



Suggesting a Teacher Assessment and Evaluation Model for Improving the Quality of English Teachers

Sabahattin Yeşilçınar ¹, Abdulvahit Çakır ²

Abstract

Teacher quality is *sine qua non* of student achievement; however, choosing a qualified teacher is possible only when a valid and fair teacher assessment and evaluation (TAE) model is in place. The present mixed-method research aimed to develop a TAE model that can tell apart qualified English prospective teachers (who have required knowledge, skills, and attitudes) from unqualified ones. The qualitative data used in this research were collected from 78 stakeholders, whereas quantitative data were obtained from 271 in-service English teachers. Stakeholders' perceptions of testing were gathered as they are informative in terms of consequential validity of exams. The sample of prospective teachers and in-service teachers involved graduates of both English Language Teaching (ELT) and other departments such as English Language and Literature (ELL) since they could become English language teachers by obtaining a teacher certificate. Semi-structured and focus group interview techniques were used to obtain qualitative data. Regarding quantitative data, the English Teachers' Attitudes towards Recruitment System (ETARS), a valid and reliable scale developed by the researcher, was applied. SPSS was used to analyse quantitative data, and content analysis was used for qualitative data. The codes and categories were formed primarily by the researcher, and NVivo 12 was used to prevent data loss. Quantitative data showed that English teachers had a negative attitude towards the current TAE model. The qualitative findings were supported by quantitative data and indicated that the current TAE model is ineffective and unfair as it has both construct under-representation and construct-irrelevant variance (favouring graduates of ELL in terms of scoring). In this context, a new TAE model that can evaluate applicants' knowledge, skills, and attitudes has been developed. This model is considered to be helpful in selecting qualified teachers, thus improving the quality of foreign language education.

Keywords

Teacher Assessment
Teacher Recruitment
English in Turkey
Consequential Validity
Washback Effect

Article Info

Received: 01.28.2019
Accepted: 11.20.2019
Online Published: 04.04.2020

DOI: 10.15390/EB.2020.8463

¹ Muş Alparslan University, Faculty of Education, English Language Teaching, Turkey, sycinar@yahoo.com

² Gazi University, Gazi Faculty of Education, English Language Teaching, Turkey, vahit@gazi.edu.tr

Introduction

Experts regard teacher quality as a determinant of the quality of any educational system (Ingersoll & Merrill, 2010). Consequently, considerable importance should be attached to teacher assessment and evaluation (TAE) policies, because there is a strong relationship between the effectiveness of an education system and its teacher selection and recruitment techniques (Jacob, 2016). To maximize the impact of the TAE process, it is necessary 'to actively look for, identify, and hire teacher applicants who exhibit what research indicates to be the qualities of effective teachers' (Stronge & Hindman, 2006, p. 1). A poor recruitment system, on the other hand, can result in a decline in the quality of education for years to come (Mason & Schroeder, 2010). Although defining 'quality' and 'how to measure it' is debatable (Cohen, 2010; Kennedy, 2008), a qualified teacher is considered to perform 'both successful teaching, which realizes intended outcomes, and good teaching, which is morally worthwhile' (Schussler, Stooksberry, & Bercaw, 2010, p. 351). Multifaceted evidence of teacher practice may be required to measure teacher quality (Darling-Hammond, 2014).

To determine the required competencies of English teachers, researchers analysed views of foreign language teachers and/or students in many countries such as Turkey (Çelik, Arıkan, & Caner, 2013), Vietnam (Nghia, 2015), Israel (Brosh, 1996), Korea (Barnes & Lock, 2010), and Thailand (Wichadee, 2010). Desired qualities can be listed as rapport, lesson delivery, fairness, knowledge and credibility, and organization and preparation in Korea (Barnes & Lock, 2010); English proficiency, pedagogical knowledge, organization and communication skills, and socio-affective skills in Thailand (Wichadee, 2010); pedagogy-specific knowledge, personality traits, professional skills, and classroom behaviours in Turkey (Çelik et al., 2013). Despite differences across countries, there seems to be a consensus that English teachers should have the required (a) knowledge (e.g. literacy, numeracy, subject knowledge, knowledge of curriculum, etc.), (b) skills (ability to perform a lesson, to use technology appropriately, to speak target language fluently, assessment literacy, classroom management, communication, etc.), and (c) attitudes (attitude towards teaching, motivational disposition, etc.). However, prospective teachers' acquisition of such qualities depends largely on teacher assessment (Kılıçkaya & Krajka, 2013).

Klassen et al. (2017) list three necessary reasons for selection in any jurisdiction:

- a) to make decisions about 'selecting in' when the number of applicants outweighs the number of available places, b) to make decisions about 'selecting out' in order to identify those candidates who may be unsuitable, and c) to provide a profile of candidates' strengths and weaknesses for future development (p. 97).

The current Turkish TAE model basically depends on the first criterion; therefore, *teacher assessment and evaluation* and *teacher selection and recruitment* are interchangeable in this study. To ensure each student has a quality teacher, educational leaders must recruit better teachers (Rose, English, & Finney, 2014). Recruiting more teachers will not solve the problems related to quality in education if the rationale behind the selection and recruitment process is not considered as finding the person who best fits the needs of the students, schools, and finally quality education (Ingersoll, 2001). Therefore, this process is of importance for many administrators who possess the ultimate chance to influence both instructional practices and student learning via the teacher selection process (Mertz, 2010). In the Teacher Strategy Paper published by the Ministry of National Education (MEB), it is emphasized that policymakers should attach utmost importance to the TAE process for better quality education and greater student achievement (MEB, 2017). This confirms that a carefully designed TAE model is a must to improve the quality of education and teachers, and enhance student achievement (please see Figure 1).

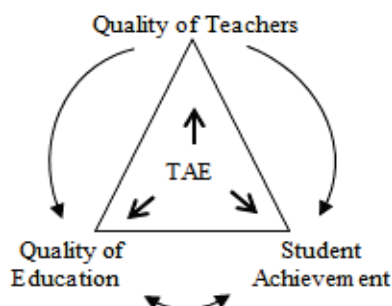


Figure 1. The impact of the TAE model on teachers, students, and education.

Teacher Assessment and Evaluation in Turkey compared to International Contexts

It is widely accepted that high-stakes tests have consequences at both the micro level of language teaching and learning inside the classroom and the macro level of education and society as a whole (Bachman & Palmer, 1996). Tests are used for many purposes, such as selection, employment, promotion, immigration, citizenship, monitoring the performance of students and schools, “implementing educational policies, reforming educational systems, and deciding on the distribution of funding” (Cheng, 2005; Eades, 2005; McNamara, 2012, as cited in Pan & Roever, 2016, p. 1). The TAE examinations in Turkey naturally aim for similar consequences. Available international literature on TAE is totally different from that of Turkey. The former generally aims to reveal school administrators’ opinions regarding vital factors in the TAE process (Liu, Liu, Stronge, & Xu, 2016). In addition, schools or districts are mostly autonomous and can hire their teachers, developing their own TAE policies so that they can eliminate ineffective teachers (Adnot, Dee, Katz, & Wyckoff, 2017). However, TAE in Turkey is highly centralized and those who want to be English teachers in public schools are assessed through KPSS (Public Personnel Selection Exam), which is conducted in four phases: general culture and general skills (120 questions), educational sciences (80 questions), ÖABT (Teaching Field Knowledge Test, 50 questions), and Interview. The first three phases are carried out through multiple-choice exams. Successful candidates are interviewed, and the interview score is determinant in deciding whether the candidate will be selected. Lacking scoring rubrics, lacking field experts, and being run in Turkish are the main limitations of the Interview phase. In contrast to Turkey, countries which achieved notable success in PISA 2015 (such as Finland, Japan, South Korea, and New Zealand) utilize various assessment tools (e.g. written exams and performance exams) (Kang & Hong, 2008; OECD, 2016; Ramírez Carpeño, & Mekochi, 2015). To examine whether the TAE model of a country positively affects its student outcomes (Rivkin, Hanushek, & Kain, 2005), in addition to Turkey, three countries were selected and compared in terms of their PISA 2015 reading scores (see Table 1). PISA divides countries into three categories: (1) countries whose students have high results, (2) countries whose students are in the middle, and (3) countries whose students have low results. South Korea, Spain, and Thailand were chosen accordingly so that they could represent all three categories in the present study. For a reasonable comparison, three criteria were considered. The first criterion is the nature of English (it should be a foreign language rather than second language in the selected countries). Second, the languages of the selected countries should not belong to the English language family. The final criterion is that there should not be a huge difference between the population of the selected country and that of Turkey.

Table 1. The Relationship between Countries’ TAE Model and Their Quality of Education

	Turkey	South Korea	Spain	Thailand
Teaching status	permanent and respected	permanent and respected	permanent and respected	permanent and respected
The need for teacher certificate	√	√	√	√
Competitive employment exam	√	√	√	√
Assessing teaching performance	-	√	√	-
PISA 2015 reading scores	below average	above average	average	below average

In all the four countries, a teacher certificate is a must, and all candidates have to pass competitive examinations held nationwide to become a teacher in public schools. In South Korea, teacher candidates need to pass a competitive test that assesses their knowledge, skills, and dispositions in three phases (Ingersoll, 2007). One of the exams consists of teaching performance and interview. Unlike Turkey, regional authorities employ teachers; therefore, their scores are applicable in the province they apply, and the recruitment process has high standards since the candidates are well assessed before being admitted to a teacher training programme (Kang & Hong, 2008). In Spain, nationwide competitive examination has two phases. The first involves a practical exam that intends to evaluate prospective teachers' specific knowledge of teaching, learning, and techniques, whereas the second phase measures pedagogical knowledge of the prospective teachers by asking them to 'present a teaching programme based on the first exam' and to 'do an oral presentation of a teaching unit' (Ramírez Carpeño, & Mekochi, 2015, p. 117). Thailand's recruitment process is similar to that of Turkey in a lot of ways; there is little correlation between the number of vacant teaching positions and the number of teachers to be employed; therefore, the needs of individual schools are not taken into account (Mattavarat, Viseshsiri, & Siribanpitak, 2017). The allocation of new teachers is based on the results of a competitive employment exam. The prospective teacher who scores the highest in the exam gets to make the first choice of the positions available. After making their choice, teachers must also pass a selection test, which resembles the employment exam (OBEC, 2015). Staff in the ESA (Educational Service Area) may also interview candidates. The ESA has been established under the authority of the Basic Education Commission in response to the decentralization of authority for educational administration envisaged in the *National Education Act* and is responsible for tasks such as establishment, overseeing, monitoring, and evaluation of basic education institutions (International Business Publications, 2008). It seems that the employment exam conducted by ESA fails to meet the 'acceptable standards of validity, fairness, and reliability' (OECD/UNESCO, 2016, p. 237). These data back the literature that countries performing well in education have a selective TAE model that takes into account the knowledge, skills, and attitudes of prospective teachers through multiple sources of information, and that countries performing poorly in education usually depend on knowledge-based multiple choice exams (Darling-Hammond, 2010).

Limitations of the current TAE Model

Shohamy (2001) coined the term Critical Language Testing (CLT) for defining the uses and consequences of tests in education and society. Shohamy (2017) advocates that the existing research on CLT tries to explain how the misuse of high-stakes tests and their impact move beyond learning and teaching. *High-stakes tests* refer to examinations whose results are used for making significant decisions that affect students, teachers, administrators, communities, schools, and districts (Madaus, 1988, as cited in Hatipoğlu, 2016). They affect both teaching paradigms and educational systems in several countries (Loumbourdi, 2014), such as 'narrowing' of the curriculum (Cheng, 2005), shifting of teaching methods (Spratt, 2005), and individual learning styles (Shih, 2008). In CLT, examining the perception of stakeholders towards education policy and how testing is used to develop educational outcomes for all students is about test validity, because any knowledge of reality (e.g. test validity, fairness, and impact) does not remain independent of how people experience it (Cohen, Manion, & Morrison, 2011, as cited in Kumar, Roberts, Bartle, & Eley, 2018). Considering *consequential validity*, *construct under-representation* and *construct-irrelevant variance* are two major threats to validity (Messick, 1989). The former happens when it is not possible for test takers to give their best in a given examination because of the failure of the test in capturing and presenting vital aspects of the construct that it aims to measure, while the latter involves factors that are irrelevant to the construct being measured and occurs because of the effects of those irrelevant factors on scores. Test taker's psychology, personality characteristics, background knowledge, general intellectual or cognitive ability, subject area knowledge, and test-taking strategies can all be potential sources of construct-irrelevant variances (Cheng, 2014). Construct under-representation and construct-irrelevant variance need to be examined to improve the test quality and thus promote positive washback. *Washback* refers to the effect of language testing on teaching and learning (Cheng, 2014). It can be negative (harmful) or positive (beneficial). While the previous one refers to a narrow definition of language ability, the latter results in 'good' teaching practice. Washback

becomes a well-known concept because of the misuses of tests (Cheng, 2014). The literature on washback is abundant in Turkey (Hatipoğlu, 2016; Özmen, 2011; Sayın & Aslan, 2016). When the aforementioned studies are examined, it is seen that they mostly focused on attitudes and perceptions of students towards exams.

KPSS is a high-stakes test and it may have intended/unintended and positive/negative effects on the stakeholders. Regarding test validity, it is vital to determine whether these exams are useful and to investigate their results. When negative consequences occur, test developers and policymakers should investigate and solve the problem (Hitchcock, Onwuegbuzie, & Khoshaim, 2015). Such a problem exists in ÖABT, which consists of Field Knowledge Test (language proficiency, linguistics, and literature) and Field Training Test (methods and approaches, and teaching language skills). Although the introduction of ÖABT aimed to select qualified teachers, the fact that the graduates of the department of ELL can apply for KPSS may lead to the occurrence of construct-irrelevant variance in ÖABT, because they can solve literature questions easily while graduates of ELT have difficulty answering them. This affects the scores, thus reducing the test validity. In other words, the current TAE model displays unintended consequences, which diminishes its fairness and validity.

Importance of the Study

To improve English teaching quality, the MEB initiated various reforms in the ELT curriculum, specifically in 1997, 2006, and 2012. These reforms are a landmark in Turkish history (e.g. introducing the communicative approach into ELT and starting to teach English in primary schools) (Kırkgöz, 2007). However, despite these reforms, being fluent in English, especially being able to communicate in English, is still a questionable issue. Although a report prepared in partnership with British Council/TEPAV (2013) claims that the failure in teaching English in Turkey is not due to English teachers, it contains contradictory findings about the language proficiency of English teachers (Enginarlar, 2014). For example, the report claims that more than 80% of the observed teachers have professional competence and language level; however, in the same report, English teachers are found to consider themselves inefficient, and about 50% of the students find the lessons boring and not effective. In all schools visited, it was observed that English is not seen as a language of communication and that some teachers used 100% Turkish with occasional English words. This targets not only English teachers but also the TAE model, because the quality of teachers mirrors the quality of the TAE model (Jacob, 2016; Stronge & Hindman, 2006, p. 17). Instead of using only one method of assessment (multiple-choice questions), the need is to adapt policies to facilitate the process of assessing and evaluating quality teachers with the required knowledge, skills, and attitudes. No study has been conducted to offer a TAE model for English language teachers in a centralized EFL context. Thus, this study aims to develop such a TAE model in which the assignment process is done centrally, criticizing the 'one-size-fits all' approach. This consideration seems to be vital for the future employability of English teachers in EFL contexts where the recruitment takes place centrally.

The study aims to develop a fair, reliable, and valid TAE model for English language teachers by:

- examining the perceptions and attitudes of education stakeholders about the current TAE model in Turkey
- revealing the impact of the current TAE model on prospective teachers, in-service English teachers, and education.

To achieve the above-mentioned aims, the following research questions were asked:

1. Which competencies should a qualified English language teacher have?
2. What are the stakeholders' perceptions of the current TAE model?
3. What are the attitudes of in-service EFL teachers towards TAE model?
4. What might an appropriate TAE model look like?

Method

Research Design

Data were collected through qualitative and quantitative methods. Mixed methods is a research methodology that benefits from the strengths of both qualitative and quantitative research by combining them within a single study in the full process of research, including philosophical assumptions, research design, data collection methods, data analysis methods, and the interpretations of results (Creswell & Plano Clark, 2011). Thus, the burden on the shoulders of the researcher may get lighter when approaching research question using strengths of mixed methods research approach. More specifically, an exploratory sequential mixed methods design was used because the researcher first collected and analysed the qualitative data, and then developed a quantitative tool (ETARS) with the emerging themes to further explain the research problem (Onwuegbuzie, Bustamante, & Nelson, 2010).

The qualitative design is based on grounded theory in which the occurring theory is grounded in empirical data and derived from inductive analysis (Strauss & Corbin, 1990). Grounded theory is the generation of theories from data and simply the discovery of emerging patterns in data (Walsh et al., 2015). In grounded theory, data analysis is an ongoing process that starts with the very first interview and continues until the study is completed (Kvale, 1996). Thus, the researchers followed data saturation (Morse, 2004) i.e. the researchers continued analysing data until no new data appeared and all codes as well as concepts were well-developed.

Participants and Sampling

The quantitative phase had 271 English teachers (please see Table 2). English teachers (from primary to upper secondary schools) were selected according to their convenience. The scale was sent to participants via e-mails, and each participant was asked whether s/he could forward the mail to one or more of her/his friends to ensure an adequate number of participants. Thus, the quantitative part included both convenience sampling, which is used for 'choosing settings, groups, and/or individuals that are conveniently available and willing to participate in the study', and snowball sampling, in which 'participants are asked to recruit individuals to join the study' (Onwuegbuzie & Collins, 2007, p. 286). Two criteria were considered while choosing the sample of English teachers: those who a) were working in public schools, and b) had ÖABT experience.

Table 2. Demographic Features of Quantitative Sample

		N	%
Sex	Male	83	30.6
	Female	188	69.4
	Total	271	100
Grades they taught	Primary	53	19.6
	Lower secondary	124	45.7
	Upper secondary	94	34.7
	Total	271	100
Field of Study	ELT	218	80.4
	Formation	53	19.6
	Total	271	100

To make the study more effective, 78 participants in the qualitative sample were selected purposefully (Tongco, 2007), including prospective teachers (n=31), English teachers (n=25), teacher trainers (n=9), education experts (n=8), and policymakers (n=5) from various institutions such as Council of Higher Education (YÖK), MEB, and universities (please see Table 3).

Table 3. Demographic Features of Qualitative Sample

Prospective teachers (ST)	Sex	Male	13
		Female	18
	Department	ELT	15
		Others	16
In-service teachers (T)	Sex	Male	10
		Female	15
	Department	ELT	14
		Others	11
Teacher trainers (TT)	Sex	Male	3
		Female	6
	Title	Full Professor	4
		Associate Professor	5
Education experts (EE)	Sex	Male	7
		Female	1
	Title	General Director	2
		Full Professor	5
		Associate Professor	1
	Affiliation	MEB, YÖK, University	
Policymakers (PM)	Sex	Male	5
		Female	-
	Title	General Director	3
		Full Professor	2
Affiliation	MEB, YÖK		

The sample of prospective teachers involved 31 senior students of ELT or other departments (e.g. ELL, American Culture and Literature, Department of Translation and Interpreting, Department of English Linguistics). The imperative criterion for choosing prospective teachers from departments other than ELT was pedagogical formation training. In this context, Gazi University, Hacettepe University, and Ankara University were contacted, and ethical permissions were obtained in October 2017. English teachers sample (n=25) included graduates of ELT or of other departments. Two criteria were considered while choosing the sample: a) presently working in public schools (the teachers of private schools were not chosen since private schools determine their own assessment criteria), and b) having ÖABT and/or interview experience. Thus, teachers working in all districts of Ankara, İzmir, Konya, Muş, and Van were interviewed, and ethical permissions were obtained from MEB in May 2018. For ethical concerns, both teachers and prospective teachers were asked to read and sign the Participant Consent Form.

The sample of teacher trainers consisted of 9 interviewees who had at least ten years' experience. The interviewed teacher trainers were either associate professors or full professors from five different universities (Çukurova, Gazi, Hacettepe, METU, and Sakarya). All teacher trainers were experts in English language teaching. The sample of education experts included 8 participants from different institutions, such as YÖK, ÖSYM, MEB, and universities. The salient feature of these interviewees was that they were all invited to the workshops on ÖABT, and that they contributed to the content and process of the TAE model in Turkey. Finally, the sample of policymakers was composed of 5 participants. They were executive persons either working as general directors in the Ministry of National Education or working as members of the Council of Higher Education and are closely involved in the education policy of the country. An informative e-mail was sent to teacher trainers, education experts, and policymakers in advance to request an appointment. Interviews were held in their rooms between February and May 2018 according to their convenience. All conversations were audio-recorded.

Data Collection Tools

Several data collection techniques were used to examine the phenomenon in question. The qualitative data collection techniques involved semi-structured interviews, focus groups, and documents, while ETARS was used to collect the quantitative data.

Semi-Structured Interviews

To obtain in-depth information from the participants, semi-structured interviews, which are more flexible and common interview types (Rubin & Rubin, 2005), were conducted. The semi-structured interview forms consisted of two parts. The first part was about participants' demographic information (sex, university and department where they graduated, etc.). The second part consisted of questions and sub-questions that aimed to reveal participants' perceptions about the TAE model (e.g. its usefulness, its functionality, whether it should be revised, etc.). Interviews were conducted in Turkish to raise the quality and quantity of the data gathered (Mackay & Gass, 2005).

Focus Group Interviews

Focus group interviews require groups of approximately six to twelve people at a time. Focus group interviews were organized because the researchers aimed to generate lively discussions among participants so as to understand the perceptions of different stakeholders towards the TAE policy and observe whether stakeholders arrived at a consensus on controversial topics over time (Onwuegbuzie, Dickinson, Leech, & Zoran, 2010). The focus group with prospective teachers included 12 participants (six graduates of ELT and six graduates of other departments). The in-service teacher focus group consisted of five ELT and four non-ELT graduates. Focus groups were selected with the assumption that they might particularly help us understand the perceptions of ELT and non-ELT prospective teachers and graduates about the TAE model, how they addressed such a conflicting issue, and whether a change occurred in their perceptions.

Document Review

Documents help researchers find information easily. They are 'social facts', which are generated, shared, and utilized in socially organized ways (Atkinson & Coffey, 1997, p. 47). They can take a variety of forms, including advertisements, minutes of meetings, manuals, diaries and journals, organizational or institutional reports, survey data, and various public records, which can provide valuable information and insights (Bowen, 2009). Therefore, the government's development plans and the reports prepared by the national education councils were examined to analyse Turkish education policies from 1923 until today. In addition, for determining a framework of teacher competencies that an effective English language teacher should have in Turkey, both local competencies (e.g. MEB, YÖK) and global competencies (e.g. CAEP, CEFR, EPOSTL, Education Council, TESOL) were reviewed.

English Teachers' Attitudes towards Recruitment System Scale (ETARS)

ETARS is a valid and reliable attitude scale developed by the researcher (Yeşilçınar, 2018). It is in Likert-type with two parts. The first part includes demographic questions (such as sex, the grades s/he teaches, field of study, reason for choosing the job, and desired profession). The second part consists of three factors together with 23 items. The first factor (*positive attitude towards TAE*) is about the positive attitude of English teachers towards the selectiveness and fairness of the TAE model, while the third one (*negative attitude towards TAE*) displays their negative attitudes. The second factor (*being interested in TAE*) aims to reveal whether English teachers are curious about the way they are assessed and evaluated while being assigned to public schools. Cronbach's alpha (α) internal consistency coefficient for the scale is .758 and the reliability coefficients of dimensions are between .703 and .918. The scale is considered reliable since '.70 and preferably higher' is an acceptable value for reliability (Büyüköztürk, 2011). The total variance of the three components of the scale was found to be 56.610, and it was also noted that the first, second, and third components explained 29.909%, 20.180%, and 6.520% of the scale, respectively.

Analysis

The qualitative data were initially analysed manually as a pilot process (Welsh, 2002) so that researchers could observe, control, and manage the data, because it is not the computer but the researcher who analyses qualitative data; the computer merely designs the data (Merriam, 1998). Then, NVivo 12 was used to combine old (manual) and new (computer) methods for best results (Welsh, 2002). NVivo helps the user import documents directly from a pdf, word, or voice source and code them easily on the screen. It makes strips visible on the screen, so that the researcher can see all codes and themes at a glance. Dörnyei's (2007) stages were considered: a) transcribing, (b) pre-coding/coding, (c) growing ideas, and (d) interpreting the data and drawing conclusions (p. 246). Content analysis was used to analyse the qualitative data. First, semi-structured and focus-group interviews were transcribed, since transcription is the first step of analysing and deducing themes (Rubin & Rubin, 2012). Researchers omitted some words (e.g. hmm, uh, huh etc.) by adopting 'edited transcription' (Hansen, 2003, p. 136) in order to ease the analysis process. Second, the researchers read the data for obtaining a general idea and for checking pre-coding. Then, the data were examined and categorized. Finally, the same transcriptions were sent to an independent researcher in the field of Measurement and Evaluation, and the researcher was requested to label the transcriptions. The inter-rater reliability coefficient was found to be satisfactory (96%) since it exceeded the threshold value (90%) (Miles & Huberman, 1994).

Data obtained from the ETARS were analysed using SPSS. Descriptive statistics were presented, including measures of frequency, central tendency, and variability. The t-test is used when the population means of only two groups is to be compared; however, ANOVA is preferred when means of 3 and more groups are to be compared (Büyüköztürk, 2011). Therefore, the t-test (Independent Samples Test) and ANOVA tests were used for the demographic variables.

Results

The results were introduced with reference to the research questions.

Findings Related to the Required Competencies for a Qualified EFL Teacher

The coding process revealed three categories of competencies (knowledge-related competencies, skill-related competencies, and attitude-related competencies) (please see Table 4). In other words, although stakeholders listed a lot of competencies for English teachers, the competencies were noted to be related to either knowledge, or skills, or attitudes of prospective teachers. Some representative quotes are:

Knowledge-Related Competencies

A qualified teacher must possess the required knowledge of the subject s/he teaches. (T6)

We suggest prospective teachers to spend more time on techniques, approaches, and principles in language teaching. Such knowledge helps them conduct lessons in accordance with learners' needs and levels so that their learners do like English lessons. (TT3)

Content knowledge and pedagogical content knowledge are a must for a qualified teacher. Teachers cannot teach their subjects if they do not have such knowledge. (PM1)

Skill-Related Competencies

Without speaking skills, you cannot use the language, so speaking skill is of importance. Then... other skills such as listening, reading, and writing. (ST28)

First, a qualified English teacher should use target language and enjoy teaching it. Then, s/he can create a good learning environment in which students join the lesson and are willing to spend time. (T20)

Having a good level of knowledge-related competencies does not guarantee quality teaching. Teachers need certain skills to be able to transfer what they know to their students. They worry about how to teach, so they take into account students' cognitive, linguistic, social, emotional, and physical needs while preparing materials, activities, and assessment tools. (TT4)

Attitude-Related Competencies

...Teachers should love their profession. S/he should love his/her job, has a smiling face, and must be open to different views. (T19)

It is very important that an English language teacher be patient, has emphatic thinking, and considers learners' needs. (EE8)

Students do not want a serious and surly teacher. Instead, if they face a teacher who has sense of humour, they take a positive attitude towards school, teacher, coursework and each other. I believe that since humour increases students' motivation, it enriches not only the learning environment but also students' learning opportunities. (ST25)

Table 4. The Required Competencies for a Qualified EFL Teacher

Categories	Codes	f
Knowledge-related competencies	Has knowledge of the subject matter	70
	Has knowledge of underlying theories of teaching (Has a repertoire of teaching techniques and approaches)	62
	Has knowledge of learners' development (cognitive, linguistic, social, emotional, and physical)	58
	Has knowledge of curriculum	33
	Has knowledge of legislation	18
Skill-related competencies	Speaks English well	78
	Transfers his/her knowledge to students	71
	Conducts enjoyable lessons	70
	Uses technological tools	66
	Prepares and uses appropriate assessment and evaluation tools	65
	Understands native and/or non-native conversation	59
	Manages students and the learning environment	55
	Prepares, uses, and evaluates materials	53
Demonstrates a coherent writing	25	
Attitude-related competencies	Loves his/her job	71
	Considers learners' needs	65
	Demonstrates a fair attitude towards their students	55
	Treats students as individuals	54
	Being open to professional development	52
	Being aware of the latest ELT developments (as a researcher)	44
	Has warmth and tolerance	40
	Has a good sense of humour	39
	Establishes effective interaction, communication, and cooperation with stakeholders	38
Being democratic and humanist	33	

Findings Related to Stakeholders' Perceptions of Current TAE Model

The second research question aimed to examine the perceptions of educational stakeholders towards the TAE model and reveal its impact on education and stakeholders. Stakeholders stated their perception of the Turkish TAE model and mentioned how this model affected prospective teachers, in-service teachers, and education. Data showed that prospective teachers postponed the acquisition of some skills until after they had PASSED the examinations while in-service teachers thought that it was too late by then (PAST) to learn these skills appropriately. PASS refers to 'Priority of System Requirements', 'Anxiety', 'Skill Delay', and 'Shortcomings', while PAST is an abbreviation used for 'Practice', 'Assessment', 'Speaking', and 'Technology Literacy' (please see Figure 2). Prospective teachers stated that they allocated their time to study topics asked in the exam for fulfilling the *system requirements*. However, there are lots of topics and it creates *anxiety*. Therefore, they started to *delay skills* that were not gauged through KPSS. Consequently, they graduated with some *shortcomings*. Similar to prospective teachers, in-service teachers found the current TAE model inadequate as it did not include and gauge the teaching problems they encountered in real classrooms. Thus, they felt inadequate in *practice, assessment literacy, speaking/pronunciation, and technology literacy*. Other stakeholders (teacher trainers, education experts, and policymakers) stated that the current TAE model prevents teachers from being CEOs of their profession. CEO refers to 'Competencies', 'Effective Teaching', and 'Openness to Personal and Professional Development' (please see Figure 2). Some representative examples were as follows:

I do not believe in the fairness of KPSS as it does not gauge all competencies. To be honest, I do not allocate time to topics that are not on the exam. I can handle them later; I mean when I become an English teacher. At the moment, it is time to study topics of ÖABT, general culture, general ability, and educational sciences. (ST30)

Our language proficiency should be assessed via examinations such as TOEFL and IELTS. There is a positive correlation between our language proficiency and the TAE model. I have problems with my proficiency, especially with pronunciation. To me, after graduation, it is not easy to overcome it. We should have managed it during undergraduate period. (T15)

The current TAE model aims to find some solutions for the existing complaints and reactions. There are a lot of candidates, and you just allege the results for the rationale behind the selection. That is, this model does not help choose prospective teachers with required assessment and technology literacy; who know how to prepare and use materials appropriately; who always look for new ELT developments; who have the required information, ability and attitude regarding teaching English; who have the required information of curriculum and legislation. From this point of view, the current TAE model should be revised as soon as possible so that it could take into account the aforementioned competencies. (EE8)

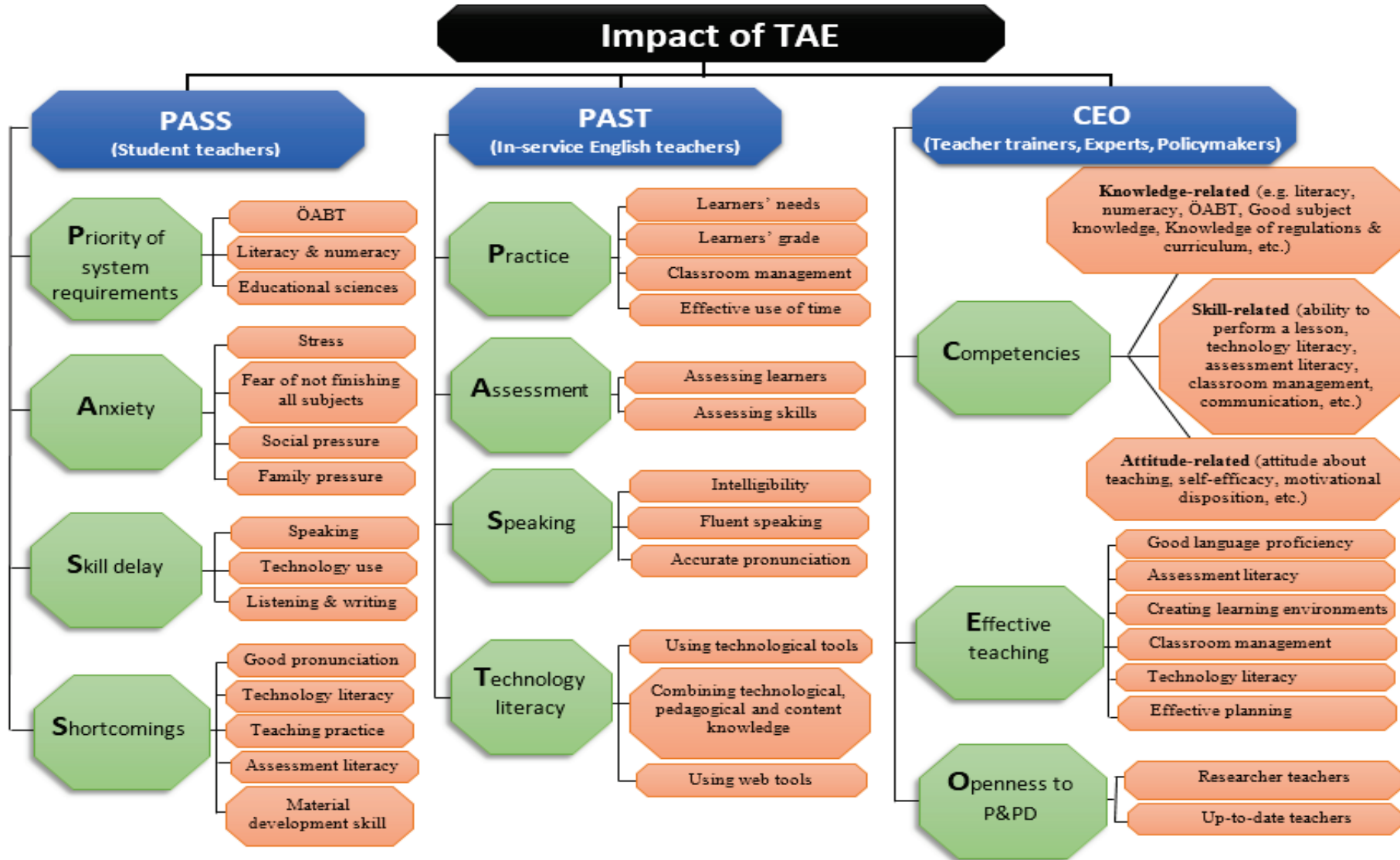


Figure 2. The impact of TAE

Findings Related to In-Service EFL Teachers' Attitudes towards TAE Model

The attitudes of English teachers were examined in terms of their sex, field of study, and reason for choosing that field. Table 5 shows whether there was a significant difference between sex variable and item total score of ETARS.

Table 51. Scores Related to the Effect of Sex Variable on ETARS

Dimensions	Sex	N	Mean	Std. Deviation	t	df	p
Total	male	82	68.9634	10.47094	-.182	268	.856
	female	188	69.2128	10.28458			

As clearly seen in Table 5, no significant difference was found between the sex variable and the item total score of ETARS $p = .856$ ($p > .05$). This shows that, in Turkey, EFL teachers' attitudes towards the current recruitment system did not differ according to their sex.

As presented in Table 6, t-Test was utilized to investigate whether there was a significant difference between the field of study variable and item total score of ETARS.

Table 6. Scores Related to the Effect of Field of Study Variable on ETARS

Dimensions	Group	N	Mean	Std. Deviation	t	Sd	Sig.
Total	ELT	218	68.9174	10.13333	-.661	269	.509
	others	53	69.9623	11.05226			

No significant difference was noted between the field of study variable and the item total score of ETARS $p = .509$ ($p > .05$). Accordingly, EFL teachers' attitudes towards the way they are recruited did not differ according to their field of study in Turkey.

In Table 7, results of one-way analysis of variance (ANOVA) was applied for the relationship between the reason for choosing one's field of study variable and the item total score of ETARS.

Table 7. Scores related to the Effect of Reason for Choosing Field of Study Variable on ETARS

Dimension	Group	N	Mean	Std. Deviation	Sum of Squares	df	Mean Square	F	Sig.
Total	love	172	69.9651	10.27169	Between Groups	374.566	3	124.855	1.178 .319
	guided	43	67.5349	9.09303	Within Groups	28306.415	267	106.017	
	English teacher	32	68.4688	12.49254	Total	28680.982	270		
	other	24	66.7917	9.21709					

According to the results, no significant difference was noted between the grade taught variable and the item total score of ETARS $p = .319$ ($F = 1.178$, $p > .05$). This shows that, in Turkey, EFL teachers' attitudes towards the way they are recruited do not differ according to the reasons for choosing their field of study.

Findings Related to the Appropriate TAE Model

The data suggested an IDEA to the authorities about how to develop a fair, valid, and reliable evaluation process for recruiting English language teachers to the public schools. The acronym IDEA emerged from the initial letters of the four features that a good TAE model should have; namely, *inclusive, distinguishable, equitable/fair, and accurate*.

Inclusive

The term inclusive refers to the features an appropriate TAE model should have. Education stakeholders emphasized that not only must an appropriate TAE model include issues related to prospective teachers' knowledge and skills of the teaching profession, but also their attitudes towards it. The stakeholders emphasized the use of different sources for assessing and evaluating prospective teachers so as to select the right person for the teaching profession. Some of the participants criticized the current model for not considering information about practice while others suggested using practicum and micro-teaching process as informative data for the evaluation of prospective teachers' teaching abilities and their personality traits. Here are some representative statements:

The current system pays attention only to prospective teachers' memorizing ability. Thus, we face teachers who cannot speak English or who have terrible pronunciation. In my opinion, an appropriate TAE model must ask prospective teachers to prove their language proficiency via TOEFL or IELTS equivalent exams. Since the current TAE model does not take into account the prospective teachers' language proficiency level, neither English language teaching programs nor prospective teachers pay attention to it. (T8)

The TAE model must be revised to give information about the prospective teachers' skills and attitudes towards teaching. In this sense, practicum process can be very informative. Likely, universities can collaborate with the ministry by using the information collected during that process. They may also use prospective teachers' marks of micro-teaching lessons. As the ministry responsible, we have revised the practicum system by adding a form, which will be filled out by both mentor and supervisor. The form will be saved electronically and utilized during interview phase. (PM1)

Distinguishable

The term distinguishable here refers to the capacity of a TAE model to distinguish effective teachers and ineffective teachers. Stakeholders stated that the current model is inadequate in gauging the skills and attitudes of effective English teachers because of its multiple choice nature. They suggested revising the current TAE model as soon as possible so that it can measure prospective teachers' skills and attitudes along with their knowledge. Some representative statements are as follows:

Examinations should be revised so that they can select prospective teachers who are good at writing, listening and speaking, and those who can do justice to teaching profession. If a person is good at English language and culture, s/he can teach it easily. Personality, conscience, and psychology are important terms for teaching. A teacher should take into account students' age and level, if not, then, s/he cannot teach English appropriately. If the practicum process is well evaluated, the whole aforementioned features can be achieved. (T11)

Knowing what to teach and how to teach is sine qua non of an effective English language teacher. In addition, personality, psychology, loving teaching job, conscience, assessment literacy, communication and collaboration, and technology literacy are necessary terms. If we want our children to be taught by language teachers who possess these features, we must assess those characteristics. Once we ask prospective teachers to demonstrate the aforementioned aspects, they will attach importance to them. Let me give you a specific example, after the introduction of ÖABT, the final year students rejoin the linguistics courses with freshmen. They gain a chance to reconsider the first year subjects from the last year viewpoint. In the case they cannot study, they may prefer to come and listen to lessons if the weekly course schedules are also available. This situation constitutes a good model for freshmen. They realize that, even after graduation, they will need the information they learned in the courses they are taking. (TT3)

Equitable/Fair

Stakeholders voiced that educational policies must consider whether evaluation practices reflect and promote fairness. The term equitable/fair refers to a test's unbiased decisions. In other words, for a fair evaluation, the test should first represent the construct; then, precautions must be taken to prevent certain groups of test takers (e.g. graduates of ELL) from being more affected by a construct-irrelevant source of variance than other groups and to avoid unintended constructs (construct-irrelevant sources), because they may cause incorrect inferences about those who take tests and, therefore, decrease validity. Stakeholders stated that literature related questions require specific knowledge, so the graduates of ELT have difficulty in answering them correctly. This provides advantages to graduates of ELL departments:

I sometimes have difficulty in understanding the present assessment and evaluation system. I was so glad when ÖABT was introduced, because for the first time there was an attempt to choose prospective teachers who are good in the field. However, it does not serve that purpose in reality. As known, literature related questions require more specific knowledge, so answering them correctly plays a huge role in getting high scores. So, unfortunately, graduates of English language and literature have the edge on us. They spend four or five years particularly on literature and, therefore, know almost all literature related details. (ST11)

Accurate

This feature has two sides. On one hand, stakeholders called for accurate scoring, underlining the importance of using the right assessment tool for the related competency. For example, they warned that pen and paper examinations in multiple-choice format do not guarantee that the right person will be selected for the teaching profession. That is, they complained about the structure and content of the current evaluation model since it did not include skill and attitude-based scores. On the other hand, they complained about the lack of using multiple assessment and evaluation sources. In other words, various methods of assessment should be used to gauge prospective teachers' knowledge, skills, and attitudes. Here is a representative response:

I have some classmates who have communication and psychological problems. They have entered Gazi ELT department, and perhaps they will pass KPSS and become teachers of our children due to the nature of current evaluation system. No attention is paid to a prospective teacher's personality, attitude, or teaching performance. Those who spend time memorizing information that is asked in KPSS become successful in this evaluation system. However, teaching requires, at least, two or more parties. In the simplest term, a person needs students in order to teach something. So is s/he really good at teaching, collaboration and cooperation, classroom management, planning a lesson, and using the right assessment and technological tools? The current TAE model resembles the lottery. We do not know who will hit the bullseye! But it should not be like that. New components should be added and their scores should be used when recruiting someone so that those who like teaching are selected. (ST8)

Discussion

This section provided the discussion of the findings with respect to research questions.

Discussion on the Required Competencies for a Qualified EFL Teacher

Stakeholders advocated that a qualified EFL teacher should have competencies related to knowledge, skill, and attitude. This is mentioned in 'General Competencies for Teaching Profession' published by the Ministry of National Education in 2017. Mohamed, Valcke, & De Wever (2017) analysed existing frameworks of eleven countries and noted that competencies were generally clustered across three main domains: '(1) knowledge: what teachers need to know, (2) skills: what teachers are able to do and (3) attitudes or behaviours: how teachers would put their knowledge and skills in practice'. (p. 155). Regarding knowledge-related competencies, stakeholders focused on competencies,

such as knowledge of subject matter, pedagogical content knowledge, a repertoire of teaching techniques and approaches, knowledge of underlying theories of teaching, knowledge of learners' development (cognitive, linguistic, social, emotional, and physical), knowledge of curriculum, and knowledge of legislation. These findings are consistent with the present literature. For example, Richards, Conway, Roskvist, & Harvey (2013) indicate subject knowledge as *sine qua non* of an effective teacher, because what s/he knows has a direct impact on the classroom environment. Larsen-Freeman and Anderson (2013) pay attention to the knowledge of the method as it contributes to teachers' repertoire of teaching techniques. Pillay (2014) states that teachers' knowledge of legislation and education law constitutes confidence and self-assurance and power. This shows that there is a correlation between English language teachers' practice and their knowledge of legislation and curriculum as well as their knowledge of underlying theories of teaching. Thus, to equip the classrooms with qualified practitioners, teacher training programs should continue to take into account knowledge-related competencies.

Skill-based competencies listed by stakeholders were as following: speaking English well; being good at teaching; conducting enjoyable lessons; using technological tools; preparing and using appropriate assessment and evaluation tools; understanding native and/or non-native conversation; managing students and learning environment; preparing, using, and evaluating materials; and demonstrating a coherent writing. These results are supported by both national and international studies (e.g. Akcan et al., 2017; Freeman, 2017; Richards, 2017; Van Canh & Renandya, 2017). Having knowledge of something is not enough for a person to be an effective English language teacher; instead, the teacher should be able to use it effectively. This indicates that performance-based scores need to be considered while recruiting teachers as there is not always a positive correlation between *knowing* something and *performing* it.

Regarding attitude-related competencies, stakeholders claimed that an effective English language teacher is a person who loves his/her job; considers learners' needs; demonstrates a fair attitude towards their students; treats students as individuals; is open to professional development; is aware of the latest ELT developments (as a researcher); has warmth and tolerance; has a good sense of humour; establishes effective interaction, communication, and cooperation with stakeholders; is democratic and humanist. This confirms both national and international studies (e.g. Bağçeci, Yıldırım, Kara, & Keskinpalta, 2015; Taner, 2017; Yeşilçınar & Çakır, 2018). That is, in addition to knowing something or being able to perform it, an effective English language teacher is required to have positive attitude towards teaching profession (Sakurai, 2012; Taner, 2017; Yalçın-İncik & Akay, 2015). Thus, it can be concluded that a person needs to know the answer of *why* in order to become an effective teacher in addition to being aware of *what* and *how* to teach (Stronge ve Hindman, 2006).

Discussion on Stakeholders' Perceptions of Current TAE Model

The perceptions revealed that all stakeholders find the current Turkish TAE model inappropriate, though the reason behind its inappropriateness differs from person to person. In addition, findings displayed the impact of the current TAE model on prospective teachers, in-service teachers, and education. Due to the structure of the current TAE model, prospective teachers believe that they will deal with the neglected competencies when the examination process PASS; that is, they delay some skills till they become teachers. They feel obliged to restrain the curriculum in order to save time for studying issues that are asked in the current TAE model. This confirms studies in other countries where high-stakes regimes were found to have an influence on constraining the curriculum and narrowing students' educational experiences (i.e. Polesel, Rice, & Dulfer, 2014). For example, Polesel et al. (2014) investigated the impact of NAPLAN (National Assessment Program – Literacy & Numeracy) on Australian schools and students. They found that high-stakes examinations can deform instructional practices and narrow down the curriculum. Regarding CLT, Shohamy (2017) states that the high-stakes tests cause both test takers and educational systems to alter their instructional behaviours.

English teachers stated that they graduated with lots of shortcomings due to the current TAE model, and that both time and opportunities stayed at PAST to compensate these shortcomings. Teachers mentioned that all these competencies should have been covered in the undergraduate period, because once prospective teachers let these competencies get out of control, it seems to be too difficult to manage them after graduation. Moreover, previous research (i.e. Allen, 2016; Özmen, 2011) shows the relationship between the type of washback (e.g. positive, negative) and the nature of the exam. For instance, Allen (2016) examined the consequential validity of IELTS, specifically focusing on washback upon learners' test preparation strategies and determined that 'IELTS Test created positive washback on learners' language ability and test preparation strategies, specifically regarding productive skills, which learners in the study context had neglected in their previous language study' (p. 1). It may be due to the fact that IELTS Test also includes productive skills. However, examining the washback effects of a high-stakes test on candidate academics in Turkey, Özmen (2011) found harmful and negative washback effects. He attributed these results to the content and style of the high-stakes test which consists of a multiple-choice test.

Finally, for other stakeholders (teacher trainers, education experts, and policymakers), the current TAE model prevents teachers from becoming the CEO of their profession. They voiced that it fails to distinguish effective teachers from ineffective ones. Unless these deficiencies are not overcome, it seems impossible to equip each classroom with a qualified/effective teacher. Education experts noted that although interview phase has been introduced to assess prospective teachers' communication skills, it failed to reach its goal due to reasons, such as taking place in Turkish, lacking good scoring rubrics, and field experts. Thus, it can be said that an oral examination is a must for training prospective teachers to speak fluent English. In this context, this study shares similar findings with Garcia Laborda, Litzler, Gonzalez Such, Bakieva, & Otero de Juan (2012) who found Spanish PAU (University Entrance Examination) insufficient in describing their language use as it is based on written aspects of the language. Therefore, if it is desired to train teachers who have the necessary knowledge, skills, and attitudes, components that can measure these characteristics should be added to the TAE model.

Qualitative data revealed that questions related to English literature require specific knowledge of that subject. While graduates of ELL could easily answer those questions, ELT graduates found them too specific to answer. Thus, it can be concluded that the current TAE model displays unintended consequences, which diminishes its fairness and validity. Henning (1987) argued that the misuse of tests or misinterpretation of score meaning are the most apparent threats to validity, which is called as construct-irrelevant variance (Messick, 1989). He expressed construct-irrelevant variances as variables unrelated to the construct being measured, which occurs when factors irrelevant to the construct influence the test scores. The test taker's personality, background knowledge, test-taking strategies, as well as general intellectual or cognitive ability may be construct-irrelevant (Bachman, 1990), and therefore effort should be made to minimize their influences. Thus, it can be interpreted that the current model is not construct relevant, and shows a tendency of favouring graduates of ELL programs because of their background knowledge.

Discussion on Attitudes of In-Service EFL Teachers towards TAE Model

The findings of ETARS showed that teachers did not approve the current TAE model regardless of demographic variables such as sex, field of study, and desired profession. That is, English language teachers considered the current TAE model inadequate, unfair and invalid, and they favoured some other additional tests that could determine their skills and attitudes in order to select effective candidates. This confirms other studies related to ÖABT and KPSS (Uyulgan & Akkuzu, 2015; Yilmaz & Yaşar, 2016). Uyulgan and Akkuzu (2015) investigated 180 prospective teachers' views on teacher selection and recruitment. The content analysis revealed that they had negative opinions regarding ÖABT as the exam made them feel anxious and inadequate. In addition, they felt that it was not fair to recruit teachers only on the basis of the scores of knowledge-based exams. Yilmaz and Yaşar (2016) examined 87 prospective teachers' perceptions of KPSS and stated that while there was no significant

difference between sex and their perceptions, a significant difference was found to be among departments where they were studying.

The aforementioned results call for performance-based scores. It is clear that schools that are interested in quality education and teachers generally pay attention to performance indicators. Jacob, Rockoff, Taylor, Lindy, and Rosen (2016) listed four stages of evaluation: general essays, subject-specific assessments, interview, and teaching audition. They examined the relationship between applicants' characteristics, hiring outcomes, and teacher performance. Findings showed that both background characteristics and a demonstration lesson are informative about teacher effectiveness, and that 'there exists considerable scope for improving teacher quality through the selection process' (p. 37). The current TAE model with multiple choice questions does not distinguish the candidates who can teach from those who cannot. This causes the participants to exhibit a negative attitude towards the current TAE model.

Discussion on an Appropriate TAE Model

The last research question revealed that education experts generally valued a TAE model that gauges not only prospective teachers' knowledge but also their skills and attitudes. When the emerging codes were scrutinized, four categories were observed: *inclusive, distinguishable, equitable/fair,* and *accurate*. As can be noticed from the initial letters of the categories, the data suggested an IDEA to the authorities about the characteristics that an appropriate TAE model need to have while recruiting English language teachers to the public schools (please see Figure 3).

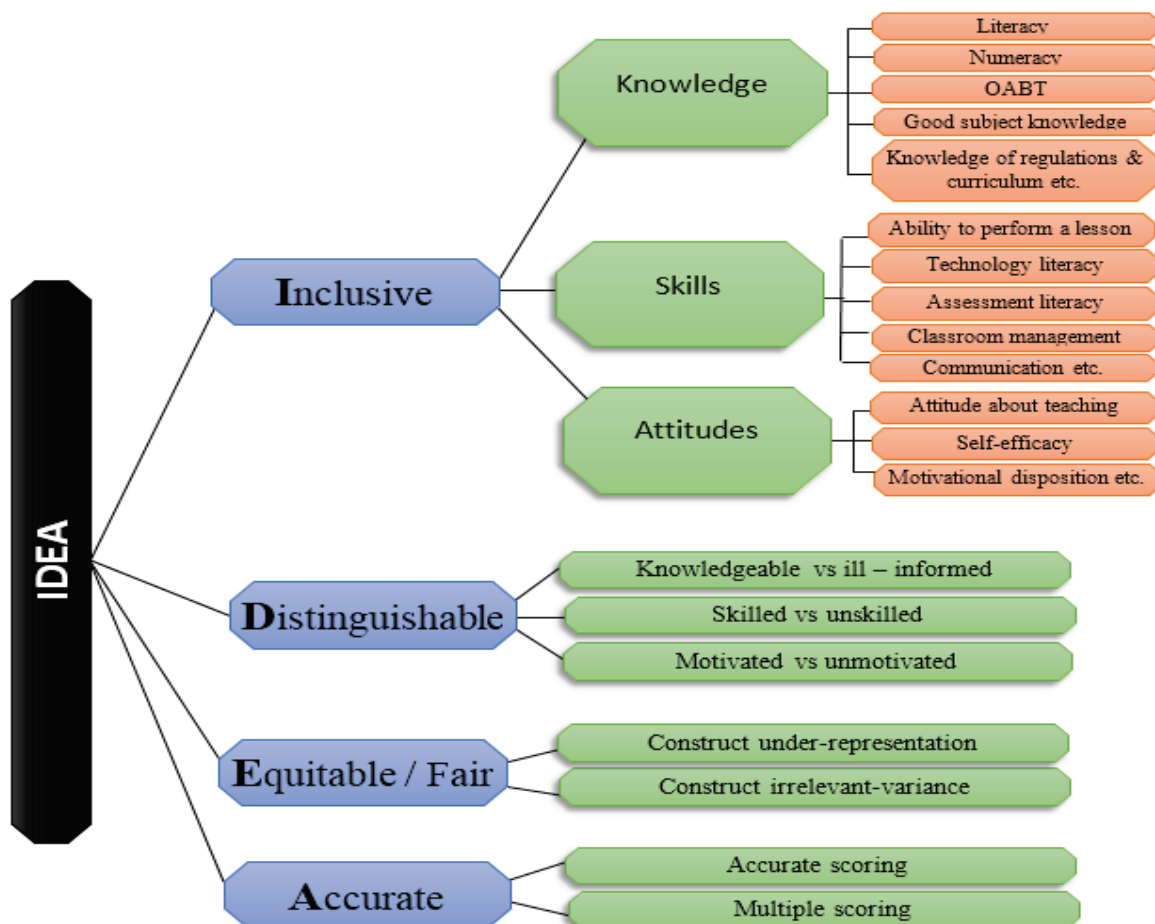


Figure 3. The suggested TAE model: IDEA

Inclusive

Education stakeholders state that the evaluation process should include competencies related to prospective teachers' knowledge, skills, and attitudes in order to recruit an effective language teacher. So, it is useful to identify what makes a teacher effective. It has been a subject of importance for many educational scholars (Brown, 2001; Harris & Sass, 2014; Noddings, 2013; Pettis, 1997; Walsh, 2002). For instance, Pettis (1997) believes that a professionally competent teacher, at first, needs to be principled, knowledgeable, skilful, and open to developing while Walsh (2002) sees discipline knowledge, pedagogical knowledge, knowledge of context, and personal knowledge as a must for an effective English language teacher. Some indicated that an effective language teacher should have good language proficiency, personality, and language-teaching skills (Brown, 2001). Others considered personal teacher dispositions, such as caring, as a vital factor in teacher effectiveness (e.g. Noddings, 2013). According to Harris and Sass (2014), the link between teacher skills and productivity can be the best indicator of teacher performance. That's why, a lot of countries usually make use of interviews including demonstration lessons. For example, in Japan, each Prefectural Board of Education is empowered to recruit teachers for their public schools through a recruitment test which generally includes both a paper test and an interview including a teaching demonstration (Ramírez Carpeño, & Mekochi, 2015). In Spain, a nationwide competitive examination, which has two phases, is conducted. Prospective teachers have a practical exam in the first part, and the second part measures their pedagogical knowledge by asking them to 'present a teaching programme based on the first exam' and to 'do an oral presentation of a teaching unit' (ibid., p. 117). In Korea (see Kang & Hong, 2008 for more details), prospective teachers are required to pass NTET, which assesses their knowledge, skills, and dispositions via three phases. In the last phase, the candidate has a chance to show her/his teaching performance followed by an interview. Thus, it can be concluded that countries having good educational outcomes pay great attention to their recruitment model since ensuring that each classroom has a quality teacher starts with the use of an efficient teacher evaluation model (Rivkin et al., 2005). Classroom practices-based teacher-quality measures 'are more closely linked to professional standards and therefore assess teachers from expectations that are known and shared by the teaching community' (Danielson, 2007, as cited in Santelices, Valencia, Gonzalez, & Taut, 2017). The perceptions of education stakeholders of this study confirm findings of Leshem and Bar-Hama (2008) who focused on how teachers should be assessed and provided three options as an answer: (1) by observable behaviour and measurable components, (2) taking into account the practicum, or (3) by their formal knowledge.

Distinguishable

An appropriate TAE model must distinguish effective teachers from ineffective ones. There is not a unique definition of a qualified or effective teacher (Ingersoll, 2001, p. 42); some researchers state that 'many kinds of teacher knowledge and experiences may contribute to teacher effects, including teachers' general academic and verbal ability, subject matter knowledge, knowledge about teaching and learning, teaching experience, and the set of qualifications measured by teacher certification' (Darling-Hammond & Wei, 2012, p. 614). In *Good to Great*, Collins (2001) states that 'when in doubt, don't hire—keep looking' (p. 54). However, the question to be considered, as Stronge and Hindman (2006) asked, is how we distinguish effective applicants from less or ineffective ones. Moreover, how do we ensure that we select a good teacher? Some researchers focus on four information sources that are referenced by principles in order to distinguish candidates with desired teaching qualities from others: '(a) past experience, (b) job interview, (c) professional references, and (d) resume' (Grove, 2008; Harris, Rutledge, Ingle, & Thompson, 2010; Kersten, 2008 as cited in Ellis, Skidmore, & Combs, 2017, p. 453). Considering how to measure teaching quality, literature lists three main lines of research (as cited in Santelices et al., 2017):

- (a) the study of observable teacher characteristics that may impact student performance (Hanushek & Rivkin 2006), often referred to as *inputs* (Goe et al., 2008); (b) the analysis of teaching quality, defined as professional practices based on professional standards and usually measured through class observation guidelines, focusing on *process* (Danielson, 2007; Darling-Hammond, 2012; Papay, 2012); and (c) estimates of value

added based on student performance on standardized tests (Millman, 1997), also known as *outputs* and more traditionally associated with the concept of teacher effectiveness (p. 114, emphasis in original).

Very recent literature has exhibited a different approach to the discussion on how to measure teaching quality, emphasizing the amalgamation of different sources of information on teacher performance as well as its overall application in teacher performance evaluation (Darling-Hammond, 2012). From this perspective, the aforementioned three research strands represent subcomponents of a single measure, instead of referring to different measures of teaching quality. The TAE process should involve 'observable teacher characteristics, professional practices, and the impact on student performance' (Santelices et al., 2017, p. 114) so as to distinguish effective teachers from ineffective ones. Also, Tucker and Stronge (2005) underlined the importance of observation, indicating classroom observations as the primary instruments utilized for evaluating teachers.

Equitable/Fair

Messick (1989) warns that two major threats to validity (construct under-representation and construct-irrelevant variance) need to be examined to improve test quality and thus promote positive washback. The former occurs when the construct has some missing points, but the latter happens when components that are irrelevant to the construct of focus influence test scores. These factors diminish test validity and affect the fairness of the test if groups are affected differentially. Thus, to establish a fair evaluation system, it is necessary to prevent those two factors and, specifically, to prevent certain groups of test takers from being more affected than other groups by a construct-irrelevant source of variance and to avoid unintended constructs (construct-irrelevant sources), because they may cause incorrect inferences about those who take tests and, therefore, decrease validity. This information confirms ST19 who asserted '*... We, students of ELL, are successful in answering questions both related to literature and those related to language proficiency since we are always asked to read and write essays during our four-year education*'. Thus, the current TAE model militates in favour of ELL graduates due to their background, rendering construct-irrelevant variance. Cho and So (2014) studied interaction of primary students learning English with test materials and their response to English language test questions. They stated that the complexity of language used in test questions might influence students' ways of demonstrating skills on the constructs targeted in the test questions. They added that the perception of test difficulty may be affected by memory load and ambiguity in task descriptions.

Another finding was that the current TAE model does not represent the construct in terms of language proficiency. ÖABT tries to measure the student teachers' language proficiency through multiple choice questions. These questions are not enough because there are no questions testing the candidates' listening comprehension, speaking, or writing skills. According to the participants, the interview is ineffective in measuring communication skills as it is conducted in Turkish, thus causing construct under-representation. This is also mentioned by Chen, Li, and Chason (2018), who alleged the possible occurrence of construct under-representation when goals of assessment are deficiently considered. These researchers believe that low-level items of cognitive function require only rote memorization in order to recall isolated facts, but they may not reflect the knowledge to represent the performance for real-world situations. Thus, it can be concluded that validity is *sine qua non* of a high-stakes test. Therefore, a well-prepared test should be far away from threats to validity, namely construct-irrelevant variance and construct under-representation, in order to prevent it from having unintended, negative, or no consequence (Green, 1998).

Accurate

Accurate has two dimensions. The first dimension refers to using the right assessment tools for the related competencies (accurate scoring). As a matter of fact, stakeholders criticized the inappropriateness of the current TAE model by emphasizing that even answering all questions does not guarantee the selection of effective teachers. That is, to what extent does a prospective teacher's score on ÖABT represents 'real or appropriate' measure of his/her English teaching ability since multiple-choice questions cannot give enough information about an individual's classroom performance. The second dimension is naturally related to the first one, and it is about the necessity of using multiple assessments and evaluation means so that it could provide enough information about the prospective teachers' knowledge, skills, and attitudes. The use of only multiple-choice pen and paper examinations would not give us the information needed about prospective teachers' teaching ability. This was also emphasized by James Popham et al. (2014) who stated that scores obtained from standardized tests are inappropriate for evaluating teacher quality. Current tests are not supportive, because standardized tests

... do not reflect and embody the skills students need, they cannot inform instruction that aims to fulfil these purposes. Additionally, multiple-choice tests do not offer any information about students' thinking or reasoning, their misconceptions, or ability to express their ideas. Thus they provide little information to help teachers shape instruction. Last but not least, since these assessments are primarily summative, they cannot inform continuous instructional improvement, a goal facilitated by a rich array of assessments within a comprehensive state assessment system. (Parsi & Darling-Hammond, 2015, p. 5)

These explanations prove how some policymakers (for example, PM4) were right in warning that '*Unless the system terminates the measurement of teaching quality through questions in multiple-choice format, it becomes impossible to equip each classroom with an effective language teacher.*' This should be a call for the amalgamation of standardized tests and classroom-embedded performance assessments. Correspondingly, examining the selection and recruitment process in Turkey and Poland, Kılıçkaya and Krajka concluded that

It is natural that an exam is conducted when there are more applicants than vacancies in any profession. However, these exams should test what teacher candidates are expected to perform in class. In other words, we cannot talk about the reliability or validity of these tests such as KPSS in Turkey... (Kılıçkaya & Krajka, 2013, p. 101)

In recent years, an interview phase has been introduced, but education stakeholders found it inaccurate for two reasons. First, it lacks a detailed, valid, and fair scoring rubric. The use of well-developed scoring rubrics 'is likely to reduce errors in the interview process... Rating scales may reduce bias and enhance interviewers' consistency of judgments about candidates' responses ...' (Stronge & Hindman, 2006, p. 25). Thus, these data suggest that the authorities should either terminate the interview phase or redesign the process in which experts and well-developed rubrics are common. In a nutshell, literature displays evidence that using multiple means of assessment results in the recruitment of more highly qualified teachers.

Conclusion and Suggestions

This research has indicated that the current TAE model has not been appreciated and valued in education society as it cannot measure knowledge, skills, and attitudes that a qualified English teacher should have. This confirms studies that found ÖABT (in particular) and KPSS (in general) as inappropriate examinations in terms of selecting qualified teachers and called for alternative examinations by using performance-based scores and multiple assessment tools (e.g. Kılıçkaya & Krajka, 2013; Uyulgan & Akkuzu, 2015). Furthermore, since there is a positive relationship between the effectiveness of teacher evaluation (or teacher selection and recruitment) and teacher's quality and job satisfaction (e.g. Liu & Johnson, 2006), the current model has been accused of not contributing to the process of bringing quality English teachers to the classrooms because of its inability to distinguish effective language teachers from ineffective ones. Since the current TAE model measures candidates with multiple-choice questions, both teachers and prospective teachers have been observed to delay their performance-based competencies. These problems caused teachers to have negative attitudes towards the recruitment system and required an investigation of the features that should be in an ideal recruitment model because bringing effective teachers in classrooms is only possible with a well-established TAE model. In a nutshell, the IDEA has been developed to make the process of teacher assessment and evaluation more valid and fair, with the aim of enhancing the quality of teaching and learning. It is assumed if the four pillars (inclusive, distinguishable, equitable/fair, and accurate) of the IDEA are implemented appropriately by the authorities during the process of recruiting language teachers, it will not be impossible to provide an effective English language teacher to each student.

TAE is a multidimensional process with a variety of factors to be taken into consideration. According to the IDEA, taking into account the prospective teachers' teaching ability, their attitudes, and their personality traits is a *sine qua non* for the recruitment or assignment of an effective English language teacher. Hence, two kinds of question come to the fore: (1) *what to assess* and (2) *how to assess*. The first one is about the features of an effective English teacher, while the second refers to the means or ways that may be employed for the assessment and evaluation process (see Figure 4). The IDEA utilizes various sources of information for the assessment and evaluation of English language teachers and emphasizes their productive skills, particularly speaking, and indicates that an effective English language teacher needs to be fluent in the target language. Besides, it advocates the use of multiple scores which cohere with the related construct. Thus, it can be concluded that the IDEA aims to provide each student with a qualified language teacher.

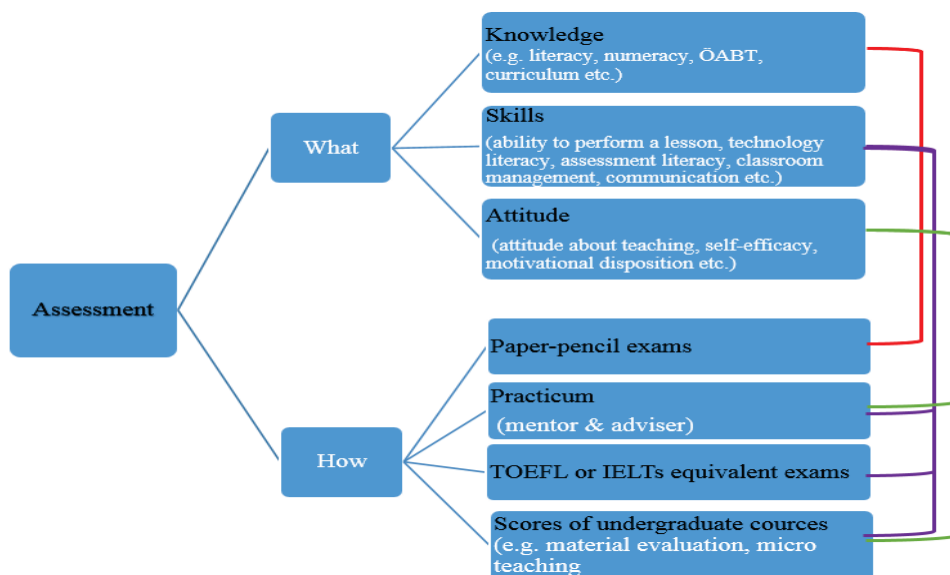


Figure 4. Competencies to be Assessed and Possible Ways of Assessment

Recommendations for Knowledge

The IDEA suggests that a qualified English teacher must have a deep knowledge of subject/content curriculum, legislation, child development, testing, and underlying theories of language teaching and learning. As those issues contain a mass of knowledge, for practicality concerns, the suggested model recommends the use of pen and paper exams for knowledge-related issues.

Recommendations for Skills

The most problematic area of the current TAE model is its lack of performance-based scoring. The suggested model proposes TOEFL- or IELTS-like exams that will be conducted by ÖSYM to measure prospective teacher's proficiency in the four language skills. To make it practical, the time for the exam should be flexible. That is, prospective teachers should be given the opportunity to apply for such exams anytime they want (before or after graduation).

To decrease the subjectivity level of the scores obtained from performance-based assessments, IDEA suggests the use of scores from several sources, rather than only a few, for the identification of candidates with greater teaching ability. That is, it advocates, for example, the use of candidates' practicum scores and scores they get in certain courses or practices (e.g. micro-teaching of school experience, material development, and evaluation) instead of depending on a one-time observation-based exam (for example, a sample lesson). However, three problems may arise:

1. scores given by universities which are more concerned with the money its students pay as tuition (some universities may be more generous in terms of scores as in the case of private high schools to be able to attract more students),
2. consistency of scores (i.e. does a score of 80 obtained in university A refer to the same level of competency as a score of 80 obtained in university B?), and
3. pedagogical formation (teacher certificate).

For the first two problems, there are two solutions: the Ministry could either introduce a common exit exam or take into account prospective teachers' university entry exam scores when evaluating them for recruitment. The practice of pedagogical formation is an enigmatic issue in Turkey. It is beyond the scope of the current study to discuss whether the pedagogical formation system for non-ELT graduates is good. The president of the country stated that the pedagogical formation requirement for admission to the teaching profession would be terminated (MEB, 2018). However, if it continues, the IDEA suggests that the non-ELT graduates should spend at least one semester with senior ELT students. Experiencing some courses and practicum processes with senior ELT students will contribute to pedagogical knowledge of non-ELT graduates.

Recommendations with regard to Attitudes

Prospective teachers' attitudes, motivation, and personality traits determine their teaching quality. Thus, an effective teacher should love his/her profession, care for the needs of learners, respect each student as an individual, and have a sense of humour. Although it is not easy to measure personality traits objectively, direct observations may be informative. That is, micro-teaching of school experience lessons and practicum processes can be instrumental in deciding whether the individual is right for the teaching profession. This case is also mentioned as the fifth action (Enabling students to switch to other programs from teacher training programs) in the Teacher Strategy Paper (MEB, 2017). There are inventories, for example, the Myers-Briggs Type Indicator (Myers & McCaulley, 1985), which allow individuals to picture their learning preferences, teaching styles, and personality characteristics (Mamchur, 1996). However, mentors' and lecturers' observation of practicum and some lessons (e.g. micro-teaching of school experience, material development, and evaluation) is vital. Thus, using the scores from more sources will improve reliability and conformability, thereby contributing to objective scoring (Lincoln & Guba, 1985, as cited in Korstjens & Moser, 2018).

This study is limited to accessible prospective teachers, teachers, teacher trainers, education experts, and policymakers. Thus, their statements may not represent those of stakeholders in Turkey. A more embracing study based on the offered model needs to be conducted by reaching out to more policymakers, especially the ones in the parliament's National Education Commission, because such a comprehensive model needs to be analysed and evaluated by the related members of the parliament. Although every effort was made to develop a fair, reliable, and valid TAE model and various stakeholders were given voice for creating reasonable methods and ways of assessing and evaluating English teachers, it is believed that a further study needs to investigate whether the IDEA works. This would lead to small-scale piloting of the model with the government support.

References

- Adnot, M., Dee, T., Katz, V., & Wyckoff, J. (2017). Teacher turnover, teacher quality, and student achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1), 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.
- Allen, D. (2016). Investigating washback to the learner from the IELTS test in the Japanese tertiary context. *Language Testing in Asia*, 6(7), 1-20.
- Atkinson, P. A., & Coffey, A. (1997). Analysing documentary realities. In D. Silverman (Ed.), *Qualitative research: Theory, method and practice* (pp. 45-62), London: Sage.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University.
- Bağçeci, B., Yıldırım, İ. Kara, K., & Keskinpaltı, D. (2015). Pedagojik formasyon ve eğitim fakültesi öğrencilerinin öğretmenlik mesleğine yönelik tutumlarının karşılaştırılması. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 17(1), 307-324.
- Barnes, B. D., & Lock, G. (2010). The attributes of effective lecturers of English as a foreign language as perceived by students in a Korean University. *Australian Journal of Teacher Education*, 35(1), 139-152.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40.
- British Council/TEPAV. (2013). *Turkey national needs assessment of state school english language teaching*. Retrieved from https://www.britishcouncil.org.tr/sites/default/files/turkey_national_needs_assessment_of_state_school_english_language_teaching.pdf
- Brosh, H. (1996). Perceived characteristics of an effective language teacher. *Foreign Language Annals*, 29(2), 125-138.
- Brown, H. D. (2001). *Teaching by principles: An interactive approach to language and pedagogy*. White Plains, NY: Longman.
- Büyüköztürk, S. (2011). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni SPSS uygulamaları ve yorum*. Ankara: Pegem Akademi.
- Chen, Y. H., Li, I. Y., & Chason, W. (2018). Construct underrepresentation. In B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation*. Thousand Oaks, CA: SAGE.
- Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: CUP.
- Cheng, L. (2014). Consequences, impact and washback. In A. J. Kunnan (Ed.), *The companion to language assessment: Evaluation, methodology and interdisciplinary themes* (pp. 1130-1145). Singapore: Wiley Blackwell.
- Cho, Y., & So, Y. (2014). *Construct-irrelevant factors influencing young English as a foreign language (EFL) learners' perceptions of test task difficulty* (Research Memorandum No. RM-14-04). Princeton, NJ: Educational Testing Service.
- Cohen, D. K. (2010). Teacher quality: An American educational dilemma. In M. M. Kennedy (Ed.), *Teacher assessment and the quest for teacher quality: A handbook* (pp. 375-401). San Francisco, CA: Jossey-Bass.
- Collins, J. (2001). *Good to great: Why some companies make the leap and others don't*. Colorado: Harper Business.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

- Çelik, S., Arıkan, A., & Caner, M. (2013). In the eyes of Turkish EFL learners: What makes an effective foreign language teacher?. *Porta Linguarum*, 20, 287-297.
- Darling-Hammond, L. (2010). *Performance counts: Assessment systems that support high-quality learning*. Washington, DC: Council of Chief State School Officers. Retrieved from <https://files.eric.ed.gov/fulltext/ED543057.pdf>
- Darling-Hammond, L. (2012). *Powerful teacher education: lessons from exemplary programs*. San Francisco: JosseyBass.
- Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1), 4-13.
- Darling-Hammond, L., & Wei, R. C. (2012). Teacher preparation and teacher learning: A changing policy landscape. In *Handbook of education policy research* (pp. 629-652). London: Routledge.
- Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford: Oxford University.
- Ellis, C., Skidmore, S. T., & Combs, J. P. (2017). The hiring process matters: The role of person-job and person-organization fit in teacher satisfaction. *Educational Administration Quarterly*, 53(3), 448-474.
- Enginarlar, H. (2014). Yabancı diller eğitimi bölümlerinin (YDE) dünü, bugünü ve yarını. In Y. Bayyurt & S. Akcan (Eds.), *Türkiye'deki yabancı dil eğitimi üzerine görüş ve düşünceler* (pp. 31-44).
- Freeman, D. (2017). The case for teachers' classroom English proficiency. *RELC Journal*, 48(1), 31-52.
- Garcia Laborda, J., Litzler, M. F., Gonzalez Such, J., Bakieva, M., & Otero de Juan, N. (2012). Washback factors and inference predictors in the Spanish university examination: The OPENPAU Project (FFI2011-22442). *Online Submission*, 46, 5914-5918.
- Green, D. R. (1998). Consequential aspects of the validity of achievement tests: A publisher's point of view. *Educational Measurement: Issues and Practice*, 17(2), 16-19.
- Hansen, K. (2003). *Writing in the social sciences: A rhetoric with readings*. Boston: Pearson Custom.
- Harris, D. N., & Sass, T. R. (2014). Skills, productivity and the evaluation of teacher performance. *Economics of Education Review*, 40, 183-204.
- Hatipoğlu, Ç. (2016). The impact of the university entrance exam on EFL education in Turkey: Pre-service English language teachers' perspective. *Procedia-Social and Behavioral Sciences*, 232, 136-144.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Rowley, MA: Newbury House.
- Hitchcock, J. H., Onwuegbuzie, A. J., & Khoshaim, H. B. (2015). Examining the consequential validity of standardized examinations via public perceptions: A review of mixed methods survey design considerations. *International Journal of Multiple Research Approaches*, 9(1), 24-39.
- Ingersoll, R. (2007). *A comparative study of teacher preparation and qualifications in six nations*. Philadelphia: Consortium for Policy Research in Education.
- Ingersoll, R., & Merrill, L. (2010). Who's teaching our children? *Education Leadership* 67(8), 14-20.
- Ingersoll, R. M. (2001). *Teacher turnover, teacher shortages, and the organization of schools* (Document R-01-1). Seattle: University of Washington, Center for the Study of Teaching and Policy.
- International Business Publications. (2008). *Thailand education system and policy handbook*. USA: Int'l Business Publications.
- Jacob, B. A. (2016). The power of teacher selection to improve education. *Evidence Speaks Reports*, 1(12), 1-7.
- Jacob, B. A., Rockoff, J. E., Taylor, E. S., Lindy, B., & Rosen, R. (2016). *Teacher applicant hiring and teacher performance: Evidence from DC public schools* (No. w22054). National Bureau of Economic Research.
- James Popham, W., C. Berliner, D., M. Kingston, N., H. Fuhrman, S., M. Ladd, S., Charbonneau, J., ... Chatterji, M. (2014). Can today's standardized achievement tests yield instructionally useful data? Challenges, promises and the state of the art. *Quality Assurance in Education*, 22(4), 303-318.

- Kang, N. H., & Hong, M. (2008). Achieving excellence in teacher workforce and equity in learning opportunities in South Korea. *Educational Researcher*, 37(4), 200-207.
- Kennedy, M. M. (2008). Sorting out teacher quality. *Phi Delta Kappan*, 90(1), 59-63.
- Kılıçkaya, F., & Krajka, J. (2013). A review of language teacher selection examination and recruitment in Turkey and Poland. *International Journal of Research Studies in Language Learning*, 2(5), 93-104.
- Kırkgöz, Y. (2007). English language teaching in Turkey: Policy changes and their implementations. *RELC Journal*, 38(2), 216-228.
- Klassen, R. M., Durksen, T., Kim, L., Patterson, F., Rowett, E., Warwick, J., ... Wolpert, M. A. (2017). Developing a proof-of-concept selection test for entry into primary teacher education programs. *International Journal of Assessment Tools in Education*, 4(2), 96-114.
- Korstjens, I., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, 24(1), 120-124.
- Kumar, K., Roberts, C., Bartle, E., & Eley, D. S. (2018). Testing for medical school selection: What are prospective doctors' experiences and perceptions of the GAMSAT and what are the consequences of testing?. *Advances in Health Sciences Education*, 23(3), 533-546.
- Kvale, S. (1996). *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks: SAGE Publications.
- Larsen-Freeman, D., & Anderson, M. (2013). *Techniques and principles in language teaching* (3rd ed.). Oxford: Oxford University Press.
- Leshem, S., & Bar-Hama, R. (2008). Evaluating teaching practice. *ELT journal*, 62(3), 257-265.
- Liu, E., & Johnson, S. M. (2006). New teachers' experiences of hiring: Late, rushed, and information-poor. *Educational Administration Quarterly*, 42(3), 324-360.
- Liu, S., Liu, C., Stronge, J., & Xu, X. (2016). Teacher characteristics for success in the classroom: Chinese principals' perceptions for hiring decisions. *Asia Pacific Education Review*, 17(1), 107-120.
- Loumbourdi, L. (2014). *The power and impact of standardised tests: Investigating the washback of language exams in Greece*. Peter Lang GmbH, Internationaler Verlag der Wissenschaften.
- Mackay, A., & Gass, S. (2005). *Second language research: Methodology and design*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Mamchur, C. (1996). *A teacher's guide to cognitive type theory and learning style*. Alexandria, VA: ASCD.
- Mason, R. W., & Schroeder, M. P. (2010). Principal hiring practices: Toward a reduction of uncertainty. *The Clearing House*, 83(5), 186-193.
- Mattavarat, S., Viseshsiri, P., & Siribanpitak, P. (2017). Proposed policy for preparation of high-quality primary school teachers in Thailand. *Kasetsart Journal of Social Sciences*, 38(2), 105-110.
- Ministry of National Education. (2017). *Teacher strategy paper*. Ankara: MEB. Retrieved from https://oygm.meb.gov.tr/meb_iys_dosyalar/2018_05/25170118_Teacher_Strategy_Paper_2017-2023.pdf
- Ministry of National Education. (2018). *2023 eğitim vizyonu*. Ankara: MEB. Retrieved from <https://www.meb.gov.tr/2023-egitim-vizyonu-aciklandi/haber/17298/tr>
- Merriam, S. B. (1998). *Qualitative research and case study applications in education: Revised and expanded from case study research in education*. CA: Jossey-Bass.
- Mertz, N. T. (2010). Teacher selection and school leader effects. *Journal of School Leadership*, 20(2), 184-207.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: Macmillan.
- Miles, M. B., & Huberman, M. (1994). *An expanded sourcebook: Qualitative data analysis* (2nd ed.). Newbury Park, CA: SAGE.

- Mohamed, Z., Valcke, M., & De Wever, B. (2017). Are they ready to teach? Student teachers' readiness for the job with reference to teacher competence frameworks. *Journal of Education for Teaching*, 43(2), 151-170.
- Morse, J. M. (2004). Theoretical saturation. In M. S. Lewis-Beck, A. Bryman, & T. F. Liao (Eds.), *The Sage encyclopedia of social science research methods* (pp. 1123). Thousand Oaks, CA: Sage. Retrieved from <http://sk.sagepub.com/reference/download/socialscience/n1011.pdf>
- Myers, I. B., & McCaulley, M. H. (1985). *Manual: A guide to the development and use of the Myers-Briggs type indicator*. PaloAlto, CA: Consulting Psychologist.
- Nghia, T. L. H. (2015). Vietnamese students' perception of English teacher qualities: Implications for teacher professional development. *International Journal of Academic Research in Education and Review*, 3(1), 7-19.
- Noddings, N. (2013). *Caring: A relational approach and moral education* (2nd ed.). Berkley, CA: University of California Press.
- OBEC. (2015). *Number of people who applied for teaching positions in OBEC schools, 2010-2014*. Bangkok: Office of the Basic Education Commission.
- OECD/UNESCO. (2016). *Education in Thailand: An OECD-UNESCO perspective, reviews of national policies for education*. Paris: OECD. doi:10.1787/9789264259119-en
- Onwuegbuzie, A. J., & Collins, K. M. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report*, 12(2), 281-316.
- Onwuegbuzie, A. J., Bustamante, R. M., & Nelson, J. A. (2010). Mixed research as a tool for developing quantitative instruments. *Journal of Mixed Methods Research*, 4(1), 56-78.
- Onwuegbuzie, A. J., Dickinson, W. B., Leech, N. L., & Zoran, A. G. (2010). Toward more rigor in focus group research in stress and coping and beyond. In K. M. T. Collins, A. J. Onwuegbuzie, & Q. G. Jiao (Eds.), *Toward a broader understanding of stress and coping. Mixed methods approaches* (pp. 243-285). United State of America: Information Age.
- Özmen, K. S. (2011). Washback effects of the inter-university foreign language examination on foreign language competencies of candidate academics. *Novitas-ROYAL Research on Youth and Language*, 5(2), 215-228.
- Pan, Y. C., & Roever, C. (2016). Consequences of test use: A case study of employers' voice on the social impact of English certification exit requirements in Taiwan. *Language Testing in Asia*, 6(1), 1-21.
- Parsi, A., & Darling-Hammond, L. (2015). *Performance assessments: How state policy can advance assessments for 21st century learning*. Retrieved from <https://edpolicy.stanford.edu/sites/default/files/publications/performance-assessments-how-state-policy-can-advance-assessments-21st-century-learning.pdf>
- Pettis, J. (1997). Developing our professional competence: Some reflections. *TESL Canada Journal*, 16(2), 67-71.
- Pillay, N. (2014). *Teachers' knowledge of legislation and education law specifically and its influence on their practice* (Unpublished doctoral dissertation). University of Pretoria, Güney Afrika. Retrieved from <https://repository.up.ac.za/handle/2263/43224>
- Polesel, J., Rice, S., & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy: A teacher perspective from Australia. *Journal of Education Policy*, 29(5), 640-657.
- Ramírez Carpeño, E., & Mekochi, Y. (2015). Initial teacher education in Japan and Spain. A comparative study. *Revista Española de Educación Comparada*, 25, 101-127.
- Richards, H., Conway, C., Roskvist, A., & Harvey, S. (2013). Foreign language teachers' language proficiency and their language teaching practice. *The Language Learning Journal*, 41(2), 231-246.
- Richards, J. C. (2017). Teaching English through English: Proficiency, pedagogy and performance. *RELC Journal*, 48(1), 7-30.

- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458.
- Rose, D., English, A., & Finney, T. (2014). *Hire better teachers now: Using the science of selection to find the best teachers for your school*. Cambridge: Harvard Education Press.
- Rubin, H. J., & Rubin, I. S. (2005). *Qualitative interviewing: The art of hearing data* (2nd ed.). Thousand Oaks, CA: Sage.
- Rubin, H. J., v Rubin, I. S. (2012). *Qualitative interviewing: The art of hearing data*. Los Angeles: SAGE.
- Sakurai, Y. (2012). Learners' perceptions of "good" foreign language teachers: A quantitative analysis between native and non-native teachers. *Electronic Journal of Foreign Language Teaching*, 9(1), 46-60.
- Santelices, M. V., Valencia, E., Gonzalez, J., & Taut, S. (2017). Two teacher quality measures and the role of context: Evidence from Chile. *Educational Assessment, Evaluation and Accountability*, 29(2), 111-146.
- Sayın, B. A., & Aslan, M. M. (2016). The negative effects of undergraduate placement examination of English (LYS-5) on ELT students in Turkey. *Participatory Educational Research*, 3(1), 30-39.
- Schussler, D. L., Stooksberry, L. M., & Bercaw, L. A. (2010). Understanding teacher candidate dispositions: Reflecting to build self-awareness. *Journal of Teacher Education*, 61(4), 350-363.
- Shih, C. M. (2008). Critical language testing: A case study of the General English Proficiency Test. *English Teaching Journal* 32(3), 1-34.
- Shohamy, E. (2001). Democratic assessment as an alternative. *Language Testing*, 18(4), 373-391.
- Shohamy E. (2017) Critical language testing. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (3rd ed.). Springer, Cham.
- Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language teaching research*, 9(1), 5-29.
- Strauss, A., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA, US: Sage.
- Stronge, J. H., & Hindman, J. (2006). *The teacher quality index: A protocol for teacher selection*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Taner, G. (2017). *Investigating perceived competencies of English language teachers in Turkey with regard to educational background and experience* (Unpublished doctoral dissertation). Middle East Technical University, Ankara.
- Tongco, M. D. C. (2007). Purposive sampling as a tool for informant selection. *Ethnobotany Research and Applications*, 5, 147-158.
- Tucker, P., & Stronge, J. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Uyulgan, M. A., & Akkuzu, N. (2015). Öğretmen adaylarının gözünden nitelikli öğretmen seçimi: Ortaöğretim fen ve matematik alanları eğitimi bölümünde okuyan öğrencilerin durumu. *KÜ Kastamonu Eğitim Dergisi*, 23(3), 917-940.
- Van Canh, L., & Renandya, W. A. (2017). Teachers' English proficiency and classroom language use: A conversation analysis study. *RELC Journal*, 48(1), 67-81.
- Walsh, K. (2002). Positive spin: the evidence for traditional teacher certification, reexamined. *Education Next*, 2(1), 79-85.
- Walsh, I., Holton, J. A., Bailyn, L., Fernandez, W., Levina, N., & Glaser, B. (2015). What grounded theory is... a critically reflective conversation among scholars. *Organizational Research Methods*, 18(4), 581-599.
- Welsh, E. (2002). Dealing with data: Using NVivo in the qualitative data analysis process. *Forum: Qualitative Social Research*, 3(2).

- Wichadee, S. (2010). Defining the effective English language teacher: Students' and teachers' perspectives. In A. M. Stoke (Ed.), *JALT 2009 Conference Proceedings*. Tokyo: JALT.
- Yalçın-İncik, E., & Akay, C. (2015). Eğitim fakültesi ve pedagojik formasyon sertifika programlarında öğrenim gören öğretmen adaylarının öğretmenlik mesleği yeterliklerine yönelik görüşleri. *Journal of Kirsehir Education Faculty*, 16(2), 179-197.
- Yeşilçınar, S. (2018). *An evaluation of EFL teacher assessment and evaluation: A suggested model* (Unpublished doctoral dissertation). Gazi University, Ankara.
- Yeşilçınar, S., & Çakır, A. (2018). Continuing professional development process: English language teachers' views on the role of teacher research. *Journal of Language and Linguistic Studies*, 14(2), 61-75.
- Yılmaz, M., & Yaşar, Z. (2016). Öğretmen adaylarının Kamu Personeli Seçme Sınavına (KPSS) yönelik algıları. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 5(2), 644-651.