Ilkogretim Online - Elementary Education Online, 2020; 19 (3): pp. 1548-1564
http://ilkogretim-online.org.tr
doi:10.17051/ilkonline.2020.733191

Development and validation of the English teachers' attitudes towards recruitment system scale*

Sabahattin Yeşilçınar, *Muş Alparslan University,* Turkey, *sycinar@ yahoo.com* ORCID: 0000-0001-6457-0211
Abdulvahit Çakır, *Gazi University,* Turkey, *vahit@gazi.edu.tr* ORCID: 0000-0001-8563-6830

Abstract. Teacher assessment and evaluation (TAE) of a nation is an important determinant of its teacher quality and student achievement. However, the literature lacks a scale for displaying English teachers' attitudes towards TAE model in an English as a Foreign Language (EFL) context where teachers are assessed and evaluated via high-stakes tests based on multiple choice questions. The English Teachers' Attitudes towards Recruitment System Scale (ETARS) has been developed based on the existing literature, and data were collected from various stakeholders (policymakers, education experts, teacher trainers, teachers, student teachers). The data were piloted with 319 teachers through exploratory factor analysis. Then, a confirmatory factor analysis was utilized with 260 teachers. Results indicated the ETARS, comprised of 23 items and three dimensions, to be a valid and reliable instrument. Finally, the ETARS was applied to 260 English teachers to reveal their attitudes towards Turkish TAE model in terms of various variables. Findings displayed that English teachers had negative attitudes towards the current TAE model. This research fills the gap in the literature by providing the ETARS.

Keywords: Teacher assessment and evaluation, teacher recruitment, foreign language in Turkey, Teacher quality, attitude scale

Received: 14.06.2019	Accepted: 22.01.2020	Published: 15.06.2020

INTRODUCTION

Academic studies and reports focusing on the processes related to teacher assessment and evaluation (Hereafter, TAE) (Al-Issa, Al-Bulushi, & Al-Zadjali, 2017; Brutti & Sánchez, 2017; Darling-Hammond & Rothman, 2011; Guha, Hyler, & Darling-Hammond, 2017; Herrmann, 2018) have revealed that a selective system is important for selecting and recruiting qualified teachers. Teacher recruitment and teacher quality are two intertwined concepts (Rice, 2008), and the issue of qualified teachers has become one of the most important agenda items for nations (Beauchamp, Clarke, Hulme, & Murray, 2013; European Commission, 2013b; OECD, 2013). It may be due to the fact that, among all school resources, the most important determinant of student achievement is qualified teachers (Koedel, Mihaly, & Rockoff, 2015; Rivkin, Hanushek, & Kain, 2005; Stronge, 2010). From that point of view, the effectiveness of any educational system and its TAE model are interrelated (Johnstone, 2004; OECD, 2005). Hence, any disorder, failure, or malfunction in the teacher evaluation process may affect the functionality of the entire system negatively and prevent effective teaching and learning environments (OECD, 2009). For example, a poor recruitment decision may not only decrease student achievement but also causes poor morale among colleagues (Clement, 2009). In order to recruit teachers, it is necessary "to make decisions about 'selecting in' when the number of applicants outweighs the number of available places" (Klassen, et al., 2017, p. 97); therefore, TAE refers to teacher selection and recruitment in the present study.

Maximizing the impact of hiring processes is vital for selecting effective teachers (Stronge & Hindman, 2006, p. 1) because a poor hiring decision may decrease quality of education (Mason & Schroeder, 2010). In addition, quality of any educational system depends on the efficiency of teachers (Eurydice, 2012 p. 9; Ingersoll & Merrill, 2010). Each student is not taught

* This article is derived from Sabahattin Yeşilçınar's PhD dissertation entitled "An evaluation of EFL teacher assessment and evaluation: A suggested model", conducted under the supervision of Abdulvahit Çakır.

by a qualified teacher; so, in order to provide high-quality teaching to all students, recruiting qualified teachers is a must for educational leaders (Rose, English, & Finney, 2010). That is, assessing and evaluating effective teachers and providing every child with high-quality education should be one of the most important issues for all nations, because making the best selection and recruitment choices is paramount to better teaching and greater student success. This purpose cannot be achieved without effective teachers since an effective teacher reduces students' anxiety (Ghasemi & Hashemi, 2011; Çelik, Arıkan, & Caner, 2013), treats students with warmth and kindness and provides an environment involving positive student-teacher interaction (Telli, den Brok, & Çakıroğlu, 2008).

Selecting and recruiting teachers does not only refer to employing a sufficient number of teachers, but also equipping teaching profession with qualified teachers (Allen, 2005 p.10), indicating that poorly trained teachers do the most harm to students (Mitchell & Barth, 1999). That is, the quality of an education system can be increased only if teachers are trained with necessary skills, knowledge and attitudes. As emphasized in the Teacher Strategy Paper published by the Ministry of National Education (MoNE) in 2017 (MEB, 2017a), policymakers should pay special care to the TAE process for better quality education and greater student achievement. In a nutshell, carefully designed TAE model is a must for ensuring quality education and teachers, and enhancing student achievement.

Qualified Teachers

Recent years have witnessed an increasing interest in policies that promote teacher quality (Adnot, Dee, Katz, & Wyckoff, 2017; European Commission, 2013a; OECD, 2013), because the most significant factor contributing to student success is considered to be teacher quality (Araujo, Carneiro, Cruz-Aguayo, & Schady, 2016; Canales & Maldonado, 2018). Although there is a lack of agreement on the definition of 'quality' and the best way to gauge it (Cohen, 2010; Kennedy, 2010), research shows that "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" may contribute to teacher effects (Darling-Hammond & Wei, 2012, p. 614). Moreover, to make the evaluation system more fair and accurate, it requires collecting multiple sources of evidence, and considering the characteristics and practices that a qualified teacher should have. A qualified teacher refers to a person who has "both successful teaching, which realizes intended outcomes, and good teaching, which is morally worthwhile" (Schussler, Stooksberry, & Bercaw, 2010 p. 351).

Teacher quality and effectiveness have been pointed out in various reports (e.g. OECD 2005; 2009) and research (e.g. Chamberlain, 2013; Goldhaber, Grout, & Huntington-Klein, 2014; Goldhaber & Hansen, 2013; Morgan & Hodge, 2014). They emphasize that one of the most important prerequisites of student achievement is a teacher who is qualified and effective. Therefore, both national and international agendas have started to delve into teacher education and teacher competencies with the aim of designing professional standards, strengthening teacher education and certification requirements, and transforming roles for teachers (Darling-Hammond, 2010).

Competency is a term used for the required knowledge, skills, attitudes, values, motivations and beliefs that teachers should have to be successful in the teaching profession (Gupta, 2011, p. 24). The Association of Teachers of English to Speakers of Other Languages (TESOL) (2010) asserts five domains for training English teachers: culture, assessment, instruction, language, and professionalism. The Common European Framework of Reference for Languages: Learning, Teaching, Assessment (Council of Europe, 2011) does not include standards per se. However, domains (which have structure, knowledge and understanding, strategies and skills, and values) make it possible to organize the Framework of Reference. The 'structure' domain is about the organization of a teacher preparation program. The 'knowledge and understanding' domain resembles 'language' and 'culture' of TESOL, whereas the 'strategies and skills' resembles TESOL's Instruction domain. In Turkey, the MoNE has revised teacher competencies by adding general qualifications and knowledge of field education to the generic teacher competencies rather than specifying a separate field-specific teacher competence so that

a single and integral text has been created to cover the competencies of each teacher in his/her field (MEB, 2017b). Consequently, the new generic teacher competencies consist of three interdependent and complementary domains: professional knowledge, professional skills, attitudes and values.

The characteristics of qualified teachers may vary from one country to another. For instance, the qualities that an English teacher needs to have can be listed as good knowledge of the target language, and taking into account students' needs and motivation in Israel (Brosh, 1996); compatibility, lesson delivery, justice, knowledge and credibility, and organization and preparation in Korea (Barnes & Lock, 2010); ability to teach, personality, and teacher-student relationship are required in Taiwan (Chen & Lin, 2009); pedagogical knowledge, English proficiency, and socio-affective skills, organization and communication skills in Thailand (Wichadee, 2010); personality traits, pedagogical knowledge, professional skills and classroom behaviors in Turkey (Çelik, Arıkan, & Caner, 2013); teaching methods, English competence, and socio-affective skills in Vietnam (Nghia, 2015). In a nutshell, although the required qualities are context-bound, there seems to be a consensus on the underlying principles of qualified English teachers, such as having

1- knowledge (e.g. literacy, numeracy, subject knowledge, knowledge of curriculum etc.)
2- skills (ability to perform a lesson, to use technology appropriately, to speak target language fluently, assessment literacy, classroom management, communication etc.)
3- attitudes (attitude about teaching, self-efficacy, motivational disposition etc.).

Teacher assessment and evaluation in Turkey

The available international literature on the context of TAE is totally different from that of Turkey. In the former, schools or districts are mostly autonomous to hire their teachers, developing their own teacher evaluation policies so that they could eliminate ineffective teachers (Adnot, Dee, Katz, & Wyckoff, 2017). However, the latter offers no autonomy to schools or districts due to its highly centralized structure. Since foundation of the Turkish republic, various TAE models have been applied such as employing graduates of Village Institutes (till 1950s), Reserve Officer Teachers (1960), Substitute Teacher (1961), Peace Volunteers (1962, US experts), Teaching Formation (1970), Teacher Training by Letters (1974), Intensive Teacher Training Program (1975), those who did their military service as teachers (1987), and appointing all university graduates as teachers in mass without any conditions (1996) (Akyüz, 2001, p. 352; Çılgın, 2009 p. 28; Öztürk, 2005, pp.195-199). In 1985, for the first time, student teachers took a central examination in order to be an English teacher (see Çılgın, 2009 for more details). However, in 1992, due to the teacher shortage, the examination requirement for teachers was removed again. In 2001, student teachers started to be re-evaluated under the name of KMS (Central Qualifying Examination for Institutions) which was renamed as Kamu Personeli Seçme Sınavı (Public Personnel Selection Examination, hereafter KPSS) in 2002. Currently, those who want to be English teachers in public schools are selected through KPSS, which is conducted in four phases: general culture and general skills (120 questions), educational sciences (80 questions), Öğretmenlik Alan Bilgisi Testi (Teaching Field Knowledge Test, hereafter ÖABT, [50 questions]), and Interview. The first three phases are based on multiple-choice exams. Unfortunately, the interview has some vital limitations such as being run in Turkish, and the lack of scoring rubrics and field experts.

Fairness of teacher assessment and evaluation model in Turkey

Considering the qualities that an English teacher needs to have, a fair evaluation system requires "gathering multiple sources of evidence about teacher practice", (Danielson, 1996, 2007; Peterson, 2000 as cited in Isoré, 2009, p.20). Two main principles are important for test fairness; (1) *The Principle of Justice* (treating every person with equal respect), and (2) *The Principle of Beneficence* (not being harmful or detrimental to society) (Kunnan, 2005, p. 235). Fairness in testing has various definitions, but the current study adopts *ETS Standards for Quality and Fairness*' definition of fairness: "the extent to which the inferences made on the basis of test scores are valid for different groups of test takers." (Educational Testing Service, 2014, p.

19). According to Messick (1989), construct-irrelevant variance and construct under-representation are two major threats to validity and they need to be examined to improve the test quality and thus promote positive washback. In the former, construct has some missing points, whereas the latter happens when irrelevant components influence test score variance. These two factors both diminish test validity and affect the fairness of the test if groups are affected differentially. In other saying, in order to evaluate student teachers fairly, firstly, the test should represent the construct; then, precautions must be taken to prevent some groups of test takers (e.g., student teachers of English language and literature) from being more influenced 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. Studies have revealed that ÖABT and KPSS are inappropriate examinations in terms of selecting qualified teachers and suggested a call for alternative examinations by using performance-based scores and multiple assessment tools (e.g., Karaca, 2011; Kılıçkaya & Krajka, 2013; Sezgin & Duran, 2011; Uyulgan &Akkuzu, 2015).

Developing a valid, reliable and fair TAE model has become a must as the current model has reliability and validity concerns. To the best of the researchers' knowledge, no scale has been developed to measure English teachers' attitudes towards the way they were recruited to public schools in a centralized system. In light of this apparent gap, this study aims at developing a fair, valid and reliable instrument for assessing English teachers' attitudes towards the recruitment system strongly based on fairness in testing (Kunnan, 2005; Messick, 1989); teacher competencies suggested by the Ministry (MEB, 2017b), TESOL (2010) and CEFR (Council of Europe, 2011); and qualities of effective teachers (Stronge, 2007). To sum up, this research aims to report the development of the ETARS, an instrument that measures the attitudes of English teachers towards the way they were recruited in a centralized system. The ETARS is considered to be a practical tool for researchers seeking to investigate the appropriateness of TAE model for English teachers.

METHODS

Research Design

The goal of this descriptive study is to develop a fair, valid and reliable tool for measuring English teachers' attitudes towards recruitment system. Hinkin's (2005) scale development process was used as a general framework. In addition, the scale was tested for internal consistency and split half reliability. During both piloting and replication phases, the Cronbach's alpha coefficients were calculated. To determine the reliability of the results, the Spearman Brown split half reliability coefficient was calculated.

Sample

During the development and validation of the ETARS, data were collected from in-service English teachers who work at public schools (from primary to upper secondary school) in different parts of Turkey. As the participants were selected according to their convenience, English teachers in Ankara were requested to join the study. However, since this scale was created in an online format and sent to participants via e-mail, we asked each participant to disseminate the mail to ensure the adequate number of participants. That is to say, the quantitative part consisted of both convenience sampling, which is used for selecting those "that are conveniently available and willing to participate in the study" and snowball sampling in which teachers are requested to "recruit individuals to join the study" (Onwuegbuzie & Collins, 2007, p. 286). The researcher took into account two criteria while choosing the sample of in-service English language teachers who would join the quantitative part of this paper: They were those a) who were working in public schools, and b) who had ÖABT experience. The pilot sample for the exploratory factor analysis (EFA) involves 319 English teachers whereas the replication sample for confirmatory factor analysis (CFA) includes 260 English teachers. A sample of at least 100 subjects is suggested to be sufficient for scale development (Kline, 1994)

and a minimum sample size of 200 for confirmatory factor analysis (McCallum, Widaman, Zhang and Hong, 1999 as cited in Hinkin, 2005). Therefore, the sample size can be considered suitable in the present study.

Measures

Demographic information form

This form asks teachers to answer questions about demographic variables such as sex, graduated department, grade taught, reason for choosing department, and desired profession

English Teachers' Attitudes towards Recruitment System Scale

The 23-item ETARS consists of three dimensions/factors, namely, *positive attitude towards TAE, being interested in TAE*, and *negative attitudes towards TAE*.

Item Generation and Refinement

In order to develop ETARS, the researcher joined a PhD scale development course for one term. The instructor of the course was an expert (full professor) who had experience in developing scales. First, an 81-item pool was created. However, based on the instructor's suggestions, some of the items were omitted and the number of items at the piloting stage was decreased to 68. Before piloting the scale, the 68-item form was sent to ten in-service English language teachers who worked at different types of schools such as primary, secondary, etc. They were requested to respond to each item carefully. All items were revised based on their feedbacks and the necessary changes were done. Then, the piloting was conducted with 319 English language teachers. The analysis generated a 32-itemscale. However, the factor load values of nine items (1, 2, 7, 12, 16, 25, 26, 28, and 31) were observed to be smaller than 0.45, which is the lower limit, and these items were excluded from the scale in line with the expert opinions. The items were re-analyzed through the promax technique. The new 3-component scale with 23 items was administered to 260 English language teachers. Data were analyzed via LISREL, and the CFA confirmed the ETARS as a reliable scale for examining teachers' attitudes about the way they are recruited.

Analysis

Data obtained from the ETARS were analyzed via the statistical analysis software, namely SPSS 20.00. EFA was conducted through the package program where as LISREL was applied for CFA. In addition, descriptive statistics were presented including measures of frequency, central tendency, and variability. The normality distribution of the relationships between demographic variables and dimensions together with the item total score of ETARS was analyzed. The p value calculated in the test being lower than $\alpha = .05$ is a sign of normal distribution (Field, 2009; Howitt & Cramer, 2011). Kolmogorov-Smirnov normality tests were applied, and the data show normal distribution. Therefore, t-test (Independent Samples Test) was used for demographic variables consisting of two categories, while one-way analysis of variance (ANOVA) tests were applied for the demographic variables consisting of three and more categories.

RESULTS

Findings of the Exploratory Factor Analysis

As a first step, items with negative meaning were reverse-scored in SPSS 20. To obtain better results from the factor analysis, the promax technique was applied to the scale items as a rotation method and it was decided to classify the scale under three factors. According to Table 1, the KMO value was found to be .864, and KMO and Barlett statistic confirmed the appropriateness of the sample size (KMO = .864; $\chi 2 = 2729.617$, p 0.000). The KMO value relates to the suitability of the sample. If a value is higher than .60, it is suitable for factor analysis (Tabachnick & Fidell, 2013). High Barlett test scores also indicate the feasibility of factor analysis, and high correlations between items (Şeker & Gençdoğan, 2006 as cited in Günüç, 2009). When Table 1 was examined, the total variance of the three components of the scale was

found to be 52.724, and the first, the second and the third components explained 28.14%, 17.88% and 6.68% of the variation, respectively.

Table 1. *Findings of KMO Barlett test values and factor analysis*

Kaiser-Meyer-Olkin (KMO)		.864
Bartlett's Test of Sphericity	Chi-Square (χ^2)	2729.617
	Sd	253
	P	.000

	Initial Eigenvalues			**Total Variance Explained** Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings
Item	Total	Variance (%)	Cumulative (%)	Item	Variance (%)	Cumulative (%)	Total
1	6.474	28.148	28.148	6.474	28.148	28.148	5.949
2	4.114	17.888	46.036	4.114	17.888	46.036	4.454
3	1.538	6.688	52.724	1.538	6.688	52.724	4.108
4	1.066	4.636	57.360				
5	1.000	4.347	61.707				
6	.914	3.972	65.679				
7	.851	3.701	69.380				
8	.774	3.363	72.743				
9	.724	3.147	75.890				
10	.660	2.869	78.760				
11	.632	2.749	81.509				
12	.544	2.363	83.872				
13	.516	2.244	86.116				
14	.476	2.068	88.184				
15	.392	1.705	89.889				
16	.371	1.614	91.502				
17	.351	1.527	93.029				
18	.327	1.423	94.452				
19	.291	1.263	95.715				
20	.273	1.186	96.902				
21	.258	1.122	98.024				
22	.251	1.090	99.114				
23	.204	.886	100.000				

n=319

Table 2 shows that the reliability coefficients of dimensions were between .703 and .918, and the factor loads of all items were greater than .40. The load value of the first component was between .593 and .858; the load value of the second component was between .646 and .801; the load value of the third component was between .513 and .814.

Table 2. *Information on the structure of the scale*

Dimension	N. of Items	Items	Name of the component	α	Item load value
D1	10	i2, i4, i6, i7, i9, i10, i12, i15, i18, 21	Positive attitude towards TAE	.918	between .593 and .858
D2	8	i8, i11, i13, i16, i19, i20, i22, i23,	Being interested in TAE	.883	between .646 and .801
D3	5	i1, i3, i5, i14, i17	Negative attitude towards TAE	.703	between .513 and .814

n=319

As it is seen, the mean was found as 69.121. The lowest total score was 42 and the highest total score was 99.

Table 3. *Descriptive statistics of the item total score*

	Mean	Std. Deviation	Minimum	Maximum
Total score	69.121	10.306	42	99

Findings of the Confirmatory Factor Analysis

Although a sample consisting of 200 participants is enough for carrying out CFA (Shah & Goldstein, 2006), CFA was carried out with 260 English teachers to compare the three-factor model's fit, and to confirm the ETARS's construct validity. Figure 1 presents the path diagram of the emerged model.

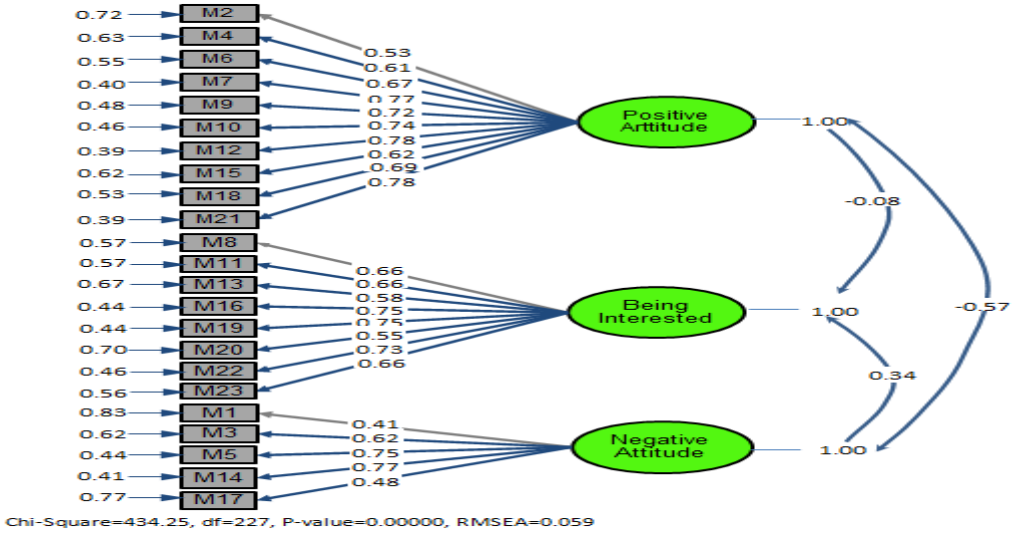

FIGURE 1. *Standard coefficients and error variances for CFA model*

As is seen in Figure 1, *the Positive Attitude towards TAE* includes 10 items. Standardized regression coefficients of these items ranged from 0.53 to 0.78. All these coefficients were significant at 0.05 level. Considering the error variance, it varied between 0.39 and 0.72. *Being Interested in TAE* has 8 items. The standard coefficients of these items ranged from 0.55 to 0.75 and all these coefficients were significant at 0.05 level. The error variances of the items of this factor varied between 0.44 and 0.70. *The Negative Attitude towards TAE* consists of 5 items whose standardized regression coefficients ranged from 0.41 to 0.77, and they were significant at 0.05 level. The error variances of the items of this factor ranged from 0.41 to 0.83. CFA was applied to determine the model data fit and to find evidence for construct validity, and the fit indices of the CFA results were given in Table 4. Many fit statistics were utilised for evaluation of the emerged model fit. The values of chi-square (χ^2), degree of freedom (*df*), Root Mean Square Error of Approximation (RMSEA) (Steiger, 1990), Normed Fit Index (NFI) (Bentler & Bonnet, 1980), and Non-Formed Fit Index (NNFI) (Bentler & Bonnet, 1980) , Comparative Fit Index (CFI) (Bentler, 1990), Goodness of Fit Index (GFI) (Jöreskog & Sörbom, 1989; Tanaka & Huba, 1984) were interpreted to assess model fit. Since the chi-square value is generally affected by sample size, χ^2 / *df* was considered. The value of CFI, NFI, NNFI and GFI being higher than 0.90 indicates that the model is compatible with the data (Hu & Bentler, 1999; Tabachnick & Fidel, 2001) even if χ^2 / *df* ratio is smaller than 5, and the RMSEA value is less than 0.08.

According to Table 4, the RMSEA value was 0.059, and the confidence interval at 90% level was between 0.051 and 0.068. The values of CFI, NFI, NNFI, GFI and χ^2 / *df* were 0.96, 0.93, 0.96, 0.84, and 1.91, respectively. All values, except the GFI, showed that the model fits well with the data. Cheung and Rensvold (2002) state that GFI can be affected by the sample size and distribution, and in these cases, values such as CFI and NNFI can be reported. Based on these findings, it can be said that the three-factor model with 23 items fits the data.

Table 4. *Fit statistics of the three-factor scale*

χ^2 (*df*)	χ^2 / *df*	RMSEA	CFI	NFI	NNFI	GFI
434,25 (227)*	1,91	0,059	0,96	0,93	0,96	0,84

*p <0,05

In addition to the aforementioned evidences, to see the reliability of measurement results obtained for the construct validity of the developed scale, Cronbach Alpha (α) reliability coefficient and Spearman Brown split half reliability of three factors were calculated and the values were presented in Table 5.

Table 5. *Reliability of the scores obtained from the sub-factors*

	Cronbach Alpha (α)	Spearman Brown Split Half Reliability
Positive Attitude towards TAE	0,90	0,87
Being Interested in TAE	0,87	0,85
Negative Attitude towards TAE	0,73	0,74

As is seen in Table 5, considering the reliability of the scores obtained from all sub-factors, the coefficients obtained for both Cronbach Alpha and Spearman Brown Split Half methods varied approximately from 0.73 to 0.90 and from 0.74 to 0.87, respectively. These values indicate that the sub-dimensions' scores were reliable.

Findings of the ETARS

The ETARS was applied to 260 English teachers in order to reveal their attitudes towards Turkish TAE model in terms of some demographic variables such as sex, department teacher graduated, grade taught, reason for choosing department, and desired profession. As it is clear in Table 6, t-Test was applied for checking the relationship between sex variable and dimensions D1, D2, D3 together with item total score of ETARS. No significant difference was noted in terms of sex variable, and the item total score of ETARS was calculated (t= -.652; p>.05), D1 sub-dimension (t= .510; p>.05), D2 sub-dimension (t= -1.475; p>.05) and D3 sub-dimension (t= -.486; p>.05). That is, in Turkey, EFL teachers' attitudes towards the current recruitment system did not differ according to sex variable.

Table 6. *Scores related to the effect of gender variable on ETARS*

	Sex	N	Means	Std. Deviation	t	df	p.
Total	male	76	68,3684	10,49424	-,652	257	,515
	female	183	69,2787	10,13122			
D1	male	76	20,7105	8,50696	,510	255	,611
	female	181	20,1381	8,08824			
D2	male	76	27,4737	7,12268	-1,475	257	,141
	female	183	28,7923	6,29936			
D3	male	76	20,1842	4,26368	-,486	255	,628
	female	181	20,4586	4,07631			

As presented in Table 7, t-Test was utilized to investigate if a significant difference occurred between graduated department variable and dimensions D1, D2, D3 together with item total score of ETARS. No significant difference was noted between graduated department variable and the item total score of ETARS (t= -.608; p>.05), D1 sub-dimension (t= .004; p>.05), D2 sub-dimension (t= -.372; p>.05), D3 sub-dimension (t= -.761; p>.05). According to these findings, EFL teachers' attitudes towards the way they are recruited did not differ according to graduated department variable in Turkey.

Table 7. *Scores related to the effect of graduated department variable on ETARS*

Dimensions	Group	N	Mean	Std. Deviation	t	Sd	Sig.
Total	ELT	211	68,8104	9,92266	-,608	258	,544
	others	49	69,7959	11,44367			
D1	ELT	209	20,3014	7,98075	,004	256	,997
	others	49	20,3061	9,09946			
D2	ELT	211	28,3270	6,49851	-,372	258	,710
	others	49	28,7143	6,84349			
D3	ELT	209	20,2775	4,26406	-,761	256	,447
	others	49	20,7755	3,44157			

Considering Table 8, ANOVA was applied for the relationship between grade taught variable and dimensions D1, D2, D3 together with item total score of ETARS. According to the results of ANOVA, no significant difference was noted to be between grade taught variable and the item total score of ETARS (F= .827 p>.05), D1 sub-dimension (F= .656; p>.05), D2 sub-dimension (F=.873 p>.05), D3 sub-dimension (F=1.274; p>.05). Regarding Table 8, Turkish EFL teachers' attitudes towards the current TAE did not differ according to grade taught variable.

Table 8. *Scores related to the effect of grade taught variable on ETARS*

Dimen-sions	Group	N	Mean	Std. Deviation		Sum of Squares	df	Mean Square	F	Sig.
Total	primary	52	70,3846	9,76017	Between Groups	172,694	2	86.347	.827	,438
	lower secondary	115	69,0783	11, 14352	Within Groups	26828.302	257	104.390		
	upper secondary	93	68,1183	9, 21583	Total	27000.996	259			
D1	primary	50	20,0200	7,66516	Between Groups	88.198	2	44.099	.656	.520
	lower secondary	115	21,9391	8,51764	Within Groups	17134.221	255	67.193		
	upper secondary	93	19,6667	8,17517	Total	17222.419	257			
D2	primary	52	29.4038	5.87871	Between Groups	75.034	2	37.517	.873	.419
	lower secondary	115	27.9565	6.76341	Within Groups	11047.366	257	42.986		
	upper secondary	93	28.3871	6.65130	Total	11122.400	259			
D3	primary	51	21.1765	3.07705	Between Groups	43.149	2	21.575	1.274	.281
	lower secondary	114	20.2632	4.39842	Within Groups	4317.130	255	16.930		
	upper secondary	93	20.0645	4.24471	Total	4360.279	257			

Table 9 shows that ANOVA was applied for the relationship between reason for choosing department variable and dimensions D1, D2, D3 together with item total score of ETARS. According to the results of ANOVA, no significant difference was found between Reason for Choosing Department variable and the item total score of ETARS (F=.896 p>.05), D1 sub-dimension (F= .064; p>.05), D2 sub-dimension (F=1.874 p>.05) and D3 sub-dimension (F=.591; p>.05). This finding shows that, in Turkey, EFL teachers' attitudes towards the way they are recruited do not differ according to reason for choosing department variable.

Table 9. *Scores related to the effect of reason for choosing department variable on ETARS*

Dimension	Group	N	Mean	Std. Deviation		Sum of Squares	df	Mean Square	F	Sig.
Total	love	162	69.6481	10.32964	Between Groups	280,622	3	93.541	.896	.444
	guided	43	67.5349	9.09303	Within Groups	26720,375	256	104.376		
	English teacher	31	69.3226	11.71150	Total	27000,996	259			
	other	24	66.7917	9.21709						
D1	love	160	20.2250	8.97582	Between Groups				.064	.979
	guided	43	20.1628	6.91420	Within Groups					
	English teacher	31	20.9032	6.61994	Total					
	other	24	20.2917	6.84283						
D2	love	162	28.9877	6.63511	Between Groups	239.037	3	79.679	1.87	.134
	guided	43	26.8372	5.28218	Within Groups	10883.363	256	42.513	4	
	English teacher	31	28.8387	7.69890	Total	11122.400	259			
	other	24	26.6667	6.06248						
D3	love	162	20.4815	4.27650	Between Groups	30.216	3	10.072	.591	.622
	guided	42	20.7619	3.63459	Within Groups	4330.063	254	17.047		
	English teacher	30	19.6667	4.60385	Total	4360.279	257			
	other	24	19.8333	3.15769						

According to Table 10, ANOVA was applied for the relationship between desired profession variable and dimensions D1, D2, D3 together with item total score of ETARS. According to the results of ANOVA, no significant difference was found to be between grade taught variable and the item total score of ETARS (F= 1.773 p>.05), D1 sub-dimension (F= .471; p>.05), D2 sub-dimension (F=2.513 p>.05), D3 sub-dimension (F=1.191; p>.05). According to these findings, whether or not in-service English language teachers perform the job they desired does not affect their attitudes towards the current TAE.

Table 10. *Scores related to the effect of desired profession variable on ETARS*

Dimension	Group	N	Mean	Std. Deviation		Sum of Squares	df	Mean Square	F	Sig.
Total	teacher	145	69,7793	10,39395	Between Groups	367,56	2	183.78		
	academician	66	69,0455	10,32314	Within Groups	26633,43	257	103.63	1.773	.172
	other	49	66,6122	9,29789	Total	27000,99	259			
D1	teacher	144	20,7431	8,67014	Between Groups	63,35	2	31.67		
	academician	66	19,7273	8,21358	Within Groups	17159,06	255	67.29	.471	.625
	other	48	19,7708	6,56308	Total	17222,41	257			
D2	teacher	145	29,0414	6,58056	Between Groups	213,32	2	106.66		
	academician	66	28,3030	6,26337	Within Groups	10909,07	257	42.44	2.513	.083
	other	49	26,6327	6,65111	Total	11122,40	259			
D3	teacher	143	20,0699	4,51391	Between Groups	40,36	2	20.18		
	academician	66	21,0152	3.67316	Within Groups	4319,91	255	16.94	1.191	.306
	other	49	20,3878	3,38389	Total	4360,27	257			

DISCUSSION, CONCLUSION, and IMPLICATIONS

The current study aims at developing an instrument for gauging the attitudes of English language teachers towards the way they are assessed and evaluated in the process of recruitment. The results of EFA indicate ETARS to be a multidimensional construct. The sub-dimensions are *positive attitude towards TAE*, *being interested in TAE*, and *negative attitude towards TAE*. The first sub-dimension examines whether the teacher recruitment system can select competent and qualified English language teachers (Council of Europe, 2011; MEB, 2017b; Stronge, 2007; TESOL, 2010). The competencies are tools of teacher evaluation; that is, they could be employed as criteria for judging and evaluating teachers (Bailey, 2006; Daresh, 2001, p. 281; Quirke, 2007, p. 2). The competencies and features (that qualified teachers should have) may vary from country to country. The items included in the first dimension confirm the

findings of previous research conducted at home (Çelik, Arıkan, & Caner, 2013; Taner, 2017; Yeşilçınar & Çakır, 2018) and abroad (Brosh, 1996; Chen & Lin, 2009; Pantić & Wubbels, 2010; Sakurai, 2012; Yalçın İncik & Akay, 2015; Wichadee, 2010), indicating that qualified English teachers are those who

- love teaching profession,
- are highly motivated to teach English at every level (e.g. primary, lower secondary, upper secondary),
- can communicate with stakeholders (e.g. students, colleagues, parents),
- can constantly improve themselves,
- have required technological knowledge, professional and field knowledge.

The second sub-dimension refers to English teachers' interest in recruitment system. In other words, it reveals whether teachers are curious about new developments (local or global) related to the recruitment system through printed (e.g. books, magazines, newspaper) or electronic sources (e.g. blogs, forums). Whether the teacher is interested in new developments related to teacher recruitment is important because it shows an English teacher's interest and attitudes towards the latest experience. Moreover, individuals who are open to experiences define themselves as independent and original, and they are generally creative, liberal, have a wide area of interest, and like change (McCrae & Costa, 1987).

The final sub-dimension searches whether the recruitment model has validity and fairness (Kunnan, 2005; Messick, 1989) concerns. For example, one of the items of this sub-dimension (i.e. *I believe that the current TAE model selects also the student teachers who cannot teach*) indicate that the current model is not valid (as it does not distinguish qualified teachers from others) and fair (as it ignores performance-related indicators). That is, the model has two major threats to validity: construct under-representation and construct-irrelevant variance (Messick, 1989). According to Chen, Li and Chason (2018), when the goals of assessment are deficiently considered, construct under-representation seems to occur. Alavi and Masjedlou (2017) investigated the validity of IELTS Academic Writing Task One (IAWTO) and found it to be decayed and inadequate as it showed the construct underrepresented, i.e., reducing and narrowing the construct.

The last sub-dimension occurs as a result of the first sub-dimension whose items show whether the current TAE model has construct-irrelevant variance. If a model is not construct-relevant, it militates in favour of some groups due to the test taker's personality, background knowledge, test-taking strategies, as well as general intellectual or cognitive ability. This was confirmed in Cho and So's (2014) study. They examined the interaction of primary students learning English with test materials as well as their response to test questions and found that the use of complex language in test questions might influence the way they demonstrate their skills on the constructs targeted in the test questions, and that perception of test difficulty may be affected by memory load and ambiguity in task descriptions.

Additionally, the internal consistency of the scale was assessed through the Cronbach's alpha, which is usually preferred by researchers for measuring internal consistency in Likert-type scales (Huck, 2012, p. 74). The item total score of the 3-dimention model (69.121%) confirms the suggestions of Hair, Black, Babin, Anderson, and Tatham (2014). First component consists of 10 items and its reliability is .918. Item load value of first component is between .593 and .858. Second component has 8 items, with .883 reliability coefficient. Item load value of second component is between .646 and .801. Finally, third component includes 5 items and its reliability coefficient is .703. Item load value of third component is between .513 and .814. Cronbach alpha (α) internal consistency coefficient for the scale is .758. The reliability coefficients of dimensions are between .703 and .918. Therefore, the scale is considered to be reliable since .70 and preferably higher is an acceptable value for reliability (Fraenkel, Wallen, & Hyun, 2011, p. 157).

Based on the findings of confirmatory factor analysis, it is fair to say that the three-factor model with 23 items fits the data (see Table 4). Furthermore, considering the reliability of the scores obtained from all sub-factors, the reliability coefficients for the scale have been found to

be .90, .87, and .73 for positive attitude towards TAE, being interested in TAE, and negative attitude towards TAE, respectively. In addition, the coefficients obtained for Spearman Brown Split Half methods vary approximately from 0.74 to 0.87. These values display that the scores obtained from the sub-dimensions of the scale were reliable. That is, these values prove that ETARS demonstrates a good level of internal consistency (Cronbach, 1951; George & Mallery, 2003; Lance, Butts, & Michels, 2006). Taking into account the reliability analyses of the scale, the ETARS is regarded to be reliable in terms of measuring the attitudes of English language teachers towards the way they are evaluated before being recruited to public schools. To enhance the reliability of the scale over time, further studies with different samples are required.

As a final step, ETARS was applied to English teachers to display their attitudes towards current TAE. Findings showed that they do not have positive attitudes towards the way they are assessed and evaluated (regardless of demographic variables such as sex, graduated department, grade taught, reason for choosing department, and desired profession). This shows that in-service English language teachers considered the current TAE model inadequate, unfair and invalid, and that they favored some other additional tests that could determine their skills and attitudes in order to reach the most appropriate candidates for the profession. This confirms Karaca (2011) who examined the attitudes of 258 student teachers studying in different departments in 2009 towards KPSS. The research findings displayed that student teachers generally had negative attitudes towards Turkish TAE model because it failed to gauge the quality of teachers. Sezgin and Duran (2011) conducted a study to determine student teachers' perceptions about Turkish TAE model, whether it affected their lives, and whether it influenced student teachers' views on teaching profession and the content of undergraduate courses. The content analysis reported that participants found TAE model as an inappropriate examination in terms of selecting qualified teachers.

Since no significant differences were observed between item total scores of ETARS and demographic variables, semi-structured interviews were conducted to learn the reason behind it. They stated that the current model requires memorizing lots of topics and subjects and asks student teachers respond multiple choice tests; however, there is an urgent need for an information rich TAE model. That is, the new TAE model should provide information about student teachers' knowledge, skills and attitudes. The aforementioned results suggest a call for performance-based scores. It is clear in the literature that schools which are interested in quality education and teachers generally pay attention to performance indicators. For example, Jacob, Rockoff, Taylor, Lindy and Rosen (2016) listed four stages of evaluation that are considered by Washington DC Public Schools while recruiting teachers: general essays, an interview, subject-specific assessments, and a 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 the selection process can be used for improving quality of teachers (Jacob et al., p. 37).

No attitude scales were available in the literature regarding English teachers' attitudes towards teacher evaluation or recruitment model in EFL contexts where the teacher recruitment is actualized centrally. Thus, the researchers developed ETARS to reveal in-service English teachers' attitudes towards the way they are recruited. Both EFA and CFA display that the scale is reliable psychometrically and demonstrates adequate factorial validity. The creation of ETARS is of paramount importance, because although attitudes can be investigated qualitatively, there is still a generalization concern. In this sense, using ETARS will help reach a wider population as Hitchcock, Onwuegbuzie and Khoshaim (2015 p.5) state that such instruments display whether the findings of qualitative research held over large groups of respondents.

By using the ETARS scale, it is possible to explore whether English language teachers have positive or negative attitudes towards teacher assessment and evaluation, whether they are interested in it. From that point of view, it can be deduced that even private institutions also can utilize ETARS in order to evaluate the fairness of their selecting system. This will help

authorities observe the functionality of their evaluation process so that they can change or revise their TAE model in order to bring effective teachers in classrooms.

Acknowledgement

We want to thank Prof. Dr. Adnan Kan for his valuable contributions during the process of scale development.

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.

Akyüz, Y. (2001). *Başlangıçtan 2001'e Türk Eğitim Tarihi*. Genişletilmiş sekizinci basım. İstanbul: Alfa.

Alavi, S. M., & Masjedlou, A. P. (2017). Construct under-representation and construct irrelevant variances on IELTS academic writing task 1: is there any threat to validity? *Theory and Practice in Language Studies, 7*(11), 1097-1107.

Al-Issa, A., Al-Bulushi, A., & Al-Zadjali, R. (2017). Student teachers' perceptions on the selection criteria of English language teachers in the Sultanate of Oman: Implications for policy reconsideration. *Journal of Language and Linguistic Studies*, *13*(2), 235-259.

Allen, M. B. (2005). Eight questions on teacher recruitment and retention: what does the research say?. *Education Commission of the States (NJ3)*.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics*, *131*(3), 1415-1453.

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

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.

Beard, C., & Wilson, J. P. (2013). *The Power of experiential learning: A Handbook for education, training and coaching* (3rd ed.). London: Kogan Page.

Beauchamp, G., Clarke, L., Hulme, M., & Murray, J. (2013). *Research and teacher education: The BERA-RSA inquiry. Policy and practice within the United Kingdom*. Project report. London, UK: British Educational Research Association.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Brosh, H. (1996). Perceived characteristics of an effective language teacher. *Foreign Language Annals, 29(2)*, 125-138.

Brutti, Z., & Sánchez, F. (2017). Does better teacher selection lead to better students? evidence from a large scale reform in Colombia. Documento de trabajo CEDE .

Canales, A., & Maldonado, L. (2018). Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics. *International Journal of Educational Development*, *60*, 33-50.

Chamberlain, G. E. (2013). Predictive effects of teachers and schools on test scores, college attendance, and earnings. *Proceedings of the National Academy of Sciences*, 201315746.

Chen, Y.-H., Li, I. Y, & Chason, W. (2018). Construct underrepresentation. In Frey, B. (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation.* Thousand Oaks, CA: SAGE Publications.

Chen, Y.-J., & Lin, S.-C. (2009). Exploring characteristics for effective EFL teachers from the perceptions of junior high school students in Tainan. *STUT Journal of Humanities and Social Sciences, 2*, 219-249.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural equation modeling*, *9*(2), 233-255.

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*.

Clement, M. C. (2009). Hiring highly qualified teachers begins with quality interview. *Kappan, 9*(2), 22–24.
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.
Council of Europe (2011). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University.
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.
Çılgın, A. (2009). Öğretmen seçim politikalari ve aday öğretmenin işe kabulü için sahip olmasi gereken nitelikler (teacher selection policies and qualifications required for teacher recruitment). Unpublished Dissertation. Yeditepe University
Daresh, J.C. (2001). *Supervision as a proactive leadership (3rd ed.).* Prospect Heights, IL: Waveland.
Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, *61*(1-2), 35-47.
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., & Rothman, R. (2011). Lessons learned from Finland, Ontario, and Singapore. In L. Darling---Hammond and R. Rothman (Eds.), *Teacher and leader effectiveness in high performing education systems* (p. 1---11). Washington: Alliance for Excellent Education and Stanford: Stanford Center for Opportunity Policy in Education.
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). Routledge.
Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service.
European Commission (2013a). Supporting Teacher Competence Developments for Better Learning Outcomes. Retrieved from http://ec.europa.eu/dgs/education_culture/repository/education/policy/school/doc/teachercomp_en.pdf
European Commission. (2013b). *EU performance and first inferences regarding education and training policies in Europe*. Brussels, Belgium: Author. Retrieved from http://ec.europa.eu/dgs/education_culture/repository/education/policy/strategic-framework/doc/pisa2012_en.pdf
Eurydice. (2012). *Key data on teaching languages at school in Europe, Brussels: Education, audiovisual and culture executive agency*. Retrieved from http://eacea.ec.europa.eu/education/eurydice/documents/key-dataseries/143en.pdf
Field, A. (2009). *Discovering statistics using SPSS (and sex and drugs and rock 'n' roll) (Third edition)*. London: SAGE.
Fraenkel, J., Wallen, N., & Hyun, H. (2011). *How to design and evaluate research in education.* (8th ed.) New York, NY: McGraw-Hill.
George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.
Ghasemi, B., & Hashemi, M. (2011). The study of the characteristics of successful English language teachers from the view point of the English language students of Islamic Azad University, Hamedan Branch, in *Procedia-Social and Behavioral Sciences*, *28*, 411-415.
Goldhaber, D., Grout, C., & Huntington-Klein, N. (2014, December). Screen twice, cut once: Assessing the predictive validity of teacher selection tools (CEDR Working Paper No. 2014-9). Seattle, WA: Center for Education Data and Research.
Goldhaber, D., & Hansen, M. (2013). Is it just a bad class? Assessing the long-term stability of estimated teacher performance. *Economica, 80*(319), 589–612.
Guha, R., Hyler, M. E., & Darling-Hammond, L. (2017). The teacher residency: A practical path to recruitment and retention. *American Educator*, *41*(1), 31.
Gupta, K. (2011). *A practical guide to needs assessment*. John Wiley & Sons.
Günüç S. (2009), *İnternet bağımlılık ölçeğinin geliştirilmesi ve bazı demografik değişkenler ile internet bağımlılığı arasındaki ilişkilerin incelenmesi.* Unpublished master's dissertation. Yüzüncü Yıl University.
Hair, J. F. Jr., Black, W. C., Babin, B. J., Anderson R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Prentice Education, Inc.
Herrmann, Z. (2018). Rethinking teacher recruitment. *Educational Leadership*, *75*(8), 18-23.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–124.
Hinkin, T. R. (2005). Scale development principles and practices. In R. A. Swanson and E. E. Holton (Eds.), *Research in organizations: Foundations and methods of inquiry* (pp. 161–179). San Fransisco, CA: Berrett-Koehler Press.
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.
Hu, L.T., & Bentler, P.M.(1999). Cut off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.
Huck, S. (2012). *Reading statistics and research. Boston*, MA: Pearson
Ingersoll, R. & Merrill, L. (2010). Who's teaching our children? *Education Leadership 67*(8), 14 -20.
Isoré, M. (2009). Teacher evaluation: Current practices in OECD countries and a literature review", *OECD Education Working Papers*, No. 23, OECD Publishing. Retrieved from http://dx.doi.org/10.1787/223283631428
Jacob, B., 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.
Johnstone, R. (2004). Language teacher education. In A. Davies & C. Elder (eds.), *The handbook of applied linguistics* (pp. 649-671). Malden, MA: Blackwell Publishing Ltd. http://dx.doi.org/10.1002/9780470757000.ch26
Jöreskog, K. G., & Sörbom, D. (1989). LISREL 7 user's reference guide. Chicago, IL: SPSS Publications.
Karaca, E. (2011). Öğretmen adaylarinin kamu personeli seçme sinavi'na (KPSS) yönelik tutumlari (Student teachers' attitudes towards the public service selection examination). *Akademik Bakış Dergisi. Vol. 1* pp. 1-18
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
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*, 96-114.
Kline, P. (1994). *An easy guide to factor analysis*. New York, NY: Routledge.
Koedel, C., Mihaly, K., & Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, *47*, 180-195.
Kuhlman, N., & Knezevle, B. (2013). The TESOL guidelines for developing EFL standards. Retrieved from https://www.tesol.org/docs/default-source/pdf/the-tesol-guidelines-for-developing-efl-standards.pdf?sfvrsn=0
Kunnan, A. J. (2005). Towards a model of test evaluation: Using the test fairness and test context frameworks. In *Multilingualism and assessment: Achieving transparency, assuring quality, sustaining diversity. Proceedings of the ALTE Berlin Conference* (pp. 229-251).
Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*, 202–220.
Mason, R.W., & Schroeder, M.P. (2010). Principal hiring practices: toward a reduction of uncertainty. *The Clearing House, 83*(5), 186-193.
MEB, (2017a). Teacher strategy paper. Ankara: MEB. Retrieved from https://oygm.meb.gov.tr/meb_iys_dosyalar/2018_05/25170118_Teacher_Strategy_Paper_2017-2023.pdf
MEB, (2017b). General competencies for teaching profession. http://oygm.meb.gov.tr/meb_iys_dosyalar/2018_06/29111119_TeachersGeneralCompetencies.pdf
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.
Mitchell, R., & Barth, P. (1999). Not Good Enough: A Content Analysis of Teacher Licensing Examinations. How Teacher Licensing Tests Fall Short. *Thinking K-16*, *3*(1), 1-20.
Morgan, G. B., & Hodge, K. J. (2014). The stability of teacher performance and effectiveness: Implications for policies concerning teacher evaluation. *Education Policy Analysis Archives*, 22(95), 1-21.

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.
OECD, (2005). Teachers matter: Attracting, developing and retaining effective teachers. Paris: Organisation for Economic Co-operation and Development. Retrieved from http://www.oecd.org/edu/school/48627229.pdf
OECD, (2009). Education at a glance. OECD Indicators. *OECD Publishing*. doi: 10.1787/eag-2009-en
OECD, (2013). Lessons from PISA 2012 for the United States: Strong performers and successful reformers in education. Paris, France: Author. http://dx.doi.org/10.1787/9789264207585-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.
Öztürk, C. (2005). Türkiye'de dünden bugüne öğretmen yetiştiren kurumlar. *İstanbul: Millî Eğitim Bakanlığı Yayınları*.
Pantić, N., & Wubbels, T. (2010). Teacher competencies as a basis for teacher education–Views of Serbian teachers and teacher educators. *Teaching and Teacher Education*, *26*(3), 694-703.
Pillsbury, P. (2005). Only the best: Hiring outstanding teachers. *Leadership, 35*(2), 36–38.
Quirke, P. (2007). *A coherent approach to faculty appraisal*. In C. Coombe, M. Al-Hamly, P. Davidson, and S. Troudi (Eds.), Evaluating teacher effectiveness in EFL/ESL contexts (pp. 89-106). University of Michigan.
Rice, J. K. (2008). From highly qualified to high quality: An imperative for policy and research to recast the teacher mold. *Education Finance and Policy*, *3*(2), 151-164.
Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417-458.
Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*(2), 247-252.
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, MA. Harvard Press.
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
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.
Sezgin, F., & Duran, E. (2011). Kamu personeli seçme sınavı'nın (KPSS) öğretmen adaylarının akademik ve sosyal yaşantılarına yansımaları. Paper presented at 5. Ulusal Sosyal Bilimler Kongresi
Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management, 24*(2), 148–169.
Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioural Research*, *25,* 173–180.
Stronge, J. (2007). *Qualities of effective teachers* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development
Stronge, J. H. (2010). Teacher effectiveness = student achievement: What the research says. Larchmont, NY: Eye On Education.
Stronge, J. H., & Hindman, J. (2006). *The teacher quality index: A protocol for teacher selection* (Association for Supervision and Curriculum Development Alexandria, Virginia, USA).
Tabachnick, B. G., and Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Allyn& Bacon.
Tanaka, J. S., & Huba, G. J. (1984). Confirmatory hierarchical factor analyses of psychological distress measures. *Journal of Personality and Social Psychology, 46,* 621–635.
Taner, G. (2017). I*nvestigating perceived competencies of English language teachers in Turkey with regard to educational background and experience.* Unpublished Doctoral dissertation, Middle East Technical University.
Telli, S., Den Brok, P., & Çakiroglu, J. (2008). Teachers' and students' perceptions of the ideal teacher. *Egitim ve Bilim*, *33*(149), 118.
TESOL International Association (2010). *TESOL P–12 professional teaching standards* (2nd ed.). Alexandria, VA: Author.
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.

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., & 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.