

RESEARCH

Open Access



Assessing Iranian EFL teachers' educational performance based on gender and years of teaching experience

Haleh Mashhadlou and Siros Izadpanah* 

* Correspondence: cyroszadpanah@yahoo.com

Department of English Language Teaching, Zanjan Branch, Islamic Azad University, Zanjan, Iran

Abstract

In recent years, there has been an increasing interest in the assessment of Iranian English teachers' performance. Besides, it was aimed to examine and compare the performance of teachers based on their gender differences and teaching experience. To the first aim, the Delphi technique was used to develop a questionnaire and the reliability of the Delphi questionnaire based on Cronbach's alpha was .982. In the first round, 25 experts including university lecturers and experienced instructors in the field of English teaching were asked to answer open-ended questions regarding important issues in the evaluation of an English teacher. Then, the related themes emerged. Using emerged themes, a questionnaire including 100 questions was designed and measured on a linear scale (1 = not important to 5 = absolutely essential). After calculating the frequency of each item, the results were resent to the panel to rate the questions. In the last phase, three criteria including 1—the mean 4 and more, 2—standard deviation less than 1, 3—less than 10% of the participants do not answer to the item were considered to decide on the final questions and components of the questionnaire. The questionnaire was distributed using Google forms. One hundred and fifty questionnaires were filled correctly and analyzed using SPSS 22. Then, the validity of questionnaire was checked. Overall, it was seen that there was not any significant difference between teachers' performance based on gender difference and teaching experience. The findings of the present study might have some implications for researchers, instructors, language teachers, school administrators, and the ministry of education.

Keywords: Evaluation, EFL teacher, Gender, Performance, Teaching experience

Introduction

One of the methods used to acquire group knowledge is the Delphi technique; it is a structured process for anticipating and assisting decision-making during survey rounds, gathering information, and finally, group consensus (Wilpers et al., 2021; Yazdimoghaddam et al., 2021). While most surveys try to answer the question of "what is," Delphi answers the question of "what can/should be" (Gruetzemacher et al., 2021).

Education is one of the main issues that each country needs to invest it due to its key role in human teaching and social development. Efficient and qualified education



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

systems make efforts to instruct educated and skillful individuals (Argudo Serrano et al., 2021). Every education program is developed to create a quality learning environment for students and learners using high academic standards (Pishghadam et al., 2021). High academic standards require effective teaching and learning which can be achieved through safe learning environments in schools and institutes, well-established academic standards, regular attendance, effective teachers, update facilities, and so on (Jackson Public Schools (JPS), 2013).

Among all of the factors, teachers' performance is considered one of the important factors which play a pivotal role in effective teaching and quality learning (Challob, 2021; Khaksefidi, 2015). In other words, teachers' role is so important that can lead to academic standards. Effective, high-skilled, and motivated teachers can improve teaching and learning (Merati et al., 2021; Santiago & Benavides, 2009). According to Galluzzo (2005), the quality of teachers is a key element in student learning. Teacher performance and quality in teaching is a matter of importance in different areas of education and student learning. One of the challenging roles of teaching in education relates to language teaching in general and English language teaching specifically. Borg (2006) and Mazandarani and Troudi (2021) stated that being a foreign language teacher is different from other areas of teaching due to the content. Some research studies emphasized the role of language teachers and indicated that teachers' pedagogical skills, creativity, behaviors, enthusiasm, management, and fairness are considered as key characteristics of effective language teachers (Al-Thumali, 2011; Çelik et al., 2013; Looney, 2011; Martin et al., 2006). Regarding the importance of effective teaching and the key role of teachers in education, there should be some standards and criteria that measure and evaluate the efficiency of teachers' performance.

One common saying about teaching is that most experts agree that no factor in student learning is more important than the quality of the classroom teacher (Mahvelati, 2021). The most important factor in determining the success or failure of students is the class teacher (Johnson et al., 2021). They believe that the effects of teachers on students' learning are few signs that the succeeding effective teachers can overcome the effects of weak and unqualified teachers.

Given what has been said, teacher evaluation cannot be ignored. The situation of educational evaluation and the establishment of performance accountability systems in many countries around the world are to increase the achievements of quality—educational justice (Pishghadam et al., 2021). Building mutual trust between principals, teachers, and students can be a means to improve teaching and thus student learning (Argudo Serrano et al., 2021). To have effective instructors, we cannot just rely on pre-service training programs or employment, and recruitment mechanisms, but in the age of information bombardment, teachers need to be able to adapt to continuous change through adaptation and continuous learning, thus developing teachers' knowledge, a key component of quality education (Firoozjahantigh et al., 2021). Their evaluation is another necessary mechanism to ensure the competence and conscientiousness of teachers. Also, in many English language programs, systematic evaluation of teachers is rare (Yan et al., 2021). It is common and irregular to be done by managers and senior teachers who do not have enough time and knowledge to do so, so teaching English according to many experts, there is an urgent need for community evaluation for teachers (Mazandarani & Troudi, 2021). A review of the relevant literature shows that advances

have been made in the theoretical and practical areas of teacher evaluation in the field of teacher evaluation. In other words, teacher evaluation should create opportunities for teachers to become more professional, but it seems that the progress that has been made in the areas of educational evaluation in the evaluation system is not enough in Iran.

Reviewing the literature on teacher performance shows that there is an absence of an objective instrument that evaluates teacher quality comprehensively. In other words, one of the most prevalent ways of teacher evaluation is observation all over the world. But it should be taken into account that there are many different strategies such as portfolios, self-evaluation, student evaluation of teachers, peer evaluation, and parent evaluation that can be used to evaluate teacher performance (Borg, 2018; Firoozjahan-tigh et al., 2021). In addition, using different sources for collecting evidence provides a more comprehensive tool of evaluation.

The next need for conducting the present study is that: In the ministry of education in Iran, there is not any written standard for EFL teacher evaluation; however, there is a structure of general standards for all teachers that do not appear to fulfill performance evaluation of EFL teachers (Ministry of Education, 2018). Since each subject matter of study in schools is different from each other, it is necessary for principals, policymakers, school principles to propose appropriate and thorough instrument in order to evaluate teacher performance of each school subject distinctively. Furthermore, it was found by the researchers that Teacher Education Center in Iran does not have any course, planning, workshop, or training on the standards of effective and quality teacher, i.e., teacher students are not exposed to standards and criteria that show the quality of effective teachers, especially EFL teachers. Therefore, assessment or an instrument that can present a rubric for teacher evaluation seems to be crucial. EFL teaching includes various important aspects affecting student learning which require to be addressed in teacher evaluation process. Additionally, there are some studies that emphasize the importance of the academic development of teachers. Studies on teacher evaluation stated that teaching is the core performance by teachers and effective teaching is not just instruction and testing (Khaksefidi, 2015; Vinhais & Abelha, 2015). Similarly, Al-Thumali (2011) emphasized that quality of teachers could improve the academic performance of students.

It can be stated that education in all fields, especially EFL teaching and learning requires some factors which can result in effective instruction and skillful educators. To this end, teachers and instructors play a pivotal role. However, teachers by themselves need to be educated and assessed in order to perform their best for better results in training and instructing successful students and learners. Thus, evaluating the performance of the EFL teachers can be regarded as an effective way to meet this requirement.

The present study aimed to assess Iranian EFL teachers in the context and compare the performance of teachers based on their gender and teaching experience.

Research questions

Regarding the purpose of the study, the following questions were raised and investigated.

Qualitative research question:

1. What are the most notable areas for assessing Iranian EFL teachers' performance in the light of quality standards?

Quantitative research questions:

2. Are there any statistically significant differences between Iranian EFL teachers' performance based on their gender?
3. Are there any statistically significant differences between Iranian EFL teachers' performance based on years of teaching experience?

Literature review

Teacher education

Education presents a significant function in training and preparing individuals for better living in societies. In other words, it has been stated that education includes teaching and learning methods utilized in schools and educational institutes similar to schools. Additionally, it is thought that the education system is a way of transmitting values and knowledge of societies (Mukerhji, 2019). As it can be inferred, teacher education requires both pre-service and in-service development. It is important for teachers, students to be instructed about theories and practices of teaching before starting their profession. After graduation and while serving as teachers, it is required not only to gain the recent changes in approaches, methods, materials, assessment, and evaluation approaches for teachers but also to track their own teaching and evaluate their own performance based on quality standards. Rahman et al. (2011) believed that if teachers could not keep pace with changes in educational and pedagogical developments, they will not perform effectively. In other words, they stated that teacher training provides knowledge of the subject matter, skills of teaching, scientific methods, and academic qualification. Similarly, Irvine (2018) emphasized that the impact of individual teachers on the academic achievement of students is undeniable; therefore, it is of utmost importance that teachers be prepared for such a critical duty. In another finding, it was proposed that teacher professional knowledge and development including pre-service preparation and ongoing development strongly correlates with student achievement (Burroughs et al., 2019). They also stated that a good command of professional knowledge or the same subject matter and curricular knowledge influences teaching quality in the classroom.

English language teacher education

Some studies emphasized the need for effective and qualified teachers for better academic achievement of students (Jyoti Sankar, 2018; Khanjani et al., 2017). Then, it was argued that effective teachers should spend pre-service and in-service education in order to be ready to provide effective instruction and teaching program in the classroom (Rahman et al., 2011). Language teachers, especially English language teachers like all teachers should be prepared for teaching in language classroom.

English teacher education program in Iran

In Iran, teacher education program is held in universities (Teacher Training University) and higher education institutes. Content of education program for preparing teachers is designed by Ministry of Education. English textbooks are designed by the Ministry of Education, as well (Atai & Mazlum, 2013). Teachers attend this pre-service program to gain related knowledge to teach in schools and institutes. Aghaalikhani and Maftoon (2018) stated that autonomy and creativity have not been paid enough attention in teacher education programs in Iran; however, there is not a consistent education program that provides standardized curricula and methodology for student teachers and teachers. Among various fields of study, English teacher education program has been neglected not only in training autonomous and creative teachers but also in involving teachers in program development and planning (Baniasad-Azad et al., 2016). Also, Atai and Mazlum (2013) stated that Iranian teachers believed that their personal experiences are more useful than what they gain from the in-service programs while education planners believed that the mentioned programs are not sufficient and helpful for teachers.

Teacher evaluation techniques

In order to evaluate teacher performance, it is necessary to collect related data. For this aim, several techniques and instruments have been presented and used by researchers so far. Studies have pointed out that using different techniques of data collection not only for teacher evaluation but also in all research studies results in high validity, reliability, and fairness (Arabzadeh, 2016; Mackey & Gass, 2015).

The need to evaluate the performance of teachers

The results of a meta-analysis of more than 5000 studies by Waters et al. (2003) showed the impact of evaluation and management activities in school on success of students. There is a very important relationship between evaluation and student success. Numerous studies have also shown that the quality of teacher teaching is the most important factor in students' success. Since the results of numerous studies, some of which have been shown that teachers play a key role in the success or failure of students, schools and education systems have become more concerned with helping to improve the quality of teachers' teaching by establishing appropriate evaluation systems (Johnson et al., 2021). According to Mazandarani and Troudi (2021), teacher evaluation is an effective strategy for improving the quality of education in school. Another factor that led to increasing attention to teacher evaluation was the increase in public demands for accountability in educational systems. Therefore, educational systems tried to have an effective teaching staff, in addition to hiring qualified teachers, by establishing effective evaluation systems, identifying the strengths and weaknesses of teachers, and providing the community and teachers with professional growth.

Criticisms of existing evaluation systems

Many researchers have criticized current models of teacher evaluation; claiming that these models are ineffective and contribute little to the development of education (Cullen et al., 2021; Donaldson & Firestone, 2021). Scriven (1981) called teacher evaluation

a “disaster” and “classroom observation” a disgrace, and urged researchers in the field to design a more effective process for teacher evaluation by using field advances. Many studies, however, have emphasized that the purpose of teacher evaluation is to ensure quality and professional growth of teacher (Danielson, 2001a, 2001b; Donaldson & Firestone, 2021; Jones et al., 2021). McLaughlin (1990) argues that current evaluation systems achieve none of these goals. Danielson and McGrill (2000) states six major weaknesses of systems of the current evaluation of teachers as follows:

1. Using old and limited criteria
2. Lack of common understanding of quality teaching
3. Inaccuracy in performance evaluation
4. One-sided and top-down view of evaluation
5. Do not differentiate between evaluating experienced and novice teachers
6. Lack of expertise of principals in teacher evaluation

The existence of these shortcomings has made the evaluation systems unable to achieve their goals as expected.

Method

Design of the study

This study has exploited both qualitative and quantitative methods. In other words, a mixed-method design was used to collect data. In a qualitative method, data were collected using the Delphi technique. In a quantitative method, participants were asked to fill in a questionnaire, and results were compared based on gender and teaching years' differences. Also, the present study included research variables. Independent variables of the study were gender and years of teaching experience. The dependent variable was the evaluation of Iranian EFL teacher performance. It means that EFL teachers performance was compared and examined based on their gender and teaching experience.

Participants

In the present study, the population was Iranian EFL teachers. Participants were selected using a purposive sampling method. The reason for selecting participants by purposive sampling was due to the design of the study, especially the qualitative section which required multiple rounds of questioning and interviewing. Since the present study used the Delphi technique to collect qualitative data, it was important to make sure of experts' availability for different rounds of designing questionnaires. Therefore, the accessibility of the participants was taken into account. In addition, limitation in time was another factor that resulted in using purposive sampling in the study. In the first phase of the study, 25 experts participated in the study. The experts participating in the qualitative study included university professors, instructors, and experienced teachers.

In a quantitative study, after designing the questionnaire, 162 EFL teachers filled in the questionnaires. It is worth mentioning that questionnaire was designed using Google Survey and the link of the questionnaire was sent to EFL teachers by email and

social media communities such as Linked in. Overall, 150 questionnaires were found correctly and completely filled with the participants.

Number of specialists

There are no strong and explicit rules about how to select and number of specialists, and the number of them depends on factors: whether the sample is homogeneous or heterogeneous, the Delphi technique or the scope of the problem, the quality of the decision, the ability of the research team to manage the study, internal and external credibility, the time to collect data and available resources, the scope of the question, and the acceptance of the answer (Belton et al., 2021; Flostrand et al., 2020; Landeta, 2006).

The number of participants was usually less than 50 and usually 15 to 20 people. Although 10 to more than 2000 persons have been reported in the articles (Barabino et al., 2021; Landeta, 2006). In homogeneous groups, usually 10-15 people are sufficient and can be accepted (Lamm et al., 2021). In Delphi, heterogeneous examples are commonly used to obtain a wide range of comments, quality responses, and acceptable solutions. This sampling increases the sample size—the problems of data collection and ultimately the complexity of reaching a consensus—performing analysis and review of results will be. However, with a larger sample size, increasing the number of judgments and combining them increases confidence (Niederberger & Spranger, 2020).

Some researchers have pointed out usually 30 persons is enough to provide information, and by increasing them, the answers are repeated and no new information is added (Wei & Hui, 2019). But others write that there is little empirical evidence about the effect of the number of participants on the credibility and trust of the available consensus process (Wilpers et al., 2021).

Instruments

In order to evaluate Iranian EFL teacher performance, previous literature of teacher evaluation on the teacher effectiveness, quality education and teachers, teacher characteristics, and extant conceptual frameworks were studied. Then, the designed questionnaire was used as an instrument to examine Iranian EFL teacher performance.

Validity of questionnaire

The validity of questionnaire was conducted by designing psychometric properties, including designing items, face validity, content validity, construct validity, stability, and internal consistency.

To evaluate the qualitative face validity, the questions were checked with 10 research experts in terms of level of difficulty, appropriateness, or ambiguity. For quantitative validity, first for each of the questionnaire expressions, the 5-point Likert scale is quite important (score 5) not at all important (score 1). Then, 20 research experts were asked to review and grade each of the phrases based on their experiences. Questions with an item impact index greater than 1.5 were identified and retained as appropriate for subsequent analysis.

Content validity index (CVI) and content validity ratio (CVR) methods were used to evaluate the content validity of the research. For this purpose, a 76-item questionnaire was provided to 15 experts. CVI scores ranged from 0.6 to 1.00 and the average of this

index was 0.82 and was reported at the desired level. Also, CVR scores in the review of experts' opinions ranged from 0.86 to 1.00 and the average CVR scores were equal to 0.91 and were obtained at the appropriate level.

Procedure

The present study was conducted to present a validated and comprehensive assessment for Iranian EFL teachers' performance evaluation in the Iranian context. Critics to the current evaluation system in Iran (Navidinia et al., 2015) believe that it is essential to have a comprehensive instrument for evaluating Iranian EFL teacher performance. Therefore, assessment shows the domains and characteristics of an effective evaluation tool. First of all, the requirements of the evaluation were considered from the related literature. According to Kennedy (2010), English teacher performance evaluation should seek to perceive the reason of teacher activities in the classroom. In other words, teacher development requires to understand the exact teaching method that teachers need in their classrooms and pay attention to the teaching context completely. In the qualitative section of the study, the Delphi technique was used. It is a process used to arrive at a group opinion or decision by surveying a panel of experts. Several rounds of questionnaires are sent out to the group of experts, and the anonymous responses are aggregated and shared with the group after each round (Chisa, 2020).

Requirements for designing a teacher evaluation model

In order to design a proposed model for evaluating English language teachers, first, its design requirements were extracted from the background related to teacher evaluation and quality teaching, educational documents of the country, and the opinions of relevant experts. These requirements are the following:

1. Evaluating the performance of English teachers should seek to understand the reason and concept of activities of teachers in the classroom (Kennedy, 2010). Since one of the goals of teacher performance appraisal is to help develop their professionalism, any feedback to the teacher without understanding the reason for the activities and a thorough understanding of how he or she teaches has little effect on his or her professional development.
2. Educational supervision (evaluation of teachers' performance) should be done by experienced people (training guide official and public in the Islamic Republic of Iran) and those who are fluent in English. Obviously, principals who are not fluent in English will not be able to evaluate effectively the performance of English teachers (Gnanasekaram & Hoar, 2021).
3. Evaluation of English teachers' performance should be done in the classroom environment and with full attention to the existing context. Based on new social and cultural theories, teaching and learning of teachers is highly dependent on the existing context (Chen & Wang, 2021, April).
4. Because all the methods used to evaluate teachers' instruction—have aspects positive and negative, none of them alone can give a true picture of teachers' performance; therefore, if possible, it is necessary to consider several different methods to evaluate the performance of teachers (Almutairi & Shraid, 2021).

5. Designing a teacher evaluation model should be based on careful consideration of relevant backgrounds and opinions of researchers, teachers, and experts as well as pay attention to the requirements of relevant national documents (Hosseini et al., 2021).

Data collection

In the first round of qualitative study, 25 experts were asked to answer open-ended questions on the important characteristics of an English teacher and its duty in Iranian schools and institutes. Participants were asked to refer to as many suggestions as they can. Then, the answers of the respondents were analyzed and the repetitive items were omitted. Also, the answers were categorized and common themes were searched. Analyzing answers to open-ended questions resulted in the following themes: 1—language skills including listening, speaking, writing, and reading and language subskills including grammar and vocabulary, 2—assessment, 3—content knowledge, 4—classroom management, 5—professional development. Using emerged themes, a new questionnaire including 100 questions was designed and measured on a linear scale. After the distribution of questionnaires and calculating the frequency of each item of the questionnaire, the results and each individual's answer were resent to the panel to rate the questions based on their importance for the third and last round. In the last phase of the qualitative study, before performing the Delphi technique, any option was set to have all three of the following properties and is included in the final template as an approved option:

1—The mean 4 and more, 2—standard deviation less than 1, 3—less than 10% of the participants do not answer to the item due to its vagueness or any other reason were considered to decide on the final questions and components of the questionnaire. The researchers prepared the evaluation which includes a questionnaire. The final version of the questionnaire included 76 questions which were rated based on the Likert scale. Then, the designed questionnaire was used in the quantitative section of the study.

In the quantitative study, the researchers designed an online questionnaire using Google survey (https://docs.google.com/forms/d/1p_Z1tx5ArmwuGGmoAEFqzipXtr7OWbDMXRdZZzFZaWE/edit) and sent the link of the questionnaire to Iranian EFL teachers using email and social media such as Linked in. Then, the results from the questionnaires were compared based on gender and teaching experience of the teachers. In fact, Iranian EFL teachers evaluated their own performance. Then, their self-evaluation was compared. In the first comparison, female and male teachers' responses were compared. Secondly, the questionnaire results were compared between teachers with 5 years of teaching experience and teachers with more than 5-year experiences. Previous studies (Gatbonton, 1999; Tsui, 2003) reported that that experienced teachers are considered those who have been teaching English more than 5 years (Rezaee & Sarani, 2017).

Data analysis

In this study, a mixed-method design including qualitative and quantitative study was conducted. In the qualitative study, experts' and teachers' responses to open-ended questions were analyzed and emerged items were investigated using the Delphi

technique. In this phase, descriptive statistics including mean, standard deviation, and percent were used to gain the final questions of the assessment. In addition, the reliability of the questionnaire was calculated by Cronbach alpha.

In a quantitative study, results from the questionnaire study were submitted to the SPSS software. Then, the differences of the results based on respondents' gender and teaching experience were analyzed using independent samples *t* test.

Results

Descriptive results

In order to know more about research variables, it is necessary to describe data before analyzing them. Descriptive statistics including central and distribution indices were introduced for each research hypothesis separately. Also, indices such as demographic characteristics, frequency, and percentage were presented in this section.

Demographic characteristics of a sample

In this section, demographic characteristics of sample based on gender were presented first. Table 1 is related to the descriptive statistics of respondents based on their gender differences.

As Table 1 shows, there were 71 female and 79 male participants in the present study. More than half of the respondents, i.e., 53% were male, and the rest of the participants (47%) were female.

As Table 2 indicated, the number of participants with less than 5 years teaching was 83 while the number of participants with more than 5 years of teaching 67. Forty-five percent of participants had less than 5 years of teaching experience while 55% had more than 5 years of teaching experience.

Delphi technique

As it was mentioned before, the present study was conducted to assess in the form of a questionnaire for evaluating Iranian EFL teachers' performance. To this end, the Delphi technique was used to collect related data and design a questionnaire. In the first phase of the Delphi technique, 25 experts including university professors and experienced instructors in the field of English language teaching were asked to answer five open-ended questions. Questions were mainly related to the performance of English teachers and factors which might be effective in the evaluation of English teacher performance. After collecting information of the open-ended questions, the researchers read the responses to find general themes and categories. After finding and categorizing re-occurred themes, general domains of teacher performance evaluation were emerged. The general domains included language skills, designing and performing instruction,

Table 1 Descriptive statistics of sample based on their gender

| | | Frequency | Percent | Valid percent | Cumulative percent |
|-------|--------|-----------|---------|---------------|--------------------|
| Valid | Female | 71 | 47.3 | 47.3 | 47.3 |
| | Male | 79 | 52.7 | 52.7 | 100.0 |
| | Total | 150 | 100.0 | 100.0 | |

Table 2 Descriptive statistics of sample based on their teaching experience

| Frequency | Percent | Valid percent | Cumulative percent | |
|-----------|---------|---------------|--------------------|-------------------|
| 83 | 55.3 | 55.3 | 55.3 | Less than 5 years |
| 67 | 44.7 | 44.7 | 100.0 | More than 5 years |
| 150 | 100.0 | 100.0 | | Total |

assessment, familiarity with students, content knowledge, classroom management, and professional development.

In the second phase, a questionnaire was designed using emerged domains. The designed questionnaire included 100 questions. After designing a questionnaire, respondents were asked to answer the questionnaire on a Likert scale: 1, not at all important; 2, slightly important; 3, moderately important; 4, very important; 5, absolutely essential. Then, the responses were analyzed and the results were provided in (Additional file 1). Participants rated the questions based on their degree of importance. Also, the mean and standard deviation of each question were calculated. The results were resent to each participant for the final round of the Delphi technique. In the last phase of the Delphi technique, participants were asked to fill in the questionnaire while they could see their own responses and the mean of each question. The results of this phase of the technique was measured according to the following criteria: 1—the mean 4 and more, 2—standard deviation less than 1, 3—less than 10% of the participants do not answer to the item due to its vagueness or any other reason were considered to decide on the final questions and components of the questionnaire. Finally, the last version of the teacher performance evaluation questionnaire was designed (Additional file 1).

According to the results of (Additional file 1), the items which had the mean of more than 4, standard deviation less than 1, and were responded by more than 10% of the participants were included in the final version of the questionnaire. In other words, questions with a mean less than 4 and a standard deviation above 1 were removed from the questionnaire. According to the mentioned criteria, the final questionnaire included 76 questions which are presented in (Additional file 1).

Another descriptive statistic of this section was related to the mean, standard deviation, maximum, and minimum of questionnaire categories as follows:

In Table 3, the value of skewness and kurtosis are in the range of -2 - 2 ; therefore, it can be concluded that the distribution of the data is almost normal.

Reliability statistics

In order to estimate the reliability of the questionnaire, Cronbach's alpha was used. Table 4 indicates the results of the test.

Inferential statistics

This section of the results includes two parts. First, normality of the variables is examined by Kolmogorov-Smirnov test. Then, research hypotheses are tested. Delphi technique and independent t test are used to test the research hypotheses.

Table 3 Central and distribution indices of questionnaires variables

| Descriptive statistics | | | | | | | |
|--------------------------------------|----------|--------|----------------|---------|---------|----------|----------|
| | <i>N</i> | Mean | Std. deviation | Minimum | Maximum | Skewness | Kurtosis |
| Listening | 150 | 21.19 | 2.68 | 12 | 25 | -0.48 | 0.39 |
| Speaking | 150 | 25.19 | 3.31 | 14 | 30 | -0.44 | -0.06 |
| Reading | 150 | 17.03 | 2.47 | 7 | 20 | -0.84 | 0.91 |
| Writing | 150 | 17.15 | 2.23 | 10 | 20 | -0.67 | 0.06 |
| Vocabulary | 150 | 21.01 | 3.10 | 11 | 25 | -0.64 | 0.09 |
| Grammar | 150 | 16.61 | 2.48 | 8 | 20 | -0.67 | 0.32 |
| Designing and performing instruction | 150 | 87.78 | 11.92 | 47 | 105 | -0.72 | 0.26 |
| Assessment | 150 | 33.33 | 4.84 | 17 | 40 | -0.64 | 0.39 |
| Familiarity of students | 150 | 29.23 | 4.03 | 18 | 35 | -0.33 | -0.44 |
| Content knowledge | 150 | 20.83 | 3.02 | 15 | 25 | -0.26 | -0.88 |
| Classroom management | 150 | 34.52 | 3.93 | 21 | 40 | -0.54 | -0.03 |
| Professional development | 150 | 323.87 | 37.14 | 209 | 385 | -0.58 | 0.02 |
| Valid <i>N</i> | 150 | | | | | | |

Normality of variables

To check the normality of research variables using Kolmogorov-Smirnov test, the null hypothesis (H0) indicates that the variable is normal and the opposite hypothesis (H1) indicates that the variable is not normal. Table 5 is the output of this test:

Regarding the significant values of Table 5 which were $p > 0.05$, it is concluded that the null hypothesis, i.e., normal distribution of the population is confirmed.

Testing research hypotheses

H01: There are not any statistically significant differences among Iranian EFL teachers' performance based on their gender.

Because of normality of variable, independent sample *T* test was used. The results are presented in Table 6.

The results of Levene's test showed that the significance level of the variables is higher than 0.05, and the assumption of equality of variances is accepted. According to the *T* test results in Table 7, the value of Sig for listening, speaking, reading, writing, vocabulary, grammar, performing instruction, assessment, familiarity of students, content knowledge, classroom management were 0.48, 0.48, 0.31, 0.31, 0.43, 0.44, 0.78, 0.78, 0.73, 0.73, 0.45, 0.45, 0.68, 0.67, 0.49, 0.49, 0.72, 0.72, 0.68, 0.68, 0.30, 0.29, and 0.49 respectively. All of the Sig values are more than 0.05; therefore, there are not any significant differences among male and female participants' performance in each mentioned category. In the following, the results of H02 are presented.

Table 4 Reliability of the questionnaire

| Cronbach's alpha | <i>N</i> of items |
|------------------|-------------------|
| .982 | 76 |

Since $\alpha = 0.982 > 0.07$, the questionnaire has high reliability

Table 5 Kolmogorov-Smirnov test for normality of variables

| Variable | Kolmogorov-Smirnov Z | Asymp. Sig. (2-tailed) | Test result |
|--------------------------------------|----------------------|------------------------|---|
| Listening | 1.243 | 0.091 | Acceptance of the H ₀ hypothesis |
| Speaking | 1.316 | 0.062 | Acceptance of the H ₀ hypothesis |
| Reading | 1.335 | 0.058 | Acceptance of the H ₀ hypothesis |
| Writing | 1.235 | 0.098 | Acceptance of the H ₀ hypothesis |
| Vocabulary | 1.306 | 0.066 | Acceptance of the H ₀ hypothesis |
| Grammar | 1.224 | 0.105 | Acceptance of the H ₀ hypothesis |
| Designing and performing instruction | 1.152 | 0.140 | Acceptance of the H ₀ hypothesis |
| Assessment | 1.283 | 0.074 | Acceptance of the H ₀ hypothesis |
| Familiarity of students | 1.071 | 0.201 | Acceptance of the H ₀ hypothesis |
| Content knowledge | 1.327 | 0.061 | Acceptance of the H ₀ hypothesis |
| Classroom management | 1.294 | 0.070 | Acceptance of the H ₀ hypothesis |
| Professional development | 1.013 | 0.256 | Acceptance of the H ₀ hypothesis |

H02: There are not any statistically significant differences among Iranian EFL teachers’ performance and years of teaching experience. Due to the normality of data, independent samples *t* test was used. The results are presented in Table 8.

As Table 8 indicated above, the result of the Levene’s test confirmed equality of variances, i.e., Sig = 0.89 > 0.05. In addition to this, the *T* test showed that Sig = 0.13 > 0.05. Thus, there is not any significant difference between less experienced teachers (less than 5 years) with experienced teachers (more than 5 years). In the following, the results of the second hypothesis are presented for each category of the questionnaire.

In Table 9, the results of Leven’s test showed that the significance level of the variables is higher than 0.05 and the assumption of equality of variances is accepted. Also, the results of *t*

Table 6 Independent sample *T* test of teacher performance based on gender

| | | Levene’s test for equality of variances | | <i>t</i> test for equality of means | | | | | | |
|-------------|-----------------------------|---|-------|-------------------------------------|---------|-----------------|-----------------|-----------------------|---|--------|
| | | <i>F</i> | Sig. | <i>t</i> | df | Sig. (2-tailed) | Mean difference | Std. error difference | 95% confidence interval of the difference | |
| | | | | | | | | | Lower | Upper |
| Performance | Equal variances assumed | 2.599 | 0.109 | 0.149 | 148 | 0.882 | 0.910 | 6.094 | -11.134 | 12.952 |
| | Equal variances not assumed | | | 0.151 | 146.865 | 0.880 | 0.910 | 6.031 | -11.010 | 12.828 |

Table 7 Independent sample *T* test of teacher performance based on gender for questionnaire categories

| | | Levene's test for equality of variances | | t test for equality of means | | | | | | |
|---|-----------------------------|---|------|------------------------------|-----------|-----------------|-----------------|-----------------------|---|------|
| | | <i>F</i> | Sig. | <i>t</i> | <i>df</i> | Sig. (2-tailed) | Mean difference | Std. error difference | 95% Confidence interval of the difference | |
| | | | | | | | | Lower | Upper | |
| Listening | Equal variances assumed | 0.05 | 0.82 | -0.71 | 148.00 | 0.48 | -0.31 | 0.44 | -1.18 | 0.55 |
| | Equal variances not assumed | | | -0.71 | 141.57 | 0.48 | -0.31 | 0.44 | -1.19 | 0.56 |
| Speaking | Equal variances assumed | 0.99 | 0.32 | -1.03 | 148.00 | 0.31 | -0.55 | 0.54 | -1.62 | 0.51 |
| | Equal variances not assumed | | | -1.03 | 148.00 | 0.31 | -0.55 | 0.54 | -1.62 | 0.51 |
| Reading | Equal variances assumed | 0.91 | 0.34 | -0.79 | 148.00 | 0.43 | -0.32 | 0.40 | -1.12 | 0.48 |
| | Equal variances not assumed | | | -0.78 | 139.19 | 0.44 | -0.32 | 0.41 | -1.12 | 0.49 |
| Writing | Equal variances assumed | 0.54 | 0.47 | -0.28 | 148.00 | 0.78 | -0.10 | 0.37 | -0.83 | 0.62 |
| | Equal variances not assumed | | | -0.29 | 146.72 | 0.78 | -0.10 | 0.37 | -0.83 | 0.62 |
| Vocabulary | Equal variances assumed | 0.57 | 0.45 | -0.34 | 148.00 | 0.73 | -0.17 | 0.51 | -1.18 | 0.83 |
| | Equal variances not assumed | | | -0.34 | 146.82 | 0.73 | -0.17 | 0.51 | -1.18 | 0.83 |
| Grammar | Equal variances assumed | 2.58 | 0.11 | 0.75 | 148.00 | 0.45 | 0.31 | 0.41 | -0.50 | 1.11 |
| | Equal variances not assumed | | | 0.76 | 146.63 | 0.45 | 0.31 | 0.40 | -0.49 | 1.10 |
| Designing and performing instruction | Equal variances assumed | 4.45 | 0.04 | 0.42 | 148.00 | 0.68 | 0.82 | 1.96 | -3.04 | 4.68 |
| | Equal variances not assumed | | | 0.42 | 144.96 | 0.67 | 0.82 | 1.93 | -2.99 | 4.63 |
| Assessment | Equal variances assumed | 2.66 | 0.11 | 0.69 | 148.00 | 0.49 | 0.54 | 0.79 | -1.02 | 2.11 |
| | Equal variances not assumed | | | 0.69 | 147.05 | 0.49 | 0.54 | 0.78 | -1.01 | 2.09 |

Table 7 Independent sample *T* test of teacher performance based on gender for questionnaire categories (Continued)

| | | Levene's test for equality of variances | | t test for equality of means | | | | | | |
|---------------------------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|-------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean difference | Std. error difference | 95% Confidence interval of the difference | |
| | | | | | | | | | Lower | Upper |
| Familiarity of students | Equal variances assumed | 0.52 | 0.47 | 0.36 | 148.00 | 0.72 | 0.24 | 0.66 | -1.07 | 1.55 |
| | Equal variances not assumed | | | 0.36 | 147.99 | 0.72 | 0.24 | 0.66 | -1.06 | 1.54 |
| Content knowledge | Equal variances assumed | 0.73 | 0.40 | -0.42 | 148.00 | 0.68 | -0.21 | 0.50 | -1.18 | 0.77 |
| | Equal variances not assumed | | | -0.42 | 147.30 | 0.68 | -0.21 | 0.49 | -1.18 | 0.77 |
| Classroom management | Equal variances assumed | 1.78 | 0.18 | 1.05 | 148.00 | 0.30 | 0.67 | 0.64 | -0.60 | 1.94 |
| | Equal variances not assumed | | | 1.06 | 147.28 | 0.29 | 0.67 | 0.64 | -0.59 | 1.93 |
| Professional development | Equal variances assumed | 2.66 | 0.11 | 0.69 | 148.00 | 0.49 | 0.54 | 0.79 | -1.02 | 2.11 |

test showed that there is a statistically significant difference between the variables writing, vocabulary, performing instruction, assessment, familiarity with students, and content knowledge (*P* value < 0.05) in the two groups of experienced and inexperienced teachers.

Discussion

Qualitative study

First of all, this study was conducted to evaluate Iranian EFL teachers' performance. Results from the Delphi technique in different rounds showed that the most important

Table 8 Independent sample *T* test of teacher performance based on teaching experience

| | | Levene's test for equality of variances | | t test for equality of means | | | | | | |
|-------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|-------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean difference | Std. error difference | 95% confidence interval of the difference | |
| | | | | | | | | | Lower | Upper |
| Performance | Equal variances assumed | 0.02 | 0.89 | -1.54 | 148.00 | 0.13 | -0.68 | 0.44 | -1.54 | 0.19 |
| | Equal variances not assumed | | | -1.56 | 145.51 | 0.12 | -0.68 | 0.43 | -1.53 | 0.18 |

Table 9 Independent sample *T* test of teacher performance based on teaching experience for questionnaire categories

| | | Levene's test for equality of variances | | t test for equality of means | | | | | | |
|--------------------------------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|-------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean difference | Std. error difference | 95% confidence interval of the difference | |
| | | | | | | | | Lower | Upper | |
| Listening | Equal variances assumed | 0.02 | 0.89 | -1.54 | 148.00 | 0.13 | -0.68 | 0.44 | -1.54 | 0.19 |
| | Equal variances not assumed | | | -1.56 | 145.51 | 0.12 | -0.68 | 0.43 | -1.53 | 0.18 |
| Speaking | Equal variances assumed | 0.12 | 0.73 | -1.10 | 148.00 | 0.28 | -0.59 | 0.54 | -1.67 | 0.48 |
| | Equal variances not assumed | | | -1.09 | 136.29 | 0.28 | -0.59 | 0.55 | -1.68 | 0.49 |
| Reading | Equal variances assumed | 1.40 | 0.24 | -1.49 | 148.00 | 0.14 | -0.60 | 0.40 | -1.40 | 0.20 |
| | Equal variances not assumed | | | -1.52 | 147.98 | 0.13 | -0.60 | 0.39 | -1.38 | 0.18 |
| Writing | Equal variances assumed | 1.47 | 0.23 | -2.69 | 148.00 | 0.01 | -0.96 | 0.36 | -1.67 | -0.26 |
| | Equal variances not assumed | | | -2.74 | 147.78 | 0.01 | -0.96 | 0.35 | -1.66 | -0.27 |
| Vocabulary | Equal variances assumed | 0.02 | 0.89 | -2.74 | 148.00 | 0.01 | -1.36 | 0.50 | -2.35 | -0.38 |
| | Equal variances not assumed | | | -2.75 | 143.73 | 0.01 | -1.36 | 0.50 | -2.34 | -0.38 |
| Grammar | Equal variances assumed | 0.01 | 0.91 | -1.80 | 148.00 | 0.08 | -0.73 | 0.40 | -1.52 | 0.07 |
| | Equal variances not assumed | | | -1.81 | 144.71 | 0.07 | -0.73 | 0.40 | -1.52 | 0.07 |
| Designing and performing instruction | Equal variances assumed | 0.22 | 0.64 | -2.32 | 148.00 | 0.02 | -4.47 | 1.93 | -8.28 | -0.66 |
| | Equal variances not assumed | | | -2.33 | 143.63 | 0.02 | -4.47 | 1.92 | -8.27 | -0.67 |
| Assessment | Equal variances assumed | 0.94 | 0.33 | -2.58 | 148.00 | 0.01 | -2.01 | 0.78 | -3.56 | -0.47 |
| | Equal variances not assumed | | | -2.63 | 147.86 | 0.01 | -2.01 | 0.76 | -3.53 | -0.50 |

Table 9 Independent sample *T* test of teacher performance based on teaching experience for questionnaire categories (*Continued*)

| | | Levene's test for equality of variances | | t test for equality of means | | | | | | |
|---------------------------|-----------------------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|-------|
| | | F | Sig. | t | df | Sig. (2-tailed) | Mean difference | Std. error difference | 95% confidence interval of the difference | |
| | | | | | | | | | Lower | Upper |
| Familiarity with students | Equal variances assumed | 0.21 | 0.65 | -2.22 | 148.00 | 0.03 | -1.45 | 0.65 | -2.74 | -0.16 |
| | Equal variances not assumed | | | -2.23 | 143.38 | 0.03 | -1.45 | 0.65 | -2.74 | -0.16 |
| Content knowledge | Equal variances assumed | 0.00 | 0.98 | -1.84 | 148.00 | 0.07 | -0.91 | 0.49 | -1.88 | 0.07 |
| | Equal variances not assumed | | | -1.85 | 142.58 | 0.07 | -0.91 | 0.49 | -1.88 | 0.06 |
| Classroom management | Equal variances assumed | 2.67 | 0.11 | -1.05 | 148.00 | 0.29 | -0.68 | 0.64 | -1.95 | 0.59 |
| | Equal variances not assumed | | | -1.03 | 128.77 | 0.30 | -0.68 | 0.66 | -1.98 | 0.62 |

and effective domains in evaluating the performance of English teachers are language skills, performing instruction, assessment, familiarity with students, content knowledge, classroom management, and professional development. Language skills (mean 4.02) mainly included questions related to the ability of teachers in teaching listening, speaking, writing, and reading. Also, vocabulary and grammar teaching was found to be important areas for evaluation of teacher performance. This finding was consistent with previous studies in the literature (Al-Thumali, 2011; Çelik et al., 2013; Khaksefidi, 2015; Richards, 2010; Van Driel & Berry, 2012; Zaminy & Ahangar, 2016). In these studies, like finding of this study, it was emphasized that effective English teachers need to be competent in teaching language proficiency (Richards, 2010) or they should possess linguistic knowledge (Khaksefidi, 2015). Performing instruction (mean 4.12) was the second theme found important for evaluating teachers. According to this part, it was expected that an English teacher provides a proper lesson plan, be familiar with different teaching methods, use creative ways, encourage creativity in thinking, promote problem-solving, be able to use various ways of error corrections, facilitate learning, and so on. Al-Khairi (2015) mentioned this and stated that giving clear instruction is important to English teacher quality. Also, Khaksefidi (2015) mentioned creativity, lesson plan, error correction as important characteristics of an effective English teacher in line with what mentioned above. Assessment (mean 4.440) was also found to be one of the important factors which should be taken into account in English teacher performance evaluation. It was found that teachers should know about different ways of student assessment and use assessment in a way that helps learners in better learning.

These findings are in agreement with (Al-Khairi, 2015; Al-Thumali, 2011; Barnes & Lock, 2013; Çelik et al., 2013; Khaksefidi, 2015; Soleimani & Zanganeh, 2014; Van Driel & Berry, 2012; Wei & Hui, 2019; Zamani & Ahangari, 2016; Yan et al., 2021). As it was found in this study, they believed that testing, scoring, assessing, and giving feedback are factors that help learners in better learning and effective teachers should know how to run them in their classroom.

Another important domain referred to familiarity with students (4.24). Results indicated that knowing students including personality traits, their learning problems, their mental status, their background, and their mental, physical, emotional, and social qualities is an important factor which teachers should prepare themselves with it. This finding further supports the idea by Çelik et al. (2013) who believed the behavior of teachers at all educational levels assists learners to gain academic achievements. This finding was also in parallel with David and Macayan (2010) in which they believed that teacher performance is an indicator of teacher effectiveness which does not include instruction and other approaches used in the classroom, i.e., it includes the teacher role as a whole person in the community of the classroom. Similarly, Zamani and Ahangari (2016) showed that effective teachers pay attention to students' opinions, let them express themselves, and help them to be confident. The fifth domain that proved to be important in qualitative study was the role of content knowledge competency.

It was mentioned frequently that it is necessary for English teachers to have a high command of textbooks, course materials, and side materials related to teaching. This finding was in line with Al-Khairi (2015), Al-Thumali (2011), Blömeke et al. (2011), Heller et al. (2012), Hill et al. (2005), Khaksefidi (2015), Samson and Collins (2012), and Wilson (2013). Mentioned studies have shown that teachers' content knowledge play a key role in students' progress.

Similarly, Kálmán et al. (2020) emphasized that the professional development of teachers including content knowledge and subject matter mastery could result in student achievement positively. The next important category for teacher performance evaluation was classroom management (mean 4.40). Richards and Schmidt (2013) have mentioned this category as one of the areas that preservice teachers should be trained on. Teacher Education (2015) classified classroom management as one of the approaches and techniques that teachers should have enough knowledge about. Richards (2010) believes that teachers along with pedagogical and content knowledge should possess skills such as curriculum planning, assessment, reflective teaching, classroom management, teaching children, and teaching the four skills. In addition, Çelik et al. (2013) plus Al-Wreikat et al. (2010) highlighted the role of classroom management and mentioned it as a key skill for teachers.

Last but not the least domain in this study was professional development (mean 4.52) which is an important criterion for teacher evaluation. It was found that a teacher's own evaluation and reflection on the evaluation of her/his performance are the important factors that develop teacher performance and result in better learning in the language classroom. Johnson and Golombek (2011) stated that teacher development influences student improvement and academic achievement positively. Also, Borg (2018) believed that self-evaluation is a technique in which teachers evaluate their own competency and skill. It can be beneficial for teacher development and professionalism. It causes teachers to engage in the evaluation process and take responsibility as

professional individuals. Taut and Sun (2014) believed that if self-evaluation results were used for teacher development purposes, it would generate more accurate data.

Teacher Education similarly stated that teacher professional development is an important factor in teacher evaluation and teacher effectiveness. Although it was discussed that all of the domains found in the present study were in line with previous studies, it is worth noting that some of the above studies find the importance of these factors by investigating students' viewpoint (Al-Khairi, 2015; Çelik et al., 2013; Wei & Hui, 2019; Zamani & Ahangari, 2016) and some others explored students' and teachers' opinions at the same (Al-Wreikat et al., 2010; Khaksefidi, 2015; Soleimani & Zanganeh, 2014). In the following section, results related to the quantitative study were discussed.

Quantitative study

Here, the results from the questionnaire study are discussed. Iranian EFL teachers were asked to fill in a teacher performance evaluation questionnaire including 76 questions under the following categories: language skills, designing and performing instruction, assessment, familiarity with students, content knowledge, classroom management, and professional development. The data collected from the questionnaire were compared and analyzed based on gender differences and teaching experience among participants. In the following, results related to each variable are discussed separately.

Teacher performance and gender difference

The first research question in the quantitative study addressed to: Are there any statistically significant differences among Iranian EFL teachers' performance based on their gender?

Results from independent samples, *T* test showed that there is not statistically a significant difference between male and female Iranian EFL teachers' performance. Not only there was no significant difference between male and female teacher performance in general but also there were not any significant differences between male and female teacher performance based on each category of questionnaire (language skills, designing and performing instruction, assessment, familiarity with students, content knowledge, classroom management, and professional development). In other words, both male and female teachers showed that language skills including listening, speaking, writing, and reading along with subskills vocabulary and grammar are important and they showed that high competency in these skills using different methods and techniques. In other domains like instruction methods, using creative ways, assessing learners, knowing students and their personality, having a good command of course materials, managing classroom, and working on their own development, findings showed that both male and female teachers highly agreed with these factors in their teaching performance.

A possible explanation for this might be that EFL teachers evaluated their own performance and it seems possible that they consider their performance effective. Therefore, another possible explanation might be that teachers have not evaluated their own performance accurately. Borg (2018) believed that one shortcoming of teacher self-evaluation is accurate assessment of teacher competence by themselves. Similarly, Davis et al. (2006) showed that results from self-evaluation among professional teachers suggest that there are few numbers of teachers who are capable of accurate self-assessing,

especially those who are not skillful enough and those who are so self-confident. Regarding the finding of teacher performance evaluation by gender and relationship of it with previous studies, it should be mentioned that most of the previous studies (Al-Khairi, 2015; Çelik et al., 2013; Johnson et al., 2021; Wei & Hui, 2019; Zamani & Ahan-gari, 2016) in the literature investigated teacher performance and effectiveness from university and school students point of view and those few studies (Al-Wreikat et al., 2010; Khaksefidi, 2015; Merati et al., 2021; Soleimani & Zanganeh, 2014) which have evaluated teacher performance mostly focused on a specific area of teacher performance such as pedagogical knowledge, pre-service teacher training program, effectiveness of teacher training courses rather than teacher performance.

It can be said that data from studies on older language learners indicate that gender has a very weak effect on foreign language learning in middle age and adulthood, and there is no difference between male and female language learners (Rahmawati, 2021). Thus, different effects of the effect of gender on foreign language learning do not appear much in the elementary school, but in the middle school, they appear in favor of girls and disappear in the later stages, and as a result, the role of gender at different ages is different (Bensalem, 2021). Girls are more successful than boys in learning a foreign language, and from the age of 13 onwards, the ratio of unsuccessful boys to unsuccessful girls is very significant, and as a result girls who are still learning a foreign language like France at the age of 16 are significantly better than the boys act (Ha et al., 2021).

Based on some studies, since men and women are not alike in beliefs, desires, behaviors, and personalities, the fundamental communication differences between them that result from this dissimilarity are also evident (Hosseini, 2016). Knowing and being aware of individual differences, talents, motivational factors, how to encourage and punish, the personality of individuals, and similar issues among men and women, can help those involved in education to better understand people and their needs (Irajzad et al., 2017). Because without knowing the psychological and behavioral aspects of people in the organization, one cannot expect the organization to move toward success and prosperity. Based on the findings, such a claim can be made that female teachers compared to male teachers, show a more positive feeling and as a result, they are more inclined to communicate with the learners.

Researchers have not treated gender in much detail from the point of psychological perspective (Yazdanpour, 2015). Female instructors were more sensitive to receiving non-positive, non-conditional negative and conditionally negative stroke than male instructors. It also showed that there is a significant difference between the emotion of male and female teachers, which arises as a result of receiving positive non-verbal stroke, not receiving stroke and negative conditional verbal stroke (Almutairi & Shraid, 2021).

However, despite this difference, gender differences fade at older ages and the rate of foreign language learning become the same for males and females.

Teacher performance and teaching experience

Findings related to the second research question analysis is discussed. The second research question addressed to: Are there any statistically significant differences among Iranian EFL teachers' performance and years of teaching experience?

Generally, results indicated that there is no significant difference between experienced (more than 5 years of teaching) and less experienced teachers (less than 5 years experienced teaching). In other words, both groups indicated almost equal performance in their teaching process. One possible explanation for this might be that teachers evaluated their own performance; therefore, it is possible that both experienced and less experienced teachers rate their performance highly. Although it was seen that there is no significant difference between the performance of experienced and less experienced teachers, there were some slight differences between the performance of experienced and less experienced teachers in some categories of the questionnaire. Results from the independent samples *T* test showed that both experienced and less experienced teachers rated their performance almost equally in domains of listening, speaking, reading, content knowledge, classroom management, and professional development. However, the values of $\text{Sig} = 0.01, 0.01, 0.02, 0.01, 0.03 < 0.05$ for writing, vocabulary, designing and performing instruction, assessment, and familiarity with students showed that there are differences between the performance of experienced and less experienced teachers.

It seems possible that these results are due to teachers' experiences in different areas of teaching. Another possible explanation could be related to the amount of content and pedagogical knowledge which each teacher might have. In other words, teacher education and degree (whether they have a bachelor or master degree) might influence teacher performance along with their experience which has not been taken into account in this study. So, it is possible that teacher performance differences in these domains of the study be rooted in their degree level or education.

The findings of the present study are similar to (Argudo Alfah et al., 2021; Firoozjahan-tigh et al., 2021) in that they showed a teacher having more than 15 years of experience with bachelor degree and teachers having less than 5-year experience with a master degree had a high positive performance in using quality standards in their teaching experience in comparison to teachers having less than 5-year experience with bachelor degree. Also, the finding of the study relating to performing instruction domain was in agreement with another study conducted by Al-Thumali (2011).

Similarly, he found that there was a significant difference between teachers' performance in planning and management of learning based on their teaching experience. It was found that teachers with less than 15-year experience had higher performance in comparison to teachers with more than 15 years of experience of teaching. However, it should be taken into account that the years of experience in the mentioned study was different from this study. According to the findings of the present study, it appears that experience and gender are two factors that represent valuable information to researchers and teachers in teaching English and evaluation of English teachers.

From the perspective of ideal identity items, it should be said that almost all respondents agreed on these items and the increase in teaching years only in most of them agree with the change in the creation of statistical tests confirm this finding and perhaps it can be said that there is no significant difference between teaching years and teaching experience. Another explanation for this may be that the increase in the number of years of teaching only affects the knowledge and belief system that is one of the components that shape identity from Clark (2008)'s perspective, although this claim needs further research.

In order to identify the requirements of designing the target model for teacher performance evaluation, the present study was limited to investigation of teacher evaluation background, teachers and experts' opinions, and requirements of educational documents in the country. Another limitation is related to data collection method which is questionnaires. This method of data collection with small number of the participants affects generalizability of data. It should be noted that using convenience sampling method is another limitation of the study.

Conclusion

This study aimed to evaluate Iranian EFL teachers' performance. Also, it was intended to compare EFL teachers' performance based on two criteria of their gender difference and teaching experience (less than 5-year experience with more than 5-year experience teachers). To the first aim, the Delphi technique was used to develop a questionnaire. In the first round, 25 experts including university lecturers and experienced instructors in the field of English teaching were asked to answer open-ended questions regarding important issues in the evaluation of an English teacher. Then, the related themes emerged. Using emerged themes, a questionnaire including 100 questions were designed and measured on a linear scale (1 = not important to 5 = absolutely essential). After calculating the frequency of each item, the results were resent to the panel to rate the questions.

In the last phase, three criteria including 1—the mean 4 and more, 2—standard deviation less than 1, 3—less than 10% of the participants do not answer the item were considered to decide on the final questions and components of the questionnaire. Returning to the question posed at the beginning of this study, it is now possible to show the most important domains for evaluation of Iranian English teachers' performance. The most obvious finding from this study was that language skills, performing instruction, assessment, familiarity with students, content knowledge, classroom management, and professional development were considered important categories that should be taken into account in the teacher evaluation process.

The final version of the questionnaire was used to examine teacher performance based on their gender and teaching experience. The questionnaire was distributed using Google form. One hundred fifty questionnaires were filled correctly and analyzed using SPSS 22. It was found that gender cannot be considered a key factor in teacher performance. In other words, it was found that male and female teachers rated their performance almost equally. The last part of the study which compared teacher performance based on teaching experience showed slight differences in some domains of teaching between experienced and less experienced teachers.

Abbreviations

JPS: Jackson Public Schools; ESL/EFL: English as a second/foreign language; TTU: Teacher Training University

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40468-021-00140-7>.

Additional file 1: Appendix A. Second Phase of Delphi Technique: Descriptive Statistics of Experts Opinions on Questionnaire. Appendix B. The Results of Questionnaire Based on Defined Criteria. Appendix C. Final Version of Questionnaire.

Acknowledgements

Not applicable.

Authors' contributions

Both authors of the research had more or less equal contributions in the process of conception, design, acquisition of data, analysis, and interpretation of data. They have all been involved in revising the manuscript critically to the same extent. All take public responsibility for the whole content. All are equally accountable for all aspects of the work. All authors read and approved the final manuscript.

Authors' information**Siros Izadpanah**

1. Ph.D in TEFL and Faculty member of the university from 1997
2. Publishing 70 articles in different journals in the world
3. <https://publons.com/dashboard/summary/>
4. Web of Science Researcher ID AAG-8887-2020
5. https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=siros+izadpanah&oq=sir

Haleh Mashhadlou

1. MA in English language teaching
2. She has 10-year experience in teaching English language

Funding

The study has not received any funding.

Availability of data and materials

The data will be available upon requesting from corresponding author—cyrosizadpanah@yahoo.com

Declarations**Competing interests**

The authors declare that they have no competing interests.

Received: 29 June 2021 Accepted: 11 September 2021

Published online: 02 November 2021

References

- Aghaalkhani, H., & Maftoon, P. (2018). English teacher education programs and professionalism: The case of Iranian novice/experienced teachers. *Research in English Language Pedagogy (RELP)*, 6(1), 78–94. <https://doi.org/10.30486/RELP.2018.538776>.
- Al-Khairi, M. (2015). Qualities of an ideal English language teacher: A gender-based investigation in a Saudi context. *Journal of Education and Practice*, 6(15), 88–98 Retrieved from <https://files.eric.ed.gov/fulltext/EJ1079989.pdf>.
- Almutairi, T. S., & Shraid, N. S. (2021). Teacher evaluation by different internal evaluators: Head of departments, teachers themselves, peers and students. *International Journal of Evaluation and Research in Education*, 10(2), 588–596.
- Al-Thumali, A. D. A. (2011). Evaluating EFL intermediate teachers' performance in the light of quality standards in Saudi Arabia (Master's thesis, Taif university, Taif, Saudi Arabia). Retrieved from <https://files.eric.ed.gov/fulltext/ED525726.pdf>
- Al-Wreikat, Y. A. A. S., Abdullah, B., & Kabilan, M. K. (2010). An evaluation of Jordanian EFL teachers' in-service training courses teaching techniques effectiveness. *English Language Teaching*, 3(4), 18–27 Retrieved from <http://files.eric.ed.gov/fulltext/EJ1081970.pdf>.
- Arabzadeh, Z. (2016). Comparison of differences in performance evaluation of EFL teachers by students and teachers' self-evaluation (Unpublished Master Thesis, University of Tabriz, Tabriz, (X)). Retrieved from [https://ganj.\(X\)doc.ac.ir/viewer/6036038b358ace8320e186ee7610c6bd](https://ganj.(X)doc.ac.ir/viewer/6036038b358ace8320e186ee7610c6bd)
- Argudo Serrano, J. C., Fajardo-Dack, T. M., Abad Celleri, M. P., & Cabrera-Tenecela, H. P. (2021). Students' perceptions on their EFL teacher efficacy: A study on EFL teachers' language proficiency and their self-efficacy. *MEXTESOL Journal*, 45(1), 99–121.
- Atai, M. R., & Mazlum, F. (2013). English language teaching curriculum in (X): Planning and practice. *The Curriculum Journal*, 24(3), 389–411. <https://doi.org/10.1080/09585176.2012.744327>.
- Baniasad-Azad, S., Tavakoli, M., & Ketabi, S. (2016). EFL teacher education programs in (X). *The absence of teachers' involvement*, 19(2), 61–86. Retrieved from <https://ijal.khu.ac.ir/article-1-2728-en.pdf>. <https://doi.org/10.29252/ijal.19.2.61>.
- Barabino, S., Aragona, P., di Zazzo, A., Rolando, M., & with the contribution of selected ocular surface experts from the Società Italiana di Dacriologia e Superficie Oculare. (2021). Updated definition and classification of dry eye disease: Renewed proposals using the nominal group and Delphi techniques. *European Journal of Ophthalmology*, 31(1), 42–48, DOI: <https://doi.org/10.1177/1120672120960586>
- Barnes, B. D., & Lock, G. (2013). Student perceptions of effective foreign language teachers: A quantitative investigation from a Korean University. *Australian Journal of Teacher Education*, 38(2), 18–36. <https://doi.org/10.14221/ajte.2013v38n2.2>.
- Belton, I., Wright, G., Sissons, A., Bolger, F., Crawford, M. M., Hamlin, I., ... Vasilichi, A. (2021). Delphi with feedback of rationales: How large can a Delphi group be such that participants are not overloaded, de-motivated, or disengaged? *Technological Forecasting and Social Change*, 170, 198.

- Bensalem, E. (2021). Classroom enjoyment and anxiety among Saudi undergraduate EFL students: Does gender matter? *Vigo International Journal of Applied Linguistics*, 1(18), 9–34.
- Borg, S. (2006). The distinctive characteristics of foreign language teachers. *Language Teaching Research*, 10(1), 3–31. <https://doi.org/10.1191/1362168806lr182oa>. <https://doi.org/10.1191/1362168806lr182oa>.
- Borg, S. (2018). *Teacher evaluation: Global perspectives and their implications for English language teaching: A literature review*. Delhi: British Council.
- Burroughs, N., Gardner, J., Lee, Y., Guo, S., Toutou, I., Jansen, K., & Schmidt, W. (2019). *A review of the literature on teacher effectiveness and student outcomes*, In *Teaching for Excellence and Equity* (pp. 7-17) (). Cham: Springer.
- Çelik, S., Arikan, A., & Caner, M. (2013). In the eyes of Turkish EFL learners: What makes an effective foreign language teacher? *Online Submission*, 20(2), 287–297 Retrieved from <https://files.eric.ed.gov/fulltext/ED549671.pdf>.
- Challob, A. I. (2021). The effect of flipped learning on EFL students' writing performance, autonomy, and motivation. *Education and Information Technologies*, 3, 1–27. <https://doi.org/10.1007/s10639-021-10434-1>.
- Chen, X., & Wang, Y. (2021). The existing problems of English teachers' classroom evaluation language and its countermeasures in junior middle school. In *2021 2nd Asia-Pacific Conference on Image Processing, Electronics and Computers* (pp. 162-165).
- Chisa, K. (2020). Shaping policy and practice in South Africa's heritage institutions through expert opinion: A Delphi method study. In *Handbook of Research on Connecting Research Methods for Information Science Research* (pp. 192-207). IGI Global.
- Cullen, J. B., Koedel, C., & Parsons, E. (2021). The compositional effect of rigorous teacher evaluation on workforce quality. *Education Finance and Policy*, 16(1), 7–41.
- Danielson, C., & McGrill, T. L. (2000). Teacher evaluation to enhance professional practice. *Learning and Instruction*, 44, 74–86.
- Danielson, C. (2001a). "New trends in teacher evaluation". *Educational Leadership*. 58. pp. 12-15
- Danielson, C. (2001b). "New trends in teacher evaluation". *Educational Leadership*. 58. pp. 12-15.
- Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E., & Perrier, L. (2006). Accuracy of physician self-assessment compared with observed measures of competence: A systematic review. *Journal of the American Medical Association*, 296(9), 1094–1102. <https://doi.org/10.1001/jama.296.9.1094>.
- Donaldson, M. L., & Firestone, W. (2021). Rethinking teacher evaluation using human, social, and material capital. *Journal of Educational Change*, 1, 1–34.
- Firoozjahantigh, M., Fakhri Alamdari, E., & Marzban, A. (2021). Investigating the effect of process-based instruction of writing on the IELTS writing task two performance of Iranian EFL learners: Focusing on hedging & boosting. *Cogent Education*, 8(1), 1881202. <https://doi.org/10.1080/2331186X.2021.1881202>.
- Flostrand, A., Pitt, L., & Bridson, S. (2020). The Delphi technique in forecasting—A 42-year bibliographic analysis (1975–2017). *Technological Forecasting and Social Change*, 150, 119773. <https://doi.org/10.1016/j.techfore.2019.119773>.
- Galluzzo, G. R. (2005). Performance assessment and renewing teacher education the possibilities of the NBPTS standards. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 78(4), 142–145. <https://doi.org/10.3200/TCHS.78.4.142-145>.
- Gatbonton, E. (1999). Investigating novice and experienced esl teacher differences: Implications for teacher training. *Japan Association for Language Teaching (JALT) Journal*, 23(6), 83-104.
- Gnanasekaram, G., & Hoar, A. (2021). Audit on availability, quality and frequency of clinical and educational supervision. *BJPsych Open*, 7(1), S79–S79. <https://doi.org/10.1192/bjo.2021.249>.
- Gruetzemacher, R., Dörner, F. E., Bernaola-Alvarez, N., Giattino, C., & Manheim, D. (2021). Forecasting AI progress: A research agenda. *Technological Forecasting and Social Change*, 170, 195. <https://doi.org/10.1016/j.techfore.2021.120909>.
- Ha, X. V., Murray, J. C., & Riazi, A. M. (2021). High school EFL students' beliefs about oral corrective feedback: The role of gender, motivation and extraversion. *Studies in Second Language Learning and Teaching*, 11(2), 235–264. <https://doi.org/10.14746/ssllt.2021.11.2.4>.
- Heller, J. I., Daehler, K. R., Wong, N., Shinohara, M., & Miratrix, L. W. (2012). Differential effects of three professional development models on teacher knowledge and student achievement in elementary science. *Journal of Research in Science Teaching*, 49(3), 333-362. 10.1002/tea.21004
- Hill, H. C., Rowan, B., & Ball, D. (2005). Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42(1), 371–406. <https://doi.org/10.3102/00028312042002371>.
- Hosseini, S. (2016). *EFL teachers' perceptions of stroke and their application in class: A qualitative study*. (Unpublished Master's thesis). Mashhad, Iran: Ferdowsi University of Mashhad.
- Hosseini, S. M., Shahhosseini, M. A., & Ahmadzahi, A. (2021). Developing and designing competency-based performance management elevation model on high school teacher in Iran: A mixed-method approach. *Journal of Management and Planning In Educational System*, 13(2), 195–236. <https://doi.org/10.52547/mpes.13.2.195>.
- Irajzad, F., Pishghadam, R., & Shahriari, H. (2017). Examining the stroking behavior of English, Persian, and Arabic school teachers in Iran: A mixed-methods study. *International Journal of Instruction*, 10(2), 219–236. <https://doi.org/10.12973/iji.2017.10114a>.
- Irvine, J. (2018). Relationship between teaching experience and teacher effectiveness: Implications for policy decisions. *Journal of Instructional Pedagogies*, 22(1), 1–19 Retrieved https://www.researchgate.net/publication/331635835_Relationship_between_teaching_experience_and_teacher_effectiveness_implications_for_policy_decisions.
- Jackson Public Schools (JPS). (2013). *Teacher performance evaluation handbook*. Stakeholders committee members. Retrieved from <https://www.nctq.org/dmsView/70-07>
- Johnson, E. S., Zheng, Y., Crawford, A. R., & Moylan, L. A. (2021). The relationship of special education teacher performance on observation instruments with student outcomes. *Journal of learning disabilities*, 54(1), 54–65. <https://doi.org/10.1177/0022219420908906>.
- Johnson, K. E., & Golombek, P. R. (2011). *Research on second language teacher education: A sociocultural perspective on professional development*. New York: Taylor & Francis. <https://doi.org/10.4324/9780203844991>.
- Jones, E., Bergin, C., & Murphy, B. (2021). Principals may inflate teacher evaluation scores to achieve important goals. *Educational Assessment, Evaluation and Accountability*, 2, 1–32.
- Jyoti Sankar, P. (2018). Innovations and initiatives in teacher education. *ZENITH International Journal of Multidisciplinary Research*, 8(5), 116–124 Retrieved from <http://www.indianjournals.com/ijor.aspx?target=ijor&volume=8&issue=5&artitle=011>.

- Kálmán, O., Tynjälä, P., & Skaniakos, T. (2020). Patterns of university teachers' approaches to teaching, professional development and perceived departmental cultures. *Teaching in Higher Education*, 25(5), 595–614. <https://doi.org/10.1080/13562517.2019.1586667>.
- Kennedy, M. M. (2010). Approaches to annual performance assessment. In M. M. Kennedy (Ed.), *Teacher Assessment and the Quest for Teacher Quality a Handbook* The American Association of Colleges for teacher Education.
- Khaksefidi, S. (2015). Foreign language teaching in (X): A model for effective EFL teaching in the (X)ian context. *Theory and Practice in Language Studies*, 5(5), 1060–1071. DOI:<https://doi.org/10.17507/tpls.0505.22>
- Khanjani, A., Vahdany, F., & Jafarigohar, M. (2017). EFL teacher education in (X): Does it promote trainees' pedagogical content knowledge? *Research in Applied Linguistics*, 8(2), 159–186 Retrieved from <https://www.sid.ir/FileServer/JE/51013020170208>.
- Lamm, K. W., Randall, N., Lamm, A. J., & Sapp, L. R. (2021). Extension Organizational Strengths: A Delphi Analysis. *Journal of Agricultural Education*, 62(2), 53–65.
- Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, 73(5), 467–482.
- Looney, J. (2011). Developing high-quality teachers: Teacher evaluation for improvement. *European Journal of Education*, 46(4), 440–455. <https://doi.org/10.1111/j.1465-3435.2011.01492.x>.
- Mackey, A., & Gass, S. M. (2015). *Second language research: Methodology and design*. Routledge, New York. <https://doi.org/10.4324/9781315750606>.
- Mahvelati, E. H. (2021). Learners' perceptions and performance under peer versus teacher corrective feedback conditions. *Studies in Educational Evaluation*, 1, 70–97. <https://doi.org/10.1016/j.stueduc.2021.100995>.
- Martin, N. K., Yin, Z., & Mayall, H. (2006). Classroom Management Training, Teaching Experience and Gender: Do These Variables Impact Teachers' Attitudes and Beliefs toward Classroom Management Style? *Online Submission*, 1, 2–15.
- Mazandarani, O., & Troudi, S. (2021). Measures and features of teacher effectiveness evaluation: Perspectives from Iranian EFL lecturers. *Educational Research for Policy and Practice*, 2, 1–24. <https://doi.org/10.1007/s10671-021-09290-0>.
- McLaughlin, B. (1990). The relationship between first and second languages: Language proficiency and language aptitude. *The development of second language proficiency*, 1, 158–178.
- Merati, H., Ghonsooly, B., & Alavi, S. M. (2021). Emotional intelligence and cultural quotient as predictors of pragmatic performance in EFL. *International Journal of Language Studies*, 15(2), 45–68.
- Mukerhji, S. N. (2019). Education. *Journal of Baltic Science Education*, 19(1), 36–49 Retrieved from <https://www.britannica.com/topic/education>.
- Navidinia, H., Reza Kiani, G., Akbari, R., & Ghaffar Samar, R. (2015). EFL teacher performance evaluation in Iranian high schools: Examining the effectiveness of the status quo and setting the groundwork for developing an alternative model. *The International Journal of Humanities*, 21(4), 27–53. Retrieved from <http://hsmmsp.modares.ac.ir/article-27-6297-en.pdf>.
- Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in public health*, 8, 457–479. <https://doi.org/10.3389/fpubh.2020.00457>.
- Pishghadam, R., Derakhshan, A., Zhaleh, K., & Al-Obaydi, L. H. (2021). Students' willingness to attend EFL classes with respect to teachers' credibility, stroke, and success: A cross-cultural study of Iranian and Iraqi students' perceptions. *Current Psychology*, 2, 1–15.
- Rahman, F., Jumani, N. B., Akhter, Y., Chisthi, S. U. H., & Ajmal, M. (2011). Relationship between training of teachers and effectiveness teaching. *International Journal of Business and Social Science*, 2(4), 150–160 Retrieved from http://ijbssnet.com/journals/Vol_2_No_4%3B_March_2011/18.pdf.
- Rezaee, A., & Sarani, A. (2017). Job performance of Iranian English teachers: Do teaching experience and gender make a difference? *Iranian Journal of English for Academic Purposes*, 6(2), 13–22 Retrieved from http://journalscnu.sinaweb.net/article_62750.html.
- Richards, J. C. (2010). Competence and performance in language teaching. *RELC Journal*, 41(2), 101–122. <https://doi.org/10.1177/0033688210372953>.
- Richards, J. C., & Schmidt, R. W. (2013). *Longman dictionary of language teaching and applied linguistics*. Routledge. <https://doi.org/10.4324/9781315833835>.
- Samson, J. F., & Collins, B. A. (2012). Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness. *Center for American Progress*, 3(2), 1–29 Retrieved from <https://files.eric.ed.gov/fulltext/ED535608.pdf>.
- Santiago, P., & Benavides, F. (2009). Teacher evaluation: A conceptual framework and examples of country practices, paper presented at the OECD-Mexico workshop "Towards a Teacher Evaluation Framework in Mexico: International Practices, Criteria and Mechanisms", Mexico City, 1-2 December 2009.
- Scriven, M. (1981). Summative teacher evaluation. In J. Millman (Ed.), *Handbook of Teacher Evaluation*. Beverly Hills: Sage Publications Inc..
- Soleimani, H., & Zanganeh, H. (2014). A survey on knowledge, dispositions, and performances of pre-service EFL teacher training programs in (X). *Procedia-Social and Behavioral Sciences*, 98(33), 1802–1810. <https://doi.org/10.1016/j.sbspro.2014.03.609>.
- Tsui, A. (2003). *Understanding expertise in teaching: Case studies of second language teachers*. Cambridge University Press.
- Van Driel, J. H., & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational researcher*, 41(1), 26–48. <https://doi.org/10.3102/0013189X11431010>.
- Vinhais, P., & Abelha, M. (2015). Supervision and appraisal of foreign language teachers' performance. *Procedia-Social and Behavioral Sciences*, 174(4), 783–790. <https://doi.org/10.1016/j.sbspro.2015.01.615>.
- Waters, T.; R.J. Marzano & B. McNulty (2003). *Balanced leadership: What 30 years of research tells us about the effect of leadership on student achievement*. Retrieved January 20, 2011, from Mid-Continent Research for Education and Learning web site: http://eric.ed.gov/ERICDOCS/data/ericdocs2sql/content_storage_01/0000019b/83/al.pdf
- Wei, W., & Hui, S. K. F. (2019). Evaluating teacher performance in language learning classes: The gap between students and department administrators. *Teachers and Teaching*, 25(4), 486–500. <https://doi.org/10.1080/13540602.2019.1625763>.
- Wilpers, A., Bahtiyar, M. O., Wall, D., Kobler, K., Sadler, L. S., Dixon, J. K., & Kennedy, H. P. (2021). Modified Delphi study on nursing practice and science in fetal care. *Journal of Obstetric, Gynecologic & Neonatal Nursing*, 50(1), 55–67. <https://doi.org/10.1016/j.jogn.2020.09.158>.

- Wilson, S. M. (2013). Professional development for science teachers. *Science*, 340(6130), 310–313. <https://doi.org/10.1126/science.1230725>.
- Yan, Z., Li, Z., Panadero, E., Yang, M., Yang, L., & Lao, H. (2021). A systematic review on factors influencing teachers' intentions and implementations regarding formative assessment. *Assessment in Education: Principles, Policy & Practice*, 2, 1–33.
- Yazdanpour, H. (2015). *Constructing and validating a Teacher Stroke Scale and examining its relationship with burnout (Master's thesis)*. Mashhad, Iran: Ferdowsi University of Mashhad.
- Yazdimoghaddam, H., Samadipour, E., Ghardashi, F., Borzooee, F., Akbarzadeh, R., Zardosht, R., ... Khalili, S. (2021). Designing a comprehensive clinical competency test for operating room technology student: Using Delphi technique and CIPP model evaluation. *Journal of Education and Health Promotion*, 1, 10–34.
- Zamani, R., & Ahangari, S. (2016). Characteristics of an effective English language teacher (EELT) as perceived by learners of English. *International Journal of Foreign Language Teaching and Research*, 4(14), 69–88 Retrieved from https://www.researchgate.net/profile/Arvind_Singh56/post/What_are_the_qualities_of_the_good_English_language_teacher/attachment/5b91801b3843b0067538216f/AS%3A667946112270336%401536262171019/download/1.pdf.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
