

**Special Issue: Teaching, Learning, Assessing, and Researching L2
Pragmatics, in Honor of Prof. Zohreh R. Eslami**

May 2021 – Volume 25, Number 1

**L1 Pragmatic Cultural Schema and Pragmatic Assessment:
Variations in Non-Native Teachers' Scoring Criteria**

Ali Dabbagh

Gonbad Kavous University, Iran
<alidabbagh@gonbad.ac.ir>

Esmat Babaii

Kharazmi University, Iran
<babai@khu.ac.ir>

Abstract

The present study investigated non-native speaker (NNS) teachers' data-driven criteria in scoring multiple-rejoinder written discourse completion test tasks (MR-WDCT) and their points of (mis)match with expert-driven criteria. Specifically, this study scrutinized factors considered by NNS experienced and novice teachers in evaluating responses to MR-WDCT containing examinees' first language (L1) pragmatic cultural schemas. To this end, 10 experienced and 10 novice male and female NNS teachers participated in two rounds of semi-structured interviews wherein they elaborated on their scoring criteria and commented on the expert-driven criteria. Content analysis of the results revealed that: a) pragmalinguistic and sociopragmatic factors were considered for scoring, though in different patterns, by both the experienced and novice NNS teachers; b) observing perceived native speaker (NS) norms for pragmatic cultural schemas was found to be of more significance for the experienced teachers than for the novice ones and in varying severity; c) following L1 pragmatic cultural schemas was considered to be acceptable only if it would not lead to misunderstanding; and d) both NNS experienced and novice teachers' data-driven scoring criteria partially matched with the expert-driven criteria. The findings highlight the role of L1 pragmatic cultural schemas in English as a lingua franca (ELF) pragmatic assessment and the need to train NNS teachers in rating such tests.

Keywords: Pragmatic Cultural Schemas, MR-WDCT, ELF Pragmatic Assessment, Non-native Teachers, Scoring Criteria

Rating concerns have recently grasped the attention of researchers in pragmatic assessment with regard to the norms that raters should adhere to while rating various types of pragmatic tests (Cohen, 2020; Liu & Xie, 2014; Taguchi, 2011; Youn, 2015; Youn & Bogorevich, 2019 among others). However, quite recently, the nature of the norms in pragmatic assessment rating has been reconceptualized with the introduction of English as a lingua franca (ELF) pragmatics (Seidlhofer, 2011; Taguchi, 2017) via transforming our understanding of a successful pragmatic act from “demonstrating native-like pragmalinguistic and sociopragmatic knowledge” to “calibrating and adjusting one’s own pragmalinguistic and sociopragmatic resources, as well as other linguistic and semiotic resources, to the interlocutor and context” (Taguchi & Ishihara, 2018, p. 88). In other words, raters of pragmatic assessment are now encountering possible myriads of pragmatic features produced by non-native speakers (NNS), which might not be correct based on native speaker (NS) norms but are considered as appropriate according to ELF pragmatic characteristics of creativity and adaptability within the act of intercultural communication. Such variations might be rooted in respondents’ cultural background. More precisely put, in answering pragmatic assessment tasks, test-takers might use, as a pragmatics resource, their own cultural conceptualizations (Sharifian, 2015, 2017), particularly their cultural schemas. These cultural conceptualizations may not hamper communication despite being different from the way NSs conceptualize that particular pragmatic situation. Such divergence might be considered as a source of variability among pragmatic assessment raters as to whether to accept pragmatic features based on the test-takers’ variations of first language (L1) cultural conceptualizations. As Cohen (2020) asserted, “an area for [pragmatic] assessment could be that of determining the extent of L1 cultural overlay taking place when learners are performing their TL [target language] pragmatics, especially in foreign as opposed to L2 contexts” (p. 4).

Previous studies on rater factor in pragmatic assessment have emphasized variability among NNS raters (Alemi et al., 2014; Alemi & Khanlarzadeh, 2016; Tajeddin & Alemi, 2014; Tajeddin & Alizadeh, 2015), NS raters (Alemi & Khanlarzadeh, 2015; Taguchi, 2011; Tajeddin & Alemi, 2013), and between NS and NNS raters comparatively (Alemi & Rezanejad, 2014; Alemi & Tajeddin, 2013; Sunnenburg-Winkler et al., 2020; Walters, 2007). However, the focus of these studies was mostly raters’ views on and practices in assessing NNS’ application of NS pragmalinguistic and sociopragmatic norms in their pragmatic assessment performance. What is missing in the literature is the extent to which raters of pragmatic assessment tasks appreciate the use of ELF pragmatic features, especially in terms of using English colored with L1 cultural conceptualizations. The aforementioned issues gave rise to the present study that aims at investigating NNS teachers’ (as raters) views and practices with regard to test-takers’ use of their L1 cultural conceptualizations. Of particular interest is the extent to which the experienced and novice teachers recognize and accept the respondents’ sense of agency (Cohen, 2020), while answering multiple-rejoinder written discourse completion tasks (MR-WDCT) focused on assessing pragmatic cultural schema aspects of speech acts. To do so, the following research questions were formulated:

1. To what extent do non-native speaker experienced and novice teachers differ in terms of rating responses to multiple-rejoinder written discourse completion tasks?

More specifically:

2. What factors do non-native speaker experienced and novice teachers attend to in rating responses containing EFL learners' L1 pragmatic cultural schemas in multiple-rejoinder written discourse completion tasks?
3. How do experienced and novice teachers compare with regard to their ratings as to the acceptability of EFL learners' multiple-rejoinder written discourse completion tasks responses containing L1 pragmatic cultural schemas?
4. To what extent do factors mentioned by non-native speaker experienced and novice teachers in their data-driven criteria in rating responses to multiple-rejoinder written discourse completion tasks match those stated in expert-driven criteria?

Literature Review

Pragmatic Competence: Expanding the Construct to ELF Pragmatics

According to Kasper and Roever (2005), pragmatic competence refers to an individual's "ability to act and interact by means of language" (p. 317). This ability is anchored in three principles of meaning, interaction, and context (Timpe Laughlin et al., 2015), which can be manifested via two intersecting components of *sociopragmatics* and *pragmalinguistics*. While sociopragmatics refers to learners' knowledge of social norms and conventions or the "sociological interface of pragmatics" (Leech, 1983, p.10), pragmalinguistics deals with "the particular resources which a given language provides for conveying particular illocution" (p.11).

The current transnational and transcultural world necessitates revisiting the construct of pragmatic competence within the scope of intercultural communication wherein ELF speakers bring their own L1-based experience as well as their shared experience with their interlocutors to their interactions, thereby establishing new, mutually acceptable norms. Such a reconceptualization of pragmatic competence to ELF pragmatics enables us to "go beyond the traditional scope of pragmatic competence focused on how learners perform a communicative act in the L2 and extend the concept to an understanding of how learners successfully participate in intercultural interaction" (Taguchi, 2017, p. 157). In other words, ELF pragmatics "is ultimately about interactional effectiveness, rather than proximity to native speaker norms" (Taguchi & Ishihara, 2018, p. 87).

Research in English as a foreign language (EFL) pragmatics has revealed prioritization of speech acts as co-constructed and negotiated sequences – either pragmalinguistically (Jenks, 2013; Schnurr & Zayts, 2013) or sociopragmatically (Knapp, 2011; Park, 2017) –, communicative effectiveness strategies (House, 2013), and accommodation/rapport building strategies (Zhu, 2017). Such features motivate ELF pragmatics to embrace the discursive approach to pragmatic competence in which one's sociopragmatic and pragmalinguistic resources are adjusted to those of the interlocutor and the context so that pragmatic action is jointly constructed (Ross & Kasper, 2013).

Features of such discursive-based ELF pragmatics might reveal themselves, on the one hand, when EFL learners are engaged in pragmatic assessment so much so that "respondents may wish to exercise their agency by refraining from engaging in the called-for pragmatic behavior out of a sense that it is inconsistent with their self-identity" (Cohen, 2020, p. 186). Yet, on the other hand, respondents might show their agency not by refusing to produce the intended

speech act but by basing their production on their own cultural background, especially on their own cultural schemas.

Pragmatic Cultural Schemas

As Cohen (2020) asserted, since NNSs are accustomed to particular L1 culture-specific speech act strategies, they might transfer these strategies into the target language. The reason can be sought in the deep engagement of NNSs in their L1 cultural schemas. According to Sharifian (2015), cultural schemas, as an element of the analytical framework of cultural conceptualizations within Cultural Linguistics (see Sharifian, 2011, 2017), are a subclass of cognitive schemas that are “culturally constructed [and] are abstracted from the collective cognitions associated with a cultural group, and therefore to some extent based on shared experiences, common to the group, as opposed to being abstracted from an individual’s idiosyncratic experiences” (Sharifian, 2015, p. 478). Cultural schemas emphasize values, belief systems, and behavior expectations germane to myriads of human experience (Sharifian, 2003) that represent culturally-constructed encyclopedic meaning and are instantiated in lexical items related to various concrete and abstract concepts, such as *ta’arof* (Sharifian, 2014) and *shekasteh-nafsi* (Sharifian, 2005) in Persian. The internalization of a particular cultural schema is not imprinted equally in the minds of individuals; rather, it is “to some extent collective and to some extent idiosyncratic” (Sharifian 2015, p. 478). In other words, cultural schemas are *heterogeneously distributed* across members of a speech community (see Figure 1).

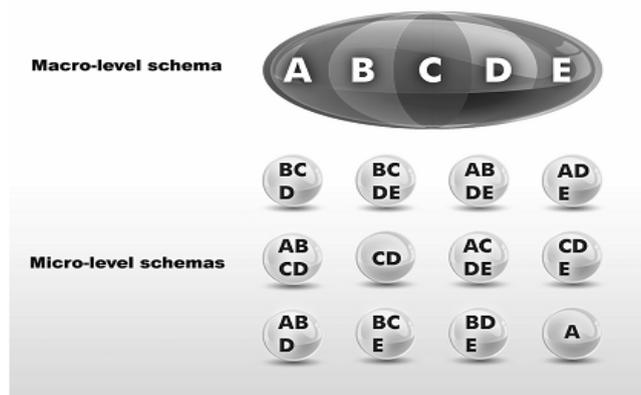


Figure 1. Heterogeneous distribution of cultural schemas (Sharifian, 2017, p. 61).

The uptake and enactment of speech acts to a great extent depend on cultural schemas as they supply a resource for pragmatic meaning. Such cultural schemas are referred to as ‘pragmatic cultural schemas’ (Sharifian, 2016) that are often considered as a possible source of shared knowledge by interlocutors. Drawing upon this ‘common ground’ (Sharifian, 2017), the performance of a speech act might be caused by a feeling of agency, which NNSs prefer to practice their L1 pragmatic cultural schema while interacting with NSs (Cohen, 2020). However, this by no means should be inferred as a deterministic agency. Rather, “the choice of particular pragmemes [...] are at the total discretion of speakers [...] to exercise creativity and remain intelligible” (Sharifian, 2016, p. 517).

Assessment of Pragmatics: Discourse Completion Tasks

The interest in (re)theorizing pragmatic competence has led scholars to design measures to assess this construct. These attempts resulted in seven measures (Hudson et al., 1995; Walters,

2007; Yamashita, 2008), among which discourse completion tasks (DCTs) have attracted the attention of both language testing researchers and practitioners. In DCTs, test-takers are given a situation to provide an appropriate speech act, routine, or implicature according to the topic of conversation, context, and the interlocutors' role relationships depicted in the instruction.

DCTs can be administered to a large number of test-takers. In addition, using DCTs, test designers can systematically alter the contextual variables of power, distance, and degree of imposition (Yamashita, 2008). Apart from these advantages, previous studies have revealed a number of negative characteristics of DCTs. First, there is a possibility that test-takers provide longer or shorter answers than expected (Edmonson & House, 1991). Second, Golato (2003) and Yamashita (2005) found a dissimilarity between the intended speech acts and the produced speech acts by respondents while answering DCTs. Third, the written description of the situation might lead to misunderstandings among test-takers as to contextual features (Yamashita, 1998). However, perhaps DCTs were chiefly criticized for their single-response format, which "under-represent[s] the discursive side of pragmatics" (Roever, 2011, p. 469). To compensate for these shortcomings, at least partially, different scholars have proposed adjustments to the original DCT, including Barron's (2003) free discourse completion task (FDCT), Schneider's (2008) dialogue production task (DPT), and Cohen and Shively's (2002/2003) multiple-rejoinder DCT. In the latter modification, which is the measure used in this study, respondents supply one or more responses in view of the rejoinders provided for the DCT situation. Such an adaptation allows respondents to produce the intended speech act within a turn-taking sequence, consistent with a salient feature of natural conversation.

Rating Scales for DCT and Their Challenges

The extensive use of DCTs in pragmatic assessment has brought forth considerable challenges in scoring a learner's performance. This is specifically so when it comes to DCTs, which require open-ended as opposed to standard responses due to reliability and validity considerations (Knoch, 2009). Rating criteria for such DCTs can either be developed before administering the test by "an individual or committee who are perceived to be experts in the teaching and assessment of the construct of interest" (Fulcher et al., 2011, p. 7) or be driven qualitatively out of the data collected from learners' performance in answering the test (Chen & Liu, 2016; Grabowski, 2009; Youn, 2015).

Most attempts in designing scoring criteria for DCTs have followed a priori approach grounded on a pre-existed rating scale, needs analysis, or a teaching syllabus. For instance, Hudson et al. (1995) developed a five-point scale from *Very satisfactory* to *Completely appropriate*, including the following six dimensions: the ability to use the correct speech act, amount of speech and information, typical expressions, directness, levels of formality, and politeness. In another study, Sasaki (1998), evaluated NSs pragmatic production with reference to the three criteria of appropriateness, grammar/structure, and fluency/pronunciation on a scale from *Excellent* to *Very poor*. Moreover, Ishihara (2010) suggested alternative ways to assess pragmatic performance in DCTs, including holistic and analytic assessment, peer and self-assessment, and focused assessment.

These intuitive methods in developing scoring criteria, however, suffer from lack of correspondence to learner performances and might be void of pedagogical insights for teachers and learners. Therefore, attempts have been made to apply the data-driven approach to designing scoring criteria in order to assess DCTs. For instance, Chen and Liu (2016)

constructed and validated a scoring criterion to evaluate the speech acts of request and apology targeted in a WDCT designed for Chinese intermediate EFL learners. Qualitatively analyzing the raters' comments, they set the emerged scale from *Excellent* to *Very poor*, attending the two main features of content and form. Their data revealed that 'content criteria' included the amount of information, politeness, clarity, and relevance while 'form criteria' consisted of grammar, phrasing, and word choice. More recently, concerning the appreciation of interactional competence in language assessment (e.g., Roever & Kasper, 2018), more interactive criteria have been introduced to design assessment scales for DCTs, including language use, sensitivity to the situation, content delivery, engagement in interaction, turn organization, and discourse markers (Youn, 2018).

Quite often, the benchmark for rating in these two approaches reviewed above is to compare the performance of NNSs with that of NSs. However, Taguchi (2011) highlights that NSs may fail to reach a consensus on the pragmatic appropriateness of utterances in relation to a particular speech act within a speech situation. Specifically, she asserted that geographical location, social class, and education level are the factors underlying variation in NSs' judgement of pragmatic fit of response to any given situation. The challenge lies in the difficulty of finding a systematic way to determine utterances that are unquestionably inappropriate to NSs but less so to NNSs (Youn & Bogorevich, 2019). Another challenge is related to features of ELF pragmatics and the agency of respondents to DCTs in rejecting NS norms if incompatible with their identity. In other words, as Youn and Bogorevich (2019) maintain, pragmatic assessors need to "determine the degree of appropriateness in pragmatic performance in considering varied pragmatic norms specific to contexts and cultures [instead of] benchmarking varied dimensions of pragmatic against native speakers' pragmatic norms" (p. 318). These challenges might surface in scoring DCTs whose respondents might refer to their L1 pragmatic cultural schemas.

Previous Studies on Rater Criteria in Assessing DCT

Considering the challenges mentioned above, there is a dearth of research on rater factors in scoring a DCT. Putting the few studies on NS teachers' criteria aside (e.g., Alemi & Khanlarzadeh, 2015; Taguchi, 2011; Tajeddin & Alemi, 2013), most of the previous studies have investigated either the NNS teachers' scoring criteria or the comparison of the criteria mentioned by NNS and NS in assessing DCT, which to some extent can shed light on ELF pragmatic assessment when the raters are being taken into account. In a pioneering study, Alemi and Tajeddin (2013) compared NNS and NS criteria in assessing second language refusal production via WDCT. Their quantitative and qualitative analysis revealed 'reasoning/explanation' and 'politeness' as leading assessment criteria for NS and NNS, respectively. In addition, NNS teachers were found to be more lenient in scoring WDCT than NS teachers were. This line of inquiry was followed by Alemi and Rezanejad (2014) in assessing the speech act of compliment. Results of the rating speech act questionnaire showed the following factors in NNS and NS criteria alike: 'politeness', 'strategy use', 'interlocutors' relationships', 'affective factors', 'linguistic accuracy', 'sincerity', 'fluency', 'authenticity', and 'cultural issues'.

With regard to investigating NNS criteria, reference can be made to Alemi et al. (2014) who found similar criteria as to the studies previously mentioned. These researchers specifically focused on the professional background of the NNS teachers and their gender, for which no significant quantitative difference was revealed. Similar results were found by Alemi and

Khanlarzadeh (2016) in assessing request speech act via video prompts. The role of NNS teachers' professional experience in the use of pragmatic assessment scoring criteria was also scrutinized in Tajeddin and Alizadeh (2015), which focused on monologic and dialogic assessment of role-plays. Despite both the experienced and novice teachers participated in their study referred to the criteria of appropriateness (general criterion) and pragmalinguistic, due attention was not paid to sociopragmatic factor by the more experienced and less experienced teachers. In addition, the raters were found to be different in the application of the three criteria mentioned above when they conducted the rating monologically, while this difference became limited to sociopragmatic criteria when they performed the assessment dialogically.

Mostly, in these studies, L1 cultural factors remained unexpanded and underexplored. However, recently, Sunnenburg-Winkler et al. (2020) explored the role of L1 (NNS and NS) in rater variation with regard to self and peer assessment of DCTs. Their results revealed raters focus on different dimensions of speech acts in assessing pragmatic appropriateness. Specifically, L1 backgrounds of the raters played a significant role to the extent that "there was notable similarity in ratings of those from the same L1 background" (p. 79) in peer -assessment but not in self-assessment.

It seems that though the role of L1 culture is indispensable to the nature of ELF pragmatics, consideration of L1 pragmatic cultural schema in ELF pragmatic assessment has been under-researched. Specifically, as was reviewed above, most previous studies on rater criteria for assessing DCT focused on NNS and NS scoring criteria variation. What awaits investigation is NNS rater variation regarding appreciating, or lack thereof, learners' adherence to norms for L1 pragmatic cultural schema in ELF pragmatic assessment, especially when raters and learners share common L1 pragmatic cultural schemas.

Method

Participants

To collect the data, 70 EFL learners and 20 EFL teachers, who filled out a consent form prior to data collection, participated in the study following convenience sampling. The participant learners were both male ($n=37$) and female ($n=33$) whose age ranged between 19 and 25 years old ($Mean = 21.8$, $SD = 2.12$) and had at least five years of experience in learning English as a foreign language at institutes. These EFL learners spoke Persian as their first language and were exposed to the culture of English-speaking countries via their EFL textbooks, English movies, and documentaries. Among the participant learners, 37, who were estimated to be at the intermediate level via the results of the administered Michigan Test of English Language Proficiency, were asked to answer the discourse completion tasks. The rationale for selecting these EFL learners was their ability to read and understand the settings stated in the pragmatic test items and complete the gaps using the appropriate linguistic form in written language and their background in being exposed to the English language and culture.

The participant teachers held B.A. or M.A. degrees in teaching English as a foreign language or English language and literature. They were between 28 and 52 years of age ($Mean = 32$, $SD = 8.16$) and spoke Persian as their mother tongue. Quite like the case for EFL learners, the participant teachers' familiarity with the culture of native speakers of English come from the EFL textbooks, English movies, and documentaries, and their education in English. Therefore, in the present study, their understanding of native speakers' culture is referred to as 'perceived pragmatic cultural schemas'. Following Farrell (2012), the participant teachers who were

“within three years of completing their teacher education program” (p. 437) were designated as novice teachers ($n=10$), and those above this level were designated as experienced teachers ($n=10$). Table 1 provides demographic information of the participant teachers.

Table 1. Participant Teachers’ Demographic Information.

| No. | Experienced/Novice | Age | Gender | Years of teaching experience |
|-----|--------------------|-----|--------|------------------------------|
| 1 | Novice | 28 | male | 2 |
| 2 | Novice | 23 | female | 1 |
| 3 | Novice | 27 | male | 3 |
| 4 | Novice | 24 | female | 2 |
| 5 | Novice | 31 | female | 1 |
| 6 | Novice | 26 | female | 1 |
| 7 | Novice | 26 | female | 3 |
| 8 | Novice | 26 | male | 3 |
| 9 | Novice | 24 | female | 2 |
| 10 | Novice | 25 | male | 2 |
| 11 | Experienced | 39 | female | 12 |
| 12 | Experienced | 41 | female | 18 |
| 13 | Experienced | 35 | female | 13 |
| 14 | Experienced | 42 | male | 15 |
| 15 | Experienced | 32 | female | 12 |
| 16 | Experienced | 48 | female | 20 |
| 17 | Experienced | 33 | female | 14 |
| 18 | Experienced | 51 | male | 32 |
| 19 | Experienced | 31 | male | 13 |
| 20 | Experienced | 34 | female | 10 |

Instruments

The Michigan Test of English language proficiency. This test was administered to ensure that the participants were at the intermediate level of English language ability. The test consisted of 100 multiple-choice items of grammar (40 items), vocabulary (40 items), and reading comprehension (20 items) to which the participants were to answer in 75 minutes. However, due to managerial considerations, the writing section was not administered. Shohamy et al. (2017) reported a high validity and reliability for this version of the test. The Cronbach-alpha reliability estimated in this study was .74.

Multiple-rejoinder written discourse completion task (MR-WDCT). As a common measure to assess learners’ pragmatic production (Taguchi, 2011), an MR-WDCT was designed and administered to EFL learners. The MR-WDCT used in the current study consisted of four settings, each of which was accompanied by a gapped discourse that the EFL learner participants were asked to complete with no less than five words without consulting any kind of dictionary. The settings were designed in a way that the responses of EFL participants might

include Persian pragmatic cultural schemas. However, since the purpose of the study centered around pragmatic cultural schemas, care was taken not to sensitize the participants to the purpose of the study via setting some gapped-turns in MR-WDCT insensitive to Persian pragmatic cultural schemas. The pragmatic set associated with each scenario in the administered MR-WDCT was based on the format represented in Table 2.

To ensure the construct validity of MR-WDCT, two individuals with doctorates in applied linguistics were asked to review the instrument. Results of Phi-coefficient analysis revealed the agreement of .86 between the two test reviewers. The test situations and the gapped multiple-rejoinder discourse were adapted based on the expert feedback.

Table 2. Pragmatic Set for Each Scenario in the Administered MR-WDCTs.

| Scenario | Pragmatic schema | Speech act/event | Probable Persian pragmeme |
|----------|--|---|--|
| No. 1 | Persian cultural schema of <i>shekasteh-nafsi</i> ‘modesty’ and <i>ta’arof</i> | responding to a compliment on an achievement | reassigning the compliment to the complimenter |
| No. 2 | Persian cultural schema of <i>sharmandegi</i> ‘feeling ashamed’ | making a request | expressing <i>sharmandegi</i> |
| No. 3 | Persian cultural schema of <i>shekasteh-nafsi</i> ‘modesty’ | responding to a compliment on the taste of food | apologies |
| No. 4 | Persian cultural schema of reassigning the compliment to God | receiving achievement news and responding to it | reassigning the achievement to God |

Interviews. To investigate teachers’ criteria for scoring the MR-WDCT, the experienced and novice EFL teachers were interviewed using a semi-structured interview protocol. The teachers were asked to list their scoring criteria and the rationale behind them twice, the first time on the basis of a sampling of completed MR-WDCTs and the second time according to an uncompleted MR-WDCT reflecting on an expert-driven scoring criteria (see below). To devise the interview questions (see appendices A & B), the components of pragmatic assessment and the role of cultural schema in assessing pragmatics were taken into consideration.

To ensure the content and face validity of the interview questions, an assistant professor of applied linguistics, with specific expertise in pragmatic assessment, was asked to examine the questions with regard to clarity and relatedness to the underlying construct, i.e., scoring discourse completion tasks. The questions were revised based on the feedback received via expert judgement.

The interviews were conducted face to face in the preferred language of the interviewees, i.e., either Persian or English. The interviews were audio-recorded, transcribed, and translated into English (if necessary) by the first author for content analysis.

Expert-driven scoring criteria. This criteria was based on Taguchi (2011) (see Table 3) regarding the ability to produce appropriate speech act in terms of linguistic appropriacy, use of semantic formula, observing interlocutors' relationship, politeness, choice of register, observing necessary formality, degree of intensity, degree of directness, naturalness, and cultural accommodation, following Hudson et al. (1999).

Procedure

Data collection. The current study was conducted in two phases. In the first phase, the proficiency level of the participant EFL learners was determined via administering the Michigan Test of English Language Proficiency. The MR-WDCT was then administered to the selected intermediate participant EFL learner in order to collect their responses to the given pragmatic scenarios. The participant EFL learners were allotted 30 minutes to complete the test. The second phase consisted of two sets of interview sessions. In the first set, the experienced and novice participant EFL teachers were given samples of completed MR-WDCTs ($n=15$) from the first phase of the study and were asked to list their scoring criteria for the test based on the answers provided by the participant EFL learners. The sample given to the participant teachers included the Persian pragmatic cultural schemas.

In the second set of interviews, the participant EFL teachers were given an expert-driven scoring criterion. The participant teachers were asked to discuss the given criteria in scoring the MR-WDCTs. Explanations of the factors mentioned in the scoring criteria were provided if necessary. They were free to challenge, adapt, and revise the criteria in any way they thought might be necessary. The second round of interviews was conducted in order to unveil the lines of (mis)match between participant teachers' data-driven scoring criteria and the expert-driven criteria.

Table 3. Scoring Criteria Provided to the Participant Teachers (Taguchi, 2011, p. 459).

| |
|---|
| 5 = Excellent |
| Almost perfectly appropriate and effective in the level of directness, politeness, and formality. |
| 4 = Good |
| Not perfect but adequately appropriate in the level of directness, politeness, and formality. Expressions are a little off from target-like, but pretty good. |
| 3 = Fair |
| Somewhat appropriate in the level of directness, politeness, and formality. Expressions are more direct or indirect than the situation requires. |
| 2 = Poor |
| Clearly inappropriate. Expressions sound almost rude or too demanding. |
| 1 = Very poor |
| Not sure if the target speech act is performed |

Data analysis. To analyze the data, the two sets of interviews were transcribed verbatim by the first researcher and then underwent content analysis (Patton, 2015). The first interview data were analyzed using inductive content analysis via open coding, category creation, and abstraction (Dey, 1993) conducted through the constant comparative procedure (Glaser & Strauss, 1967) in order to describe experienced and novice Persian EFL teachers' data-driven

criteria for scoring the given MR-WDCT. The same data analysis procedure was applied to the data from the second interview in order to explore whether the experienced and novice teachers had changed scoring criteria that they utilized in the first interview upon being provided with expert-driven scoring criteria for MR-WDCTs. In so doing, the possible (mis)matches between teachers' scoring criteria and the one developed based on expert suggestions were scrutinized.

Once the interview process and analysis was over, member checking was conducted via sharing the extracted criteria with the interviewees in order to ensure about their clarity and accuracy and see if the interviewees “agree, argue with, or want to add” to the extracted criteria (Rallis & Rossman, 2009, p. 266). The generated criteria were approved by 86% of the interviewees. To improve the confirmability of the study, an outside researcher audited the entire data analysis procedure via discussion sessions between the auditor and the researchers. The lines of disagreement were debated and resulted in compromise after adaptation.

Results

NNS Teachers' Data-driven Scoring Criteria

Content analysis of the experienced and novice participant teachers' first-round interviews revealed similar data-driven scoring criteria for rating cultural pragmatic speech acts in the administered MR-WDCTs, though with different distribution among teachers of each group. Table 4 illustrates teachers' criteria (to observe ethical considerations, the participant teachers are referred to as NT for novice teachers and ET for experienced teachers, for short, while reporting the results).

Table 4. Experienced and Novice Teachers' Scoring Criteria.

| Orientation | Criteria | Percentage | |
|-------------------------|---|--------------------------|---------------------|
| | | Experienced NNS teachers | Novice NNS teachers |
| Sociopragmatic | (In)formality | 30 | 60 |
| | Observing interlocutors' social status | 10 | 30 |
| | Politeness | 20 | 20 |
| | Observing required speech act | 70 | 90 |
| | Following NS norms for pragmatic cultural schemas | 80 | 60 |
| Pragmalinguistic | Spelling and punctuation | 50 | 40 |
| | Vocabulary and lexical chunks | 40 | 50 |
| | Grammatical accuracy | 100 | 80 |
| Other | Hand writing | 10 | 20 |

As the Table above shows, the experienced and novice participant teachers considered both pragmalinguistic and sociopragmatic aspects in their scoring criteria. In the pragmalinguistic category, ‘grammatical accuracy’ attracted the attention of all the experienced participant teachers and most novice ones as the main factor in their scoring criteria. In addition, ‘vocabulary and lexical chunks’ side by side ‘spelling and punctuation’ were received by the

experienced and novice teachers alike as the second frequent factor in the pragmalinguistic category. One of the novice teachers mentioned that:

NT4:

I will attend to the appropriate and proper use of words in my scoring regarding, ehm, the situation and the context students operate. By this, I mean the relevance and appropriacy of words in specific situations and sentences in general with regard to, how do you say it, ehm, the prepared circumstances, let's say.

Beside these factors, most of the teachers in both groups emphasized sociopragmatic orientation, the most frequent of which were observing required speech acts targeted in the MR-WDCT and the degree of L1 cultural influence. As the Table above reveals, the former factor seems to be more important for the novice teachers (90%) than for the experienced ones (70%). Referring to two of the answered MR-WDCTs, one of these teachers stated that:

NT6:

This language learner could figure out how to respond to the statements brought about in this test so that, ehm, s/he looks talking in a way that is related to the topic of the test. Actually, ehm, he could easily find out what to say, actually, he knows how to communicate in the different given situations. On the other hand, this other test-taker, used, you know, some weird sentences in his answers. It seems he does not understand what is going on in the given conversation!

The following section elaborates on the results with regard to the reaction of participant teachers towards test-takers' use of NS pragmatic cultural schemas or lack thereof.

Treating Answers Containing L1 Pragmatic Cultural Schemas

In answering the administered MR-WDCTs, some of the participant test-takers used their L1 pragmatic cultural schemas, including the Persian cultural schema of TA'AROF or SHEKASTEH NAFSI (i.e., self-lowering) as in *don't reject my hand* in offering food and *I did nothing dear professor and it was all because of your help* in responding to the speech act of compliment, respectively. Stating it differently, these respondents did not follow norms for NS pragmatic cultural schemas in answering the MR-WDCTs. As Table 4 above shows, observing NS norms for pragmatic cultural schemas, as 'perceived pragmatic cultural schemas' by the NNS raters, was shown to be of more significance for the experienced NNS teachers (80%) than for the novice ones (60%) in terms of frequency. By perceived pragmatic cultural schemas, it is meant what NNS raters consider to be NS schemas, no matter how incomplete these perceptions could be. However, two of these novice teachers were more severe than other experienced and novice teachers in commenting on this factor in their stated scoring criteria. In other words, not only did they mention following norms for L1 pragmatic cultural schemas as a negative factor to be considered in scoring, but also they referred to the exact meeting of their perceived norms for NS pragmatic cultural schemas, as are mentioned in the following quotes:

NT4:

Absolutely students should not transfer Persian expressions [into English] because it is meaningful only in our context [i.e., Iran] but not in, let's say, the target language context, I suppose.

NT8:

Learners used their mother tongue in some cases as a, how do you call it, yeah, a frame of reference. That's funny! It is not acceptable and even not appropriate as I understand native speaker way of using English language.

On the other hand, not all the experienced participant NNS teachers did perceive such pragmatic cultural schema interference as purely a negative factor that should result in score reduction. Two of these teachers postulated that:

ET10:

Students are expected to answer in such a way that it sounds somewhat like a natural dialogue in a similar way as native speakers use language, ehm, by using, ehm, some common expressions. But I never reduce their mark if they do not follow this. To me, a short note for them is fine.

ET1:

It is possible that we ignore this, I mean, as a negative point. In my opinion, this is a natural thing. Inevitably, the responses of a Persian learner are different from, say, an Arab or a European English learner. This is because they convey similar meanings in different ways or even convey concepts unique to their own culture and this naturally affect their responses.

As can be inferred from the last two quotes above, some of the experienced participant teachers assumed the use of L1 pragmatic cultural schemas in answering MR-WDCTs as a natural and even inevitable feature of the current use of English, which is in congruence with principles of ELF pragmatics. In dealing with such responses in MR-WDCTs, the two novice teachers mentioned above (i.e., NT4 and NT5) and three experienced (i.e., 30%) NNS teachers expressed use of L1 pragmatic schemas as a negative factor in their first-round interview though they were different in terms of the severity they assigned to this negativity. The following are some of the ideas of the experienced and novice teachers who stated that they would not assign a score to such answers.

NT4:

The answers must be according to the culture of native speakers of English since we are teaching their language. When we are teaching English, one of our aims is to teach English culture, as well. Of course as much as we know about it. When we, ehm, administer a test like this, we are after checking [i.e., scoring] whether students are familiar with cultural differences between Persian and English.

ET6:

I think there would be no score for such answers at all, you know, we do not have such expressions and frame of thought in English, as much as I know. These answers are word-by-word translation from Persian into English. Ehm, English people, they do not understand such sentences. After all, learners are studying English language and the framework is and should be the way native speakers of English think and use language.

ET2:

Even if the translated answers were correct in terms of grammar and vocabulary, I would subtract the whole score since such an answer would not be effective and meaningful for a native speaker of English. I do this so that, ehm, the student understands that something is, eh, wrong with the given answer.

However, the rest of the experienced and novice participant teachers (i.e., 70% in each group), except for one novice teacher with a positive attitude toward such answers, expressed more moderate views (in comparison to quote mentioned by NT4, NT8, ET2, and ET6 above) and considered other factors in scoring answers containing test-takers' L1 pragmatic cultural schemas. More specifically, the experienced teachers referred to 'examiner's cultural background', 'interlocutor's (mis)understanding', 'grammatical accuracy', 'lexical appropriacy', and novice teachers noted 'observing formality' while 'understanding required speech act' and 'learners' language proficiency level' were mentioned by both groups. In other words, if the given answer is successful in meeting the criteria mentioned above but not in observing perceived NS norms for pragmatic cultural schemas, these teachers will reduce only a portion of the total score, which is evident in the three extracts below:

NT1:

It is not important for me that students answer in exactly the same way as native speakers do in a similar situation. It is, ehm, because they [students] are not that much prepared at lower levels to act like a native speaker.

NT6:

In such answers I will subtract only a part of the score. It is because [the learner] understood what to say [in that situation] but he could not do so in an appropriate way. That is, he could not think like a native speaker of English and inserted his Persian thought in his sentences. However, if the learner is successful in conveying the required message, even if it is not purely said following native speaker cultural norms as I understand it, I will assign a part of the total score.

ET1:

It is a bit subjective. It depends on who scores the test. If the examiner is a non-Iranian person [in the case of the given answers in this MR-WDCTs], he or she will not understand a word of it. But I, as an Iranian examiner, I will understand the meaning of these answers and will not subtract a score since the main factor, that is understanding the situation and talking accordingly, is observed and the learner expressed that in his/her own words.

Despite these mild views, about 60% of the novice and 50% of the experienced participant teachers claimed to take into account the intensity of pragmatic cultural schema transfer from L1 to L2 in the learners' responses to MR-WDCTs while scoring the responses. More specifically, they mentioned that they would assign no score for sentences highly affected by the L1 pragmatic cultural schema of TA'AROF, such as *don't reject my hand* since it would result in misunderstanding or even offense on the side of the NS interlocutor. On the other hand, these teachers asserted that they would score the expressions partially, as perceived by them, that were common in English though not in the required speech act in the administered MR-WDCT since they were slightly influenced by L1 pragmatic cultural schemas, e.g., *it was all because of your help* mentioned above.

Quite related to the aforementioned ideas, nearly 10% of the participant teachers of both groups believed the answers containing L1 pragmatic cultural schemas might cause cultural transmission to the NS interlocutor and therefore should not be conceptualized as entirely a negative point. This idea clearly reflects the ELF pragmatics perspective that is evident in the following quote:

NT5:

This type of answer [containing L1 cultural norms] can result in, you know, transmitting culture. That is the person to whom the student is talking to in this test might ask for the meaning and, ehm, explanation of the sentence used. Of course it all depends on the relationship between the interlocutors in that are they willing to step into such cultural understanding or not.

NNS Experienced and Novice Teachers' Reactions to the Expert-driven Scoring Criteria

A comparison of the criteria expressed by the experienced and novice teachers listed in Table 4 above with the expert-driven criteria prepared for the MR-WDCT in this study demonstrated points of match and mismatch. That is, except for 'choice of register', 'degree of intensity', 'degree of directness', and 'naturalness' all other factors from the expert-driven criteria were noted by the novice and experienced participant teachers alike. It should be mentioned that content analysis of the second round of interview data revealed that six experienced teachers and one novice participant teacher accepted the factors they did not refer to from the expert-driven criteria; however, they considered them as implicit in the factors they directly stated.

ET3:

It seems that politeness is expressed in different forms [in this criterion], such as directness, intensity, and formality. I think we could include all within degree of formality.

NT5:

I considered politeness and degree of intensity and directness as one single factor.

The above quotes clearly demonstrate that both of these experienced and novice teachers could not distinguish among closely related concepts when it came to scoring responses containing the L1 pragmatic cultural schemas. The expert-driven criteria, however, might act as a consciousness raising tool for other participant teachers:

NT8:

I haven't paid attention to degree of intensity so far. I think it is so dynamic and, ehm, depends on, err, situation and context.

ET6:

I did not deal with these factors in the first interview. I think they are great and useful for me in my career.

In addition to the adaptations mentioned above, 60% of the experienced and 70% of the novice participant teachers voted for expanding the generally stated factor of 'cultural accommodation' within the expert-driven scoring criteria, so that all the examiners would have a similar understanding of this vague factor, as they called it. While the experienced participant teachers included 'use of semantic formula', 'formality', and 'grammatical accuracy' in their expansion of the culture criteria, novice teachers assumed 'observing interlocutors' relationship' and 'use of NS proverbs' as building blocks of culture to be attended to in scoring MR-WDCTs. 'Politeness', 'lexical appropriacy', and 'observing NS cultural norms [pragmatic cultural schemas]' were also mentioned in common by the two groups. All these factors were mentioned by the teachers in relation to their perceived NS norms, but the L1-based ones were claimed to be acceptable to the extent that it did not cause misunderstanding, as was reported above. In addition to this, two experienced and three novice participant teachers suggested

setting ‘cultural accommodation’ for higher scores of the given criteria in order to evaluate the performance of learners who were above intermediate proficiency level.

In summary, the experienced and novice NNS teachers noted varied factors when asked to score MR-WDCTs that incorporate pragmatic cultural schemas. The following Table encapsulates the aforementioned results at a glance.

Table 5. Experienced and Novice Teachers Treatment of Culture and Pragmatic Cultural Schemas in Responses to MR-WDCT.

| Main results regarding use of L1 PCS* | NNS teacher | |
|---|---|--|
| | experienced | novice |
| Use of L1 PCS in answering MR-WDCT | considered as negative (30%) considered as neutral (70%) | considered as negative (20%) considered as positive (10%) considered as neutral (70%) |
| Considering the intensity of PCS transfer from L1 to L2 | 60% | 50% |
| Other factors to consider in scoring responses containing L1 PCS | examiner’s cultural background understanding the required speech act interlocutor’s (mis)understanding grammatical accuracy lexical appropriacy learners’ language proficiency level | observing the required speech act observing formality learners’ language proficiency level |
| Factors to include in defining ‘cultural accommodation’ in the expert-driven scoring criteria | NS-like politeness NS-like use of semantic formula formality according to NS norms lexical appropriacy grammatical accuracy | NS-like politeness observing interlocutors’ relationship lexical appropriacy use of NS proverbs |
| Setting cultural-related factors as scoring criteria for higher proficiency levels | 20% | 30% |
| Areas of (mis)match between data- driven and expert- driven criteria | Matches <ul style="list-style-type: none"> • observing necessary formality • observing interlocutors’ relationship • politeness • cultural accommodation • linguistic appropriacy • use of semantic formula Mismatches <ul style="list-style-type: none"> • choice of register • degree of intensity • degree of directness • naturalness | |

*: pragmatic cultural schemas

Discussion

Summary of Findings

The present study aimed at investigating NNS experienced and novice teachers’ scoring criteria driven from test-takers’ responses to MR-WDCTs, on the one hand, and scrutinizing if the extracted criteria encompassed (mis)matches with expert-driven criteria developed for the similar tests, on the other. Specifically, the current study sought to examine NNS teachers’ criteria for assessing responses that are influenced by test-takers’ L1 cultural background,

especially their L1 pragmatic cultural schemas, which are assumed to be common among the raters and the test-takers.

Results of content analysis of the semi-structured interviews with 10 experienced and 10 novice NNS teachers showed that they emphasized both pragmalinguistic and sociopragmatic factors in scoring MR-WDCTs. Of the extracted criteria, ‘grammatical accuracy’ from the pragmalinguistic group and ‘observing required speech act’ on a par with ‘following norms for NS pragmatic cultural schemas’ from the sociopragmatic category were the most dominant factors besides the other ones, namely ‘observing formality’, ‘politeness’, and ‘interlocutor’s social status’. Despite this similarity, the present study revealed a considerable difference between NNS experienced and novice teachers’ scoring criteria in terms of the consideration of observing perceived NS norms for pragmatic cultural schemas. Specifically, this factor was more important for the experienced teachers than for the novice ones, though with varying severity. The raters who were less severe in taking L1 pragmatic cultural schema transfer into account remarked assigning only a portion of the score to this factor and reserving the rest for attending to other pragmalinguistic and sociopragmatic factors. Interestingly, one of the noted factors was ‘examiner’s cultural background’ in that sharing particular pragmatic cultural schema might convince the examiner as to the partial appropriateness of the given response.

Results also revealed the similarity of the data-driven scoring criteria with the expert-driven one, despite some observed differences. As was observed in content analysis of the interviews, in redefining the ‘cultural accommodation’ factor in expert-driven criteria, the experienced and novice teachers referred to observing NS norms in terms of ‘use of semantic formula’, ‘formality’, ‘observing interlocutors’ relationship’, ‘use of proverbs’, and ‘politeness’, all of which were based on how the participant NNS teachers understood NS norms. However, the L1 pragmatic cultural schema influence is acceptable for them only if it does not result in misunderstanding.

Limitations

There are inevitably certain limitations to this study that merit attention. First, this study used convenience sampling based on the accessibility of participants which constraints its transferability. Second, the number of participants includes 20 NNS teachers in two groups of 10 experienced and 10 novice teachers, which is relatively small. To improve dependability, further studies should include more participants. Third, test developers’ ideas were not taken into account in investigating the consideration of L1 pragmatic cultural schema in scoring pragmatic assessment tasks.

Interpretations

Despite the limitations mentioned above, the results can be interpreted as follows. The stated factors by NNS experienced and novice teachers for scoring MR-WDCTs are in line with the results of previous similar studies, including Alemi and Khanlarzadeh (2016), Alemi et al. (2014), Chen and Liu (2016), Taguchi (2011), and Tajeddin and Alizadeh (2015). In addition, these factors signify both experienced and novice teachers’ awareness, at least partially, of the most important factors underlying pragmatic assessment that challenges the results of most pragmatic rater research, including Alemi and Rezanejad (2014), Alemi et al. (2014), and Tajeddin and Alizadeh (2015).

The observed difference between experienced and novice teachers' scoring criteria is similar to Alemi et al. (2014) and in contrast to Alemi and Khanlarzadeh (2016). In addition, the present study added to the results of previous related studies (e.g., Alemi & Tajeddin, 2013; Sonnenburg et al., 2020) qualitatively via unearthing layers of cultural factors in teachers' criteria. The influence of raters' cultural background in scoring answers containing L1 pragmatic cultural schema was also observed by Alemi and Rezanejad (2014), who concluded that cultural similarity between test-takers and examiners might shift the behavior of assessors to the lenient side. Such leniency was observed in the present study in varying severity. This can be discussed in light of the dissimilar perception of experienced and novice teachers of a) NS norms and b) the influence of L1 pragmatic cultural schemas in conveying the intended speech act. Not only might the aforementioned leniency of some teachers in the current study represent their awareness, as unconscious as it might be, of ELF pragmatics principles, but also it can reflect a lack of any systematic way to distinguish between appropriate and inappropriate responses in relation to NS norms (Youn & Bogorevich, 2019).

The resulting similarity of the data-driven scoring criteria with the expert-driven one highlights that both the experienced and novice teachers enjoyed an acceptable pragmatic assessment literacy. However, their categorization of the closely-related pragmatic concepts under a single heading in their assessment criteria and their adaptation of the given expert-driven criteria indicate that they need more training in pragmatic assessment. It is postulated that working with such expert-driven criteria in scoring pragmatic assessment of various kinds can raise NNS teachers' awareness of the possible distinctions among pragmalinguistic and sociopragmatic factors.

The acceptability of following norms for L1 pragmatic cultural schema in responding MR-WDCTs is in agreement with ELF assessment principles suggested by Elder and Davies (2006). However, only experienced teachers classified the linguistic factors of lexical and grammatical aspects under cultural accommodation. This can be interpreted as perceiving lexical and grammatical issues inseparable from verbal aspects of culture, which was ignored by novice teachers. Since pragmatic cultural schemas can be conveyed via lexicon and grammar (Sharifian, 2016), training for novice teachers in pragmatic assessment seems to be necessary.

Suggestions for Future Research

ELF pragmatic assessment awaits further research, some of which can be proposed taking the interpretation of the findings of the present study into account. First, future research can focus on the comparison of NNS and NS scoring criteria in addressing instances of L1 pragmatic cultural schema transfer within pragmatic assessment performance. Second, further studies might be conducted on a similar topic to the one investigated in the present study centering on expert, rather than experienced teachers, as pragmatic assessment raters. Finally, the effect of rater training in pragmatic cultural schemas on scoring ELF pragmatic assessment is waiting to be investigated.

Pedagogical Implications

The results of the current study inform ELT teachers about the role L1 pragmatic cultural schemas might play in scoring ELF pragmatic assessment tasks. Specifically, the results highlight the necessity to raise raters' awareness with regard to accepting responses to pragmatic assessment tasks which contain L1 pragmatic cultural schemas as far as they do not

hamper communication. Such a view on rating pragmatic assessment tasks can be considered as the first steps in appreciating ELF pragmatic assessment.

Conclusions

It can be concluded from the findings that pragmatic cultural schemas should be perceived as an essential scoring criterion for pragmatic assessment despite the possibility of being addressed in degrees of severity by experienced and novice teachers. This consideration can shed more light on the newly established field of ELF pragmatic assessment and its rating issues. In addition, the current study further the discussions regarding the integration of culture in language assessment tasks, especially in pragmatic assessment, wherein the language and culture assessment meet.

About the Authors

Ali Dabbagh is a Lecturer of Applied Linguistics at Gonbad Kavous University, Iran, where he teaches research methodology, language assessment, teaching methodology, linguistics, and philosophy of education. His research interest centers on language and culture, especially Cultural Linguistics. He presented and published articles in national and international academic conferences and journals, including *Issues in Language Teaching*, *Teaching English Language*, *The Journal of Asia TEFL*, *The Language Learning Journal*, and *International Journal of Applied Linguistics*. ORCID ID: <https://orcid.org/0000-0002-4795-1984>

Esmat Babaii is Professor of Applied Linguistics at Kharazmi University, Iran, where she teaches research methods, language assessment and discourse analysis to graduate students. She has served on the editorial boards and/or the review panels of several national and international journals. She has published articles and book chapters dealing with issues in language assessment, systemic functional linguistics, appraisal theory, test-taking processes, discursive analysis of textbooks, and critical approaches to the study of culture and language. Her most recent work is co-authored a paper on Middle Eastern culture-specific lexicon published in *Asian Englishes* (2020). ORCID ID: <https://orcid.org/0000-0001-9998-8247>

To cite this article:

Dabbagh, Al., & Babaii, E. (2021). L1 Pragmatic Cultural Schema and Pragmatic Assessment: Variations in Non-Native Teachers' Scoring Criteria. *Teaching English as a Second Language Electronic Journal (TESL-EJ)*, 25(1). <https://tesl-ej.org/pdf/ej97/a2.pdf>

References

Alemi, M., Eslami Rasekh, Z., & Rezanejad, A. (2014). Iranian non-native English speaking teachers' rating criteria regarding the speech act of compliment: An investigation of teachers' variables. *The Journal of Teaching Language Skills*, 6(3), 21–49. <https://doi.org/10.22099/jtls.2015.2481>

- Alemi, M., & Khanlarzadeh, N. (2015). Native EFL raters' criteria in assessing the speech act of complaint: The case of American and British EFL teachers. *Iranian Journal of Applied Language Studies*, 7(2), 25–60. <https://doi.org/10.22111/IJALS.2015.2672>
- Alemi, M., & Khanlarzadeh, N. (2016). Pragmatic assessment of request speech act of Iranian EFL learners by non-native English speaking teachers. *Iranian Journal of Language Teaching Research*, 4(2), 19–34. <https://doi.org/10.30466/IJLTR.2016.20363>
- Alemi, M., & Rezanejad, A. (2014). Native and non-native English teachers' rating criteria and variation in the assessment of L2 pragmatic production: The speech act of compliment. *Issues in Language Teaching*, 3(1), 65–88.
- Alemi, M., & Tajeddin, Z. (2013). Pragmatic rating of L2 refusal: Criteria of native and nonnative English teachers. *TESL Canada Journal*, 30(7), 63–81. <https://doi.org/10.18806/tesl.v30i7.1152>
- Barron, A. (2003). *Acquisition in interlanguage pragmatics: Learning how to do things with words in a study abroad context*. John Benjamins.
- Chen, Y-S., & Liu, J. (2016). Constructing a scale to assess L2 written speech act performance: WDCT and email tasks. *Language Assessment Quarterly*, 13(3), 231–250. <https://doi.org/10.1080/15434303.2016.1213844>
- Cohen, A.D. (2020). Considerations in assessing pragmatic appropriateness in spoken language. *Language Teaching*, 53(2), 183–202. <https://doi.org/10.1017/S0261444819000156>
- Cohen, A. D., & Shively, R. L. (2002/2003). Measuring speech acts with multiple rejoinder DCTs. *Language Testing Update*, 32, 39–42.
- Dey I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. Routledge.
- Edmonson, W.J., & House, J. (1991). Do learners talk too much? The waffle phenomenon in interlanguage pragmatics. In R. Philipson, E. Kellerman, L. Selinker, M. Sharwood Smith, & M. Swain (Eds.), *Foreign language pedagogy research: A commemorative volume for Clause Faerch* (pp. 273–287). Multilingual Matters.
- Elder, C., & Davies, A. (2006). Assessing English as a lingua franca. *Annual Review of Applied Linguistics*, 26, 282–301. <https://doi.org/10.1017/S0267190506000146>
- Farrell, T. (2012). Novice-service language teacher development: Bridging the gap between preservice and in-service education and development. *TESOL Quarterly*, 46(3), 435–449. <https://doi.org/10.1002/tesq.36>
- Fulcher, G., Davidson, F., & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5–29. <https://doi.org/10.1177/0265532209359514>
- Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, 24(1), 90–121. <https://doi.org/10.1093/applin/24.1.90>

- Grabowski, K. C. (2009). *Investigating the construct validity of a test designed to measure grammatical and pragmatic knowledge in the context of speaking* [Unpublished PhD dissertation]. Columbia University.
- House, J. (2013). Developing pragmatic competence in English as a lingua franca: Using discourse markers to express (inter)subjectivity and connectivity. *Journal of Pragmatics*, 59, 57–67. <https://doi.org/10.1016/j.pragma.2013.03.001>
- Hudson, T., Detmer, E., & Brown, J. D. (1995). *Developing prototypic measures of cross cultural pragmatics* [Technical Report #7]. University of Hawai'i at Manoa.
- Ishihara, N. (2010). Assessing learners' pragmatic ability in the classroom. In D. H. Tatsuki & N. R. Houck (Eds.), *Pragmatics: Teaching speech acts* (pp. 209–228). Teachers of English to Speakers of Other Languages.
- Jenks, C. (2013). “Your pronunciation and your accent is very excellent”: Orientations of identity during compliment sequences in English as a lingua franca encounters. *Language and Intercultural Communication*, 13(2), 165–181. <https://doi.org/10.1080/14708477.2013.770865>
- Kasper, G., & Roever, C. (2005). Pragmatics in second language learning. In E. Hinkel (Ed.), *Handbook of research in second language learning and teaching* (pp.317–334). Routledge.
- Knapp, A. (2011). Using English as a lingua franca for (mis-)managing conflict in an international university context: An example from a course in engineering. *Journal of Pragmatics*, 43(4), 978–990. <https://doi.org/10.1016/j.pragma.2010.08.008>
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275–304. <https://doi.org/10.1177/0265532208101008>
- Leech, G. (1983). *Principles of pragmatics*. Longman.
- Liu, J., & Xie, L. (2014). Examining rater effects in a WDCT pragmatic test. *Iranian Journal of Language Testing*, 4(1), 50–65.
- Mondada, L. (2012). The dynamics of embodied participation and language choice in multilingual meetings. *Language in Society*, 41(2), 213–235. <https://doi.org/10.1017/S004740451200005X>
- Park, S.-Y. J. (2017). Transnationalism as interdiscursivity: Korean managers of multinational corporations talking about mobility. *Language in Society*, 46(1), 23–38. <https://doi.org/10.1017/S0047404516000853>
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Sage.
- Rallis Sh. F., & Rossman G. B. (2009). Ethics and trustworthiness. In J. Heigham & R.A. Crocker (Eds.), *Qualitative research in applied linguistics: A practical introduction* (pp. 263–287). Palgrave MacMillan.
- Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, 28(4), 463–481. <https://doi.org/10.1177/0265532210394633>
- Ross, S., & Kasper, G. (2013). *Assessing second language pragmatics*. Palgrave Macmillan.

Roever, C., & Kasper, G. (2018). Speaking in turns and sequences: Interactional competence as a target construct in testing speaking. *Language Testing*, 35(3), 331–355. <https://doi.org/10.1177/0265532218758128>

Sasaki, M. (1998). Investigating EFL students' production of speech acts: A comparison of production questionnaires and role plays. *Journal of Pragmatics*, 30(4), 457–484. [https://doi.org/10.1016/S0378-2166\(98\)00013-7](https://doi.org/10.1016/S0378-2166(98)00013-7)

Schnurr, S., & Zayts, O. (2013). “I can't remember them ever not doing what I tell them!": Negotiating face and power relations in “upward” refusals in multicultural workplaces in Hong Kong. *Intercultural Pragmatics*, 10(4), 593–616. <https://doi.org/10.1515/ip-2013-0028>

Seidlhofer, B. (2011). *Understanding English as a lingua franca*. Oxford University Press.

Sharifian, F. (2003). On cultural conceptualizations. *Journal of Cognition and Culture*, 3(3), 187–207. <https://doi.org/10.1163/156853703322336625>

Sharifian, F. (2005). The Persian cultural schema of *shekasteh-nafsi*: A study of complement responses in Persian and Anglo-Australian speakers. *Pragmatics & Cognition*, 13(2), 337–361. <https://doi.org/10.1075/pc.13.2.05sha>

Sharifian, F. (2011). *Cultural Conceptualizations and Language: Theoretical Framework and Applications*. John Benjamins.

Sharifian, F. (2014). Cultural schemas as common ground. In K. Burrige, & R. Benczes (Eds.), *Wrestling with words and meanings: Essays in honor of Keith Allan* (pp. 219–235). Monash University Publishing.

Sharifian, F. (2015). Cultural Linguistics. In F. Sharifian (Ed.), *The Routledge handbook of language and culture* (pp. 473–492). Routledge.

Sharifian, F. (2016). Cultural pragmatic schemas, pragememes, and practs: A Cultural Linguistics perspective. In K. Allan, A. Capone, & I. Kecskes (Eds.), *Pragememes and theories of language use* (pp. 505–519). Springer.

Sharifian, F. (2017). *Cultural Linguistics: Cultural conceptualizations and language*. John Benjamins.

Schneider, K. (2008). Small talk in England, Ireland and the U.S.A. In K. Schneider, & A. Barron (Eds.), *Variational pragmatics: A focus on regional varieties in pluricentric languages* (pp. 97–139). John Benjamins.

Shohamy, E., Iair, G., & May, S. (2017). *Language testing and assessment* (3rd ed.). Springer.

Sonnenburg-Winkler, S. L., Eslami, Z. R., & Derakhshan, A. (2020). Rater variation in pragmatic assessment: The impact of linguistic background on peer-assessment and self-assessment. *Lodz Papers in Pragmatics*, 16(1), 67-85.

Taguchi, N. (2011). Rater variation in the assessment of speech acts. *Pragmatics*, 21(3), 453–471. <https://doi.org/10.1075/prag.21.3.08tag>

Taguchi, N. (2017). Interlanguage pragmatics. In A. Barron, P. Grundy, & G. Yueguo (Eds.), *The Routledge handbook of pragmatics* (pp. 153–167). Routledge.

- Taguchi, N., & Ishihara, N. (2018). The Pragmatics of English as a lingua franca: Research and pedagogy in the era of globalization. *Annual Review of Applied Linguistics*, 38, 80–101. <https://doi.org/10.1017/S0267190518000028>
- Tajeddin, Z., & Alemi, M. (2013). Criteria and bias in native English teachers' assessment of L2 pragmatic appropriacy: Content and FACETS analyses. *Asia-Pacific Education Researcher*, 23(3), 435–434. <https://doi.org/10.1007/s40299-013-0118-5>
- Tajeddin, Z., & Alemi, M. (2014). Pragmatic rater training: Does it affect non-native L2 teachers' rating accuracy and bias? *Iranian Journal of Language Testing*, 4(1), 66–83.
- Tajeddin, Z., & Alizadeh, I. (2015). Monologic vs. dialogic assessment of speech act performance: Role of nonnative L2 teachers' professional experience on their rating criteria. *Journal of Research in Applied Linguistics*, 6(1), 3–27. <https://doi.org/10.22055/RALS.2015.11257>
- Timpe Laughlin, V. T., Wain, J., & Schmidgall, J. (2015). *Defining and operationalizing the construct of pragmatic competence: Review and recommendations*. ETS Research Report Series, 1–43. <https://doi.org/10.1002/ets2.12053>
- Walters, F.S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155–183. <https://doi.org/10.1177/0265532207076362>
- Yamashita, S. (1998). *What do DCTs and roleplays assess?* Paper presented at the meeting of the American Association for Applied Linguistics, Seattle, WA.
- Yamashita, S. (2005) 'I'm sorry, but could you be a bit more careful?': Pragmatic strategies of Japanese elderly people. In D. Tatsuki (Ed.), *Pragmatics in language learning: Theory and practice* (pp. 119–136). The Japan Association for Language Teaching (JALT).
- Yamashita, S. (2008). Investigating interlanguage pragmatic ability: What are we testing? In E. Alcon Soler, & A. Martinez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 201–223). Multilingual Matters.
- Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, 32(2), 199–225. <https://doi.org/10.1177/0265532214557113>
- Youn, S. J. (2018). Task-based needs analysis of L2 pragmatics in an EAP context. *Journal of English for Academic Purposes*, 36, 86–98. <https://doi.org/10.1016/j.jeap.2018.10.005>
- Youn, S. J., & Bogorevich, V. (2019). Assessment in L2 pragmatics. In N. Taguchi (Ed.), *Routledge handbook of SLA and pragmatics* (pp. 308–321). Routledge.
- Zhu, W. (2017). How do Chinese speakers of English manage rapport in extended concurrent speech? *Multilingua*, 36(2), 181–204. <https://doi.org/10.1515/multi-2015-0112>

Appendix A: First interview questions

1. Based on the answers provided by intermediate students for the given test items, what factors will you take into consideration while scoring these tests? Please provide your reasons behind the factors you focus on.
2. As you see, some of the students' answers do not match the way a native speaker use language in the given scenario. What factors are important for you in scoring such answers? Why?
3. To what extent is following native speaker norms significant for you in scoring these items?

Appendix B: Second interview questions

1. Which of the factors in the given criteria should be considered in scoring this test? Which of the factors do you think are not necessary? Why?
2. How would you adapt the given criteria if you were asked to score this test?

Copyright rests with authors. Please cite TESL-EJ appropriately.