Listening and Speaking for Real-World Communication: What Teachers Do and What Students Learn From Classroom Assessments

Melissa H. Yu¹, Barry Lee Reynolds¹,², and Chen Ding¹

Abstract
Even though standardized international communication tests have been frequently studied, very little research has explored how teachers planned listening and speaking classroom assessments or which classroom-based tests were more beneficial for teaching real-world English communication. A qualitative inquiry was undertaken to investigate these assessment issues among five English as foreign language teachers and 24 of their students through the collection and analysis of classroom observation and post assessment interview data. While teachers tended to draw on textbook listening and speaking activities to assess those skills, how they graded students focused heavily on the students’ communicative competence as listeners and speakers of English rather than on their ability to answer comprehension questions correctly in the classroom assessments. Students identified a mismatch between classroom instruction and assessments and also a mismatch between the English used in assessments and the English used in real-world communication.

Keywords
EFL classrooms, classroom assessment, real-world communication

Introduction
Scholars have often used the concept of English as an international language (EIL) to criticize standardized language tests. Their discussions have centered on how inadequately measurement descriptors have been defined (e.g., Jenkins & Leung, 2017) and how international, regional, or national test systems have influenced English as a foreign language (EFL) teaching and learning (Hamid, 2014; Kim, 2019; McNamara, 2014; Zhang & Elder, 2014). In contrast, classroom-based tests and feedback on student language performance have not received equal attention in the EIL field (Nguyen et al., 2020; Yu, 2018). Accordingly, the classroom perspective of this study represents a new aspect of how EFL teachers assess students’ English.

Very few studies have scrutinized how EFL teachers have planned and implemented classroom assessment activities, how students have used language in assessment activities, or how students have perceived the English used in assessments (Hill & McNamara, 2011). “Teachers are involved at all stages of the assessment cycle” (Davison & Leung, 2009, p. 401), so documenting how EFL teachers develop tests and assess student performance is important. Most students’ English is tested by teachers with teacher-constructed assessments (Nation, 2013).

In traditional EFL classrooms, assessments are often regarded as a way to facilitate learning, but not necessarily for communication. For example, there is considerable discussion about the positive or negative washback effects of tests on learning English for international communication (e.g., Mckinley & Thompson, 2018). From an EIL perspective, McKay and Brown (2016) think that “unlike standardized tests, classroom tests are typically used to assess what students know or can do in the language with reference to what is being taught in a specific classroom or program” (p. 80). The current study aims to explore the relationship between assessments and language learning for real-world communication, positing that assessment should focus on what students know about English for real-world communication and what students can do in English as it relates to what they have been taught in the classroom.

¹University of Macau, Faculty of Education, Taipa, Macau SAR, China
²University of Macau, Centre for Cognitive and Brain Sciences, Taipa, Macau SAR, China

Corresponding Author:
Barry Lee Reynolds, Faculty of Education, University of Macau, Room 1017, E33, Av. da Universidade, Taipa, Macau SAR, China.
Email: BarryReynolds@um.edu.mo
Lowenberg (1993) was the first to question the use of native-speaking (NS) standards to test EFL students’ language. In 2006, Jenkins and Taylor continued the debate about the well-accepted NS normative approach to measure students’ language performance. Jenkins (2006) first suggested accepting certain non-native speaking (NNS) English pronunciation and grammatical uses within the NS normative approach of language testing. However, Taylor (2006, p. 52) questioned Jenkins, asking whether it is appropriate to suggest which NNS features are acceptable without first clarifying “which NS” variety was being discussed. After a decade, the NS, monolingual normative approach remains the central issue to be addressed in studies on language assessments (e.g., A. Brown, 2013; McNamara, 2014; Saito, 2019).

Regarding the content of assessments, World Englishes or EIL scholars recommended that varieties of English should be incorporated into English language tests (e.g., J. D. Brown, 2004; Yamaguchi, 2019). Nevertheless, Canagarajah (2006, p. 236) disapproved of these previous suggestions because this may reinforce or encourage the use of “a single variety” of English. In a dilemma over what English should be included in classroom assessments (Tomlinson, 2010), one pedagogical concern for EFL teachers to address is whether or not to incorporate varieties of English into classroom assessments. The second concern for EFL teachers is how they should grade the variety of English that students use during assessment activities (Ambele & Boonsuk, 2020). Finally, from the students’ point of view, EFL teachers should consider which elements of classroom assessments are relevant to real-world communication (e.g., Ambele & Boonsuk, 2020; Lee & Ahn, 2020).

To address the concerns and questions arising from the literature reviewed above, three questions guided this research:

**Research Question 1 (RQ1):** On which test resources did teachers rely when planning speaking-related assessment activities?

**Research Question 2 (RQ2):** What were the main criteria teachers used to rate their students’ speaking skills?

**Research Question 3 (RQ3):** How did students perceive English language use in the assessment activities, as it contrasted with English for real-world communication?

**Method**

To answer the questions above, a qualitative inquiry was carried out. This section covers how participant sampling was undertaken, followed by data collection and analysis.

**A Qualitative Inquiry**

Convenient sampling (Creswell & Poth, 2018) was employed to recruit five teachers and 24 of their students from three universities in Taiwan. Teacher 1 (T1) taught *Oral Training* courses to the first- and second-year English majors in University 1 (U1); the other teachers taught the first-year non-English majors in University 2 (U2) and University 3 (U3) *Freshman English* courses. T1 from U1 had taught English for more than 10 years whereas the teachers from U2 and U3 had only taught English for 2.5 years. U1 students had received an International English Language Testing System (IELTS) score of at least 7 (The Common European Framework of Reference for Languages [CEFR] C1 equivalent). U2 students had received IELTS scores between 5 and 6 (CEFR B1-B2 equivalent). Half of the U3 students had passed the elementary level of the General English Proficiency Test (GEPT) (CEFR A2 equivalent) while the other half had passed the intermediate level of the GEPT (CEFR B1 equivalent). Students received 4 hr of *Oral Training or Freshmen English* per week. The main reason for recruiting these teachers and their students was the teachers claimed a teaching objective of enhancing students’ English for communication. Other methodological reasons for sampling these participants are given below.

First, the content and foci of assessment events may vary according to tests of different types, test writers, the content, process, or contexts of testing, students’ levels of English, etc. (Cheng et al., 2008). In this case, the diversity of assessment activities required the present study to provide a broader scope through which classroom assessments were planned, implemented, and marked by different teachers for different courses and how the assessments related to real-world communication. Second, without exploring, comparing, and contrasting cases of classroom assessment events, there would have been insufficient evidence for a researcher to judge which assessment activity introduced more EIL elements than the others. Third, assessment activities were carried out across teaching contexts. Under different circumstances, we need to attend to the “local situations, raters, and norms” in each classroom, as Canagarajah (2006) suggested. By doing so, this study yielded a better understanding of cause–effect relationships between planning listening and speaking classroom assessments and assessment results. Finally, the validity of this qualitative inquiry was enhanced through multiple perspective mixing. For instance, this research considered classroom assessments from the viewpoint of test planners (i.e., teachers) versus test takers (i.e., students), one teacher versus the others, traditional EFL versus EIL, and classroom versus standard assessments (Cohen et al., 2018).

T1 required students to give three oral presentations as well as participate in class discussions related to these presentations to assess the students’ oral English. While T1 specifically spelled out particular behaviors that were expected to receive participation and discussion points (40% of the course grade), the requirements for the three oral presentations were not explicitly stated. T1’s assessment approach could be considered as communicative due to the emphasis she placed on the discussion about presentations and her role in encouraging critical thinking and
Table 1. Summary of Assessment Activities Arranged by Teacher.

| Teacher/Activity element | Assessment activities | T1 | T2 | T3 | T4 | T5 |
|--------------------------|----------------------|----|----|----|----|----|
| Assessment activities    | Short speech, individual presentation, group discussion for group report, debate | Individual presentation (mid-term exam) | Individual presentation (final exam), group discussion, and report (whole-class assessments) | Giving a summary (whole-class assessment) | Teacher and student Q&A |
| Duration                 | 1,200 min (480 min and 720 min) | 200 min | 50 min | 100 min | 100 min |
| Class size               | 17^a and 13^b | 55 | 12 | 33 | 32 |
| References used for planning assessment activities | Textbook, TOEFL iBT, IELTS | Textbook, “own ideas” |

Note. GEPT = General English Proficiency Test; TOEFL iBT = Test of English as a Foreign Language Internet-Based Test; IELTS = International English Language Testing System.

^aOral Training I course. ^bOral Training II course. ^cNot analyzed. ^dDuration includes time from the beginning to the end of the teachers’ assessments, which at times included small breaks given to students.

Active listening. T4 assessed learners in ways similar to T1. T4 would ask students to volunteer to provide feedback on student presentations by first summarizing the presentation and then asking follow-up questions to the presenter. T4 used linked-skill activities to build background knowledge on presentation topics before asking students to present on a topic. For example, students might read articles and watch videos on a topic before being expected to give a presentation on the topic.

Teacher 2 (T2), Teacher 3 (T3), and Teacher 5 (T5)’s assessment approach could be considered structuralist. T2 provided learners with a rubric that clearly indicated the five components that would be evaluated when students gave 3-min presentations (see Online Appendix 4). The speech components were evaluated with questions that could receive 1 to 5 points. Students did not know what the topic of their presentations would be until the day of the presentation, a sort of impromptu speech style. T3 provided rather structured lessons in which students were lectured on the structure of a presentation that would be expected for the Test of English as a Foreign Language Internet-Based Test (TOEFL iBT) exam and given time to prepare a self-selected presentation topic outside of class before giving their presentation in class. T5 was similar to T3 in that lessons were structured around how to give a speech based on the requirements of the GEPT. T3 also encouraged students to select topics that were similar to those that appear on the GEPT. T2, T3, and T5 taught in a manner that aligned more with standardized exams by framing their assignments to mimic the type of tasks that appeared on these exams. However, the content was not always the same as what would appear on such exams. In this regard, the teachers still provided opportunities for students to express themselves in their own way about topics of their own choosing with the caveat of structuring the presentations to fit the structure of the standardized exams. Using the standardized exams as a structure or framework for instruction would not force language use beyond students’ current language proficiency. As it was only the structure and not the content that was aligned with the exams, cultural bias should not have influenced students’ interests or abilities.

Data Collection and Analysis

Classroom observation and post-assessment interviews were conducted to collect data. Recognizing how classroom observation is useful for capturing the reality of teaching and learning behaviors (Cohen et al., 2018), the first author conducted three-and-half-months of participant observation. Altogether, 57.5 hr of class time was audio-recorded, but not all the content of that 57.5 hr involved assessment events. This article only focused on the 27.5 hr of listening and speaking classroom assessments. Table 1 provides an overview of the assessment activities, duration, class size, and assessment references used by the teachers.

Echoing Blommaert and Dong’s (2010) concern about the inherent pitfalls of de-contextualized interviews, the interviews with teachers focused on the classroom assessment activities that teachers and/or students actually engaged in, rather than imagined ones for interviewees to elaborate on. A similar strategy was also applied while interviewing students about their perceptions of the English presented in assessment activities and in real-world communication. Online Appendices 1 and 2 present the interview questions used with teachers and their students.

Verbatim transcriptions of audio-recorded assessment activities and interviews were carried out. Pseudonyms were used to ensure the confidentiality of participants. Online Appendix 3 illustrates the transcription conventions. The qualitative content analysis coding began with exploring the existing themes derived from the research questions and literature review, re-reading the transcripts to quantify and identify the most-frequently cited themes, and identifying the common themes between the former and the latter (Schreier, 2012). The reliability (i.e., consistency) of the coding was checked using an intra-coder approach. According to Schreier (2012), for qualitative content analysis, the most
suitable coefficient is the simple percentage of agreement. Dividing the number of agreements between Time 1 and Time 2 coding yielded an intra-rater reliability rating of .99, an acceptable level (Miles & Huberman, 1994). The remaining inconsistencies were reviewed by the second author and then discussed with the first author before finalizing the coding. Based on the initial themes, an index framework was constructed by identifying the linkage between themes and subthemes. From exploring the themes to establish an index framework, the content analysis of data resources was carried out (Richie et al., 2003). Within the established index framework, passages with reference to the themes were labeled and sorted. The last stage of coding included thematic charting and summarizing indexed categories.

Following this process of coding, the index framework of this study was established. It consisted of two main themes and six subthemes used to answer the research questions. The first theme considered what assessment activities were developed and used, such as how teachers planned assessment activities and how students’ speaking skills were assessed and rated. The second theme explored students’ perceptions of English language use presented in the assessment activities and the real-world communicative language use. All the direct citations from participants were selected based on their relevance to the research questions.

### Findings and Discussion

**RQ1: On Which Test Resources Do Teachers Rely When Planning Speaking-Related Assessment Activities?**

Table 2 and the discussion below reveal where the five teachers obtained the idea of planning assessment activities. Table 2 illustrates that textbook-based speaking activities were the predominant resource teachers used to plan assessment activities. T1 indicated, “There are many activities in the textbooks. I will select some interesting activities for students to engage in.” The other teachers expressed similar opinions about selecting or adapting textbook activities to assess students’ speaking skills rather than planning speaking assessments on their own.

Teachers also planned speaking assessment activities based on ideas from the national and international English language tests (Table 2). Three out of five teachers (T2, T3, and T5) indicated that some ideas they incorporated into their speaking assessment activities came from their experiences taking standardized English language proficiency tests. For instance, T2 and T3 indicated that some of their assessment activities were inspired by the TOEFL iBT and IELTS, while T5 crafted final exam speaking activities similar to those in the GEPT. Drawing on ideas from the TOEFL iBT, T2’s students were given “a statement” and “15 s to take notes and prepare for the speaking test”; she required her students “to present their agreement or disagreement with the statement.” Similarly, T3 used an idea from the IELTS speaking tests, indicating that her “students will be given a topic, 1 min to prepare, and 2 min for speaking.”

These examples echo Berger’s (2012) observation that some in-service teachers drew on their experience of taking standardized tests when designing assessment activities. Furthermore, McKinley and Thompson (2018) took the perspective of international communication to argue against NS and monolingual approaches to design standardized examinations. While it appeared that these local teachers also drew on their experiences to design tests for their students, the interview data indicated that the local teachers simply used the standardized examinations as guides for selecting item formats or assessment techniques.

### Which Main Criteria Did EFL Teachers Use to Assess Language Skills?

Now the discussion focuses on how students engaged in the speaking assessment activities and teachers’ grading. Extract 1 below exemplifies a classroom assessment activity planned and implemented by T1.

**Extract 1.** T1’s student’s presentation (T1S1 as presenter; T1S2 as audience to give comments or ask questions).

1. **T1S1** . . . thank you.
2. 1 . . . [T1 made comments on T1S1’s presentation first.] So, what do you guys think?
3. **T1S2**. what do you think about her [T1S1’s] speech?
4. 4 **T1S2**. Hm, () I like her first picture, World Map. Uh (;) however, I think <<she can>>, Um (), she make eye contact . . . and conclusion. . . . I prefer using words, key words.
5. 6 **T1** . . . She [T1S1] used the whole text [in the Conclusion slide]. So, she can use bullet points . . .

As Online Appendix 4 shows, one criterion T1 used when grading presentations was “You present yourself as a critical thinker with a good command of the English language.” When asked about it, T1 pointed out that “the content” [of
students’ presentations, such as] “the relevance of reference resources to their presentation topic,” as the main criteria to decide whether students presented themselves as critical thinkers. Then, she added, “The teacher is a listener and should focus on the content of what the student is saying.” Regarding how she judged whether students “have a good command of the English language” for real-world communication, she said, “In fact, a unit of our textbook indicates that speakers are listeners as well, suggesting students should be active listeners and how to do so.” She went on to explain.

I ask students questions to provide a communicative context. I also engage students as they listen to each other talk by raising questions to the presenters or making comments on their partners’ words. By doing this, I can get students involved. At the same time, I can get students involved in the speech by providing listeners opportunities to speak.

She concluded, “When students get involved in each other’s presentations, they are engaging in real-world communication contexts.”

Extract 2 is another example of giving a presentation as the major speaking classroom assessment.

Extract 2. T2’s student speaking during a presentation.

T2S1 . . . hello, everybody. Today, I will talk about student association, which I enjoy, eh, (?) I join in. My name is . . . My major is . . . Civil Engineering. And this is my outline, first, I will(,) tell you who we are, and second, . . . and the last, our principle. And this is our . . . this my um our clothes (he pointed at the pictures of where the student association is based and the association uniform), . . . The meaning of two words (Chinese characters), 服務, service /ˌsəˈvaɪs/ (service should be pronounced as /səˈvaɪs/ according to NS norms). Do you know how to explain the Good Life service /ˌsəˈvaɪs/? ((audience: @@ (some learning partners nodded, others laughed, and others said yes)) . . . look at . . .

As illustrated, T2S1’s pronunciation revealed linguistic deviation from NS linguistic norms but T2 did not correct T2S1’s pronunciation during or after the exam. T2’s other students also listened and responded to T2S1. During this 1-hr oral exam, T2’s other students’ presentations also presented English pronunciation and grammatical use beyond the NS normative scope. However, T2 maintained her stance against correcting students’ English according to NS linguistic norms. Responding to the researcher’s question about how T2 felt about the pronunciation of her other student who intended to pronounce the word camp /kæmp/ but ended up pronouncing can’t /kænt/ in the exam, T2 replied that Um, . . . that male student, his word, camp, he pronounced as can’t /kænt/, but this (.) [she showed the researcher this student’s score on the screen], I did not deduct his point. @@ . . . Many students had similar problems. I did not deduct their points, either. I only paid attention to the main structure.

The second author looked at the screen, pointed out one student who got a perfect score and said “wow, 100 points.” T3 replied, “Yes, we teachers are very tolerant of students’ language use@@.” The results showed that T2’s grading had more to do with how students structured their speeches, and less to do with whether students’ English language use conformed to NS-based accuracy (see Online Appendices 4 and 5). T2’s grading and descriptors represented a local way to measure students’ English language use and locally developed grading rubrics (see Online Appendices 4 and 5). Clearly, this way differed from the NS normative approach of developing test descriptors or measuring students’ linguistic accuracy, which has been criticized by Rose and Syrbe (2018).

Teachers recommended that speaking be graded according to what students say/content (T1, T2, T3, T4, and T5), how confident they are (T2, T3, T4, and T5), how comfortably they are with making mistakes (T1, T2, T4, and T5), and how they negotiate meaning through interaction (T1, T2, T3, T4, and T5). Online Appendix 4 illustrates the examples of the teachers’ assessment descriptors. The following paragraph discusses examples of each teacher’s grading priority or major grading principle.

T1’s grading method prioritized what students said and whether they “got involved in each other’s presentations.” In T2’s case, she emphasized “how students structured what they said” or “whether they raised questions to interact with the audience” as the foci of her grading. In the interview, T4 commented, “Without a doubt, those students whose English was good will receive high scores, but this does not mean that students will receive low grades because of their poor English.” T4 added, “If students do not say anything, I cannot give them grades . . . If [students’] English is not good, as long as they studied hard to prepare for speaking, grading will not focus much on how good their English is.” When the researcher asked T5 about his students’ use of non-native English in assessments, T5 repeated the word accept three times to emphasize that he had no problem with his students’ English lacking NS qualities or linguistic accuracy. He added, “You were there too, you saw it . . . I helped [the examinees] answer my questions when they did not understand.” The results showed that none of the five teachers’ grading criteria encouraged NS linguistic accuracy or NS normative conformity. While scholars have claimed that EFL students should achieve NS linguistic accuracy when assessed by international or national exams (Mckinley & Thompson, 2018; Rose & Syrbe, 2018), the current study’s findings indicated that the local teachers did not expect NS accuracy inside their classrooms.

The assessment criteria prioritized by these five teachers when grading students’ spoken English were generally
congruous. First of all, NS linguistic quality and accuracy were not prioritized or used as essential criteria when grading students’ English. This finding does not support A. Brown’s (2013) observation that international English assessments prioritize NS English use. Second, students’ use of NNS English did not necessarily equate with lower grades. Third, as Jenkins (2000) suggested, these teachers accepted students’ NNS English on the proviso that it is interactive, conveys meanings, and/or helps other students engage in speaking assessment activities. This finding also aligned with the advice given in the previous literature, indicating that interaction/performance-based test activities facilitated students’ communicative language use and skills (Rea-Dickins, 2008; Waring, 2008). Overall, the NNS deficient perspective, NS-NNS dichotomous approach, and NS linguistic accuracy were not prioritized in these teachers’ speaking assessment activities and grading criteria.

RQ3: How Do Students Perceive English Language Use in the Assessment Activities in Contrast With English for Real-World Communication?

Regarding student perception, the first author asked students to reflect on any perceived contrasts among the English they, their teachers, or others use. All agreed that the English presented in classroom assessment activities was different from “their own English” or “the English that they had heard in communicative contexts.”

T1 indicated that getting students “involved in each other’s presentations” was a kind of real-world communication. Holding a different point of view, T1S4 pointed out, “When I went back to the dormitory talking to my Portuguese friend, I did not speak like I did when I presented or gave a speech or had a conversation or discussion in the classroom.” However, she did not deny that her English language proficiency had become better and she was “more able to communicate with others fluently” in non-language assessment contexts. Similarly, T2S2 indicated that when conversing with his grandmother’s Filipino health assistant, he found her English different from classroom English, yet T2S2 could communicate with her with few problems. As T3 and T4 taught a mixed group of local and international students, T3S1 exemplified that “the English spoken by international students is different from the teachers’ English.” T4S1 held a similar view, indicating his English was different from internationals. T3S2 said, “Those international students gave presentations with some incorrect English, but they still spoke so well. Even though their English had grammatical mistakes, I understood what they were saying when we conversed for group assignments.” Regarding T5’s Q&A assessment activity, T5S1 emphasized that the “teacher’s English sounds like the English that is used to converse with people and doesn’t sound like a listening test.” Nevertheless, T5S1 insisted, “What was tested was still different from what was used for communication.”

The general consensus among students’ views on English presented in assessment activities and communication was that they were not the same. However, English produced for classroom assessments did not necessarily hinder students from communicating with others for non-assessment purposes despite the differences (e.g., T1S4, T2S2). Most students agreed that English presented in speaking assessment activities, including their peers’ and teachers’ English, was different from the English used for the other real-world communication activities in which they were involved. They felt that it was like using English to do two different things (e.g., T1S1). Moreover, in some cases, the similarities and differences between English presented in classroom assessments and the English used in real-world communication co-existed.

All English majors involved in this study agreed that the teachers’ and peers’ English presented in the classroom assessments was as natural as the English they heard each other use in other communicative contexts. Most non-English majors involved in this study agreed that teachers and students adjusted their English in order to participate in speaking assessment activities. These adjustments included “speaking English slowly” (e.g., T3S3, T5S4) and “pronouncing with extraordinary clarity” (e.g., T3S4, T4S3). In this case, the non-English majors said that students tried to use English as best as they could in order to get better assessment results and that teachers adjusted their English in classroom assessment activities in order to help students understand what to do in the assessment activities. At the same time, students highlighted the real-world communicative nature of teachers’ and peers’ English by exemplifying how they did Q&A assessment activities. When answering, they used pragmatic skills such as clarification (e.g., T2S4) to communicate with each other, “explaining things to each other in different ways” (e.g., T5S3) and commenting on how teachers’ and students’ English represented “different accents” (e.g., T4S4). This corresponds to McKay and Brown’s (2016, p. 80) point about developing EIL classroom tests: These assessment activities should “typically [be] used to assess what students know or can do in the language with reference to what is being taught in a specific classroom or program.”

While most students expressed doubt about the linguistic authenticity of the teachers’ and peers’ English in classroom assessment activities, they associated each other’s English with real-world communicative language in terms of applying pragmatic skills (McKay & Brown, 2016) and representing varying accents (Jenkins, 2000). This further reflected the students’ ability to separate language used for assessments or feedback delivery from language for communication even though they did find some assessment aspects were connected to real-world communication. Obtaining knowledge about real-world communication through
engaging in assessment activities confirms the notion that language assessments can facilitate learning (Waring, 2008). This finding also indicates that the washback effects of English language assessments on learning for international communication are not simply either positive or negative but represent further opportunities for students to use their English differently. Finally, this research suggests that learning English to communicate and pass assessments did not necessarily present two separate entities but rather, they might occur simultaneously.

Based on the findings of this research, it is reasonable to say that using communication-examination binary opposition to conceptualize classroom assessments creates the impression that students use English either for communication or for assessments. This view of English presented the relationship between English for communication and assessments as completely different or at least two-way. Instead, classroom assessment activities yielded a space for students to participate in the activities both as users who spoke English for communication and as examinees whose English was being tested. In other words, EFL students just used their English to complete assessments and for other non-assessment communication activities. The result shows the necessity of a wider scope or a more open approach to conceptualizing local speaking classroom assessments from a communicative perspective. This wider scope or open approach will prevent researchers from assuming that EFL students learn English either for assessments or for communication.

Conclusion

The teachers involved in this study used textbook resources for assessment content and some of them also used standardized language assessment formats to structure classroom assessment activities. Although the teachers claimed different standards for assessing students’ oral language proficiency, none of them assessed students using NS norms. The students involved in this study obtained knowledge about real-world communication by engaging in speaking assessment activities. For instance, they understood the linguistic diversity of teachers’ and peers’ English within both communicative and assessment contexts. This understanding was obtained by parsing interactions among various groups of English speakers both in and outside the classroom. As Pennycook (2017, p. xi) argued, “We need to understand the diversity of what English is and what it means in all these contexts.”

Examining the classroom assessments in this research also raised questions, pointing out some limitations for future research to consider. First, the teachers in this research had autonomy regarding what to assess and how, but that might be not the case in other research contexts. This calls for future studies to explore whether teacher autonomy has been universally developed and if so, what is its impact on classroom assessments.

Teachers cannot predict what future communicative or assessment English needs their students might have. Therefore, it is reasonable to suggest that assessment activities should be designed to allow students to explore useful resources and ways to operate in English as appropriately as they can for different purposes in different contexts. This would be far more beneficial than prescribing assessment resources or activities that scholars assume will be relevant to students’ language use. Therefore, teachers are suggested to provide avenues for students to reflectively elaborate on what they have learned through various activities and assessments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was financially supported by the University of Macau (MYRG2018-00008-FED) for the research reported in this article.

ORCID iD

Barry Lee Reynolds https://orcid.org/0000-0002-3984-2059

Supplemental Material

Supplemental material for this article is available online.

References

Ambele, E. A., & Boonsuk, Y. (2020). Voices of learners in Thai ELT classrooms: A wake up call towards teaching English as a lingua franca. *Asian Englishes*. Advance online publication. https://doi.org/10.1080/13488678.2020.1759248

Berger, A. (2012). Creating language-assessment literacy: A model for teacher education. In J. Huttner, B. Mehlmauer-Larcher, S. Reichl, & B. Schiffrin (Eds.), *Theory and practice in EFL teacher education: Bridging the gap* (pp. 57–82). Multilingual Matters.

Blommaert, J., & Dong, J. (2010). *Ethnographic fieldwork: A beginner’s guide*. Multilingual Matters.

Brown, A. (2013). Multicompetence and second language assessment. *Language Assessment Quarterly, 10*(2), 219–235.

Brown, J. D. (2004). What do we mean by bias, Englishes, Englishes in testing, and English language proficiency? *World Englishes, 23*(2), 317–319.

Canagarajah, A. S. (2006). Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly, 3*(3), 229–242.

Cheng, L., Rogers, W. T., & Wang, X. (2008). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education, 33*(1), 9–32. https://doi.org/10.1080/02602930601122555

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge.

Yu et al.
Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design* (4th ed.). Sage.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415. https://doi.org/10.1002/j.1545-7249.2009.tb00242.x

Hamid, M. O. (2014). World Englishes in international proficiency tests. *World Englishes*, 33(2), 263–277. https://doi.org/10.1111/weng.12084

Hill, K., & McNamara, T. (2011). Developing a comprehensive, empirically based research framework for classroom-based assessment. *Language Testing*, 29(3), 395–420.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford University Press.

Jenkins, J. (2006). The spread of English as an international language: A testing time for testers. *ELT Journal*, 60, 51–60. https://doi.org/10.1093/elt/cci080

Jenkins, J., & Leung, C. (2017). Assessing English as a lingua franca. In E. Shohamy, I. G. Or, & S. May, (Eds.), *Language testing and assessment* (pp. 1–15). Springer International.

Kim, E.-Y. J. (2019). Utility and bias in a Korean standardized test of English: The case of i-TEPS (Test of English proficiency developed by Seoul National University). *Asian Englishes*, 21(2), 114–127. https://doi.org/10.1080/13488678.2018.1463346

Lee, E. S. Y., & Ahn, H. (2020). Overseas Singaporean attitudes towards Singlish. *Asian Englishes*. Advance online publication. https://doi.org/10.1080/13488678.2020.1795783

Lowenberg, P. H. (1993). Issues of validity in tests of English as a world language: Whose standards? *World Englishes*, 12(1), 95–106.

McKay, S. L., & Brown, D. J. (2016). *Teaching and assessing EIL in local contexts around the world*. Routledge.

Mckinley, J., & Thompson, G. (2018). Washback effect in teaching English as an international language. In J. I. Liontas, M. DelliCarpini, & S. Abrar-ul-Hassan (Eds.), *TESOL encyclopaedia of English language teaching* (pp. 1–12). John Wiley. https://doi.org/10.1002/9781118784235.eelt0655

Saito, J. (2019). The (re)construction of ideologies of English-speaking elites at a global corporation in Japan. *Asian Englishes*, 22(2), 163–178. https://doi.org/10.1080/13488678.2019.1680913

Schreier, M. (2012). *Qualitative content analysis in practice*. Sage.

Taylor, L. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal*, 60(1), 51–60.

Tomlinson, B. (2010). Which test of which English and why? In A. Kirkpatrick (Ed.), *The handbook of world Englishes* (pp. 599–616). Routledge.

Waring, H. Z. (2008). Using explicit positive assessment in the language classroom: IRF, feedback, and learning opportunities. *The Modern Language Journal*, 92(8), 377–594.

Yamaguchi, T. (2019). Multi-competence, expressivity, non-native variants: An investigation into Japanese English. *Asian Englishes*, 22(2), 112–124. https://doi.org/10.1080/13488678.2019.1654724

Yu, H. M. (2018). Using textbooks to teach and learn English for lingua franca communication: An insight into classroom autonomy. *The Asia-pacific Education Researcher*, 27(4), 257–266.

Zhang, Y., & Elder, C. (2014). Investigating native and non-native English-speaking teacher raters’ judgements of oral proficiency in the college English test-spoken English test (CET-SET). *Assessment in Education: Principles, Policy & Practice*, 21(3), 306–325.