Native and Non-native Language Teachers' Perspectives on Teacher Quality Evaluation

Zia Tajeddin
Tarbiat Modares University
<tajeddinz@modares.ac.ir>

Zari Saeedi
Allameh Tabataba'i University
<saeedi.za@gmail.com>

Hamideh Mozaffari
Allameh Tabataba'i University
<hamideh_mozaffari@yahoo.com>

Abstract

Various features of teacher instruction underpin the criteria used for the evaluation of teacher quality. The current study sought to explore whether nativeness/non-nativeness affects the criteria teachers consider for teacher quality evaluation. To this end, the participants were provided with five video clips of teaching, each presenting a 10-min lesson taught in a real classroom environment. They were requested to rate the quality of the teachers and to point out and describe the criteria they used to rate the teachers. Content analysis of the data indicated that preparation, caring, classroom management, and instruction constituted the general criteria the native and non-native teachers employed to evaluate teacher quality. Considerable differences, however, were observed between the two groups regarding a few of the criteria. The native teachers valued teachers' efficient use of learners' L1 more than the non-native teachers, while teachers' linguistic accuracy and fluency of speech were highlighted by more non-native teachers. Besides, issues related to caring, management, and instruction grabbed the attention of both native and non-native teachers, while preparation received substantially less attention. It can be concluded that the use of video-mediated peer observation can provide a platform to uncover the implicit beliefs teachers hold toward teacher quality.

Keywords: teacher quality evaluation, evaluation criteria, native English-speaking teachers, non-native English-speaking teachers

Teacher quality has been long regarded as one of the most significant contributing factors to student learning (Ingersoll, 2020). It is, then, incumbent on those in the educational sphere to be able to evaluate teacher quality most reliably and validly. There has been mounting debate in the literature about whether the quality of a teacher should be measured based on (1) teacher input such as credentials, (2) teaching process, i.e., teacher practices often measured through classroom observations, (3) teaching output, i.e., the impact of teaching on student achievement, or a combination of these elements (Stronge et al., 2011). Although teacher evaluation systems in several educational contexts still rely upon teacher input and output, it is teaching performance that is now regarded as the most significant indicator of teacher quality (Gergen & Gill, 2020; Olsen, 2021). This, in turn, has resulted in a serious question: What teacher attributes and practices can best define quality in teachers' performance and hence should be considered in teacher evaluation? Some attempts have been made by recent scholars to introduce such features (Danielson, 2013; Marzano, 2013; Stronge, 2018). However, it is also necessary to know how teachers, given their recent inclusion and recognition as plausible evaluators in various contexts, evaluate teacher quality.

Against this backdrop, it seems crucial to explore what criteria teachers employ to evaluate teacher quality. Prior research has addressed the impact that variables such as teaching experience and discipline have on the criteria that teachers apply for evaluating teachers (e.g., Chand Dayal & Alpana, 2020; McCoy & Lynam, 2021; Nguyen & Pham, 2020; Torres et al., 2017). Torres et al. (2017), for instance, investigated the criteria teachers from different educational disciplines employ for evaluating their peers. The results indicated no major differences between the teachers from various disciplinary fields. The teachers from all fields consistently emphasized communication and interaction with students and students' engagement in the class. Scant research in L2 education, however, has examined how teacher nativeness/non-nativeness might influence teacher evaluation.

The debate over native versus non-native English-speaking teachers has long dominated the English language teaching profession. Non-native speaker teachers are often defined against native speaker teachers, implying a privileged status bestowed on teachers for whom English is the native language. This privilege rests on the argument that a native speaker teacher can best model the fluent and appropriate use of language, is well aware of the cultural connotations of language, and can best judge the acceptability of any piece of language. Phillipson (1992, p.195), however, called this "the native speaker fallacy" and highlighted that none of these abilities are beyond the capability of a non-native speaker of a language and can be well developed through teacher education. Such criticisms along with a sudden increase in the number of non-native speakers of English worldwide and the need for effective intercultural interaction led ELT scholars, almost from the early 1980s, to question the adequacy of the century-old native speaker model. This is aligned with the rise of English as an International Language, English as a Lingua Franca, and World Englishes, which emphasize the importance of non-native varieties of English, the significance of intelligibility in recently prevalent NNS-NNS communication, and the non-exclusive ownership of English. These new strands challenged the raciolinguistic ideologies pervading the ELT industry, that is the idea that competent teachers are white native speakers of English (Ramjattan, 2019; Rosa & Flores, 2017). Despite this acknowledgment in both professional and academic literature, non-native speaker teachers of English continue to be disprivileged in several educational contexts (Kubota, 2022; Rosa, 2016; Sekaja et al., 2022). Given the significance of this issue and the fact that teacher evaluation criteria, as primary indicators of teacher quality, have not been

examined from native and non-native teachers' perspectives, the study aimed to explore the criteria native and non-native English language teachers apply for teacher quality evaluation.

Literature Review

Teacher Quality Evaluation

Teacher quality evaluation refers to a process employed to measure teacher effectiveness which primarily attempts at enhancing student learning (Danielson, 2001). The method of conducting this seemingly easy process, however, has undergone major shifts over time. Measures of student learning, particularly student achievement scores, had been traditionally relied on by various stakeholders to decide upon the quality of teachers in several educational contexts. However, this method, which is commonly referred to as value-added modeling (VAM), has been the target of several criticisms as to its reliability and validity (Leon & Thomas, 2015). Theoretical and empirical evidence suggests that VAM scores suffer from inconsistency, are influenced by the characteristics of students, and do not show any significant relationship with other more accepted methods of teacher evaluation (Amrein-Beardsley & Geiger, 2020; Briggs & Domingue, 2011; Collins, 2014).

In view of these criticisms, researchers began to reconsider the century-old question of how teachers' quality can be reliably and validly evaluated. The most plausible response, according to the literature (Gergen & Gill, 2020), lies primarily in the very performance of teachers. Instead of employing student achievement to judge teacher quality, evaluation systems should rely on teachers' practices so as to foster student learning. The method of using teachers' performance as the criterion for examining the quality of their work has been called the performance-based/standards-based evaluation model. Although there exists no one best method to evaluate teachers, literature offers positive evidence as to the validity and reliability of employing teacher performance for evaluating teacher quality (e.g., Coady et al., 2020; Patrick et al., 2020).

Teacher Criteria for Teacher Quality Evaluation

Any teacher evaluation system, irrespective of the educational level, attempts to achieve two main goals: to exhibit accountability to various stakeholders and to foster teacher professional development (Lillejord & Bort, 2020). Not all evaluation programs, however, manage to achieve these purposes. An effective teacher evaluation requires careful planning of the whole process, namely the frequency, nature, and use of evaluation, as well as the composition and qualifications of evaluators (Marzano et al., 2020). The evaluators, therefore, play a vital role in the success of any teacher quality evaluation system. Given skepticisms about the validity of principals'/administrators' evaluation of teachers and concerns about the reliability of relying on judgments from a single evaluator, it is recommended in the literature that multiple evaluators inform the evaluation of teacher quality, with experienced teachers as a key evaluator (Darling-Hammond, 2015; Gordon, 2020). Although top-down approaches toward teacher quality evaluation still dominate ELT in several contexts (where evaluation is done by administrators, policy makers, etc.), teachers have been recently acknowledged as plausible evaluators either in the form of self-evaluation or peer evaluation (Borg & Edmett, 2018).

Whatever their advantages and disadvantages, teacher self-evaluation and peer observation are currently employed in several educational programs worldwide. Teachers are often provided with a pre-specified set of criteria based on which they should evaluate their own or their peers'

teaching (O'Leary, 2020). Probably due to the popularity of such methods, we know little about the criteria that teachers themselves utilize if required to evaluate teaching practices. Meanwhile, recent studies have mostly turned toward exploring the benefits, challenges, and obstacles involved in the process of self-/peer-evaluation (e.g., Hendry et al., 2021; Visone, 2022). It should be noted that the majority of the studies in this area, as the following review reveals, have been conducted in general education contexts and in very few cases in ELT. Nonetheless, teachers, according to the available evidence, often use three kinds of criteria for evaluating their or other teachers' practices: (a) learners' classroom behavior and experience, (b) teachers' classroom practices, and (c) the output of teaching; that is, student learning. Among these, student learning experience has proved to be a key criterion that teachers employ for evaluation. In his study, Madsen (2005) relied on interviews to identify the criteria that novice teachers typically employ for evaluating their practices. Student classroom participation, interest, and autonomy have been reported as the major criteria that the teachers use to evaluate themselves. Nevertheless, teachers' actual self-evaluation practices need a more valid picture of the issue. To develop a self-evaluation questionnaire, more recently, Torres et al. (2017) addressed the issue from the perspective of lecturers from different educational disciplines. Student-related behavior such as participation recurred in the data. Torres et al., however, did not specify the teaching experience of the participants to ensure whether their perspectives were valid enough to inform a questionnaire. The extent to which the students showed interest in the lesson was further considered for self-evaluation by the primary student teachers in McCoy and Lynam's (2021) study.

Student learning has further emerged, albeit in a few studies, as a major criterion for self-evaluation among both pre-service and in-service teachers. The inexperienced and award-winning faculty members (Dunkin, 1991), the student elementary teachers (Dunkin et al., 1996), and both novice and expert primary teachers (Madsen, 2005) have referred to student immediate learning of a particular lesson, the ability to apply the learning in a later lesson, and long-term learning for evaluating themselves. Besides, many pre-service teachers in Olson et al.'s (2004) study used student learning measured through achievement scores to judge the quality of their teaching. Success in formative and summative tests was also reported as a self-evaluation criterion by a few teachers in Kremer and Ben-Peretz (2006) who used interviews to examine the issue among 90 teachers from various levels and schools in the northern district of Israel. However, this criterion has received scant attention in more recent studies addressing the issue, probably due to the recent recognition that student learning is affected by several factors other than teachers' pedagogical knowledge and expertise.

In addition to student-related factors, teacher attributes and skills have recurrently been mentioned as one of the evaluation criteria in prior studies. Prospective primary teachers in Madsen's (2005) study considered their affective (e.g., patience), cognitive (e.g., content knowledge), and behavioral (e.g., classroom management) features important for evaluating their practices. However, the author, as mentioned earlier, employed interviews rather than a more revealing method such as examining teachers' actual self-evaluation practices. This methodological limitation was improved by subsequent studies that used video-based reflection/observation to investigate the issue. Nonetheless, except for Gröschner et al.'s (2018) study which did not specify the participants' teaching experience, the teachers inquired were restricted to pre-service ones. Although not directly exploring teacher self-evaluation criteria, Ibrahim et al.'s (2012) study of teachers' reflective practices revealed that content knowledge, management of time and student behavior, and instructional issues (e.g., clarity of

explanations) entailed the main features the chemistry student teachers considered in reflecting on the quality of their practices. In Gröschner et al.'s (2018) study, the ability to engage students in learning was the major issue on which the teachers relied for self-evaluation. More recently, using a video-mediated approach, the secondary pre-service teachers in Chand Dayal and Alpana's (2020) study employed teaching-related issues such as planning and classroom management, as well as teacher personal characteristics such as confidence and enthusiasm, to evaluate their teaching. Similarly, the primary student teachers in McCoy and Lynam's (2021) study valued both affective and behavioral aspects of their practices. The key evaluation criteria included managing time, maintaining discipline, instructional clarity, monitoring learning, creating an interactive environment, engaging students, and encouraging.

These criteria also emerged in peer observations, with studies incorporating experienced in-service teachers. Hammersley-Fletcher and Orsmond (2005), for instance, examined the observation experience of teachers with a range of experience across two schools during 2002-2003 and found that the evaluation concentrated on mechanics of teaching (e.g., lesson structure) and student experience of learning (the degree to which they enjoyed learning). Promoting and encouraging student participation and engagement through activities such as group discussion entailed the key criteria the student teachers in Harford and MacRuairc's (2008) and the in-service teachers in Shortland's (2010) and Hendry et al.'s (2014) research employed to evaluate their peers. Teachers' pedagogical practices (e.g., monitoring) further recurred in Torres et al.'s (2017) investigation of the issue across different educational disciplines. To develop a peer observation instrument, Dillon et al. (2020) further explored criteria STEM faculty members from a university in Portland employed to evaluate peers. The use of group work and student engagement were the most important areas for the faculty. Similarly, in Hendry et al.'s (2021) more recent study, although focusing on the advantages of peer observation, teachers' skills in engaging students and creating a motivating learning environment were the major quality indicators experienced physics academics used in peer observation.

Although not directly addressing the criteria teachers apply for teacher evaluation, the findings from a few studies suggest that ESL teachers draw on some common attributes/skills to judge their own/their peers' practices. To investigate peer observation opportunities for professional learning, Sarfraz (2019), for instance, examined experienced non-native ESL academics' observation practices. Teachers' content knowledge and teaching practices, as the observation notes revealed, constituted two general criteria the participants highly valued, although they emphasized teachers' mastery of subject matter over their pedagogical practices arguing that it will foster learner enthusiasm in the classroom. In a similar vein, Nguyen and Pham's (2020) exploration of the problems 10 English lecturers encountered in conducting observation of their colleagues in a university in Vietnam indicated that teachers' knowledge of the subject and their pedagogical methods were used as two significant criteria for teacher evaluation. More recently, Yuan et al. (2022) explored the extent to which pre-service non-native language teachers' reflective practices can contribute to their professional learning. Instructional strategies, teacher language, classroom interaction, and catering for student diversity entailed the major instructional criteria used in their reflections. Although these studies are illuminating, none of them investigated the issue in relation to native speaker teachers.

Other frequent criteria that teachers, according to the literature, use to evaluate their teaching, include judgments and comments of others such as students, colleagues, supervisors, and

parents gained either through formal evaluation instruments or through some informal feedback. Approximately half of the pre-service and experienced teachers in Macleod (1988), several novice teachers and a few award-winning professors in Dunkin (1991), all the student teachers in Dunkin et al. (1996), and several faculty members in Wood and Su (2017) used feedback from external sources including supervisors, principals, colleagues, and student parents in evaluating themselves. A recent meta-review by Harrison et al. (2022) further reported student feedback as a key source of some teacher evaluations.

Notwithstanding this body of research that has addressed the criteria that teachers with differing educational backgrounds, teaching experience, and from varying disciplines employ in teacher evaluation, scant research in L2 education has problematized nativeness in teacher quality evaluation. Given that the criteria teachers employ in evaluation reflect their perceptions of what constitutes teacher quality and that teacher beliefs, whether explicitly or implicitly held, influence their practices, an exploration of native and non-native teachers' evaluation criteria can contribute to a better understanding of the debate over native/non-native English speaker teachers. By demonstrating the extent to which native and non-native teachers' criteria converge/diverge, the results can provide some evidence to support/refute the traditional dichotomy between native and non-native teachers. The study, therefore, aimed to explore criteria native and non-native English language teachers apply for teacher quality evaluation. To fulfill the purpose of this research, the study addressed the following research question: What criteria do native and non-native English language teachers apply for teacher quality evaluation?

Method

Participants

Two groups of teachers took part in the study: native teachers (i.e., teachers who speak English as their first language/mother tongue) and non-native teachers (i.e., teachers for whom English is the second/foreign language). In total, 50 native (male = 26, female = 24) and 50 non-native teachers (male = 21, female = 29) participated in the video-mediated peer observation on a voluntary basis and conducted the observation. They were selected based on convenience sampling, that is, according to their availability and willingness to cooperate. The non-native participants included Persian-speaking EFL teachers who taught at private English institutes across the country where English is learned as a foreign language. The native participants were ESL teachers teaching English at language institutes in a range of contexts including the United States, the United Kingdom, and Australia.

The participants had different educational backgrounds, ranging from B.A. to Ph.D., in English-related fields including Teaching English as a Foreign/Second Language (native = 20, non-native = 25), English Literature (native = 4, non-native = 10), Linguistics (native = 6, non-native = 1), Translation (native = 1, non-native = 9), and Foreign Language Studies (native = 1). In a few cases, their education was in non-English fields such as Philosophy (native = 3), Education (native = 6), Psychology (native = 1), Engineering (non-native = 2), Architecture (non-native = 1), Biology (non-native = 1), Physics (non-native = 1), Finance (native = 1), Theology (native = 1), Business (native = 1), Tourism (native = 1), Law (native = 1), Distance Learning (native = 1), Criminal Justice (native = 1), and Psycholinguistics (native = 1). A number of participants (N = 7, NN = 15) further held CELTA. Moreover, they had varying teaching experience ranging from 1 to 33 years. Following Gatbonton (2008), those teachers

who had less than two years of experience were regarded as novice and those having at least five years as experienced. Table 1 displays participants' demographic information.

Table 1. Participants' Demographic Information

		Frequency	Percentage
Country of origin	Iran	50	50%
	US	19	19%
	UK	17	17%
	Australia	14	14%
Gender	Male	47	47%
	Female	53	53%
Field of study	English-related fields	77	77%
	Non-English fields	23	23%

Research Design and Instrument

The current study explored the criteria that native and non-native English teachers employ to evaluate teacher quality. To identify these criteria and compare them across native and non-native speaker teachers of English, a qualitative research design was employed to gather data on the participants' behavior, in this case teacher evaluation practice. Classroom observation has been the most widely used technique for peer evaluation. However, given the foreign language context of Iran where the study was conducted, it was almost infeasible to gain access to and have native speaker teachers observe classrooms. The study, hence, relied on video-mediated observation as the main instrument to answer the research question. To this end, the participants were provided, by the third author, with five video clips of teaching. Each video captured a 10-min lesson presented in a real classroom environment.

The participants were given an open-ended structured evaluation form that contained the following sections to facilitate the process of observation: (1) a section that explained the procedure for observing and evaluating each lesson (using examples), (2) a section on major characteristics of each lesson including lesson focus, learners' proficiency level and age range, cultural context (ESL or EFL), and classroom setting (monolingual or multilingual), and (3) some space for writing the evaluations of each lesson.

The teaching videos selected for the current study included classrooms in both ESL and EFL contexts and both monolingual and multilingual settings. The learners were from a variety of age range (i.e., child, adolescent, and adult) and at different proficiency levels (i.e., beginner, lower intermediate, and intermediate). The lessons had varying teaching focus (particularly, grammar and vocabulary), and the teachers were both native and non-native English speakers. Each lesson is described below:

Video 1: A grammar lesson focusing on the phrase "I like…". The learners are beginners, the class is monolingual (Chinese), and the context is ESL, Australia.

Video 2: A vocabulary lesson focusing on the vocabulary of emotion. The learners are beginners, the class is monolingual (Turkish) and the context is EFL, Turkey.

Video 3: A grammar lesson focusing on simple present tense. The learners are lower intermediate, the class is monolingual, and the context is EFL, Pakistan.

Video 4: A grammar lesson focusing on adverbs of frequency. The learners are intermediate, the class is multilingual, and the context is ESL, the UK.

Video 5: A grammar lesson focusing on past perfect. The learners are intermediate, the class is multilingual, and the context is ESL, the UK.

Data Collection and Analysis

Data collection began by inviting the teachers to participate in the study. The native teachers were invited to take part in the study online, via advertising through social networking including LinkedIn and Instagram. The non-native teachers, however, were recruited either in person at foreign language institutes in Iran or online through the same methods. In addition to obtaining informed consent from individual participants, they were informed about the main goal of the research and were assured that the names would be kept confidential and anonymous. To this end, native and non-native teacher participants in the present study are named T1-T50. The participants were further informed that they could withdraw from the study at any time. The participants were then provided with five video clips of teaching through either Skype or Telegram and were requested to reflect on the quality of the teachers. That is, the teachers were asked to rate the quality of the teachers from ineffective to highly effective (ineffective, developing, effective, highly effective), and then to point out and describe the criteria they used to rate the teachers.

The data which included the notes from video-mediated observations were subjected to qualitative analysis, in particular content analysis. Following Creswell (2015), all the observation data were initially read several times to gain an overall sense of the data. The relevant codes were then derived. Finally, the codes that shared similar characteristics were organized into categories and then into themes. In particular, coding was conducted "inductively using three levels of *open coding*, *axial coding*, and *selective coding*" (Riazi, 2016, p. 37). Initially, the third author read the transcribed data several times, segmented the meaningful utterances, and wrote a label for each of them (open coding). Some of the extracted codes included "effective seating arrangement", "efficient use of the board", "efficient use of the available technological advances", and "proper use of the available realia in the class". Then, the extracted codes were compared and merged into broader categories in similar cases (axial coding). For example, the aforementioned codes were reduced to the category of "managing physical space". Finally, the relevant categories were grouped to come across more abstract themes (selective coding). More precisely, the above category, along with other emerged categories of "managing class time", "managing instructional strategies", and "managing student behavior" clustered under the theme of "classroom management" as each constitutes a key aspect of managing classrooms. Importantly, the specific categories and themes emerged through the dynamic synthesis of both the literature review (e.g., Danielson, 2013; Marzano, 2013; Stronge, 2018) and the data.

During content analysis of the data, it was found that some of the teacher quality criteria the participants employed were either too general or vague. As a form of member-checking the data, we, then, returned to the participants and asked for an explanation/clarification. Following content analysis, the emerged criteria were compared across native and non-native teachers. Furthermore, to ensure reliability in coding, the inter-coder reliability was calculated. Two

independent coders (the third author and a Ph.D. holder in Teaching English as a Foreign Language) coded around 10% of the data selected randomly, as recommended in the literature. Cohen's Kappa was next employed to calculate reliability coefficients. Cohen's Kappa measures inter-coder consistency and considers values of above .80 as extremely reliable, values of above .60 as sufficiently reliable, and values less than .40 as lacking reliability. Since the calculated coefficient proved to be 0.71, the coding was considered reliable.

Findings

Based on the content analysis, the teachers' criteria for evaluation were placed into four major themes. Aligned with the literature on teacher quality evaluation reviewed above, the criteria clustered under the themes of preparation, caring, classroom management, and instruction. First, knowledge of content and teaching methodology, required for effective preparation, are among the two major repertoires that teachers require to be prepared for teaching (Danielson, 2013, Marzano, 2013, Stronge, 2018). Second, teachers who attempt to create an enjoyable, comfortable, and respectful classroom environment, who show patience, enthusiasm, and fairness, and who constantly encourage their learners indeed demonstrate their ability to care for their learners (Danielson, 2013). Third, making efficient use of classroom time, maintaining order and discipline in the classroom, orchestrating the instructional strategies, and organizing physical space are regarded as the key elements of classroom management (Danielson, 2013; McLeod et al., 2003). Finally, communicating effectively with learners, engaging learners, monitoring their understanding and providing them with constructive feedback, presenting a well-organized instructional unit, and making sure to suit the instruction to learners' proficiency level all relate to the efficiency of instruction (Danielson, 2013; Marzano, 2013).

Table 2 shows the distribution of each theme and category among the native and non-native teachers. Both native and non-native teachers, as the table displays, considered certain qualities related to preparation, caring, classroom management, and instruction to evaluate teacher quality, although issues concerning preparation were comparatively less valued by both groups. The importance given to the majority of the evaluation criteria did not turn out to be greatly different among the native and non-native teachers. As can be seen in the table, most native and non-native participants (almost above 70%) considered creating an enjoyable and comfortable classroom environment, managing time, instructional strategies, student behavior, and space, communicating effectively with learners, engaging students in learning, monitoring learning and providing constructive feedback, and presenting a well-structured lesson. Besides, some teachers from both groups pointed to teacher fairness, encouragement, enthusiasm, patience, and appropriacy of the lesson for learners' proficiency level in their evaluations. Few native and non-native teachers, furthermore, employed teachers' knowledge of subject and knowledge of teaching methodology.

Despite this, as Table 2 reveals, slight differences were observed between the two groups regarding a few of the criteria. The native teachers valued creating an environment of respect and providing clear direction for activities more than the non-native teachers. With respect to three instruction-related criteria, however, considerable disparity was observed between the two groups. Table 2 shows that substantially more non-native teachers considered the accuracy and fluency of teachers' English for evaluation than the native ones. Substantially more native teachers, moreover, emphasized the efficient use of L1 in the classroom than the native teachers.

Table 2. A Comparison of Native and Non-native Teachers' Teacher Evaluation Criteria

Theme	Category		Native – Non-native teachers Frequency (percentage)
Preparation	Knowledge of subject		8 (16%) – 11 (22%)
	Knowledge of teaching methodology		3 (6%) – 5 (10%)
Caring	Creating an enjoyable and comfortable classroom environment		36 (72%) – 37 (74%)
	Creating an environment of respect and rapport		27 (54%) – 20 (40%)
	Encouraging learners		25 (50%) – 21 (42%)
	Treating learners fairly		22 (42%) – 17 (34%)
	Demonstrating enthusiasm for teaching		15 (30%) – 12 (30%)
	Exhibiting patience		10 (26%) – 11 (22%)
Classroom management	Managing classroom time		42 (86%) – 38 (76%)
	Managing instructional strategies		39 (78%) – 37 (74%)
	Managing student behavior		42 (84%) – 37 (74%)
	Managing classroom space		48 (96%) – 48 (96%)
Instruction		Providing clear instruction	50 (100%) – 50 (100%)
		Providing clear directions	30 (60%) – 24 (48%)
	Communicating effectively with learners	Establishing clear lesson objectives	9 (18%) – 10 (20%)
		Teacher's linguistic accuracy	18 (36%) – 29 (58%)
		Teacher's fluency of speech	6 (12%) – 18 (36%)
		Providing sufficient instruction	14 (28%) – 15 (30%)
	Engaging students in learning		46 (92%) – 46 (92%)
	Monitoring student learning		37 (74%) – 42 (84%)
	Providing constructive feedback		34 (68%) – 34 (68%)
	Presenting a well-structured lesson		32 (64%) – 32 (64%)
	Making efficient use of learners' L1		24 (66%) – 15 (42%)
	Using an appropriate lesson for learners' proficiency level		15 (30%) – 10 (20%)

Each of the four themes and their constituent categories (the criteria employed by native teachers [NTs] and non-native teachers [NNTs]) are described below.

Preparation

Two teacher characteristics related to preparation were identified in the native and non-native teachers' criteria for teacher quality evaluation. These criteria included:

- Knowledge of subject

- Knowledge of teaching methodology

Teachers' knowledge of content (i.e., English) emerged as a key evaluation criterion, valued, however, by few participants. Overall, 8 native and 11 non-native teachers pointed to this criterion either explicitly or implicitly. In excerpt 1, for instance, an experienced non-native teacher points to a teacher's insufficient knowledge of the difference between main adverbs of frequency.

Excerpt 1

In explaining the difference between rarely and occasionally she [the teacher] said: "Rarely we use occasionally and occasionally we use rarely!!! They're quite similar!" These sentences show that the teacher doesn't know the grammar herself. (T41: NNT)

The second preparation-related criterion which was rarely considered by the teachers of either native or non-native status concerned teacher pedagogical knowledge. In promoting student learning, very few teachers (NT = 3, NNT = 5) emphasized that mastery over the content is not sufficient. English teachers require to demonstrate a good understanding of methodologies that suit teaching English. A beginner non-native teacher highlighted the significance of teachers' knowledge of both subject and pedagogy in the following way (see excerpt 2):

Excerpt 2

She [The teacher] is a native speaker, we can't say that she doesn't know adverbs of frequency. But, I do believe that the problem is she doesn't know how to teach such a subject to her students. (T8: NNT)

Caring

Another set of teacher characteristics/behavior that the participants repeatedly mentioned for evaluating teachers can be clustered under the theme of caring. This entailed the following criteria:

- Creating an enjoyable and comfortable learning environment
- Creating an environment of respect and rapport
- Encouraging learners
- Treating learners fairly
- Demonstrating enthusiasm for teaching
- Exhibiting Patience

One set of these criteria concerned teacher behavior that indicates they care for learners. Creating an enjoyable and comfortable learning environment and creating an environment of respect and rapport are two such criteria recurring in the data. Several native and non-native teachers highlighted the significance of creating an enjoyable and comfortable learning environment for learners so that they could participate and take the potential risks involved in learning a foreign/second language. In other words, 36 native and 37 non-native teachers employed this criterion. For example, an Iranian non-native respondent teaching English for less than 2 years explains how an uncomfortable classroom atmosphere endangers learners' class participation (excerpt 3).

Excerpt 3

When the teacher asked them [the students] if there is any question, they were just looking at their papers, couldn't say or perhaps didn't dare saying anything. They didn't feel relaxed enough (T20: NNT).

Demonstrating respect for the learners in terms of behaviour, attitude, and language was also highlighted by the participants. Almost half of the native and non-native participants (NT = 27, NNT = 20) referred to this issue in their evaluations. In excerpt 4, an experienced non-native teacher refers to this issue, condemning a teacher's disrespectful behavior toward one of his learners.

Excerpt 4

At the beginning, he [the teacher] made a face at one of the students who had a hard time uttering a statement. This sounded a bit disrespectful as others laughed at his gesture. (T36: NNT)

The second set of caring-related criteria entailed personality-related traits characterizing a caring teacher. Native and non-native participants employed, to largely similar extents, the following teacher characteristics to evaluate teachers: giving learners adequate encouragement to actively participate in the classroom (NT = 25, NNT = 21), equal treatment of learners (regardless of their gender, ability, personality or sitting location) (NT = 22, NNT = 17), demonstrating enthusiasm for what they are doing (i.e., teaching) (NT = 15, NNT = 12), and exhibiting due tolerance and patience in any encounter with L2 learners (whether it be dealing with their language errors and mistakes, giving them the time that they need to use the new language or dealing with their misbehavior) (NT = 10, NNT = 11). Excerpt 5 provides an example of a beginner American native teacher's belief that teachers should distribute their attention evenly among all learners in the classroom.

Excerpt 5

The class was small enough in numbers and I did feel at some points more attention was focused on the students in the center of the group and less on the students towards the left of the class. (T12: NT)

Another novice native teacher from the UK, as can be seen in excerpt 6, further emphasizes the significance of teacher patience in the classroom, commenting that one of the teachers he was observing did not exhibit patience in dealing with a learner's error.

Excerpt 6

[The teacher] seems quite impatient with the students. E.g. during the 'I always swim' ordering task, she was not supportive when they gave incorrect answers, just moved on to someone who knew the correct answer. (T9: NT)

Classroom Management

The third set of teacher evaluation criteria that native and non-native participants employed concerned various aspects of classroom management. Teachers' description of management corresponds to the key elements of classroom management, including:

- Managing space

- Managing time

- Managing instructional strategies

- Managing student behavior

Managing space, particularly classroom layout and available physical resources, appeared as the most important skill. Almost all the native and non-native participants (NT = 48, NNT = 48) referred to at least one aspect of space management to evaluate teachers. The physical arrangement of the classroom, to many native and non-native participants, can offer opportunities to foster learning or otherwise hinder it. A non-native teacher having above 10 years' experience of teaching, for instance, explains that seating arrangement can affect classroom interaction and should be considered important for learning (see excerpt 7).

Excerpt 7

Classroom seats were not properly arranged to encourage peer learning and interaction. (T36: NNT)

Effective use of physical resources in the classroom also emerged from the data. Teachers' skill in using the board, the available technological advances, and the accessible realia to advance learning was repeatedly mentioned by both native and non-native teachers. 45 native and 39 non-native teachers considered this aspect of space management. In excerpt 8, for instance, an experienced native teacher from the UK explains how the teacher could have used the board to assist learning in two ways:

Excerpt 8

No (or very limited) use of visual aids-she [the teacher] could have used the board to 1) write down key points/vocabulary; 2) draw pictures to help convey the meaning of what she's teaching. (T34: NT)

Classroom time management was the second important management-related issue that the native and non-native participants considered for evaluation. Appropriate pacing of a lesson, devoting adequate time to each instructional strategy (e.g., presentation, pair/group activities, etc.), recurred in teacher responses. 42 native and 38 non-native participants pointed to this issue. An experienced non-native teacher points to this issue (excerpt 9), arguing that the teacher spent excessive time on instruction, leaving learners little time to practice the point of the lesson.

Excerpt 9

Not providing enough time for each exercise: The teacher's explanation was too long, students should have had more time to go through each exercise, especially because there weren't many chances to practice the new lesson. (T29: NNT)

Teachers' ability in identifying/implementing appropriate instructional strategies to promote learning was the next important evaluation criterion mentioned by more than half of the native and non-native participants: 39 native and 37 non-native teachers. Referring to a teacher's inability in this regard, an experienced native teacher from Australia states (see excerpt 10):

Excerpt 10

Instead of excessive lecturing on the grammar point, the teacher could have provided the learners with opportunists for meaningful and interactive activities to gradually absorb the point. (T43: NT)

In addition to the management of space, time, and instructional strategies, several native and non-native teachers (NT = 42, NNT = 37) valued teachers' skill in managing student behaviour in the classroom. In excerpt 11, an American experienced native teacher, for instance, explains how a teacher used verbal and non-verbal techniques to maintain order in a class of young learners:

Excerpt 11

Children are difficult to control and manage. So, he [the teacher] used various techniques, verbal & gestures, to maintain order. (T26: NT)

Instruction

A third set of evaluation criteria the teachers employed can be grouped together under the theme of instruction. These criteria are listed below:

- Communicating effectively with learners

- Engaging students in learning

- Monitoring learning and providing constructive feedback

- Presenting a well-structured lesson

- Making efficient use of learners' L1

- Using an appropriate lesson for learners' proficiency level

Teachers communicate with learners in the classroom for several related instructional purposes. Clarity and accuracy of certain communicative actions (particularly, providing instruction, establishing objectives, and providing direction for activities) were valued by both native and non-native teachers. Clarity of instruction provided to the learners emerged as the most significant criterion in this regard. All the native and non-native teachers employed this criterion in one way or another. The teachers emphasized the need for grading language, either by simplifying the vocabulary and grammar used to give instruction or by speaking more slowly, an example of which is given by an experienced non-native teacher in excerpt 12:

> **Excerpt 12**
>
> *Her [the teacher's] level adaptation is not efficient. For example, her examples for "Simple Present Tense," consist of structures and grammar higher than "Lower Intermediate.* (T45: NNT)

In addition to language, the content of instruction should be made clear enough for learners to understand. This was regarded essential by approximately all the native and non-native participants. In excerpt 13, an experienced teacher from the US explains how exemplification and explanation helped the teacher make his instruction clear.

> **Excerpt 13**
>
> *Perfect tenses are difficult for learners (and native speakers). So, the key to explaining these was through day to day examples. He [the teacher] then gave accurate explanations.* (T40: NT)

Besides ensuring instructional clarity, teachers should provide learners with sufficient instruction on a given subject matter. This evaluation criterion was stated by almost an equal number of native and non-native participants (NT = 14, NNT = 15). An experienced non-native teacher (see excerpt 14), for example, blames the teacher for not explaining different uses of simple present tense (the focus of the lesson):

> **Excerpt 14**

Simple present has three usages. She [the teacher] taught just one of them and didn't mention the rest of them and she mentioned next week we will study SIMPLE PAST. (T28: NNT)

Clarity of lesson objectives to learners was further regarded as significant, although by few native and non-native participants. Overall, 9 teachers from the native group and 10 teachers from the non-native highlighted the importance of this issue. In excerpt 15, a novice native teacher from the UK emphasizes the point that the instructional goal was not made clear to the learners:

Excerpt 15

Students' talking time limited to repeating sentences from the board but without any explanation on the part of the teacher of why they are doing this or what point they are learning. (T17: NT)

To accomplish the learning objectives, the instructional activities must, moreover, be clearly explained to the learners. 30 native and 24 non-native participants drew on this criterion while evaluating teachers. An American native teacher having above 20 years' experience points to this as a weakness in a teacher's practices (excerpt 16):

Excerpt 16

The teacher gives poorly explained directions for all the class activities. For example, when she tells them to do group work she simply tells them to stand up and find someone to talk to but that it doesn't matter who they talk to. (T28: NT)

The extent to which an English teacher, as the key language model for learners, uses English accurately in the classroom (while accomplishing the three afore-mentioned communicative purposes) proved to be a concern among native and non-native participants, although more highlighted by the non-native teachers than the native ones. In other words, 18 native and 29 non-native teachers considered this issue significant, with some highlighting that the continued inaccurate use of language (in particular, grammar, vocabulary, and pronunciation) by the teacher will result in fossilization. As excerpt 17 demonstrates, a native experienced participant from the UK negatively evaluates a teacher's inaccurate pronunciation of "finish", emphasizing the learning problem it causes (i.e., fossilization):

Excerpt 17

The pronunciation of the teacher's English is not always accurate. The teacher also asked "Finish?" instead of "Finished?" The students look at the teacher as a model of correct English. Such mistakes will be considered true by the students and may lead to fossilization. (T27: NT)

Fluency of speech, besides accuracy, was considered significant by native and non-native teachers, although much more valued by the non-native ones. The concern over fluency, moreover, was less than the accuracy of teachers' language use. In fact, 6 participants from the native group and 18 teachers from the non-native group pointed to this issue.

The following experienced non-native teacher (excerpt 18), for example, considers this point when he blames a teacher's excessive use of learners' L1 on her lack of fluency.

Excerpt 18

The teacher uses students' first language a lot because she's not fluent. (T42: NNT)

Instruction in a second language involves decisions that are primarily aimed at engaging learners in learning. Learner engagement entailed another significant evaluation criterion that caught the attention of several native and non-native participants (NT = 46, NNT = 46). Both native and non-native teachers highlighted the role that classroom practices can play in encouraging and fostering learner engagement. The comment below (excerpt 19), given by a novice non-native teacher, illustrates that the teacher employed a good mix of individual, pair work, and role play to fully engage the learners in the lesson.

Excerpt 19

She [the teacher] keeps students engaged by getting them to answer questions individually and do pair work. Students are also asked to do role play which is very effective, leading to active learning. (T11: NNT)

To assess student learning for the purposes of instruction, teachers must effectively monitor student understanding. Several native and non-native teachers (NT = 37, NNT = 42) considered it essential to properly monitor student understanding both at various stages during instruction and while students are doing individual/group activities and to provide feedback if necessary. A novice native teacher from the US (excerpt 20), for instance, criticizes that the method (i.e., asking direct questions such as "Understand?") the teacher used for monitoring learning does not elicit evidence of learners' understanding:

Excerpt 20

"Understand?" and "Any questions?" do not specifically address potential misunderstandings of students. Asking specific students about specific difficulties may help them and the overall class understand more completely. (T7: NT)

While monitoring learners' understanding during instruction, teachers further need to offer timely and appropriate feedback to learners. Both native and non-native teachers' observation notes valued teachers' ability to provide learners with constructive feedback using various related terms including constructive, good, and appropriate. 34 native and 34 non-native teachers referred to this issue in one way or another. Highlighting this point, an experienced native teacher from the US considered fossilization a consequence of not providing effective feedback (excerpt 21).

Excerpt 21

She [the teacher] did not correct a student's error. When teachers fail to provide appropriate feedback students tend to continue making the error. As time passes, these errors become ingrained and difficult to remove. (T43: NT)

Any coherent instructional unit requires a well-organized structure. Presenting a well-structured lesson, ensuring that there is a good warm-up, instruction, practice, and closure, was another key criterion on which almost half of the participants, both native (NT = 32) and non-native (NNT = 32), relied to evaluate teachers. A beginner non-native teacher, for instance, emphasizes this criterion, referring to the warm-up stage of teaching she believes the teacher missed (see excerpt 22):

Excerpt 22

She [the teacher] did not pre-teach the structure form. She did not use warm up. (T10: NNT)

To be effective, instruction should incorporate any available resources that help foster learning. The efficient use of learners' L1 emerged as a further evaluation criterion among both groups of teachers, although more highlighted by the native ones. The majority of these participants (NT = 24, NNT = 15) referred to the debilitative impact that the excessive use of L1 might have on language learning. One of these beliefs is expressed in excerpt 23 by a native beginner teacher from Australia:

Excerpt 23

The excessive use of the native language creates a "crutch" for the students which should be discouraged. (T21: NT)

A few teachers (NT = 9, NNT = 7), moreover, referred to the ways in which using learners' L1 may support learning. A novice non-native teacher, for instance, considers it beneficial to use learners' L1 to set tasks (see excerpt 24):

Excerpt 24

The teacher used students' L1 to give direction of the tasks to ensure that they have understood what to do. (T23: NNT)

The instructional content should further suit the proficiency level of the learners. This issue included the less common evaluation criterion among both native (NT = 15) and non-native (NT = 10), an example of which is presented in excerpt 25, given by a novice native teacher from the UK, arguing that the grammatical point that the teacher considered for the class suits those beginner learners' proficiency:

Excerpt 25

Teacher chose a simple language pattern, and extended it in different ways. Good for these beginner adult students. The language structure which was chosen for the lesson was appropriate for the level of the learners. (T2: NT)

Discussion

The primary purpose of this study was to explore the criteria that native and non-native English teachers apply for evaluating teachers. Video-mediated peer observations were conducted to achieve this purpose. Content analysis of the data indicated commonalities between native and non-native teachers about many criteria for evaluating teacher quality. In other words, both groups of teachers employed certain qualities related to preparation (e.g., knowledge of content), caring (e.g., fairness), classroom management (e.g., time), and instruction (e.g., monitoring learning) to evaluate teachers. Importantly, almost all of the employed criteria are among the components of evaluation frameworks such as those in Danielson (2013) and Marzano (2013), which have been considered, by both theoretical and empirical research, as valid and reliable quality indicators for teacher evaluation (Coady et al., 2020; Kettler et al., 2022; Patrick et al., 2020). The similarity of native and non-native English speaker teachers' evaluation criteria, therefore, suggests that native and non-native English teachers are aware, to largely similar extents, of the criteria which contribute to teacher quality and hence should be considered for teacher evaluation. This shared awareness may be the result of native and non-native teachers' educational background and qualities generally expected of teachers, as

several scholars have highlighted that many teacher quality indicators are common across cultures (Cochran-Smith, 2021; Olsen, 2021).

Despite this convergence, the findings showed that a few of the criteria were more emphasized by one group. Native teachers employed creating an environment of respect and rapport (a caring-related issue) and providing clear directions for activities (an instruction-related issue) more than the non-native teachers. These few differences between native and non-native teachers' evaluation criteria may originate from different cultural/educational contexts in which they were teaching (Western and Eastern, respectively). This is in line with Grant et al.'s (2021) argument that the qualities various educational stakeholders employ for teacher evaluation essentially reflect conceptions toward effective teaching, and are therefore at least in part context-bound. Empirical studies have also provided evidence on how context can influence the criteria which are used in teacher evaluation (e.g., Darling-Hammond, 2021; Min, 2021). Goodwin and Low (2021), for example, discussed how the contemporary focus on holistic education (a belief that every child wants to and can learn) in Hong Kong turned creating a caring and safe learning environment into a key component of teacher evaluation.

Besides these slight differences, considerable disparity was observed between the two groups as to three criteria. Substantially more non-native teachers considered teachers' linguistic accuracy and fluency of speech for evaluation than native teachers. Apparently, the fact that non-native speaker participants paid far more attention to English teachers' language proficiency than native ones implies that non-native teachers are more obsessed with teachers' mastery over language as a requirement for teaching than native ones. The current study has, in fact, provided further empirical support to this hypothesis. This preoccupation has been suggested in a recent body of research in ELT. Studies into the self-efficacy beliefs of EFL/ESL teachers have convincingly demonstrated a positive relationship between non-native teachers' language proficiency and their self-efficacy beliefs (Faez et al., 2021; Hoang & Wyatt, 2021; Wyatt & Dikilitaş, 2021). The evidence that the more proficient the teachers, the stronger their belief in their capabilities in teaching English as a foreign language indicates the great value non-native speaker teachers attach to English teachers' linguistic proficiency.

This belief probably originates from discriminations that non-native speaker teachers experience due to their comparatively lower L2 proficiency. These discriminations, which surprisingly recent paradigms such as World Englishes and English as an International Language could not totally eliminate, have been widely acknowledged and challenged, in both professional and academic literature (Lowe, 2020; Thompson, 2021). Despite this, except for one participant, no point was made, either explicitly or implicitly, about the supremacy of native teachers' accuracy or fluency by these participants. Only one non-native teacher having 20 years of experience appeared to consider nativeness significant for an English teacher when he noted "He enjoys native speaker accuracy and fluency". The findings, therefore, cannot be interpreted as non-native teachers' dominant disposition toward the native English-speaker model.

The findings showed that more native speaker teachers referred to the efficient use of learners' L1 in the classroom than the non-native ones. This may suggest a comparatively higher recognition by the native teachers of the fact that learners' resources, their L1 in this case, can be efficiently employed by English teachers to support student learning. This acknowledgment by comparatively more native speaker teachers is of considerable significance as recent literature has highlighted that learners' L1, if properly used, can serve not only as an effective

learning tool, but also as an invaluable pedagogical technique (e.g., Soh, 2020; Yuzlu & Dikilitas, 2022). As the native and non-native English-speaking participants in the current study included an equal number of novice and experienced teachers with similar educational degrees, this higher recognition by the native teachers can not be attributed to these background characteristics, namely teaching experience and educational degree. Furthermore, substantially more native teacher participants (N = 18) studied non-English-related fields at the university than the non-native ones (N = 5). Hence, field-relevant university education similarly cannot explain the native teachers' higher acknowledgment of this issue. Differences in the cultural/educational contexts in which native and non-native English teachers were teaching (Western and Eastern, respectively) and the subsequent differences in teacher education programs they probably experienced, as described earlier, might have contributed to native teachers' higher consideration of learners' L1.

Although issues related to caring, classroom management, and instruction grabbed the attention of both native and non-native participants, teachers' knowledge base, particularly knowledge of subject matter and pedagogy (the two preparation-related categories), was rarely considered by the two groups. It needs to be noted, however, that several of the recurrent evaluation criteria are the realization of teachers' knowledge of content and pedagogical issues (criteria such as providing sufficient instruction and engaging students in learning). Moreover, these two components of teachers' knowledge have been reported in studies investigating teachers' perspectives on effective teaching (e.g., Korkmazgil & Seferoğlu, 2021; Yuan, Mak, & Yang, 2022; Nguyen & Pham, 2020; Sarfraz, 2019). It seems implausible then to interpret this finding as the participating teachers' lack of awareness that knowledge of subject and pedagogy plays a significant role in effective teaching. Rather, the most plausible reason might be that during the observation process, the participants focused their attention on what they could see in teachers' video-recorded behavior and practices in the classroom and hence did not refer to the knowledge that lies behind such practices.

Finally, since the findings suggest very few differences between native and non-native English teachers' evaluation criteria, the assumed superiority of native speaker teachers and in essence, the long-lasting controversy over native and non-native teachers is in part questioned. The bias and discriminations that non-native speaker teachers of English are still experiencing in various educational contexts (e.g., Gerald, 2022; Ramjattan, 2019; Rosa & Flores, 2017) arise from ideological concerns rather than educational ones, that is, the raciolinguistic ideologies. The fact that the native and non-native English-speaking teachers in the current study largely concurred on the criteria that should be employed in teacher evaluation provides evidence of this issue. Given that the criteria teachers employ for teacher evaluation reflect their beliefs about teacher quality (Grant et al., 2021) and that teachers' beliefs influence their practices (e.g., López-Barrio et al., 2021), the lack of substantial differences between the criteria employed by the so-called superior native and inferior non-native teachers in the present study questions, at least in part, this assumed superiority of native speaker teachers of English and calls for more research to help decrease the status and employment discriminations from which the majority of the non-native teachers worldwide are suffering.

Conclusion and Implications

The criteria that the teachers drew on in video-mediated observations were generally consistent among both native and non-native teachers. This study demonstrated that these criteria for evaluating teacher quality were broadly consistent with widely-endorsed models of teacher

evaluation, particularly Danielson's (2013) Framework for Teaching. The teacher quality criteria considered by the native and non-native teachers in the present study, despite differences in a few cases, are in line with Danielson's argument that teachers' classroom practices should be evaluated based on the extent to which they (1) create a productive and positive learning environment, (2) properly manage classroom space, time, student behavior, and instructional strategies, and (3) implement effective instruction. The findings, however, documented that non-native English teachers are comparatively less aware of the significant role that the efficient use of learners' L1 may play in instructional quality, but are more obsessed with teachers' linguistic accuracy and fluency. Overall, then, it suggests the importance of teachers' reflection on and examination of their conceptualization of teacher quality (as reflected in the criteria they employ for teacher evaluation), which is likely to bring about sustainable changes in their classroom practices (Li, 2020; Smith, 2020).

The findings, in particular, have one important implication for teacher education programs. The criteria that native and non-native participants employed for teacher evaluation indicate teachers' implicit belief, or in Borg's (2018) words "theories-in-use" about what constitutes teacher quality. Exposure to video captures of the real classroom environment provided the participants with the opportunity to reflect on and examine the strengths and weaknesses of each teacher's practices, hence revealing what they consider indicators of quality teaching. Given that teachers' beliefs — whether explicitly or implicitly held — have a bearing on their classroom practices (e.g., Kartchava et al., 2020; López-Barrio et al., 2021) and that teacher quality has proved to be one of the most contributing factors to student learning (Calero & Escardíbulb, 2020; Canalesa & Maldonado, 2018), unravelling teachers' beliefs about teacher quality is of great significance to teacher educators. It could be reasonable to say that the use of video-mediated peer observation can provide a platform for educators to uncover the implicit beliefs teachers hold toward teacher quality.

The current study, however, had limitations that should be considered in the interpretation of the findings. First, given the EFL context where the study was conducted, it was almost infeasible to gain access to and observe native speaker teachers' classrooms. As such, future research may be focused on data collected from native teachers' classroom teaching. Second, the study relied on video-mediated observation rather than the participants' direct observation of classrooms for evaluating teacher quality. Finally, the gender and field of study of the participants were not considered as possible sources of variation in teachers' criteria for evaluating teacher quality. Accordingly, further research could be conducted to examine the impact of teachers' gender and educational background on their evaluation criteria.

About the Authors

Zia Tajeddin is Professor of Applied Linguistics at Tarbiat Modares University, Iran. His main areas of research include teacher education and L2 pragmatics. He serves as the co-editor of the Springer book series *Studies in Language Teacher Education*. He co-edits two international journals: *Applied Pragmatics* (John Benjamins) and *Second Language Teacher Education* (Equinox). He has published his studies in *Journal of Language, Identity, & Education, International Journal of Applied Linguistics, Language Testing, The Language Learning Journal,* and *Language and Intercultural Communication*, among others. He is the co-editor of *Lessons from Good Language Teachers* (Cambridge University Press, 2020), *Pragmatics Pedagogy in English as an International Language* (Routledge, 2021), and *Teacher Reflection:*

Policies, Practices and Impacts (Multilingual Matters, 2022). ORCID ID: 0000-0002-0430-6408

Zari Saeedi received her Ph.D. from the British University of Trinity College and is an Associate Professor of Allameh Tabataba'i University, Iran. She has taught various B.A., M.A., and Ph.D. courses in different universities, taking part in different national/international conferences, presenting/publishing papers/books on a range of topics including educational neuro/psycholinguistics and brain functioning, cognitive language learning and in particular, brain based language learning (BBLL), culture, technology-assisted language learning, and Role and Reference Grammar Theory of linguistics. Her published paper and book in John Benjamins and Equinox publications are among the recent ones. Her most recent publication is the dictionary of virtual education with a focus on computer assisted language learning. ORCID ID: 0000-0003-2165-8891

Hamideh Mozaffari is a Ph.D. candidate in Teaching English as a Foreign Language (TEFL) at Allameh Tabataba'i University, Iran. Her areas of interest include teacher education, English as an international language, and language skills. She has published papers and book reviews in *Language Teaching Research, Innovation in Language Learning and Teaching*, and *English Today*. ORCID ID: 0000-0003-3530-6731

To Cite this Article

References

Amrein-Beardsley, A., & Geiger, T. (2020). Methodological concerns about the education value-added assessment system (EVAAS): Validity, reliability, and bias. *SAGE Open, 10*(2), 1–15. https://doi.org/10.1177/2158244020922224

Borg, S. (2018). Teachers' beliefs and classroom practices. In P. Garrett & J. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 75–91). Routledge.

Borg, S., & Edmett, A. (2018). Developing a self-assessment tool for English language teachers. *Language Teaching Research*, *23*(5), 655–679. https://doi.org/10.1177/1362168817752543

Briggs, D., & Domingue, B. (2011). *Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles unified school district teachers.* National Education Policy Center.

Calero, J., & Escardíbul, J. (2020). Teacher quality and student skill acquisition: An analysis based on PIRLS-2011 outcomes. *Educational Studies, 46*(6), 676–692. https://doi.org/10.1080/03055698.2019.1628710

Canalesa, A., & Maldonado, M. (2018). Teacher quality and student achievement in Chile: Linking teachers' contribution and observable characteristics. *International Journal of Educational Development, 60,* 33–50. https://doi.org/10.1016/j.ijedudev.2017.09.009

Chand Dayal, H., & Alpana, R. (2020). Secondary pre-service teachers' reflections on their micro-teaching: Feedback and self-evaluation. *Waikato Journal of Education, 25*(1), 73–83. https://doi.org/10.15663/wje.v25i0.686

Coady, M., Miller, M., Jing, Z., Heffington, D., Lopez, M., Olszewska, A., Jong, E., Yilmaz, T., Ankeny, R. (2020). Can English learner teacher effectiveness be observed? Validation of an EL-modified framework for teaching. *TESOL Quarterly, 54*(1), 173–200. https://doi.org/10.1002/tesq.544

Cochran-Smith, M. (2021). Exploring teacher quality: International perspectives. *European Journal of Teacher Education, 44*(3), 415–428. https://doi.org/10.1080/02619768.2021.1915276

Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS value-added assessment system. *Education Policy Analysis Archives, 22*(98), 1–42. https://doi.org/10.14507/epaa.v22.1594

Creswell, J. (2015). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.

Danielson, C. (2001). New trends in teacher evaluation. *Educational Leadership, 58*(5), 12–15.

Danielson, C. (2013). *The framework for teaching: Evaluation instrument.* Danielson Group.

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

Darling-Hammond, L. (2021). Defining teaching quality around the world. *European Journal of Teacher Education*, *44*(3), 295-308. https://doi.org/10.1080/02619768.2021.1919080

Dillon, H., James, C., Prestholdt, T., Peterson, V., Salomone, S., & Anctil, E. (2020). Development of a formative peer observation protocol for STEM faculty reflection. *Assessment & Evaluation in Higher Education, 45*(3), 387–400. https://doi.org/10.1080/02602938.2019.1645091

Dunkin, M. (1991). Orientations to teaching, induction, experiences and background characteristics of university lecturers. *Australian Educational Researcher, 18*(3), 31–52. https://doi.org/10.1007/BF03219483

Dunkin, M. J., Precians, R. P., & Nettle, E. B. (1996). Elementary student teachers' self-evaluation: Learning criteria for judging lessons and self as teacher. *Journal of classroom interaction, 31*(1), 11–19. https://www.jstor.org/stable/23870610

Faez, F., Karas, M., & Uchihara, T. (2021). Connecting language proficiency to teaching ability: A meta-analysis. *Language Teaching Research, 25*(5), 754–777. https://doi.org/10.1177/1362168819868667

Gatbonton, E. (2008). Looking beyond teachers' classroom behavior: Novice and experienced ESL teachers' pedagogical knowledge. *Language Teaching Research, 12*(2), 161–182. https://doi.org/10.1177/1362168807086286

Gerald, G. P. B. (2022). *Antisocial language teaching: English and the pervasive pathology of whiteness.* Channel View Publications.

Gergen, K., & Gill, S. (2020). *Beyond the tyranny of testing: Relational evaluation in education.* Oxford University Press.

Goodwin, A., & Low, E. (2021). Rethinking conceptualisations of teacher quality in Singapore and Hong Kong: A comparative analysis. *European Journal of Teacher Education, 44*(3), 365–382. https://doi.org/10.1080/02619768.2021.1913117

Gordon, S. (2020). *Standards for instructional supervision: Enhancing teaching and learning.* Routledge.

Grant, L., Stronge, J., & Xu, X. (2021). *International beliefs and practices that characterize teacher effectiveness.* IGI Global.

Gröschner, A., Schindler, A., Holzberger, D., Alles, M., & Seidel, T. (2018). How systematic video reflection in teacher professional development regarding classroom discourse contributes to teacher and student self-efficacy. *International Journal of Educational Research, 90*, 223–233. https://doi.org/10.1016/j.ijer.2018.02.003

Hammersley-Fletcher, L., & Orsmond, P. (2005). Reflecting on reflective practices within peer observation. *Studies in Higher Education, 30*(2), 213–224. https://doi.org/10.1080/03075070500043358

Harford, J., & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. *Teaching and Teacher Education 24*(7), 1884–1892. https://doi.org/10.1016/j.tate.2008.02.010

Harrison, R., Meyer, L., Rawstorne, P., Razee, H., Chtkara, U., Mears, S., & Balasooria, S. (2022). Evaluating and enhancing quality in higher education teaching practice: A meta-review. *Studies in Higher Education, 47*(1), 80–96. https://doi.org/10.1080/03075079.2020.1730315

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

Hendry, G., Georgiou, H., Lloyd, H., Tzioumis, V., Herkes, S., & Sharma, M. (2021). 'It's hard to grow when you're stuck on your own': Enhancing teaching through a peer observation and review of teaching. *International Journal for Academic Development, 26*(1), 54–68. https://doi.org/10.1080/1360144X.2020.1819816

Hoang, T., & Wyatt, M. (2021). Exploring the self-efficacy beliefs of Vietnamese pre-service teachers of English as a foreign language. *System, 96*, 1–40. https://doi.org/10.1016/j.system.2020.102422

Ibrahim, N., Surif, J., Arshad, M., & Mokhtar, M. (2012). Self-reflection focusing on pedagogical content knowledge. *Procedia - Social and Behavioral Sciences, 56,* 474–482. https://10.1016/j.sbspro.2012.09.679

Ingersoll, R. (2020). Misdiagnosing the teacher quality problem. In D. Cohen, S. Fuhrman, & F. Mosher (Eds.), *The state of education policy research* (pp. 296–312). Routledge.

Kartchava, E., Gatbonton, E., Ammar, A., & Trofimovich, P. (2020). Oral corrective feedback: Pre-service English as a second language teachers' beliefs and practices.

Language Teaching Research, 24(2), 220–249. https://doi.org/10.1177/1362168818787546

Kettler, R. J., Hua, A., Dudek, C. M., Reddy, L. A., Arnold-Berkovits, I., Wiggs, N. B., Lekwa, A., & Kurz, A. (2022). Improving measurement of teacher performance: Alternative scoring for classroom-based observational systems. *Educational Evaluation, 27(*3), 269–284. https://doi.org/10.1080/10627197.2022.2088494

Korkmazgil, S., & Seferoğlu, G. (2021). Teacher professionalism: Insights from Turkish teachers of English into the motives that drive and sustain their professional practices. *Journal of Education for Teaching, 47*(3), 366–378. https://doi.org/10.1080/02607476.2021.1897781

Kremer, L., & Ben-Peretz, M. (2006). Teachers' self-evaluation --concerns and practices. *Journal of Education for Teaching: International Research and Pedagogy, 10*(1), 53–60. https://doi.org/10.1080/0260747840100104

Kubota, R. (2022). Racialised teaching of English in Asian contexts: Introduction. *Language, Culture and Curriculum.* https://doi.org/10.1080/07908318.2022.2048000

Leon, S. D., & Thomas, L. (2015). *Collaboration, rubrics and teacher evaluation.* Information Age Publishing.

Li, L. (2020). *Language teacher cognition: A socio-cultural perspective.* Springer Nature.

Lillejord, S., & Bort, K. (2020). Trapped between accountability and professional learning? School leaders and teacher evaluation. *Professional Development in Education, 46*(2), 274–291. https://doi.org/10.1080/19415257.2019.1585384

López-Barrios, M., Martín, M., & Debat, E. (2021). EFL vocabulary teaching beliefs and practices: The case of two teachers in Argentina. *TESOL Journal, 12*(1), 1–17. https://doi.org/10.1002/tesj.533

Lowe, R. (2020). *Uncovering ideology in English language teaching: Identifying the 'native speaker' frame.* Springer Nature.

Macleod, G. R. (1988). Teacher self-evaluation: An analysis of criteria, indicators, and processes used by teachers in judging their success. *International Journal of Educational Research, 12*(4), 395–408. https://doi.org/10.1016/0883-0355(88)90033-X

Madsen, A. (2005). *Where is the "self" in teacher self-assessment? An examination of teachers' reflection and assessment practices in relation to their teaching practices* (Unpublished Ph.D. dissertation). Iowa State University, Iowa.

Marzano, R. (2013). *The Marzano teacher evaluation model.* Marzano Research Laboratory.

Marzano, R., Rains, C., & Warrick, P. (2020). *Improving teacher development and evaluation: A guide for leaders, coaches, and teachers.* Marzano Resources.

McCoy, S., & Lynam, A. (2021). Video-based self-reflection among pre-service teachers in Ireland: A qualitative study. *Education and Information Technologies, 26*(1), 921–944. https://doi.org/10.1007/s10639-020-10299

McLeod, J., Fisher, J., & Hoover, G. (2003). *The key elements of classroom management: Managing time and space, student behavior, and instructional strategies.* ASCD.

Min, M. (2021). Teacher effectiveness: Policies and practices for evaluating and enhancing teacher quality in South Korea. In L. Grant, J. Stronge, & X. Xu (Eds.), *International beliefs and practices that characterize teacher effectiveness* (pp. 227–245). IGI Global.

Nguyen, P., & Pham, H. (2020). Academics' perceptions of challenges of a peer observation of teaching pilot in a Confucian nation: The Vietnamese experience. *International Journal for Academic Development, 26*(4), 448–462. https://doi.org/10.1080/1360144X.2020.1827260

O'Leary, M. (2020). *Classroom observation: A guide to the effective observation of teaching and learning.* Routledge.

Olsen, B. (2021). Teacher quality around the world: What's currently happening and how can the present inform the future? *European Journal of Teacher Education, 44*(3), 293–294. https://doi.org/10.1080/02619768.2021.1917053

Olson, J., Madsen, A., Bruxvoort, C., & Clough, M. (2004). *Where's the Teacher? Preservice teachers' difficulty in seeing the teacher's critical role.* Paper presented at the annual meeting of the National Association of Research in Science Teaching, Vancouver, Canada.

Patrick, H., French, B., & Mantzicopoulos, P. (2020). The reliability of framework for teaching scores in kindergarten. *Journal of Psychoeducational Assessment, 38*(7), 831–845. https://doi.org/10.1177/0734282920910843

Phillipson, R. (1992). *Linguistic imperialism.* Oxford University Press.

Ramjattan, V. A. (2019). Racist nativist microaggressions and the professional resistance of racialized English language teachers in Toronto. *Race Ethnicity and Education, 22*(3), 374–390. https://doi.org/10.1080/13613324.2017.1377171

Riazi, A. M. (2016). *The Routledge encyclopedia of research methods in applied linguistics.* Routledge.

Rosa, J. (2016). Standardization, racialization, languagelessness: Raciolinguistic ideologies across communicative contexts. *Journal of Linguistic Anthropology, 26*(2), 162–183. https://doi.org/10.1111/jola.12116

Rosa, J., & Flores, N. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society, 46*(5), 621–647. https://doi.org/10.1017/S0047404517000562

Sarfraz, S. (2019). *Rethinking formative assessment through peer observation and reflection: A case study of Pakistani ESL lecturers' cognition and practices* (Unpublished Ph.D. dissertation). The University of Waikato, New Zealand.

Sekaja, L., Adams, B., & Yagmur, K. (2022). Raciolinguistic ideologies as experienced by racialized academics in South Africa. *International Journal of Educational Research, 116,* 1–14. https://doi.org/10.1016/j.ijer.2022.102092

Shortland, S. (2010). Feedback within peer observation: Continuing professional development and unexpected consequences. *Innovations in Education and Teaching International, 47*(3), 295–304. https://doi.org/10.1080/14703297.2010.498181

Smith, R. (2020). *Mentoring teachers to research their classrooms: A practical handbook.* British Council.

Soh, K. (2020). *Teaching Chinese language in Singapore: Concerns and visions.* Springer.

Stronge, J. (2018). *Qualities of effective teachers.* Association for Supervision and Curriculum Development.

Stronge, J., Ward, T., & Grant, L. (2011). What makes good teachers good? A cross-case analysis of the connection between teacher effectiveness and student achievement. *Journal of Teacher Education, 62*(4), 339–355. https://doi.org/10.1177/0022487111404241

Thompson, A. (2021). *The role of context in language teachers' self-development and motivation: Perspectives from multilingual settings.* Multilingual Matters.

Torres, A., Lopes, A., Valente, J., & Mouraz A. (2017). What catches the eye in class observation? Observers' perspectives in a multidisciplinary peer observation of teaching program. *Teaching in Higher Education*, *22*(7), 822–838. https://doi.org/10.1080/13562517.2017.1301907

Visone, J. (2022). What teachers never have time to do: Peer observation as professional learning. *Professional Development in Education, 48*(2), 203–217. https://doi.org/10.1080/19415257.2019.1694054

Wood, M., & Su, F. (2017). What makes an excellent lecturer? Academics' perspectives on the discourse of 'teaching excellence' in higher education. *Teaching in Higher Education, 22*(4), 451–466. https://doi.org/10.1080/13562517.2017.1301911

Wyatt, M., & Dikilitaş, K. (2021). English language teachers' self-efficacy beliefs for grammar instruction: Implications for teacher educators. *The Language Learning Journal*, *49*(5), 541–553. https://doi.org/10.1080/09571736.2019.164294

Yuan, R., Mak, P., & Yang, M. (2022). 'We teach, we record, we edit, and we reflect': Engaging pre-service language teachers in video-based reflective practice. *Language Teaching Research, 26*(3), 552–571. https://doi.org/10.1177/1362168820906281

Yuzlu, M. Y., & Dikilitas, K. (2022). Translanguaging in the development of EFL learners' foreign language skills in Turkish context. *Innovation in Language Learning and Teaching, 16*(2), 176–190. https://doi.org/10.1080/17501229.2021.1892698