



Content list available at www.urmia.ac.ir/ijltr

*Iranian Journal
of
Language Teaching Research*



Urmia University

Pragmatic assessment of request speech act of Iranian EFL learners by non-native English speaking teachers

Mino0 Alemi ^{a,*}, Neda Khanlarzadeh ^b

^a *Islamic Azad University – Tehran-West Branch, Iran*

^b *Sharif University of Technology, Iran*

ABSTRACT

The analysis of raters' comments on pragmatic assessment of L2 learners is among new and understudied concepts in second language studies. To shed light on this issue, the present investigation targeted important variables such as raters' criteria and rating patterns by analyzing the interlanguage pragmatic assessment process of the Iranian non-native English speaking raters (NNESRs) regarding the request speech act, while considering important factors such as raters' gender and background teaching experiences. For this purpose, 62 raters' rating scores and comments on Iranian EFL learners' requests based on six situations of specified video prompts were analyzed. The results of the content analysis of raters' comments revealed nine criteria, including pragmalinguistic and socio-pragmatic components of language, which have been noted by raters differently through six request situations. Among the considered criteria, politeness, conversers' relationship, style and register, and explanation were of great importance to NNESRs. Furthermore, t-test and chi-square analysis of raters' assigned rating scores and mentioned criteria across different situations verified the insignificance of factors such as raters' gender and teaching experiences on the process of EFL learners' pragmatic assessment. In addition, the results of the study suggest the necessity of teaching L2 pragmatics in language classes and in teacher training courses.

Keywords: interlanguage pragmatics; non-native English speaking raters; pragmatic assessment; rating criteria; request

© Urmia University Press

ARTICLE HISTORY

Received: 10 Jan. 2016

Revised version received: 24 Feb. 2016

Accepted: 28 May 2016

Available online: 1 July 2016

* Corresponding author: Islamic Azad University – Tehran-West Branch, Iran
Email address: minooalemi2000@yahoo.com

Introduction

Since pragmatic knowledge has been stated explicitly in Bachman's (1990) model of Communicative Competence, numerous scholars have studied its various aspects. One of the interesting aspects of pragmatics to investigate is the area of interlanguage pragmatics (ILP) which deals with L2 learners' pragmatic knowledge about the target language. Concepts of teaching and testing ILP attracted the attention of many researchers, as several rubrics were proposed for implementing them in foreign language classes (*see* Cohen, 2008). However, there are so many unresolved issues regarding incorporating L2 pragmatics into language teaching programs, especially in assessment area which plays a major role in second language process. Occasionally, language learners produce utterances which are linguistically accurate but pragmatically unsatisfactory. Such problems should not be overlooked by teachers, but rather treated with caution in order to save learners from a great deal of embarrassment in communication with native speakers. According to Roever (2007), L2 pragmatic assessment is a new branch of language testing; however, there are not many tests in this respect. Furthermore, the study of speech act rating, an aspect of pragmatic assessment, is one of the new and underdeveloped research areas which is in rigorous need of analysis. Few researchers including Alemi and Tajeddin (2013), and Taguchi (2011) have attempted to investigate this issue so far; thus, this study aims to explore Iranian non-native English speaking teachers' assessment patterns regarding request productions of Iranian EFL learners.

Literature Review

ILP rating

According to Bachman, pragmatic assessment, like any other type of assessment, is composed of three major phases: the theoretical definition of the construct, operational definition of the construct, and observation of the learners' performance (cited in Tajeddin, 2014). Theoretical definition refers to the underlying psychological trait that is intended to be assessed. In the description of pragmatic competence, Leech (1983) and Thomas (1983) proposed two important components namely pragmalinguistics and socio-pragmatics. The former refers to linguistic resources of expressing a speech intention and the latter focuses on social constraints which have to be considered while using the linguistic resources.

In operationalizing the pragmatic competence, several studies were accomplished with the focus of extracting the pragmatic knowledge of language learners. Hudson, Brown and Detmer (1995) and Hudson, Detmer, and Brown (1992) employed different methods in testing politeness and degree of directness of learners' apology, request and refusal competencies. The instruments used included oral DCTs, written DCTs, multiple choice DCT, role plays, and self-assessment.

Later, Roever (2005, 2006, 2007) developed a web-based test of pragmatics which was different from the discussed instruments in that he tried to focus on implicatures and routine formula. Roever's instrument was also less biased and more appropriate for both Asian and European test takers. Walters (2004, 2007, 2009) criticized the previous speech act theory-based L2 pragmatic tests, and claimed that they raised validity issues due to their lack of compatibility with conversational data. Walter focused on conversation analysis in testing pragmatic comprehension, production, compliment responses, and pre-sequences responses of language learners through role plays, DCTs, and listening comprehension. However, Walters' instrument was not free from pitfalls as it lacked reliability in all of its sub-part tests.

The last step of pragmatic assessment deals with quantifying the observation of learners' pragmatic performances, which according to Bachman (1990), could be accomplished through either rating on scales or counting the correct responses. While most researchers prefer scoring based on defining appropriateness levels on scales which are dominantly employed in rating DCTs and role plays (Alemi, 2012; Hudson et al., 1995; Taguchi, 2011), the second way is commonly used for multiple choice DCTs.

Cohen (2014) stressed the necessity of considering several important factors in assessing pragmatic competence. He suggested strategies such as: using realistic situations, checking different features of the performance, asking students to compare their responses with native speakers, asking students to explain their rationales for their answers, and being strategic in assessing various aspects of the speech acts.

A new branch of pragmatic assessment enquiries is related to raters' issues such as raters' biases and criteria in rating. Rater-related issues were found to be very critical in the assessment process as they could easily undermine the validity of the test (Bachman, 2004). Moreover, as Knoch, Read, and Von Randow (2007) maintained, several sources of biases and errors might penetrate in raters' judgments which could affect the rating quality. However, there have been limited studies which aimed at investigating the interface between interlanguage pragmatic assessment and raters' criteria. Taguchi (2011) analyzed native English-speaking raters' assessment of requests and opinions produced by Japanese EFL students. The results of the introspective interview of raters regarding their rating norms as well as the analysis of their comments about their rating decisions revealed that native raters considered issues such as "politeness markers", "amount of speech", "strategies", "clarity of intention", "directness", and "content" of the EFL learners' responses. He also found that raters were divergent in their assessments, although they were all native speakers of English. Alemi and Tajeddin (2013) also focused on ILP rating; they investigated native and non-native raters' rating criteria regarding the assessment of EFL learners' production of the refusal speech act. Their analysis of raters' comments revealed six major criteria: "politeness", "brief apology", "irrelevancy of speech act", "postponing to another time", "explanation", and "statement of alternatives". Later in the same line, Alemi, Eslami-Rasekh, and Rezanejad (2015) analyzed Iranian non-native EFL teachers' rating criteria during the assessment of EFL learners' compliment productions. The content analysis of raters' justifications showed seven major criteria including: "politeness", "interlocutors' characteristics and relationship", "variety and range", "socio-pragmatic appropriateness", "sincerity", "complexity", and "linguistic appropriacy". Alemi et al. (2015) also analyzed the relationship between raters' gender and teaching experiences and found significant differences in terms of frequency of their rating criteria.

Moreover, Sydorenko, Maynard, and Guntly (2015) analyzed salient criteria employed by three raters whilst assessing extended request sequences by EFL learners. They examined raters' opinions on EFL learners' head acts and supporting moves, combined with further analysis of raters' comments about learners' requests. Consequently, it was found that criteria such as "appropriate request strategies", "repetitiveness", "cultural misunderstanding", and "intonation" were among salient factors considered by raters in scoring.

Request studies

According to Li (2000), among various speech acts investigated in several studies, request was of great importance for many researchers (e.g., Blum-Kulka, 1987; Hassall, 2004; Takahashi & DuFon, 1989; Woodfield, 2008). This importance is due to the complexity in relationships between its form, meaning, as well as pragmatics and the critical social risks involved for speakers. According to Searle's (1976) well-known classification of illocutionary acts, request is characterized as a directive speech act through which the speaker attempts to make the listener do something; or in Trosborg's

(1995) words, “an illocutionary act whereby a speaker (requester) conveys to hearer (requestee) that he/she wants the requestee to perform an act which is for benefit of speaker” (p. 187). According to Brown and Levinson (1987), request, due to its seriousness, is considered as a face-threatening speech act which can be performed either in direct manner or accompanied by mitigating devices. The seriousness of face-threatening act can be assessed through factors such as “social distance”, “degree of imposition”, and the “power” of the interlocutors.

In Blum-Kulka and Olshtain’s (1984) cross-cultural speech act realization project (CCSARP), based on the analysis of requests across eight languages, requests were divided into three levels in terms of their directness: “the most direct level”, “conventionally indirect level”, and “nonconventional indirect level”. In direct level, requests are expressed by syntactic means. In conventionally indirect level, requests are expressed indirectly by conventionalized or fixed request expressions known by the speakers of the language. Nonconventional indirect level refers to requests in forms of hints which are usually implied through contextual factors. Blum-Kulka and Olshtain further proposed different strategies employed in mentioned levels of directness in order of their directness: “mood derivable”, “explicit performatives”, “hedged performatives”, “locution derivable”, “scope stating”, “language-specific suggestory formula”, “reference to preparatory conditions”, “strong hints”, and “mild hints”. Moreover, some studies attempted to categorize some of the request strategies. Takahashi (1996) classified the preparatory expressions into four categories: “preparatory questions”, “questions regarding permission”, “mitigated-preparatory”, and “mitigated-wants”.

During recent years, request has been analyzed in forms of cross-cultural and interlanguage studies. Several studies have proved that besides its universal characteristics, the existing differences in performing and realizing request speech act necessitate teaching and testing it for the EFL learners (*see* Eslami-Rasekh, 2005; Izaki, 2000; Jalilifar, Hashemian, & Tabatabaee, 2011; Woodfield, 2008). Therefore, more studies are required in order to inform EFL teachers about various aspects of the request speech act, which, in turn, should be considered in teaching and assessment processes. This study, in response to the gap in literature regarding NNESRs’ assessment behaviors during rating request speech act, aimed at exploring different criteria employed by Iranian non-native EFL teachers as well as their rating variations. To this end, the following research questions were addressed:

1. What are the criteria considered by Iranian NNESRs during the pragmatic assessment of request speech act?
2. Is there any significant difference between female and male NNESRs’ rating scores and rating criteria?
3. Is there any significant difference in NNESRs’ rating scores and rating criteria based on their teaching experience?

Method

Participants

The main purpose of the present study was to reveal the criteria that underpinned non-native English speaking raters’ (NNERs) rating of the EFL learners’ request productions. Participants of the study consisted of 62 non-native English speaking teachers and 12 Iranian EFL learners. The group of NNESRs included English teachers from different language institutes in Iran with various teaching experiences (classified into two levels of 1-5 and 6-11). The EFL teachers were also

selected from both genders (28 males and 34 females). Moreover, all of the EFL teachers were M.A. holders or M.A. students of TEFL (Teaching English as a Foreign Language); therefore, they were familiar with the concept of L2 pragmatics and language testing. The group consisting of 12 Iranian EFL learners came from upper-intermediate to advanced levels, as they were required to understand the video prompts and produce related responses.

Instruments

A video-prompted discourse completion test (DCT) was employed for collecting the necessary data for this survey. The 6 video prompts of the DCT were selected from American movies to set the request situation for the EFL learners. The video prompts covered various degrees of formality, power, and distance between interlocutors with the focus on everyday life context, educational context, and workplace context. The six situations of the request video prompts included: *Asking for a book from a friend, Asking a neighbor/boss to turn down the volume of the music, asking a shopkeeper to show you a dress, asking one of your employers to talk in slower rate, asking a neighbor to help you with fixing your computer, and asking a stranger to be silent in public library*, respectively. In the final DCT, the transcription of each video prompt situation was prepared, for the sake of convenience in rating, followed by EFL learners' responses to each situation. In addition, a five-point Likert scale (1= very unsatisfactory, 2= unsatisfactory, 3= somewhat appropriate, 4= appropriate, and 5= most appropriate) was placed after every response for raters.

Data collection procedure

In order to elicit request productions of EFL learners, the video prompts were shown to the group consisting of 12 Iranian English learners. The responses were reviewed by the authors and only one answer was selected for each situation. The selected answers had to represent the typical pragmatic mistakes of EFL learners, for example utterances which shows that learners are not familiar with cultural norms of the target language society, and vary in the degree of appropriateness so that the raters could select different points on the Likert scale in rating answers across different situations.

During the next phase, the video prompt situations and selected responses of EFL learners were transcribed in the form of written DCT and were distributed among the 62 NNESRs. The raters rated the EFL learners' responses on the five-point Likert scale and mentioned the criteria that they considered throughout the rating process.

Data analysis

Data analysis consisted of both quantitative and qualitative procedures. Through the qualitative phase of data analysis, the criteria noted by NNESRs were analyzed and categorized. Thereon, the frequency of each criterion was calculated through quantitative procedure in order to find the dominant criteria. Moreover, t-test and chi-square were conducted to determine if there was any statistically significant difference between a) female and male NNESRs' rating scores and rating criteria, and b) NNESRs' rating scores and criteria preferences based on their teaching experience.

Results

Rating criteria

In order to answer the first research question on the various criteria that NNESRs used during pragmatic assessment of EFL learners' requests, the content analysis technique was employed. The following criteria were found in raters' comments:

- (1) *Directness*: This request criteria deals with the directness and indirectness of the EFL learners' productions. Occasionally, direct and intense requests such as "Leave me alone." or "Clean up this mess, please." (Blum-Kulka & Olshtain, 1984, p. 202) seem more effective due to contextual factors, while in other circumstances, indirect request seems more appropriate and polite to the hearers. An example of one of the raters' comment regarding this criterion is presented below:

Example: This is too direct, although the speaker is of higher status compared to the listener.

- (2) *Politeness*: This criterion which is one of the main criteria refers to the degree of politeness of the EFL learners' request. Politeness perception roots in cultural norms of the raters and EFL learners to a great extent. It is also bounded by situational features and the interlocutors. The following example illustrates the politeness notion in NNESRs' comments:

Example: I think it's not very polite. The managers should respect the teachers, especially in front of other colleagues.

- (3) *Language usage accuracy*: This criterion is not at all concerned with the pragmatic aspect of language, but it was mentioned by raters several times. It is mainly about the accuracy of the structures, grammar, and lexical items of the produced sentences. The following example indicates an instance of such a criterion:

Example: There are some grammar mistakes. For example, "it" should be replaced by its reference "the music".

- (4) *Authenticity and cultural errors*: This criterion reflects the genuineness and naturalness of the produced responses, as well as their cultural appropriateness regarding L2 society. In some situations, NNESRs found the EFL learners responses odd, unnatural or unlikely to be uttered by a native speaker. An example of one of the NNESRs comments regarding this criterion is presented below:

Example: This sentence seems odd and unnatural. Americans would never say that, especially the "go ahead" part.

- (5) *Style and register*: This criterion refers to the attention given to the use of the formal or informal style as well as appropriate language in a given situation. The following example highlights one of the raters' comments on style and register criterion:

Example: Asking your friend formally might lead to misunderstanding.

(6) *Explanation*: This criterion refers to the necessity of the brief explanation or introduction before making request which indicates the speaker's reason for requesting. One of the examples of this criterion is presented below:

Example: I think it's better to add an introduction and clarify the request.

(7) *Statement of optimal example*: Through this criterion, raters supplied various examples of the ideal request for the specified situations which might be accompanied with explanation of each employed move. Such alternatives are usually given for evaluating the EFL learners' responses with an ideal example. An instance of raters' alternatives is presented below:

Example: She/he could say: "I need that book for my assignment. Please let me borrow it for a few days if you don't need it"

(8) *Query preparatory and softeners*: This criterion refers to the importance of the use of preparatory expressions such as could you, would you, etc., as well as words or phrases which can moderate the request (i.e. please, thank you, if it's OK with you). An example of raters' comments in this respect is given below:

Example: "Pardon me" followed by the word "excuse me" is more favored.

(9) *Conversers' relationship*: Generally what we communicate with other people is based on our social relationships, as realizing, establishing, sustaining, and changing social relations are among important factors in communication (Adel, Davoudi, & Ramezanzadeh, 2016). Paying attention to interlocutors' relationship and closeness based on the contextual factors seems highly important to the NNESRs, since they noted this factor frequently in their comments. An example of such a criterion is given below:

Example: It depends on the closeness of the relationship. If it's an employee boss relationship, then the sentences are informal and not proper for this situation, whereas, it is considered proper between 2 friends.

The above mentioned criteria were used differently across six situations of the DCT, since various request situations demand different criteria to be noted. Table 1 shows the frequency and percentage of use of each criterion by NNESRs in each situation.

Table 1
Frequency of Request Criteria among Iranian NNESRs

| Situations | D | P | LUA | ACE | SR | E | SOE | QPS | CR | Total |
|------------|-------|-------|------|------|--------|--------|-------|-------|--------|-------|
| S1 | 3 | 13 | 5 | 4 | 26 | 8 | 1 | 3 | 29 | 92 |
| S2 | 5 | 20 | 2 | 1 | 17 | 18 | 5 | 6 | 22 | 96 |
| S3 | 2 | 17 | 2 | 1 | 19 | 2 | 1 | 9 | 4 | 57 |
| S4 | 8 | 28 | 1 | 1 | 12 | 4 | 0 | 5 | 30 | 89 |
| S5 | 4 | 17 | 3 | 0 | 8 | 19 | 4 | 11 | 4 | 70 |
| S6 | 18 | 34 | 1 | 0 | 9 | 3 | 9 | 9 | 11 | 94 |
| Total | 40 | 129 | 14 | 7 | 91 | 54 | 20 | 43 | 100 | 498 |
| Percentage | 8.03% | 25.9% | 2.8% | 1.4% | 18.27% | 10.84% | 4.01% | 8.63% | 20.08% | 100% |

Note. D: Directness; P: Politeness; LUA: Language Usage Accuracy; ACE: Authenticity and Cultural Errors; SR: Style and Register; E: Explanation; SOE: Statement of Optimal Example; QPS: Query Preparatory and Softeners; CR: Conversers' Relationship

As it is depicted in Table 1, “politeness” (25.9%) was the leading criteria among NNEs. Other criteria which were used abundantly were “conversers’ relationship” (20.08%), “style and register” (18.27%), and “explanation” (10.84%). Criteria such as “preparatory and softeners” (8.63%) and “directness” (8.03%) were judiciously mentioned during assessment process. On the other hand, “alternatives” (4.01%), “language usage accuracy” (2.8%), and “authenticity and cultural errors” (1.4%) were among the least frequent criteria. Table 1 also indicates that the number and types of criteria employed in each situation are varied. For example, criteria such as “authenticity and cultural errors”, as well as “statement of optimal examples” were not mentioned in all situations, while “politeness” and “style and register” were amply pointed out in every situation.

Moreover, the descriptive statistics of raters’ assigned scores in each situation is presented in Table 2, which illustrates the convergence or divergence of NNEs’ rating.

Table 2
Descriptive Statistics of NNEs’ Rating Scores

| Situations | N | Minimum | Maximum | Mean | Standard Deviation |
|------------|----|---------|---------|------|--------------------|
| request1 | 62 | 1 | 5 | 3.64 | 1.31 |
| request2 | 62 | 1 | 5 | 3.03 | 1.2 |
| request3 | 62 | 1 | 5 | 4.19 | 1.09 |
| request4 | 62 | 1 | 5 | 2.27 | 1.07 |
| request5 | 62 | 1 | 5 | 3.61 | 0.96 |
| srequest6 | 62 | 1 | 5 | 1.67 | 0.9 |
| Total | 62 | 1 | 5 | 3.07 | 1.39 |

As it can be drawn from Table 2, the total mean score of the raters’ is 3.07, which denotes that NNEs generally considered the EFL learners’ requests as “somewhat appropriate” (according to the Likert scale of the DCI). However, the mean scores were inconsistent across various situations, as the rating scores in varied from 1 to 5. The highest mean score is for the third situation, denoting a request from shopkeeper to show you a dress, which according to Table 1, devoted fewer criteria to itself compared with other situations. On the other hand, the lowest mean score is for the last situation, which was about asking a stranger to be silent in public library. In addition, the minimum and maximum ranges of the given scores, proved the variability and divergence of the rating scores in each situation. Table 2 also highlights the standard deviations of the rating scores among NNEs which are lower in the last two situations.

Gender Factor

To investigate the effect of NNEs’ gender on their rating scores and criteria in response to the second research question, descriptive statistics, independent samples t-test and chi-square analysis were employed. The result of the descriptive statistics of both genders’ rating scores is presented in Table 3. Table 3 shows that in most situations, namely 1, 2, 5 and 6, the means of female raters’ assigned scores are lower than male raters’ scores. Meanwhile, the standard deviations of male raters’ given scores in the majority of situations are higher compared to female raters’. Furthermore, in order to figure out the significance of the differences between male and female raters’ assigned scores, independent samples t-test was run which is illustrated in Table 4. The result of the independent samples t-test ($t(60) = -9.80, p > .05$) proved that there was no significant difference between the female and male NNEs’ rating of the Iranian EFL learners’ request productions.

Table 3
Descriptive Statistics of Male and Female Raters' Rating Scores

| Gender | Situations | N | Minimum | Maximum | Mean | Standard Deviation |
|---------|------------|----|---------|---------|------|--------------------|
| Males | request1 | 28 | 1.00 | 5.00 | 3.28 | 1.3 |
| | request2 | 28 | 1.00 | 5.00 | 3 | 1.21 |
| | request3 | 28 | 2.00 | 5.00 | 4.32 | .86 |
| | request4 | 28 | 1.00 | 5.00 | 2.32 | 1.15 |
| | request5 | 28 | 1.00 | 5.00 | 3.42 | 1.03 |
| | request6 | 28 | 1.00 | 5.00 | 1.64 | .91 |
| Females | request1 | 34 | 1.00 | 5.00 | 3.94 | 1.27 |
| | request2 | 34 | 1.00 | 5.00 | 3.05 | 1.2 |
| | request3 | 34 | 1.00 | 5.00 | 4.08 | 1.26 |
| | request4 | 34 | 1.00 | 5.00 | 2.23 | 1.01 |
| | request5 | 34 | 2.00 | 5.00 | 3.76 | .88 |
| | request6 | 34 | 1.00 | 4.00 | 1.7 | .9 |

Table 4
Independent Samples t-test of Male and Female NNEsRs' Rating Scores

| | Levene's Test for Equality of Variances | | t-Test for Equality of Means | | | | | | |
|---------------|---|------|------------------------------|----|-----------------|---------------------|-----------------|----------------------------|---------------------------------------|
| | F | Sig. | t | df | Sig. (2-tailed) | (2-Mean Difference) | Std. Difference | Error95% of the Difference | Confidence Interval of the Difference |
| Total Request | .094 | .760 | -.980 | 60 | .331 | -.79412 | .81074 | Lower -2.41583 | Upper .82760 |

In addition, chi-square analysis was conducted in order to analyze the significant difference between female and male raters' employed criteria. Table 5 and Table 6 indicate the chi-square test of both genders and their noted criteria.

Table 5
Chi-square Test of Male and Female NNEsRs' Employed Criteria

| | Value | df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) | Point Probability |
|------------------------------------|-------------------|----|-----------------------|----------------------|----------------------|-------------------|
| Pearson Chi-Square | .334 ^a | 1 | .563 | .590 | .298 | |
| Continuity Correction ^b | .280 | 1 | .597 | | | |
| Likelihood Ratio | .334 | 1 | .563 | .590 | .298 | |
| Fisher's Exact Test | | | | .590 | .298 | |
| Linear-by-Linear Association | .334 ^c | 1 | .563 | .590 | .298 | .033 |
| N of Valid Cases | 3348 | | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 223.10.

b. Computed only for a 2x2 table

c. The standardized statistic is -.578.

Table 6
Symmetric Measures

| | Value | Approx. Sig. | Exact Sig. |
|--------------------|-------------------------|--------------|------------|
| Nominal by Nominal | Phi | .010 | .563 |
| | Cramer's V | .010 | .563 |
| | Contingency Coefficient | .010 | .563 |
| N of Valid Cases | 3348 | | |

As the result of the chi-square indicated in Table 5, there was no significant difference between female and male raters' use of various criteria ($\chi^2(1) = 0.334, p > .05$); therefore, the gender factor did not affect the raters' preferences of the mentioned criteria.

Experience factor

According to the third research question being: "Is there any significant difference in NNESRs' rating scores and rating criteria based on their teaching experience", the intended raters' teaching experiences were divided into two categories:

Group 1: Less experienced Iranian EFL teachers whose teaching experience is from 1 year to 5 years.

Group 2: Experienced Iranian EFL teachers who have 6-11 teaching experience.

The descriptive statistics and t-test of these two groups' rating scores are presented in Table 7 and Table 8, respectively.

Table 7
Descriptive Statistics of Raters' Experience and Their Rating Scores

| Experience | N | Minimum | Maximum | Mean | Standard Deviation | |
|------------|----------|---------|---------|------|--------------------|------|
| 1-5 | request1 | 32 | 1 | 5 | 3.78 | 1.28 |
| | request2 | 32 | 1 | 5 | 2.93 | 1.16 |
| | request3 | 32 | 1 | 5 | 4.28 | 1.11 |
| | request4 | 32 | 1 | 5 | 2.28 | 1.05 |
| | request5 | 32 | 2 | 5 | 3.84 | 0.84 |
| | request6 | 32 | 1 | 5 | 1.75 | 0.95 |
| 6-11 | request1 | 30 | 1 | 5 | 3.5 | 1.35 |
| | request2 | 30 | 1 | 5 | 3.13 | 1.25 |
| | request3 | 30 | 2 | 5 | 4.1 | 1.09 |
| | request4 | 30 | 1 | 5 | 2.26 | 1.11 |
| | request5 | 30 | 1 | 5 | 3.36 | 1.03 |
| | request6 | 30 | 1 | 4 | 1.6 | 0.85 |

As Table 7 illustrates, the mean of the rating scores between Group 1 and Group 2 are very close in all situations. Moreover, it has been proved that in all situations except for situation 2, the mean scores of Group 1 are higher than Group 2; In other words, less experienced teachers were more lenient in scoring the EFL learners' responses compared to more experienced teachers. Taking this further, the results of the standard deviation calculations revealed that less experienced raters were more convergent in their ratings in situations 1, 2, 4 and 5.

In addition, an independent samples t-test was conducted in order to check the relationship between raters' teaching experience and their rating scores. According to Table 8, t-test results ($t(41.6) = 1.106, p > .05$) demonstrate that there are no significant differences between Group 1 and Group 2 raters' rating scores.

Table 8
Independent Samples t-test of NNESRs' Experience and Their Rating Scores

| | Levene's Test for Equality of Variances | | t-test for Equality of Means | | Sig. (2-tailed) | Mean Difference | Std. Error Difference | 95% Confidence Interval of the Difference | |
|---------------|---|------|------------------------------|--------|-----------------|-----------------|-----------------------|---|---------|
| | F | Sig. | t | Df | | | | Lower | Upper |
| Total Request | 11.684 | .001 | 1.106 | 41.608 | .275 | -.90833 | .82160 | -.75019 | 2.56686 |

Furthermore, the chi-square test was run in order to investigate the differences between Group 1 and Group 2 members' preferred criteria. As presented in Table 9 and Table 10, the results of the chi-square ($\chi^2(1) = 1.599$, $p > .05$) indicate that there was no significant difference in NNESRs' rating criteria based on their teaching experience.

Table 9
Chi-square Test of NNESRs' Experience and Their Criteria

| | Value | Df | Asymp. Sig. (2-sided) | Exact Sig. (2-sided) | Exact Sig. (1-sided) | Point Probability |
|------------------------------------|--------------------|----|-----------------------|----------------------|----------------------|-------------------|
| Pearson Chi-Square | 1.599 ^a | 1 | .206 | .223 | .112 | |
| Continuity Correction ^b | 1.478 | 1 | .224 | | | |
| Likelihood Ratio | 1.598 | 1 | .206 | .223 | .112 | |
| Fisher's Exact Test | | | | .223 | .112 | |
| Linear-by-Linear Association | 1.599 ^c | 1 | .206 | .223 | .112 | .017 |
| N of Valid Cases | 3348 | | | | | |

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 239.03.

b. Computed only for a 2x2 table

c. The standardized statistic is 1.264.

Table 10
Symmetric Measures

| | | Value | Approx. Sig. | Exact Sig. |
|--------------------|-------------------------|-------|--------------|------------|
| Nominal by Nominal | Phi | .022 | .206 | .223 |
| | Cramer's V | .022 | .206 | .223 |
| | Contingency Coefficient | .022 | .206 | .223 |
| N of Valid Cases | | 3348 | | |

Discussion

The issue of EFL teachers' rating criteria and patterns in ILP assessment has remained understudied, despite its great impact on the process of teaching and testing of second language. The present study was conducted with the participation of multiple non-native raters in a foreign language context, in order to examine the overall criteria use and scoring patterns of Iranian NNESRs in assessing EFL learners' requests. The study also explored the effects of gender differences and teaching experiences related to raters in the assessment process.

The primary objective of this study was to explore the dominant criteria employed by NNESRs in request rating process. The criteria were of both general and speech-act-specific types. The general criteria such as "politeness", "conversers' relationship", "statement of optimal example", "style and register", as well as "language usage accuracy" can be applied in the assessment of other speech acts and were mentioned in previous studies to some extent. As an example, "politeness" criterion was greatly stressed by various raters in Taguchi's (2011) study of assessing request and apology speech acts, Alemi and Tajeddin's (2013) analysis of refusal speech act assessment, as well as Alemi et al.'s. (2015) study of non-native raters' assessment of compliment speech acts.

Moreover, the observed criteria included both pragmalinguistic and socio-pragmatic features. From this view point, “conversers’ relationship” is among socio-pragmatics factors, while “query preparatory and softeners” refers to pragmalinguistic aspect. The importance of both aspects in either teaching or testing pragmatic knowledge in language classes was repeatedly noted in previous researches (*i.e.* Eslami-Rasekh & Eslami-Rasekh, 2008; Roever, 2007).

Regarding the scoring of the EFL learners’ production, Iranian NNESRs acted divergently, as their minimum and maximum scores in most situations ranged from 1 to 5 based on the 5 point Likert scale. In line with this argument, even when the raters were categorized based on their genders and teaching experiences, this divergence in rating scores was present in each category. Besides, high degree of standard deviations of the rating scores, as well as a number of unanswered questions in the comment section of DCT prove the lack of knowledge and awareness of NNESRs regarding pragmatic assessment. In fact, this claim was supported in several studies and in some cases the necessity of pragmatic instruction was stressed (Eslami-Rasekh, 2005; Tajeddin & Alemi, 2013).

For the second and third research questions, the effects of factors such as raters’ genders and teaching experiences on scoring and criteria use were explored. Based on the achieved results, raters’ genders and teaching experiences did not have any effect on their rating scores and criteria. In a similar analysis regarding the speech act of compliment, Alemi et al. (2015) evidenced the significance of these two factors in raters’ criteria, but not in their rating scores. Insignificance of raters’ gender and teaching experience in their assessment can be another evidence for the importance of pragmatic instruction for NNESRs. Since experience would not lead to pragmatic proficiency and awareness in raters, it can be inferred that all NNESRs regardless of their background, should be trained in order to perform appropriately in pragmatic aspects of language.

Conclusion

The study revealed nine different criteria employed by NNESRs in rating request productions. The criteria were: “politeness”, “conversers’ relationship”, “style and register”, “language usage accuracy”, “statement of optimal examples”, “authenticity and cultural errors”, “query preparatory and softeners”, “explanation”, and “directness”. Some criteria including “politeness”, conversers’ relationship”, as well as “style and register” due to the situational requirements were mentioned more frequently than others, while other criteria such as “authenticity and cultural errors” and “linguistic usage accuracy” were rarely noted. Using criteria differently in each situation implicates the fact that specific variables need to be considered in each situation based on the present contextual factors.

The results also indicated that despite some similarities among NNESRs, certain degrees of variations and inconsistency were observed in raters’ assessment. This could be due to lack of pragmatic knowledge on the part of NNESRs which is not unexpected, as cultural discrepancies between L1 and L2 cause pragmatic misunderstandings. NNESRs in EFL context were not usually trained how to teach and assess L2 pragmatics. Consistent with the results of the current study, an organized and comprehensive pragmatic course or teacher training workshop for EFL teachers could be a possible solution to this problem.

The study also has some important implications for EFL learners and material developers. As a matter of fact, a proficient language learner must have a good level of pragmatic competence. Learners’ responses in this study showed that they need pragmatics instruction as a part of their language education while most of the textbooks for language learning lack sufficient L2 pragmatic exercises or do not consider cross-cultural differences between L1 and L2 societies (*see* Alemi &

Irandoost, 2012; Alemi, Roodi, & Bemani, 2013; Safa, Moradi, & Hamzavi, 2015); meanwhile, learning and teaching pragmatics appropriately in L2 classes is not possible without any textbooks and learning material. In fact, the importance of teaching L2 pragmatic and the need for pragmatically appropriate learning materials become vivid in countries like Iran, where teachers and learners do not have an easy access to native speakers or authentic learning materials and as a result, English pragmatic awareness is insufficient.

As the final point, it has to be mentioned that ILP rating is a new area in L2 teaching studies and it has many aspects which have remained overlooked. Further studies need to be done to find EFL teachers' rating criteria in assessing unstudied speech acts such as criticism, congratulation, etc. Moreover, the present investigation did not focus on assessing nonverbal aspects of EFL learners' outputs which are among important issues in producing a speech act and need to be studied. Features such as facial expression, tones, and body movements of the learners' can be explored in feature research.

References

- Adel, M., Davoudi, M., & Ramezanzadeh, A. (2016). A qualitative study of politeness strategies used by Iranian EFL learners in a class blog. *Iranian Journal of Language Teaching Research*, 4(1), 47-62. Retrieved from [http://www.urmia.ac.ir/sites/www.urmia.ac.ir/files/\(4\)_0.pdf](http://www.urmia.ac.ir/sites/www.urmia.ac.ir/files/(4)_0.pdf)
- Alemi, M. (2012). *Patterns in interlanguage pragmatic rating: Effects of rater training, intercultural Proficiency, and self-assessment*. Unpublished Ph.D. dissertation, Allameh Tabataba'i University, Tehran, Iran.
- Alemi, M., & Irandoost, R. (2012). A textbook evaluation of speech acts: The case of English result series. *International Journal of Applied Linguistics and English Literature*, 1(6), 199-209. doi:10.7575/ijalel/v.1n.6p.199
- Alemi, M., Eslami-Rasekh, Z., & Rezanejad, A. (2015). Iranian Non-native English Speaking Teachers' Rating Criteria Regarding the Speech Act of Compliment: An Investigation of Teachers' Variables. *Journal of Teaching Language Skills*, 6(3), 21-49. Retrieved from http://jtls.shirazu.ac.ir/pdf_2481_b731ff0a586bb61f319e6915de4e0b9a.html
- Alemi, M., Roodi, F., & Bemani, M. (2013). Pragmatic investigation in three global English textbooks. *Internet Journal of Language, Culture and Society*, 37, 55-60. Retrieved from https://www.researchgate.net/profile/Minoor_Alemi/publication/235330154_pragmatic_investigation_of_three_international_textbooks_in_terms_four_speech_acts/links/0046351ecbafd420ec000000.pdf
- Alemi, M., & Tajeddin, Z. (2013). Pragmatic rating of L2 refusal: Criteria of native and non-native English teachers. *TESL Canada Journal*, 30, 63-81. Retrieved from <http://teslcanadajournal.ca/index.php/tesl/article/download/1152/978>
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F. (2004). *Statistical analyses for language assessment*. Cambridge, UK: Cambridge University Press.

- Blum-Kulka, S. (1987). Indirectness and politeness in requests: Same or different? *Journal of pragmatics*, 11(2), 131-146. doi:10.1016/0378-2166(87)90192-5
- Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied linguistics*, 5(3), 196-213. doi:10.1093/applin/5.3.196
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language*. Cambridge: Cambridge University.
- Cohen, A. D. (2008). Teaching and assessing L2 pragmatics: What can we expect from learners?. *Language Teaching*, 41(02), 213-235.
- Cohen, A. D. (2014). Towards increased classroom assessment of pragmatic ability. *Iranian Journal of Language Testing*, 4(1), 4-25. Retrieved from <http://ijlt.ir/journal/images/PDF/420-2014-4-1.pdf>
- Eslami-Rasekh, Z. (2005). Raising the pragmatic awareness of language learners. *ELT Journal*, 59(2), 199-208. doi:10.1093/elt/cci039
- Eslami, Z. R., & Eslami-Rasekh, A. (2008). Enhancing the pragmatic competence of non-native English-speaking teacher candidates (NNESTCs) in an EFL context. *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing*, 30(2), 178-197.
- Hassall, T. (2004). *Requests by Australian learners of Indonesian*. *Journal of Pragmatics*, 35, 1903-1928. doi:10.1016/S0378-2166(03)00115-2
- Hudson, T., Brown, J. D., & Detmer, E. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Vol. 7). Natl Foreign Lg Resource Ctr.
- Hudson, T., Detmer, E., & Brown, J. D. (1992). *A framework for testing cross-cultural pragmatics* (Vol. 2). Natl Foreign Lg Resource Ctr.
- Izaki, Y. (2000). Cultural differences of preference and deviations from expectations in requesting: A study of Japanese and French learners of Japanese in contact situations. *Journal of Japanese Language Teaching*, 104, 79-88.
- Jalilifar, A., Hashemian, M., & Tabatabaee, M. (2011). A cross-sectional study of Iranian EFL learners' request strategies. *Journal of Language Teaching and Research*, 2(4), 790-803. doi:10.4304/jltr.2.4.790-803
- Knoch, U., Read, J., & von Randow, J. (2007). Re-training writing raters online: How does it compare with face-to-face training? *Assessing Writing*, 12(1), 26-43. doi:10.1016/j.asw.2007.04.001
- Leech, G. (1983). *The Principles of pragmatics*. London: Longman.
- Li, D. (2000). The pragmatics of making requests in the L2 workplace: A case study of language socialization. *Canadian Modern Language Review*, 57(1), 58-87. doi:10.3138/cmlr.57.1.58

- Roever, C. (2005). *Testing ESL pragmatics*. Frankfurt, Germany: Peter Lang.
- Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229-256. doi:10.1191/0265532206lt329oa
- Roever, C. (2007). DIF in the Assessment of Second Language Pragmatics. *Language Assessment Quarterly*, 4(2), 165-189. doi:10.1080/15434300701375733
- Safa, M. A., Moradi, M., & Hamzavia, R. (2015). Iranian EFL teachers and learners perspective on potentiality of Top Notch series for intercultural competence development. *Iranian Journal of Language Teaching Research*, 3(2), 47-66. Retrieved from [http://www.urmia.ac.ir/sites/www.urmia.ac.ir/files/\(4\).pdf](http://www.urmia.ac.ir/sites/www.urmia.ac.ir/files/(4).pdf)
- Searle, J. R. (1976). A classification of illocutionary acts. *Language in society*, 5(01), 1-23.
- Sydorenko, T., Maynard, C., & Guntly, E. (2015). Rater Behaviour When Judging Language Learners' Pragmatic Appropriateness in Extended Discourse. *TESL Canada Journal*, 32(1), 19-41. Retrieved from <http://teslcanadajournal.ca/index.php/tesl/article/view/1197/1017>
- Taguchi, N. (2011). Rater variation in the assessment of speech acts. *Pragmatics*, 21(3), 453- 471. doi:10.1075/frag.21.3.08tag
- Tajeddin, Z. (2014). *Interlanguage pragmatics and language teaching*. Tehran, Iran: Jungle Publications.
- Tajeddin, Z., & Alemi, M. (2013). Criteria and bias in native English teachers' assessment of L2 pragmatic appropriacy: Content and FACETS analyses. *The Asia Pacific Education Researcher*, 23(3), 425-434. doi:10.1007/s40299-013-0118-5
- Takahashi, S. (1996). Pragmatic transferability. *Studies in second language acquisition*, 18(02), 189-223. doi:10.1017/S0272263100014881
- Takahashi, S., & DuFon, P. (1989). *Cross-linguistic influence in indirectness: The case of English directives performed by native Japanese speakers*. Unpublished manuscript, University of Hawai'i at Manoa, Honolulu. Retrieved from <http://files.eric.ed.gov/fulltext/ED370439.pdf>
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2), 91-112. doi:10.1093/applin/4.2.91
- Trosborg, A. (1995). *Interlanguage pragmatics: Requests, complaints, and apologies*. Berlin: Mouton de Gruyter. doi:10.1515/9783110885286
- Walters, F. S. (2004). *An application of conversation analysis to the development of a test of second language pragmatic competence*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.
- Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155-183. doi:10.1177/0265532207076362

- Walters, F. S. (2009). A conversation analysis-informed test of L2 aural pragmatic comprehensions. *TESOL Quarterly*, 43(1), 29–54. doi:10.1002/j.1545-7249.2009.tb00226.x
- Woodfield, H. (2008). Interlanguage requests in English: A contrastive study. In M. Pütz & J. N. V. Aertselaer (Eds.), *Contrastive pragmatics: Interlanguage and Cross-cultural Perspectives* (pp. 231-264). Berlin/New York: Mouton De Gruyter.

Mino Alemi is Assistant Professor of Applied Linguistics at Islamic Azad University, Tehran-west Branch. She is the founder of Robot-Assisted Language Learning (RALL) in Iran which she achieved as a post-doctoral researcher in Sharif University of Technology. Her areas of interest include discourse analysis, interlanguage pragmatics, and materials development.

Neda Khanlarzadeh holds an MA in TEFL from Sharif University of Technology, Tehran, Iran. She has been working on interlanguage pragmatic rating for her M.A. thesis. Her current areas of research includes cross-cultural and interlanguage pragmatic studies.