



Journal of Teaching Language Skills
(JTLs)

40(2), Spring 2021, pp. 161-192

DOI: 10.22099/jtls.2021.39747.2942

Online ISSN: 2717-1604

Print ISSN: 2008-8191

Research Paper

Teachers' Rating Criteria for Classroom Oral interviews as Influenced by their First Languages and Educational Backgrounds

Ali Sayyadi *

Sayyed Mohammad Alavi **

Hossein Karami ***

University of Tehran

Abstract

The current study attempted qualitatively to explore and compare the qualities that native and Iranian English teachers (with and without related educational backgrounds) attend to while rating their students' oral productions in the classroom context. In doing so, the perceptions of 19 native English teachers (9 graduates in TEFL and 10 graduates in other majors) along with 18 Iranian English teachers (10 graduates in TEFL and 8 graduates in other majors) were sought through semi-structured interviews. The data were collected after the outbreak of Coronavirus (COVID-19) disease in 2020, which gave the researchers no choice but to look for haphazard cases with specific features in social networks. The recorded interviews were analyzed attentively through content analysis. The findings indicated that although all native and non-native respondents focused intensively on the structural features of language in general while rating oral interviews, they had notably different views regarding some sub-features within each category. Further results showed that the native and non-native TEFL-graduate teachers, unlike their peers with unrelated educational backgrounds, also gave credits to several message-based and pragmatic aspects of oral production. The findings have practical

Received: 16/02/2021 Accepted: 02/05/2021

* Ph.D Candidate, Email: ali.sayyadi1989@gmail.com, Corresponding Author

** Professor, Email: smalavi@ut.ac.ir

*** Assistant Professor, Email: hkarami@ut.ac.ir

implications for researchers, pre-service and in-service teachers, and teacher educators.

Keywords: Classroom Assessment, Speaking, Rating Criteria, Native Teachers, Non-native Teachers, Educational Background

A classroom can potentially create a community of practice where learning occurs within a communicative process to, specifically, learners of English as a Foreign Language (EFL) who have limited opportunities to develop their communicative abilities outside classroom contexts (Downer et al., 2010; Rezvani & Sayyadi, 2015). In classroom contexts, the effectiveness of teaching and learning communicative skills might depend on the efficiency of classroom assessment, which is carried out not only for examining students' performance but also for promoting the pedagogic and learning behaviors considered in classrooms (Cumming, 2009; Davison & Leung, 2009). Hence, remarkable attention ought to be paid to the role of classroom assessment in promoting EFL language learners' oral communicative abilities.

Since the domination of communicative approaches to language teaching and learning in language learning classrooms in the 1970s and 1980s, traditional test approaches have been gradually replaced by communicative language assessment (Brown, 2004). More specifically, instead of giving limited responses to questions that mostly tap into their receptive skills and knowledge of language forms, language learners should be involved in fulfilling interactive tasks to show their command of communicative skills.

Speaking assessment is presumed to be extremely challenging to prepare, administer, and interpret as speaking tasks are not capable of isolating the single skill of oral production and involve aural or reading comprehension (Brown & Abeywickrama, 2010; Fulcher & Davidson, 2007). Besides, "

eliciting the specific criteria designated for a task can be tricky because beyond the word level, spoken language offers a number of productive options to test-takers" (Brown & Abeywickrama, 2010, p. 413). Accordingly, as Luoma (2004) and Brown and Abeywickrama (2010) argued, in assessing speaking skills, language examiners are strongly suggested to be cautious in specifying the scoring procedure and criteria in order to achieve a sufficiently reliable index.

Brown (2004) considers oral interviews as the first thing that comes to mind when speaking assessment is mentioned. They are the most typically utilized tests administered for assessing one's skill at performing what oral production is used for in actuality (Hughes, 2003; Staples, Laflair & Egbert, 2017). On the other hand, despite the wide usage in academic contexts and a relatively high degree of the face, content, and construct validity attributed to oral interviews (Bachman & Palmer, 1990; Zahedi & Shamsaee, 2012. Weir, 1990), some cautions are to be considered when it comes to interpreting the results of examinees' performances in oral interviews. Assessing one's communicative competence through oral interviews is potentially open to raters' subjective interpretations and erroneous application of scoring criteria (Correia, 2016; Fulcher, Davidson & Kemp, 2011; Kuiken & Vedder, 2014) and, consequently, to concerns about reliability. Accordingly, "rater variability as a result of differential rating perceptions is becoming increasingly important because different raters may draw on their own personalized constructs irrespective of the criteria they are given" (Tajeddin, Alemi & Pashmforoosh, 2011, p.126).

In classroom contexts, language teachers are typically in charge of developing and carrying out oral interviews and rating the performances of their students (Salaberry, 2000; Wang & Yu, 2017). However, they may find

assessing their students' oral capacities challenging owing to the difficulty ascribed to defining the components of speaking ability (Taylor, 2006) and various factors (e.g., their age, gender, first language, educational backgrounds) which might prevent them from carrying out fair judgments on their students' oral performances (Qashoa, 2012; Willis & Willis, 2007). Depending on their first language, amount of experience, and educational backgrounds and levels, they may assess their students' speaking ability based on varying criteria (Caban, 2003; Zhang & Elder, 2011). Therefore, exploring the likely effects of such factors might reflect a more comprehensive understanding of the speaking construct and raise teachers' awareness of the qualities regarded by English teachers with varying characteristics.

Inconsistent results regarding the likely effects of raters' first language on their judgments have been reported in the related literature (see, e.g., Brown et al., 2005; Kim, 2009; Plough et al., 2010; Zhang & Elder, 2011), which calls for carrying out further research in this area. Also, a question that has still remained under-researched is how comparable the rating criteria used by teachers with varying educational backgrounds are. Accordingly, the present study sought to explore and compare the rating criteria employed by native and Iranian English teachers who had and had not graduated in Applied Linguistics or Teaching English as a Foreign Language (TEFL). It is worth noting that students who take higher-education courses on various aspects of English language education in English-speaking countries and Iran should major in Applied Linguistics and TEFL, respectively. Given the close similarities between the two majors and for the sake of keeping consistency, teachers who graduated in Applied Linguistics or TEFL are referred to as *TEFL-graduate teachers* in this study.

Literature Review

Consideration of assessment in classroom context requires the adoption of an approach that "moves away from the standardized, multiple-choice type test towards approaches where the assessment task closely matches the desired performance and takes place in an authentic or classroom context" (Gipps, 1994, p. 11). Unlike high-stakes standardized tests which are developed and administered by testing experts and centers based on well-studied generalizable standards, classroom assessment is teacher-directed and context-bound (Black & Wiliam, 2004). That is, the teacher is individually in charge of developing and administering tests based on specific context-dependent objectives to boost the learning process of learners. An assessment carried out by a teacher is essentially an informal activity (Downer et al., 2010) which can include posing questions, observing activities, and eventually judging students' works in a systematic or ad hoc way (Gipps, 1994). However, as Gipps (1994) argued, repeated teacher assessment in a range of contexts over a period of time might enable him to build up a narrowly-based evaluation of his students' achievements and performances. Having observed these characteristics, Gipps (1994) regarded teacher assessment to be highly valid concerning content and construct and subscribed high consequential validity to teacher assessment in case it is used for formative purposes.

Speaking Assessment

The typical characteristics of speaking manifest it as a reciprocal skill conditioned by various factors, including the linguistic, strategic, and pragmatic competencies of speakers along with the processing, reciprocal, and contextual conditions of the communication practice, which might influence the level of speech intelligibility (Isaacs & Trofimovich 2012; Rezvani &

Sayyadi, 2015). Although intelligibility has long been equated with accurate pronunciation, recent related literature has shown that one's ability to understand second language (L2) speech is not limited to pronunciation, but also other linguistic domains such as lexicogrammatical and lexical richness, discourse measures, and fluency (temporal) variables (Trofimovich & Isaacs, 2012). In other words, plenty of other linguistic features except pronunciation can affect one's understanding of L2 speech. On the other hand, L2 accent is assumed to be "a much narrower construct that is most strongly related to the linguistic factors commonly referred to under the umbrella term *pronunciation*, including word stress, rhythm, and segmental production accuracy" (Isaacs, 2016, p.138).

To put the speaking skill into motion, language users also ought to resort to some interactive skills in order to control their oral production and perform appropriate speech acts in certain contexts (Trofimovich & Isaacs, 2012). In other words, to construct lucid interactional behaviors, conversation participants are supposed to possess capacities other than the mere knowledge of the language system or discourse to satisfy some conditions that derive from the particular nature of speech. These are what Fan (2018) refers to as interaction skills that integrate language users' basic motor-perceptive skills with their adaptability and flexibility to help interaction make sense.

Such complexities make it hard to define speaking constructs accurately and clearly to be unanimously considered and employed by all assessors while striving to judge one's oral communicative abilities. As argued by Luoma (2004), the speaking proficiency concept and constructs are not well-defined, as speakers simultaneously need to draw on various communicative skills which normally develop at different rates. Therefore, some examiners might rate test-takers' speaking performance based on the content of their message,

while others might rely on properties of examinees' oral production, which might hinder the determination of a clear definition for focal constructs and addressing potential sources of irrelevant constructs (Winke, Gass & Myford 2013). One solution suggested by Lado (1961) to tackle such a problem was to employ decontextualized, discrete-point items to measure the goodness of one's oral performance and articulation. However, subsequent research (see e.g., Larson & Hendricks, 2009) indicated that even in case of using such cases, raters would not be able to agree to a true way to make accurate judgments about the rightness or appropriateness of one test taker's oral production due to its association with raters' subjectivity. Accordingly, to evaluate the performances of speakers, a considerable number of criteria are taken into consideration (e.g., grammar, vocabulary, pronunciation, fluency, and appropriateness). Therefore, when assessing this skill, language instructors need to develop the assessment tools based on speaking proficiency constructs that are defined carefully and objectively (Kuschmann & Lowit, 2015).

Raters' Criteria for Rating Speaking

The review of the literature indicates that although several studies have been conducted to scrutinize speaking assessment, no consensus regarding the most important speaking constructs and criteria to be considered by raters has been reached. Thuy & Nga (2018), for instance, highlighted accent, grammar, vocabulary, and appropriateness as typically important elements of speaking assessment. According to Butler et al. (2000, p. 10), such elements "include accomplishment of a task, sufficiency of response, comprehensibility, adequacy of grammatical resources, range and precision of vocabulary, fluency, and cohesion". Caban (2003) also pointed out the mastery of

grammar, content, fluency, pragmatics, pronunciation, compensation techniques, and overall intelligibility as seven categories that are normally employed by teachers in classroom oral assessments. Raters in Ang-Aw and Goh's study (2011) considered two groups of criteria when assessing speaking, including criterion factors (e.g., clarity of expression, elaboration of response, and engagement in conversation) and non-criterion factors (e.g., range of vocabulary, the novelty of ideas, and inter-candidate comparison).

As teacher raters may need to rely on their intuition, knowledge, contextual understanding, and experience in the process of developing and implementing classroom oral assessment (Salaberry, 2000; Wang & Yu, 2017), it seems critical to investigate their perceptions to explore the qualities they take into account while assessing their students. Meanwhile, a close examination of the likely effects that some factors such as teachers' first language and educational backgrounds might have on their choices might also give rise to practical implications.

The studies undertaken to compare native and non-native English teachers regarding their evaluation of language learners' speaking performances have mainly reported that non-native raters mainly focused on forms of language in their evaluations and were less tolerant of errors in comparison with their native colleagues (see, e.g., Brown, 1995; Fayer & Krasinski, 1987; Kim, 2009; Zhang & Elder, 2011). Kim (2009), for instance, argued that native raters focused more on the elaborate features of speaking in pronunciation and grammar use. Zhang and Elder (2011) also found non-native teachers to be less concerned with message-focused criteria for assessing speaking. Such findings stand in contrast to those documented in Tajeddin, Alemi, and Pashmforoosh's (2011) study, where the non-native raters reported to take into consideration some non-linguistic categories of

speaking rating criteria, including fluency, topic management, affective variables, organization, and function. Hence, it seems that exploration of the qualities that native and non-native teachers opt for while undertaking oral assessment merits more investigations. Accordingly, the current study was conducted to detect and compare the criteria considered by native English teachers (NET) and Iranian English teachers (IET) when rating their students' oral performances.

Michell (2014) viewed the variability in rating criteria caused by the educational backgrounds of teachers as a threat to the quality of teacher-based assessment. Although language teachers' lack of educational qualifications has been assumed a concern in language assessment practices, there is a dearth of research on this topic (Phung, 2018). Some studies, however, have investigated the influence of training on teachers' assessment of oral productions (e.g., Caban, 2003; Kim, 2009; Khoshsima & Afiati, 2015; Tajeddin et al., 2011; Zhang & Elder, 2011). The results showed that untrained teachers focused more on the linguistic resources as criteria for their assessments and that trained teachers drew more often on the non-linguistic categories of content and fluency. However, Khoshsima and Afiati (2015) found no significant difference in the trained and untrained teachers' ratings. It seems that no consensus in this regard has been reached, and the effects of educational backgrounds on teachers' perceptions of significant speaking rating criteria are under-researched. Therefore, the current study was carried out to potentially contribute to enhancing the body of knowledge in the related literature.

Research Questions

An attempt was made in the present study to find answers to the following research questions:

1. What qualities do native and Iranian TEFL-graduate teachers focus on while assessing language learners' speaking through classroom interviews?
2. What qualities do native and Iranian non-TEFL-graduate teachers focus on while assessing language learners' speaking through classroom interviews?
3. How comparable are the speaking assessment qualities considered by the native English teachers (with and without related majors) and the Iranian English teachers (with and without related majors)?

Method

Given the nature of the research questions and the aim to come up with a profound examination of participants' perceptions about the most important speaking qualities, a qualitative research method was employed in the present study.

Participants

The study was aimed at exploring what criteria ELTs (with and without related fields of study) and IETs (with and without related fields of study) regard when they evaluate their students' oral communicative ability. In doing so, much care was exercised to seek the perceptions of teachers who had experienced teaching English language for at least 5 years. Teachers with relevant educational backgrounds were required to have completed at least their undergraduate studies in Applied Linguistics or TEFL majors. It should also be noted that teachers who graduated in other English Language majors

such as English Literature, Linguistics, and Translation were not involved in the study due to their limited exposure to approaches, techniques, and technical terms in language pedagogy and assessment.

It is worth mentioning that the data were to be collected after the outbreak of Coronavirus (COVID-19) disease in 2020, which gave the researchers no choice but to find appropriate cases through social networks. More specifically, to sample the participating language teachers, one of the researchers randomly looked for appropriate cases with the required features in some social networks such as LinkedIn, Academia, and Facebook and sent them messages containing a brief description of the objectives of the study along with formal participation requests. Those who accepted the request were required to fill in a demographic questionnaire and be interviewed. The data collected from each interview were analyzed before the conduction of the next one. This procedure was kept on up to reaching the state of data saturation and coherence.

In total, to collect the data up to the state of saturation, the perceptions of 9 native TEFL-graduate teachers (5 males and 4 females), 10 native non-TEFL-graduate teachers (3 males and 7 females), 10 Iranian TEFL-graduate teachers (4 males and 6 females), and 8 Iranian non-TEFL-graduate teachers (3 males and 5 females) were sought. All in all, they had experienced English language teaching for at least 8 years, and their ages ranged from 30 to 46. The demographic information about the native participants is illustrated in Table 1. It is worth noting that although some native participants had experienced English language teaching in different countries, only their current working place is illustrated in the table. The native non-TEFL-graduate teachers had professionally worked in countries where English was not spoken as a first language and mainly taught English for general and

specific purposes to EFL learners in language centres and colleges. The native TEFL-graduate teachers, however, mostly worked in their own countries and taught similar courses in different colleges and institutes. Both groups acknowledged that they had recurrently evaluated their students' oral abilities through classroom interviews.

Table 1.

Demographic Information About the Native Participants

No	Name*	Age	Gender	Teaching Experience	Latest Field of Study	Level of Education	Current working place
1	Olivia	46	Female	20 years	Applied Linguistics	Master	Canada
2	John	41	Male	14 years	Applied Linguistics	Master	England
3	Emily	40	Female	12 years	Applied Linguistics	Master	Qatar
4	Jacob	35	Male	10 years	Applied Linguistics	Master	Australia
5	Lily	35	Female	10 years	Applied Linguistics	Bachelor	The U.S.
6	Oscar	31	Male	9 years	Applied Linguistics	Master	China
7	Robert	35	Male	9 years	Applied Linguistics	Bachelor	England
8	Sophia	33	Female	9 years	Applied Linguistics	Bachelor	England
9	Joe	30	Male	8 years	Applied Linguistics	Bachelor	New Zealand
10	William	42	Male	15 years	Anthropology	Master	Kuwait
11	Victoria	42	Female	14 years	Archaeology	Master	Qatar
12	Jessica	39	Female	14 years	Civil Engineering	Bachelor	China
13	Tracy	37	Female	12 years	Nursing	Bachelor	Sweden
14	Barbara	35	Female	10 years	Nuclear Physics	Master	Qatar
15	Susan	32	Female	10 years	Nursing	Bachelor	Russia
16	Daniel	35	Male	10 years	Liberal Arts	Master	China
17	Linda	32	Female	8 years	Mathematics	Master	Oman
18	Michael	31	Male	8 years	Civil Engineering	Bachelor	Qatar
19	Jennifer	30	Female	8 years	Geography	Bachelor	Germany

*Note: The names are fictitious

The demographic information about the Iranian participants is also shown in Table 2. They had experienced teaching English for general purposes to EFL learners in Iranian language institutes (n=18) and universities (n=4). Two TEFL-graduate teachers had also taught to English-major students in Iranian universities. Like their native peers, the Iranian participants acknowledged using oral interviews to assess their learners' speaking abilities.

Table 2.

Demographic Information About the Iranian Participants

No	Name*	Age	Gender	Teaching Experience	Latest Field of Study	Level of Education
1	Fatemeh	40	Female	14 years	Applied Linguistics	Master
2	Ahmad	34	Male	12 years	Applied Linguistics	Ph.D.
3	Elahe	40	Female	12 years	Applied Linguistics	Master
4	Maryam	32	Female	10 years	Applied Linguistics	Bachelor
5	Narges	35	Female	10 years	Applied Linguistics	Bachelor
6	Nazanin	32	Female	9 years	Applied Linguistics	Bachelor
7	Azadeh	34	Female	9 years	Applied Linguistics	Bachelor
8	Peyman	40	Male	10 years	Applied Linguistics	Master
9	Sadegh	32	Male	8 years	Applied Linguistics	Ph. D.
10	Ali	42	Male	8 years	Applied Linguistics	Master
11	Mozhgan	35	Female	10 years	Accounting	Master
12	Abbas	31	Male	8 years	Microbiology	Bachelor
13	Najmeh	30	Female	8 years	Economics	Master
14	Marzieh	32	Female	8 years	Economics	Master
15	Negar	40	Female	7 years	Sociology	Bachelor
16	Pooya	33	Male	6 years	History	Master
17	Arezoo	30	Female	6 years	Nursing	Bachelor
18	Saman	33	Male	6 years	Management	Master

*Note. The names are fictitious

Instruments

Two instruments were employed in this study. Firstly, a demographic questionnaire was administered to the participants before undertaking the interviews. It aimed to provide information regarding the participants' gender, age, nationality, current working place, higher-education backgrounds and levels, years of experience in teaching the English language, and the types of courses they had taught. The questionnaire also included a four-point Likert scale (with "never", "hardly ever", "sometimes", and "very often" items) examining the extent to which the respondents used classroom interviews to assess their students' speaking proficiency.

Secondly, semi-structured interviews were employed in the present study to let the respondents extend, elaborate on, and provide details for their perceptions of essential qualities of oral communicative abilities in classrooms. Such a plan could address the richness and depth of the responses given by the respondents as well as the comprehensiveness of the emerging findings. The interview questions were developed in English and explore the criteria they take into consideration while assessing their students' oral abilities.

Data Collection Procedure

The participants were initially asked to fill in the demographic questionnaires they received through Emails or LinkedIn. Subsequently, the respondents were requested to be interviewed through Skype. Those who declined the request, for any reason, were sent the interview questions and were asked to record and send back their responses. It should also be noted that the respondents were requested to define and elaborate on the components they considered to mention. That is, if they were going to underline, for instance, learners' command of grammar as a determining criterion used for

assessing speaking skill, they were required to clarify what they exactly meant by *the command of grammar*.

To ensure the trustworthiness of the data, the researchers sought to avoid bias through the recommended strategies (for more details, see McMillan & Schumacher, 2006). More specifically, they employed a persistent field-work, accounted for participants' language verbatim accounts, and took hand-written notes of the key points of the interviewees' responses in addition to recording all interviews with the permission of the Participants. After each interview, the noted data were informally checked with the participants interviewed through Skype. That is, the data were returned to the respondents to check for accuracy and resonance with their experiences. Data accumulation was kept on up to reaching the state of data saturation and coherence. Eventually, interviews with 9 native TEFL-graduate teachers and 10 native non-TEFL-graduate teachers were recorded. Also, 10 interviews with Iranian teachers with related educational backgrounds and 8 interviews with Iranian teachers without related educational backgrounds were carried out. The recorded interviews were, on average, 13 minutes in length. The collected data were subsequently transcribed and prepared to be analyzed.

Data Analysis Procedure

Content analysis was employed to analyze the interviews. To carry out a constant content analysis, care was taken to define and develop proper units of analysis and coding schemes. Subsequently, the codes were transformed into categorical labels or themes when recurred as patterns in the observations or interviews. This procedure was proceeded incrementally up to data saturation and coherence and, eventually, conclusions. Finally, the emerging themes were analyzed descriptively.

Results

Non-TEFL-graduate Teachers' Perceptions

The results generated by the analysis of the interviews with the native and Iranian non-TEFL-graduate teachers indicated that both groups focused on similar qualities in speaking assessment. Each group, however, voiced the necessity of considering some speaking assessment criteria which were not taken into account by the other. The findings are discussed in detail in the following sections.

Iranian teachers

A close analysis of the Iranian non-TEFL-graduate teachers' responses indicated that they unanimously regarded learners' command of grammar, vocabulary, and pronunciation as well as their ability to speak fluently as essential qualities to focus on while assessing their speaking skill. However, they represented no consensus on the critical components of each skill. For assessing learners' pronunciation, for instance, it was unveiled that examination of their accuracy was perceived by each of the respondents to be vital, whereas the necessity to focus on their accents was voiced by only four of the respondents. Learners' ability to observe word stress was also tapped, though scantily, by the interviewees as a criterion to regard when learners' pronunciation is to be assessed. Abas, as the only one who bolded the significance of word stress, maintained that:

Another skill that I always focus on is my students' ability to use word stress correctly. In my opinion, it is something that is rarely noticed in the classrooms in our country. As you know, sometimes a shift in the place of stress can change the meaning of the word, which shows why it is so important.

The abilities to provide related responses to speaking prompts (emphasized by Najmeh, Arezoo, and Saman) and to communicate ideas clearly (voiced by Mozghan and Arezoo) were also marked by some of the IETs with irrelevant educational backgrounds. Table 3 provides a detailed representation of the major themes emerging from the Iranian non-TEFL-graduate participants' words about their interview rating criteria:

Table 3.

Major Themes and Coding Schemes on Important Oral Assessment Criteria as Perceived by the Iranian non-TEFL-graduate Teachers

Theme Codes	Sub-themes (n*)
Grammar	1. Accuracy (8)
	2. Range (7)
Fluency	1. Speed (8)
	2. Pauses (4)
Vocabulary	1. Appropriateness (8)
	2. Diversity (5)
Pronunciation	1. Accuracy (8)
	2. Accent (4)
	3. Stress (1)
Topic Management	1. Relevance (3)
Communicative Effectiveness	1. Clarity of Ideas (2)

* The number of times mentioned by the respondents

Native teachers

The results obtained from the analysis of the interviews with the native non-TEFL-graduate teachers indicated that they, like their Iranian peers,

unanimously agreed that grammar, fluency, vocabulary, and pronunciation were to be examined meticulously while assessing English language learners' oral communicative skills (see Table 4). However, closer scrutiny of Table 3 and Table 4 reveals that NETs with irrelevant educational backgrounds, in contrast to their Iranian colleagues, noticed their students' hesitations to evaluate their fluency in speaking. Their consideration of learners' ability to use idioms and expressions to convey their meaning was another noticeable point. In other words, three NETs who graduated in unrelated majors, unlike the IETs, claimed to scan their students' idiomatic use of language in speaking tests. Regarding the significance of comprehending and using idiomatic language, Jessica maintained that:

Native speakers' communications are full of idioms, expressions, and proverbs. So, I suppose it's really important to know what they mean by those idioms. I always encourage students who have just moved here to learn and use English idioms. And, even those who use idioms in their exam performances are given credits in my classes.

Further comparison of the results obtained from analyzing the words voiced by NETs and IETs with irrelevant educational backgrounds showed that students' accents were not taken as a concern to the NETs, in contrast to some of the Iranian teachers who pointed to learners' accents as a subject of examination in speaking tests. Concerning other differences between the two groups, it was found that only some of the native teachers tended to focus on their students' abilities to comprehend the oral test questions accurately and also "manage their talk efficiently within given times"(voiced by Jessica). Table 4 shows the qualities that native non-TEFL-graduate teachers normally

focus on while striving to assess speaking proficiency through classroom interviews.

Table 4.

Major Themes and Coding Schemes on Important Oral Assessment Criteria as Perceived by the Native non-TEFL-graduate Teachers

Theme Codes	Sub-themes (n*)
Grammar	1. Accuracy (10) 2. Range (8)
Fluency	1. Speed (10) 2. Pauses (4) 3. Hesitation (4)
Vocabulary	1. Appropriateness (10) 2. Diversity (5) 3. Idiomatic use (3)
Pronunciation	1. Accuracy (10) 2. Intonation (7) 3. Stress (2)
Topic Management	1. Relevance (9)
Comprehension	1. Understanding the speaking prompts (4)
Time Management	1. Timing the Talk (2)
Communicative Effectiveness	1. Clarity of Ideas (4)

* The number of times mentioned by the respondents

TEFL-graduated Teachers' Perceptions

Iranian teachers

In line with all of the participants with unrelated university degrees, the graduates in TEFL pointed, in complete agreement, to the commands of grammar, fluency, vocabulary, and pronunciation as four critical subjects of consideration and examination in speaking assessment.

A close comparison of the Iranian groups' perceptions indicated that all speaking assessment criteria voiced by IETs with irrelevant educational backgrounds were also considered by the other group (see Table 3 & Table 5). However, the TEFL graduates reported taking into account some other qualities while assessing their students' oral abilities through speaking tests, including learners' comprehension skills, time management, and confidence. Concerning the significance of learners' listening comprehension as a speaking assessment criterion, Elahe asserted that:

When I want to test my students' speaking skills, the first thing that I always pay attention to is whether they can comprehend and understand my question. If they do not understand my question, I expect them to ask me in English to repeat my question or clarify it. But anyway their misunderstandings can cause losing some score.

In the same way as three of her counterparts, Maryam perceived learners' hesitation in speaking as "an indication of poor fluency" and, unlike the Iranian teachers with irrelevant majors, cited it as an important speaking quality. Meanwhile, the results indicated that learners' ability to utter words with accurate stresses seemed more determining to the Iranian teachers who graduated in TEFL than those in other majors because more than half of them

pointed it out as a critical quality to regard in speaking assessment, whereas only one of the Iranian non-TEFL-graduate teachers tapped into it.

Along with giving relevant responses, providing adequate supportive ideas was stated by six IETs who graduated in TEFL to be required to manage a speaking topic. Narges, in this regard, declared that:

The quality of a student's speaking mainly depends on his ability to support and expand what he claims. Any speaker should be able to elaborate on his ideas and try to convince the listener or examiner by providing enough information that is related to the question that has been asked by the examiner.

Table 5 shows the main themes and coding schemes emerging from the interviews with the Iranian teachers who completed their higher educations in TEFL.

Table 5.

Major Themes and Coding Schemes on Important Oral Assessment Criteria as Perceived by the Iranian TEFL-graduate-teachers

Theme Codes	Sub-themes (n*)
Grammar	1. Accuracy (10) 2. Range (9)
Fluency	1. Speed (10) 2. Pauses (7) 3. Hesitations (4)
Vocabulary	1. Appropriateness (10) 2. Diversity (9)
Pronunciation	1. Accuracy (10) 2. Stress (6) 3. Intonation (4) 4. Accent (4)

Topic Management	1. Relevance (8) 2. Adequacy of Supportive Ideas (6)
Communicative Effectiveness	1. Clarity of Ideas (8)
Comprehension	1. Understanding speaking prompts (5)
Time Management	1. Timing the Talk (2)
Confidence	1. Stress-free speech (2)
Non-verbal behavior	1. Eye Contact (2) 2. Body Language (2)

* The number of times mentioned by the respondents

Native teachers

Grammar, fluency, pronunciation, and vocabulary were the subjects that every native TEFL-graduate teacher, like all other participants, referred to as some significant qualities to regard in classroom speaking assessment. To emphasize the significance of employing accurate language forms in carrying on a well-structured oral performance, Sophia argued that:

I strongly disagree that students should be merely judged based on their ability to convey their meaning, no matter how accurate they are. Personally, I'm not interested in listening to or even rating speaking performances that are replete with structural mistakes and errors. I prefer an accurate but non-fluent performance rather than the opposite.

However, one distinctive point found in the interviews with the native TEFL graduates was that they, in contrast to the other groups, noticed how an examinee's "repetitions might impede [her] fluency in speaking and [consequently] affect her exam score negatively" (voiced by Jacob). Besides, the application of communicative strategies, efficient organization of ideas, and affective variables were three other speaking assessment criteria that were mentioned and focused on solely by the NETs with related majors. More

specifically, six of them gave credits to test-takers' ability to "compensate for [their] lack of knowledge in words or forms through communicative strategies" (quoted from Lily) through, for instance, "circumlocution or giving definitions" (voiced by Sophia). To them, a qualified speaker was required to initiate and terminate her words efficiently and develop her talk through interconnected ideas. Unlike the Iranian teachers, they made no points indicating whether they assigned any credits to their students' native-like accents in speaking. Table 6 shows, in brief, the criteria that the NETs with related majors focus on while assessing their students in the classroom context.

Table 6.

Major Themes and Coding Schemes on Important oral Assessment Criteria as Perceived by NETs With Related Educational Backgrounds

Theme Codes	Sub-themes (n*)
Grammar	1. Accuracy (9) 2. Range (9)
Fluency	1. Speed (9) 2. Pauses (7) 3. Hesitations (4) 4. Repetition (4)
Vocabulary	1. Appropriateness (9) 2. Diversity (9) 3. Idiomatic usage (4)
Pronunciation	1. Accuracy (9) 2. Stress (5) 3. Intonation (4) 4. Rhythm (2)
Topic management	1. Relevance (7) 2. Adequacy of supportive ideas (5) 3. Coverage (4)
Communicative effectiveness	1. Clarity of ideas (8)

Communicative strategies	1. Compensation (6) 2. Avoidance (3)
Comprehension	1. Understating speaking prompts (4)
Time management	1. Timing the Talk (4)
Confidence	1. Stress-free speech (3)
Organization	1. Initiation, development, and termination (2) 2. The interconnectedness of ideas (2)
Non-verbal behavior	1. Eye contact (2) 2. Body language (2)
Affective Variables	1. Emotion and Engagement (2)

* The number of times mentioned by the respondents

Discussion

The present study sought to explore the classroom speaking rating criteria as perceived and voiced by native and non-native English teachers with related and unrelated degrees in higher education. The results showed that structural features of language, including grammar, vocabulary, and pronunciation, were the most frequently mentioned rating criteria used by the participating non-native teachers in rating oral performances. That is, all of the IETs with relevant and irrelevant educational backgrounds attended to the structural resources of language as a justification for their ratings. Such findings seem to support those revealed in previous studies (see e.g., Brown et al., 2005; Kim, 2009; Plough et al., 2010; Zhang & Elder, 2011) where non-native teachers were found to be more critically oriented towards specific linguistic features of speaking while rating their students' oral performances. Likewise, the results of the present study also showed that the participating native teachers tended unanimously to focus on examinees' masteries of grammar, vocabulary, and pronunciation while assessing their speaking ability. This is not in line with the results documented by Kim (2009) and

Zhang and Elder (2011), who found native teachers to be more concerned with message-focused criteria rather than linguistic-oriented ones.

Zhang and Elder (2011), in their study, indicated substantial differences between native and non-native English teachers regarding the saliency they attributed to the content features of speaking and, specifically, learners' ability to utter what they mean fluently. Nonetheless, the current study indicated that both NETs and IETs comparably pointed to learners' ability to converse their intentions fluently as a critical ability to focus on while assessing their oral productions.

Although all participants with varying first languages and educational backgrounds focused on similar structural rating criteria in general, the results revealed that they considered different sub-features or sub-skills for each construct. As argued by Zhang and Elder (2011, p.43), "this convergence with respect to overall standards does not indicate that the groups are interpreting the oral proficiency construct in the same way". Examining pronunciation as an example, the findings unveiled that IETs, unlike NETs, assigned credits to learners' native-like accents. The native TEFL-graduate teachers, in contrast to the other groups, were sensitive about learners' rhythms in speech. The IETs with irrelevant educational backgrounds did not attend to parodic features of pronunciation such as intonation, rhythm, and stress. Such findings are in accord with those concluded by Kim (2009), who showed in his study that native and non-native teachers tended to focus on different features of speaking in the area of pronunciation.

Except for the non-English major IETs who mainly resorted to structural criteria to rate their students, all of the participating groups extended their assessment criteria to message-focused and communicative aspects of speaking. However, it appears that training and educational backgrounds play

more important roles than *nativeness/non-nativeness* in helping the teachers form a more profound understanding of speaking construct because the native and non-native participants who graduated in TEFL took into account some factors (e.g., adequacy of supportive ideas, organization, communicative strategies, non-verbal behaviors, and confidence) which were seemingly neglected even by the native teachers with different majors. This supports the findings in some studies (e.g., Caban, 2003; Kim, 2009; Tajeddin et al., 2011; Zhang & Elder, 2011) where trained teachers were reported to attend more to the pragmatic categories of content and fluency. In brief, it appears that the training the native and non-native participants had received in universities helped them develop a communicative view of language "in which both structural and pragmatic aspects of language should be attended to" (Tajeddin et al., 2011, p. 148).

The different orientations in rating oral interviews by the TEFL-graduate and non-TEFL-graduate teachers might be associated with McNamara's (1996) distinction between the weak and strong approaches to the performance evaluation. In the weak approach, the test is solely used to judge one's mastery of linguistic knowledge, whereas in the strong approach, it is a vehicle for evaluating the examinee's capacity to fulfill a given communicative task which typically involves a representation of linguistic and pragmatic factors. It accordingly seems that doing a degree in a related field of study has induced the TEFL-graduate teachers to take a strong approach to speaking assessment. It supports Zhang and Elder (2011), who argued that raters' attention to, for instance, test-takers usage of communicative strategies is indicative of their tendency to carry out an assessment task based on candidates' ability to "meet the requirements of a communicative task that they might be required to carry out in a real-world encounter" (p. 43).

Conclusions and Implications

The current study was an investigation of the classroom speaking assessment criteria used by NETs and IETs with and without related educational backgrounds. In other words, it was an attempt to explore to what extent English language teachers with varying first languages and fields of study in higher education focused on different qualities while assessing their students' oral performances. Attentive analysis of the accumulated data suggested that both NETs and IETs (with and without related majors) assigned much weight to the structural features of language in general. However, they focused on different qualities and features when assessing their students' grammar, vocabulary, and pronunciation masteries.

The four groups of participants also pointed to student's fluency as a critical speaking rating criteria. However, some of them defined the fluency construct in different ways and attributed varying qualities to fluent speaking. For example, the native TEFL-graduate teachers, unlike the other groups, argued that they assigned negative point values to test-takers' repetitions while rating their speaking because, in their views, repetitions impede fluent oral production. Also, non-English major IETs judged students' speaking fluency solely based on their rate of speech and pauses and did not lend much weight to their hesitations or repetitions in speaking.

It was also revealed that the native and non-native teachers who had completed their higher education in TEFL had a better understanding of the speaking construct and took into account several message-based and pragmatic factors that were neglected by their non-English major colleagues. More specifically, the non-English major IETs mainly focused on linguistic aspects of oral production, and fewer pragmatic criteria were taken into consideration by NETs with irrelevant majors than the teachers with related

majors. Therefore, it appears that the English major teachers, regardless of their first languages, could represent a more communicative perspective of the language assessment through focusing on both structural and pragmatic aspects of oral production.

The results of the present study may have some practical implications. They, for instance, may raise classroom English teachers' and researchers' awareness of the qualities that language teachers with varying first language and educational backgrounds focus on while assessing their learners' oral abilities. Given the critical role of education in training more insightful language proficiency raters, as unveiled by the current study, untrained language teachers might be encouraged to do a degree in Applied linguistics or TEFL in order to grow a more profound understanding of speaking constructs and assessment. Further, the qualities which were perceived to be critical by the participants might be discussed in-depth in training programs which are concerned with teachers' assessment literacy development. Last but not the least, the results may also contribute to growing the body of knowledge in the related literature.

References

- Agasøster, S. (2015). *Assessment of Oral English. A study of Assessment Practice of Oral English at Lower Secondary Schools in Norway*. Unpublished Master's thesis, The University of Bergen.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. S. (1990). The construct of the FSI Oral Interview. *Language Learning*, 31(1), 67-86.
- Black, P., & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. *Yearbook of the National Society for the Study of Education*, 103(2), 20-50.

- Brown, A. (2004). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20, 1-25.
- Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-academic-purposes speaking tasks*. ETS Research Report No. RR-05-05. Princeton, New Jersey: Educational Testing Service.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (Vol. 10). White Plains, NY: Pearson Education.
- Butler, F. A., Eignor, D., Jones, S., McNamara, T., & Suomi, B. K. (2000). *TOEFL 2000 speaking framework*. Princeton, NJ: Educational Testing Service.
- Caban, H. L. (2003). Rater group bias in the speaking assessment of four L1 Japanese ESL students. *University of Hawai'i Second Language Studies Paper 21* (2).
- Correia, R. C. (2016). Assessing Speaking Proficiency: A Challenge for the Portuguese EFL Teacher. *e-TEALS*, 7(1), 87-107.
- Cumming, A. (2009). Language assessment in education: tests, curricula, and teaching. *Annual Review of Applied Linguistics* 29, 90-100.
- Davison, C. and Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly* 43, 393-415.
- Downer, J. T., Booren, L. M., Lima, O. K., Luckner, A. E., & Pianta, R. C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early childhood research quarterly*, 25(1), 1-16.
- Fan, J. (2018). Oral Interaction: Concept, Competence, and Assessment. In K. Soh (Ed.), *Teaching Chinese Language in Singapore* (pp. 83-97). Singapore: Springer.
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
- Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5-29.
- Gipps, C. (1994). Quality in teacher assessment. *Enhancing quality in assessment*, 71-86.
- Hughes, A. (2003). *Testing for language teachers*. Ernst Klett: Sprachen.
- Isaacs, T. (2016). Assessing speaking. In D. Tsagari & J. Banerjee (Eds.), *Handbook of second language assessment* (pp. 131-146). Berlin: DeGruyter Mouton.

- Isaacs, T., & Trofimovich, P. (2012). Deconstructing comprehensibility: Identifying the linguistic influences on listeners' L2 comprehensibility ratings. *Studies in Second Language Acquisition*, 34(3), 475-505.
- Khoshsima, H., & Afiati, L. (2015). English and Non-English major Teachers' Assessment of Oral Proficiency: a case of Iranian Maritime English Learners. *Iranian Journal of English for Academic Purposes*, 4(1), 33-45.
- Kuiken, F., & Vedder, I. (2014). Raters' decisions, rating procedures, and rating scales. *Language Testing*, 31(3), 40-61.
- Kuschmann, A., & Lowit, A. (2015). The role of speaking styles in assessing intonation in foreign accent syndrome. *International journal of speech-language pathology*, 17(5), 489-499.
- Lado, R. (1961). *Language Testing: The Construction and Use of Foreign Language Tests*. A Teacher's Book.
- Larson, J. W., & Hendricks, H. H. (2009). A context-based online diagnostic test of Spanish. *Calico Journal*, 26(2), 309-323.
- Lee, S. (2010). *Current practice of classroom speaking assessment in secondary schools in South Korea*. Unpublished MA thesis. The University of Queensland, Australia.
- Luoma, S. (2004). Assessing speaking (Cambridge language assessment). *nature*, 1, 2.
- Lumley, T., & McNamara, T. F., (1995). Rater characteristics and rater bias: implications for training. *Language Testing*, 12, 4-71.
- McNamara, T.F. (1996). *Measuring second language performance*. London and New York: Addison Wesley Longman.
- Mertler, C. A. (2016). *Classroom assessment: A practical guide for educators*. Routledge.
- Phung, d. (2018). *Variability in teacher oral English assessment decision-making* [Unpublished doctoral dissertation]. University of New South Wales, Sydney.
- Plough, I. C., Briggs, S. L., & Van Bonn, S. (2010). A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. *Language Testing*, 27(2), 235-260.
- Qashoa, S. H. (2012). Effects of teacher question types and syntactic structures on EFL classroom interaction. *The International Journal of Social Sciences*, 7(1), 52-62.

- Rezvani, R., & Sayyadi, A. (2015). Instructors' and learners' questioning: A case of EFL classroom discourse. *Journal of Teaching Language Skills*, 34(3), 141-164.
- Salaberry, R. (2000). Revising the revised format of the ACTFL Oral Proficiency Interview. *Language Testing*, 17(3), 289-310.
- Staples, S., Laflair, G. T., & Egbert, J. (2017). Comparing language use in oral proficiency interviews to target domains: Conversational, academic, and professional discourse. *The Modern Language Journal*, 101(1), 194-213.
- Tajeddin, Z., Alemi, M., & Pashmforoosh, R. (2011). Non-native teachers' rating criteria for L2 speaking: Does a rater training program make a difference? *Teaching English Language*, 5(1), 125-153.
- Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal*, 60(1), 51-60.
- Thuy, N. H. H., & Nga, T. T. T. (2018). An investigation into EFL teachers' perceptions of in-class English speaking assessment. *VNU Journal of Foreign Studies*, 34(2).
- Wang, Y., & Yu, C. (2017). Social interaction-based consumer decision-making model in social commerce: The role of word of mouth and observational learning. *International Journal of Information Management*, 37(3), 179-189.
- Weir, C. J. (1990). *Communicative language testing*. New York: Prentice-Hall.
- Winke, P., Gass, S., & Myford, C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, 30(2), 231-252.
- Willis, D., & Willis, J. (2007). *Doing task-based teaching*. Oxford: Oxford University Press.
- Zahedi, K., & Shamsaee, S. (2012). Viability of construct validity of the speaking modules of international language examinations (IELTS vs. TOEFL iBT): evidence from Iranian test-takers. *Educational Assessment, Evaluation and Accountability*, 24(3), 263-277.
- Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing*, 28(1), 31-50.

Appendix

Interview Questions

1. Do you use oral interviews to assess your students' speaking skill in classrooms?
2. What are your rating criteria while assessing your students' speaking skill through oral interviews?

