development or teacher evaluation?', in C. Coombe, L. Stepheson and S. Abu-Rmaileh (eds), *Leadership and management in english language teaching.* Dubai: TESOL Arabia, pp. 227-242.

Kane, T. J., & Staiger, D. O. (2012). *Gathering feedback for teaching: combining high-quality observations with student surveys and achievement gains* (Research paper, MET Project). Seattle: Bill and Melinda Gates Foundation.

Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). Have we identified effective teachers? Validating measures of effective teaching using random assignment. *In Research Paper*. MET Project. Bill & Melinda Gates Foundation.

Khan, G., Khan, A., Hussain, S., & Shaheen, N. (2017). Teacher evaluation: Global perspectives and lessons for Pakistan. *Dialogue (Pakistan)*, *12*(3).

Khany, R., & Darabi, R. (2014). ELT in Iran: Reflection of the principles-based and post-method pedagogy in language teaching. *Procedia-social and Behavioral Sciences*, *98*, 908-916.

Kilbourn, B., Keating, C., Murray, K., & Ross, I. (2005). Balancing feedback and inquiry: How novice observers (supervisors) learn from inquiry into their own practice. *Journal of Curriculum & Supervision*, *20*(4).

Kim, T., & Danforth, S. (2012). Non-authoritative approach to supervision of student teachers: Cooperating teachers' conceptual metaphors. *Journal of Education for Teaching*, *38*(1), 67-82.

King, M. (2015). Evaluating experienced teachers. *Teacher Evaluation in Second Language Education*, 167-179.

Knop, C. K. (1980). *Teaching a second language: A guide for the student teacher.* Language in Education: Theory and Practice. No. 28.

Koçak, A. (2003). *A study on learners' readiness for autonomous learning of English as a foreign language.* (Unpublished master's dissertation). Middle East Technical University, Ankara.

Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review, 47*, 180-195.

Koh, J. H. L., & Chai, C. S. (2016). Seven design frames that teachers use when considering technological pedagogical content knowledge (TPACK). *Computers & Education, 102*, 244-257.

Kumaravadivelu, B. (2001). Toward a postmethod pedagogy. *TESOL Quarterly, 35*(4), 537-560.

Kumaravadivelu, B. (2003). *Beyond methods: Macrostrategies for language teaching*. Yale University Press.

Kumaravadivelu, B. (2006). *Understanding language teaching: From method to postmethod*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kumaravadivelu, B. 2012. *Language teacher education for a global society*. Routledge: Taylor& Francis.

Kydd, L. (1997). Teacher professionalism and managerialism. *Professional development for educational management.* Open University Press, Buckingham, 111-117.

Lai, E. (2010). Getting in step to improve the quality of in-service teacher learning through mentoring. *Professional Development in Education, 36*(3), 443-469.

Lantolf, J. P., & Thorne, S. L. (2006). *Sociocultural theory and the sociogenesis of second language development.* New York: Oxford University Press.

Lasagabaster, D., & Sierra, J. M. (2011). Classroom observation: desirable conditions established by teachers. *European Journal of Teacher Education, 34*(4), 449-463.

Little, O., Goe, L., & Bell, C. (2009). *A practical guide to evaluate teacher effectiveness*. National Comprehensive Center for Teacher Quality.

Lupascu, A. R., Pânisoară, G., & Pânisoară, I. O. (2014). Characteristics of effective teacher. *Procedia-Social and Behavioral Sciences*, *127*, 534-538.

Malderez, A. (2001). New ELT professionals? The role and status of mentors. *English Teaching Professional, 19*, 57-58.

Mann, S. (2004). ‘Evaluation’, in H. Harnisch and P. Swanton (eds). Adults Learning Languages: A CILT Guide to Good Practice, London: *CILT, the National Centre for Languages*, pp. 113-129.

Mann, S. (2005). The language teacher's development. *Language Teaching*, *38*(3), 103-118.

Mann, S., & Walsh, S. (2015). Reflective dimension of CPD: Supporting self-evaluation and peer evaluation. *Teacher evaluation in second language education*.

Marshall, K. (2011). Principal evaluation rubrics. *Recuperadode*. https:// www. cinnaminson.com/pdf/curriculum/Principal%20Evaluation%20Rubrics.pdf

Marshall, K. (2012). Teacher Evaluation: What's Fair? What's Effective? *Educational Leadership, 70*(3), 50-53.

Marshall, K. (2013). How to make teacher evaluations accurate, fair, and consistent. *Education Week*, 1-5.

Marshall, K. (2017). The truth about teacher evaluation. *American School Board Journal*, *204*(2), 76-81.

Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction.* Alexandria, VA: ASCD.

Marzano, R. J. (2012). Teacher evaluation. *Educational Leadership, 70*(3), 14-19.

Marzano, R. J. (2013). The Marzano teacher evaluation model. *Learning Sciences International*, *1*, 1-60.

Marzano, R. J., Carbaugh, B., Rutherford, A., & Toth, M. D. (2013). Marzano center teacher observation protocol for the 2014 Marzano teacher evaluation model. *Palm Beach Gardens, Florida: Marzano Center/Learning Sciences International*.

Marzano, R. J., Frontier, T., & Livingston, D. (2011). *Effective supervision: Supporting the art and science of teaching.* Alexandria, VA: ASCD.

Marzano, R. J., & Toth, M. D. (2013). *Teacher evaluation that makes a difference: A new model for teacher growth and student achievement.* Alexandria, VI: ASCD.

Marzano, R. J., & Toth, M. D. (2014). Teaching for rigor: A call for a critical instructional shift. *Learning Sciences Marzano Center*, 1-24.

Master, B. (2014). Staffing for Success. Educational Evaluation and Policy Analysis, 36(2), 207-227. doi:10.3102/0162373713506552

Masters, G.N. (2009). A shared challenge: Improving literacy, numeracy and science learning in Queensland primary schools. *Camberwell, VIC: Australian Council for Educational Research*.

McCaffrey, D. F., & Lockwood, J. R. (2008). Value-added models: Analytic issues. In *Workshop on value-added modeling sponsored by the national research council and the national academy of education, board on testing and accountability*, Washington, DC.

McCaffrey, D. F. (2010). Value-added models. *Elsevier*, 489-494.

McIntyre, D., & Hagger, H. (1993). Teachers' expertise and models of mentoring. *Mentoring: Perspectives on School-Based Teacher Education*, 86-102.

McIntyre, J., Hobson, A. J., & Mitchell, N. (2009). Continuity, support, togetherness and trust: Findings from an evaluation of a university-administered early professional development programme for teachers in England. *Professional Development in Education*, *35*(3), 357-379.

Mercado, L. A., & Mann, S. (2015). Mentoring for teacher evaluation and development. *Teacher Evaluation in Second Language Education*, 35-54.

Mercer, J. (2006). Appraising higher education faculty in the Middle East: leadership lessons from a different world. *Management in Education*, *20*(1), 17-26.

Mertler, C. A. (2019). *Action research: Improving schools and empowering educators*. SAGE Publications, Incorporated.

Mielke, P., & Frontier, T. (2012). Keeping improvement in mind. *Educational Leadership*, *70*(3), 10-13.

Miles, M. B., Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage Publications.

Miller, L. and Young, J. (2007). 'What is in it for me? A performance management system to please everyone', in C. Coombe, M. Al-Hamly, P. Davidson and S. Troudi (eds), *Evaluating Teacher Effectiveness in ESL/EFL Contexts*. Ann Arbor: University of Michigan Press, pp. 74-88.

Mitchell, K. J., Robinson, D. Z., Plake, B. S., & Knowles, K. T. (2001). *Testing teacher candidates: The role of licensure tests in improving teacher quality*. National Academy Press, 2101 Constitution Avenue, NW, Lockbox 285, Washington, DC 2005.

Moradi, K., & Sabeti, G. (2014). A comparison of EFL teachers and EFL students' understandings of 'highly effective teaching'. *Procedia-Social and Behavioral Sciences*, *98*, 1204-1213.

Morey, A. I. (2003). Major trends impacting faculty roles and rewards: An international perspective. *Globalization and Reform in Higher Education*, 68-84.

Moskal, A. C., Stein, S. J., & Golding, C. (2016). Can you increase teacher engagement with evaluation simply by improving the evaluation system? *Assessment & Evaluation in Higher Education, 41*(2), 286-300.

Moss, J. T. (2015). *The Danielson model of teacher evaluation: Exploring teacher perceptions concerning its value in shaping and improving instructional practice*. Doctoral dissertation, Seton Hall University.

Motallebzadeh, K., Hosseinnia, M., & Domskey, J. G. (2017). Peer observation: A key factor to improve Iranian EFL teachers' professional development. *Cogent Education, 4*(1), 1-12.

Munby, H. (1989). Reflection-in-action and reflection-on-action. *Current Issues in Education, 9*(1), 31-42.

Murphey, T., & Yaode, Q. (2007). A China initiative: Portfolio-based teacher development and appraisal. *Evaluating Teacher Effectiveness in ESL/EFL Contexts*, 119-133.

Nakamura, I. (2014). A Discussion of Practitioner Research: How Are Reflective Practice, Action Research, and Exploratory Practice Different? *Bulletin of Higher Education, Okayama University, 10*, 105-114.

Nami, F., Marandi, S. S., & Sotoudehnama, E. (2016). CALL teacher professional growth through lesson study practice: An investigation into EFL teachers' perceptions. *Computer Assisted Language Learning, 29*(4), 658-682.

Nilsson, P. (2008). Recognizing the needs - student teachers´ learning to teach from teaching. *NorDiNa*, 1, 92-107.

Norman, P. J. (2011). Planning for what kind of teaching? Supporting cooperating teachers as teachers of planning. *Teacher Education Quarterly, 38*(3), 49-68.

Nunan, D. (1987). Communicative language teaching: Making it work. *ELT Journal, 41*(2), 136-145.

Ozturk, I. H. (2012). Teacher's role and autonomy in ınstructional planning: the case of secondary school history teachers with regard to the preparation and ımplementation of annual ınstructional plans. *Educational Sciences: Theory and Practice*, *12*(1), 295-299.

Oxford, R. (1990). *Language learning strategies: What every teacher should know.* New York: Newbury House Publishers.

Oxford, R. L. (2003). Language learning styles and strategies: An overview. *Learning Styles & Strategies/Oxford, GALA, 2003*, 1-25.

Pandhiani, S. M., Chandio, M. T., & Memon, S. (2016). Exploratory practice: Uses and implications. *International Research Journal of Arts & Humanities (IRJAH)*, *43*(43).

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.

Patten, M. L., & Newhart, M. (2017). *Understanding research methods: An overview of the essentials*. Taylor & Francis.

Pylman, S. (2016). Reflecting on talk: A mentor teacher's gradual release in co-planning. *The New Educator*, *12*(1), 48-66.

Quirke, P. (2015). A system for teacher evaluation. *Teacher Evaluation in Second Language Education*, 101-114.

Randall, M. (2015). Observing for feedback: A counselling perspective. *Teacher Evaluation in Second Language Education*, 55-65.

Rantanen, P. (2013). The number of feedbacks needed for reliable evaluation. A multilevel analysis of the reliability, stability and generalisability of students' evaluation of teaching. *Assessment & Evaluation in Higher Education*, *38*(2), 224-239.

Rayan, P. (2013). Peer Observation of teaching: A tool for professional development. *English for Specific Purposes World*, *14*(39), 45-54.

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, *44*, 159-178.

Reeve, J., & Cheon, S. H. (2016). Teachers become more autonomy supportive after they believe it is easy to do. *Psychology of Sport and Exercise*, *22*, 178-189.

Richards, J. C. (2001). *Curriculum development in language teaching.* Cambridge: Cambridge Language Education.

Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers: Strategies for teacher learning*. Cambridge: Cambridge University Press.

Richards, J. C., & Rodgers, T. S. (1986). Through the looking glass: Trends and directions in language teaching. *RELC, 18*(8), 45-73.

Richards, J. C., & Rodgers, T. S. (2001). *Approaches and methods in language teaching.* Cambridge: Cambridge University Press.

Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge.

Ridge, B. L., & Lavigne, A. L. (2020). Improving instructional practice through peer observation and feedback: A review of the literature. *Education Policy Analysis Archives, 28*(61), 1-32.

Riera, G. (2011). New directions in teacher appraisal and development. *Leadership and Management in English Language Teaching. Dubai: TESOL Arabia,* 49-66.

Roberts, J. R. (1993). Evaluating the impacts of teacher research. *System, 21*(1), 1-19.

Roberts, J. (1998). *Language teacher education*. London: Arnold.

Robson, S., & Turner, Y. (2007). 'Teaching is a co-learning experience': academics reflecting on learning and teaching in an 'internationalized' faculty. *Teaching in Higher Education, 12*(1), 41-54.

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review, 94*(2), 247-252.

Rockoff, J. E., & Speroni, C. (2010). Subjective and objective evaluations of teacher effectiveness. *American Economic Review, 100*(2), 261-66.

Rush, L. S., Blair, S. H., Chapman, D., Codner, A., & Pearce, B. (2008). A new look at mentoring: Proud moments and pitfalls. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 81*(3), 128-132.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*, 68-78.

Saban, A., & Kocbeker, B. N. (2007). Prospective teachers' conceptions of teaching and learning revealed through metaphor analysis. *Learning and Instruction*, *17*(2), 123-139.

Saberi, M., & Rahimi, A. (2014). On the effect of personality and efficacy on second language instructors' performance. *Procedia-Social and Behavioral Sciences*, *136*, 240-245.

Saito, E. (2012). Key issues of lesson study in Japan and the United States: A literature review. *Professional Development in Education*, *38*(5), 777-789.

Sanders, W. L., Wright, S. P., & Horn, S. P. (1997). Teacher and classroom context effects on student achievement: Implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, *11*(1), 57-67.

Sandilos, L. E., Sims, W. A., Norwalk, K. E., & Reddy, L. A. (2019). Converging on quality: Examining multiple measures of teaching effectiveness. *Journal of School Psychology*, *74*, 10-28.

Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). Rethinking teacher evaluation in chicago: Lessons learned from classroom observations, principal-teacher conferences, and district implementation. *Research Report*. Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.

Sass, T. R., Semykina, A., & Harris, D. N. (2014). Value-added models and the measurement of teacher productivity. *Economics of Education Review*, *38*, 9-23.

Savaş, H. (2016). Maintaining the efficacy of warm-up activities in EFL classrooms: An exploratory action research. *Journal of Qualitative Research in Education*, *4*(2), 65-85.

Savaş, H., & Erol, A. Y. (2015). Distribution and comparison of language learning strategies used by language learners and their status as repeating and non-repeating at language levels. *Journal of Language and Literature Education*, *3*(14), 1-22.

Savignon, S. J. (2007). Beyond communicative language teaching: What's ahead? *Journal of Pragmatics*, *39*(1), 207-220.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action.* Temple Smith, London.

Schwille, J., Dembélé, M., & Schubert, J. (2007). Global perspectives on teacher learning: Improving policy and practice. *International Institute for Educational Planning (IIEP) UNESCO*.

Shah, S. R., & Al Harthi, K. (2014). TESOL classroom observations: A boon or a bane? An exploratory study at a Saudi Arabian university. *Theory & Practice In Language Studies*, *4*(8).

Shoffner, M. (2009). The place of the personal: Exploring the affective domain through reflection in teacher preparation. *Teaching and Teacher Education, 25*(6), 783-789.

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*, 4-14.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1-22.

Simonian, J., & Robertson, M. (2002). Learner empowerment through self-evaluation. In A. Mackenzie & E. McCafferty (Eds.), *Developing autonomy* (pp. 199-205). Tokyo, Japan: The Japanese Association of Language Teachers (JALT).

Skehan, P. (1989). Language testing part II. *Language Teaching, 22*(1), 1-13.

Skinner, B., & Hou, H. (2014). The evaluation of non-native speaking English language trainee teachers' practice: unfolding university supervisors' and host teachers' perspectives on judging performance. *British Council ELT Research Papers, 14(6)*, 1-18.

Slattery, P. (2006). *Curriculum development in the post-modern era.* (2nd ed.). London: Routledge.

Smith, C. (2008). Building effectiveness in teaching through targeted evaluation and response: Connecting evaluation to teaching improvement in higher education. *Assessment & Evaluation in Higher Education, 33*(5), 517-533.

Spilková, V. (2001). Professional development of teachers and student teachers through reflection on practice. *European Journal of Teacher Education, 24*(1), 59-65.

Spratt, M., Pulverness, A., & Williams, M. (2005). *The TKT course: Teaching knowledge test*. Cambridge University Press.

Stanulis, R. N., & Bell, J. (2017). Beginning teachers improve with attentive and targeted mentoring. *Kappa Delta Pi Record, 53*(2), 59-65.

Stein, S. J., Spiller, D., Terry, S., Harris, T., Deaker, L., & Kennedy, J. (2013). Tertiary teachers and student evaluations: Never the twain shall meet?. *Assessment & Evaluation in Higher Education, 38*(7), 892-904.

Strauss, R. P., & Sawyer, E. A. (1986). Some new evidence on teacher and student competencies. *Economics of Education Review, 5*, 41-48.

Stronge, J. H. (Ed.). (1997). *Evaluating teaching: A guide to current thinking and best practice.* Thousand Oaks, CA: Corwin Press.

Stronge, H. J. (2007). *Qualities of effective teachers*. Alexandria, Virginia, USA: Association for Supervision and Curriculum Development.

Strøm, B., & Falch, T. (2020). The role of teacher quality in education production. In *the Economics of Education* (pp. 307-319). Academic Press.

Tanış, A., & Dikilitaş, K. (2018). Turkish EFL instructors' engagement in professional development. *Eurasian Journal of Applied Linguistics, 4*(1), 27-47.

Tarhan, H., Karaman, A. C., Kemppinen, L., & Aerila, J. A. (2019). Understanding teacher evaluation in Finland: A professional development framework. *Australian Journal of Teacher Education, 44*(4), 33-50.

Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research*. Sage Publishing.

Taylor, E. S., & Tyler, J. H. (2012). Can teacher evaluation improve teaching? *Education Next, 12*(4), 78-84.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-3651.

Teng, F. (2019). Teacher Autonomy: A Buzzword in Teaching English as a Foreign Language. *Autonomy, Agency, and Identity in Teaching and Learning English as a Foreign Language*, 21-36.

Thiel, T. (1999). Reflections on Critical Incidents. *Prospect, 14*(1), 44-52.

Thornbury, S. (2006). *An AZ of ELT*. Oxford: Macmillan.

Todd, P. E., & Wolpin, K. I. (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal, 113*, 3-33.

Tomlinson, B. (2003). Developing materials to develop yourself. *Humanising Language Teaching, 5*(4).

Trotman, W. (2015). Reflective peer observation accounts: What do they reveal? *Teacher Evaluation in Second Language Education*, 181-191.

Tschannen-Moran, M.., & Hoy, W. A. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*(2001) 783-805.

Tucker, P. D., & Stronge, J. H. (2005). *Linking teacher evaluation and student learning* Alexandria, VA: Association for Supervision and Curriculum Development.

Ulas, J., & Aksu, M. (2015). Development of teacher autonomy scale for Turkish teachers. *Procedia-Social and Behavioral Sciences, 186,* 344-349.

Urrieta, L., 2007. Figured worlds and education: An introduction to the special issue. *The Urban Review, 39*(2), 107-116.

Uzun, L. (2016). Evaluation of the latest English language teacher training programme in Turkey: Teacher trainees' perspective. *Cogent Education, 3*(1), 1-16.

Vodicka, D. (2006). The four elements of trust. *Principal Leadership, 7*(3), 27-30.

Voogt, J., & McKenney, S. (2017). TPACK in teacher education: Are we preparing teachers to use technology for early literacy? *Technology, Pedagogy and Education, 26*(1), 69-83.

Wan, W., Low, G. D., & Li, M. (2011). From students' and teachers' perspectives: Metaphor analysis of beliefs about EFL teachers' roles. *System, 39*(3), 403-415.

Watson-Davis, R. (2009). *Lesson observation pocketbook.* Alresford, Hampshire: Teachers' Pocketbooks.

Wei, W. (2015). Using summative and formative assessments to evaluate EFL teachers' teaching performance. *Assessment & Evaluation in Higher Education, 40*(4), 611-623.

Wermke, W., & Forsberg, E. (2016). The changing nature of autonomy: transformations of the late Swedish teaching profession. *Scandinavian Journal of Educational Research.* doi:10.1080/00313831.2015.1119727.

White, S. (2009). Articulation and re-articulation: Development of a model for providing quality feedback to pre-service teachers on practicum. *Journal of Education for Teaching*, 35 (2), 123-132.

Widodo, H. P., & Zakaria, F. (2008). Language teacher supervision: A case-based approach. *Reading*, *8*(1), 14-43.

Wright, T. (1987). *Roles of teachers and learners*. Oxford University Press, USA.

Wyatt, M., & Dikilitaş, K. (2016). English language teachers becoming more efficacious through research engagement at their Turkish university. *Educational Action Research, 24*(4), 550-570.

Yavuz, A. (2011). The problematic context of mentoring: evidence from an English language teaching department at a Turkish university. *European Journal of Teacher Education, 34*(1), 43-59.

Yıldırım, Ö. (2012). A study on a group of Indian English as a second language learners' perceptions of autonomous learning. *Turkish Online Journal of Qualitative Inquiry, 3(2),* 18-29.

Yook, C., & Lee, Y. H. (2016). Korean EFL teachers' perceptions of the impact of EFL teacher education upon their classroom teaching practices. *Asia-Pacific Journal of Teacher Education, 44*(5), 522-536.

Zeichner, K. (1999). The new scholarship in teacher education. *Educational researcher, 28*(9), 4-15.

Zhang, L. F. (2009). From conceptions of effective teachers to styles of teaching: Implications for higher education. *Learning and Individual Differences*, *19*(1), 113-118.

Zonoubi, R., Rasekh, A. E., & Tavakoli, M. (2017). EFL teacher self-efficacy development in professional learning communities. *System*, *66*, 1-12.