

Native and Non-native English Teachers' Rating Criteria and Variation in the Assessment of L2 Pragmatic Production: The Speech Act of Compliment

Minoo Alemi

*Assistant Professor, Department of TEFL, Tehran-west Branch,
Islamic Azad University, Iran*

Atefeh Rezanejad

Ph.D. Candidate of TEFL, Allameh Tabataba'i University, Iran

Received: October 4, 2013; **Accepted:** April 23, 2014

Abstract

Pragmatic assessment and consistency in rating are among the subject matters which are still in need of more profound investigations. The importance of the issue is highlighted when remembering that inconsistency in ratings would surely damage the test fairness issue in assessment and lead to much diversity in ratings. Our principal concern in this study was observing the criteria that American and Iranian EFL/ESL teachers consider when rating Iranian EFL learners' pragmatic productions regarding the speech act of compliment. The instrument utilized in this study was WDCTs and a speech act rating questionnaire administered to sixty American and sixty Iranian EFL/ESL teachers. In order to come up with the criteria, the reasoning and justifications of the raters when rating learners' pragmatic productions were analyzed carefully through content analysis. The results showed that overall the raters considered nine general criteria when rating. They included: "Strategy use", "Affective factors", "Politeness", "Interlocutors' relationships", "Linguistic accuracy", "Sincerity", "Authenticity", "Fluency", and "Cultural issues". Also, the most frequent criterion among the native and non-native raters was "Strategy use" and "Politeness" respectively. Finally, it was concluded that due to some inconsistencies and variations in the ratings and criteria of both native and non-native raters, it seems that both groups are in need of pragmatic workshops and training sessions. The results of this study can have important implications for EFL/ESL teacher educators who are considerate of the importance of pragmatic training and instruction.

Keywords: pragmatic assessment; rating; speech act of compliment; American and Iranian EFL/ESL teachers

Authors' emails: minooalemi2000@yahoo.com; rezanejad_a85@yahoo.com

INTRODUCTION

The peculiar status of interlanguage pragmatics (ILP), considered as the study of the development and use of strategies for linguistic action by non-native speakers, has gained more attention in second language research. Blum-Kulka and Kasper (1993, p. 3) define interlanguage pragmatics as “a non-native speaker’s use and acquisition of linguistic action patterns in a second language”. Interlanguage pragmatics, or ILP, is known as the study of second language pragmatics. According to Kasper (1998, p. 184), ILP is “the study of nonnative speakers’ comprehension, production, and acquisition of linguistic action in L2, or put briefly, ILP investigates how to do things with words in a second language”.

Although the concept of interlanguage pragmatics has been investigated from different perspectives, it should be pointed out that the issue of pragmatic rating is still a new topic in interlanguage pragmatic studies and deserves much more attention. In view of this need in the literature, this study aimed at investigating what criteria underpin native and non-native EFL/ESL teachers’ rating criteria with a focus on the speech act of compliment.

REVIEW OF LITERATURE

Assessment of second language pragmatics is “a relatively recent part of L2 testing, and not many tests exist” (Roever, 2007, p. 165). Oller (1979) was the first who introduced the concept of pragmatic assessment. According to Oller, a pragmatic proficiency test is

any procedure or task that causes the learners to process sequences of elements in a language that conform to the normal contextual constraints of that language, and which requires the learners to relate sequences of linguistic element via pragmatic mapping to extralinguistic context (p. 38).

Still a more recent topic in assessment remains to be the issue of rating and raters’ criteria. Some researchers have studied rating criteria in the assessment of language tasks (e.g. Ang-Aw & Goh, 2011; Du, Wright, & Brown, 1996; Eckes, 2008; Galloway, 1980; Johnson & Lim, 2009; Lim, 2011; Orr, 2002; Plough, Briggs, & Van Bonn, 2010; Roch, 2007; Schaefer, 2008; Wigglesworth, 1994; Zhang & Elder, 2011). However, most of these studies were focusing on the

assessment of speaking or writing. For example, in a study by Son (2010), rater bias in elicited imitation ratings was examined. The main focus in this study was put on the raters' language background. The results showed that no bias could be observed in relation to raters' first or second language background.

A comparative study on ratings of speaking and oral performance by native and non-native teachers was done by Zhang and Elder (2011). 20 non-native and 19 native English speaking teacher raters' comments were analyzed in this study. The results indicated that there was no significant difference among the raters regarding their overall judgment. Nevertheless, the raters emphasized various features of oral proficiency in their ratings.

Concerning the ratings of writing, Eckes (2005) did a study on raters' consistency in rating speaking and writing performance. The participants showed more consistency in their ratings than their rating criteria. Furthermore, a longitudinal study on novice and professional raters in rating writing was done by Lim (2011). It was focused on observing how novice teachers' rating quality developed over time. The findings showed that consistency in ratings among novice and experienced teachers was not always different.

Some other studies (e.g. Brown, 2003, 2005; Ducasse, 2009; Johnson & Lim, 2009; May, 2007, 2009) have as well been done to investigate raters' perspectives and points of view. However, not much has been done concerning pragmatic rating and rater variation in rating pragmatic productions. The previous studies mentioned above have mainly used introspective verbal protocols to examine how raters' characteristics—e.g. gender, language background, and experience— influenced their evaluation of L2 oral or written productions. The main finding of these studies was that a great amount of variation and discrepancies could be observed among the raters because each of them came from a different background.

Walter (2007) was one of the first researchers who conducted a study on pragmatic rating. 42 learners of English took part in his study, accompanied by a native English-speaking rater for a 10 to 15-minute activity. The results of the study showed that the raters had differing viewpoints toward the productions and rated them differently based on the specific type of interpretation that they had. That is to say the raters, native or non-native, were much influenced by the things that were important to them.

To be more specific, Taguchi (2011) did a study on rater variation in the assessment of speech acts of request and opinion. The study used introspective verbal protocol to investigate four native raters' rating criteria, reasoning, and overall evaluation of appropriateness of learners' productions. 64 speech act productions in total were randomly collected from eight students. In the next stage, the raters were asked to rate each speech act and then explain their reasons for their rating. The results of the study showed that the native raters had some dissimilarity in their ratings. That's why at the end the researcher concluded that the native raters and their specific criteria cannot always be used as an unquestionable benchmark.

Moreover, Tajeddin and Alemi (2012) explored whether a training program focused on pragmatic rating would have a significant effect on the accuracy of non-native English speaking (NNES) raters' ratings of refusal production as measured against native English speaker (NES) ratings and whether NNES rating difference reduces after training. They concluded that pragmatic rater training can positively influence non-native ratings by getting them closer to those of natives and making them more consistent across raters.

Lee (2012) investigated rating behavior between Korean and native English-speaking raters (NES). The results revealed Korean raters' sense of inferiority in measuring linguistic components. They were more severe in scoring grammar, sentence structure, and organization, whereas the NES raters were stricter toward content and overall scores.

Moreover, Alemi (2012) explored patterns and variations in native and non-native interlanguage pragmatic rating of refusal and apology speech acts. To find out patterns and variations in ratings of native and non-native raters, the content of the raters' justifications and reasoning were analyzed carefully. The analysis of the raters' comments revealed five apology criteria (expression of apology, explanation/reasoning, repair offer, promise for future, politeness) and eleven refusal criteria (brief apology, statement of refusal, offer suitable consolation, irrelevancy of refusal, explanation/reasoning, cultural problem, dishonesty, thanking, postponing to other time, statement of alternative, politeness). The results of this study also indicated that the non-native raters were more linear in rating than native raters.

In a most recent study, Tajeddin and Alemi (2013) focused on native English raters' criteria for the assessment of EFL learners' pragmatic competence. They first tried to find the criteria for rating the speech act of apology in L2 by native English teachers. Then they investigated whether there was rater bias in native English teachers' rating of apology. To this end, 51 educated native English teachers, from the U.S., the U.K and from Australia, New Zealand, and Canada, rated six different pragmatic situations for an apology discourse completion task (DCT) which were accompanied by an L2 learner's response to each situation. The raters were also asked to describe the way they rated the response to each DCT situation. The content analysis of raters' justifications revealed five criteria they mostly applied in their rating: expression of apology, situation explanation, repair offer, promise for future, and politeness. They also used FACETS to locate the rater bias. Results indicated that raters showed different ratings and were not much consistent in their ratings. They finally concluded that native criteria cannot always be regarded as a benchmark.

But still the gap in the literature on pragmatics and more specifically interlanguage pragmatic rating is an investigation of rating criteria concerning the speech act of compliment. To the best of the researchers' knowledge, although some different studies have been done in the domain of pragmatic rating (Tajeddin & Alemi, 2013; Alemi, Eslami, & Rezanejad, 2014a; Alemi, Eslami, & Rezanejad, 2014b), no study could be found in the literature which addressed the criteria that native and non-native raters considered during rating compliment productions of EFL learners.

PURPOSE OF THE STUDY

The main purpose of this study was to investigate variations in interlanguage pragmatic assessment among native and non-native English teachers. Since pragmatic knowledge is an indispensable part of language proficiency as defined by Bachman (1990), finding the patterns native and non-native raters use in their assessment of learners' pragmatic performance, the main concern of this study, is very important.

In this study we asked Iranian and American EFL/ESL teachers of English (NES and NNES raters) to rate Iranian EFL learners' pragmatic productions regarding the speech act of compliment. To be more precise, we were concerned with finding patterns and variations

in the ratings of native and non-native raters in relation to the speech act of *compliment*. To this end, the following research questions were raised:

1. What criteria are used by native and non-native English speaking raters in rating the speech act of compliment produced by EFL learners?
2. Is there any significant difference between native and non-native English speaking raters in rating the speech act of compliment produced by EFL learners?

METHOD

Participants

The participants in this study included two main groups. The first group of participants was native American English speaking EFL/ESL teachers. The group comprised of sixty educated American native English speaking teachers (NES raters), 16 males and 44 females, with field of studies related to education and teaching English. The second group constituted sixty non-native English speaking Iranian teachers (NNES raters), 34 females and 26 males, who were selected from different language centers in Iran. They all had MA in TEFL (for the sake of homogeneity) and have taught English in different levels. Both the native and non-native groups were asked to rate the Iranian learner's pragmatic productions based on the Likert scale ranging from 1 (highly inappropriate) to 5 (most appropriate) and also comment on their appropriateness.

Instrumentation

In order to collect the data for this research study, the researchers developed a speech act rating questionnaire (see the appendix). The heart of this questionnaire comprised of Iranian EFL learners' responses to seven compliment WDCT (Written Discourse Completion Test) situations. That is to say, the questionnaire was prepared based on the learners' answers to those WDCTs. The learners were supposed to read seven different situations in which one would compliment someone and write exactly what they would say in that situation. It needs to be mentioned here that in the selection of the situations, the three variables of relative power, social distance, and

degree of imposition introduced by Brown & Levinson (1987) played an important role.

After collecting all the answers from the EFL learners, one answer for each situation was selected and rechecked by 2 pragmatic specialists to be added to the speech act rating questionnaire. It was tried to select a variety of different answers in order to ensure that the answers are varied enough in their degree of pragmatic appropriateness. In the next stage, the raters were asked to first rate the EFL learners' responses based on a 5-point Likert scale ranging from 1 to 5: (1) highly inappropriate, (2) inappropriate, (3) somewhat appropriate, (4) appropriate, and (5) most appropriate. Once they finished with rating the responses, they provided their criteria for their pragmatic rating. In fact, the raters were not provided with any option in the criteria section to choose from, but were asked to produce their own set of criteria. Below in Table 1 a sample of the questionnaire item can be found.

Table 1: Speech act rating questionnaire item sample

Compliment Situation: You have just read a new, outstanding article on an interesting topic by one of your professors. When you meet your professor how would you compliment him on that article?

EFL learner Answer: I just read your new interesting article and I'm so curious to know more about it.

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

.....

Data Collection Procedure

The data collection stage in this study comprised of two levels. In the first phase thirty Iranian upper-intermediate EFL learners completed seven WDCTs on the speech act of compliment. After some sessions of discussions among the researchers, one answer for each situation was selected to be rated by the NES and NNES raters. In order to collect data both from the native and non-native teachers, the researchers used both printed form of the questionnaire and also the electronic administration. However, in the case of native speakers, most of the data was collected through email.

Data Analysis

This study focused on the criteria that native and non-native EFL/ESL teachers consider while rating Iranian pragmatic productions regarding the speech act of compliment. The main procedure utilized was content analysis. Moreover, the study also drew on descriptive statistics and inferential statistics to analyze the data.

RESULTS

NES-NNES Raters' Compliment Criteria

Our first research question in this study was:

(1) What criteria are used by native and non-native English speaking raters in rating the speech act of compliment produced by EFL learners?

In order to answer our first research question, that is investigating the criteria important to the raters, the reasoning and justifications of the teachers for the appropriateness of the EFL learners' compliment productions were analyzed carefully. To do so, the qualitative procedure of content analysis was employed in order to discover the major themes in the data. The researchers perused the comments produced by the raters regarding each WDCT response. After eliciting more than nearly a hundred micro criteria, the next step was labeling the micro criteria with the intention of coming up with some macro criteria. After several observations, readings, and discussions among the researchers and also consulting with a panel of expert who were knowledgeable figures in the field of pragmatics, nine macro criteria on the whole could be extracted from the native and non-native raters' comments. From among these nine general criteria, seven of them were analogous among the two groups of raters. However, two criteria were specifically elicited from the native raters. The nine macro criteria are presented below (those present among native speakers are accompanied with a NES mark next to them and those present among non-native raters, have a NNES sign).

- (1) Politeness (NES & NNES)
- (2) Interlocutors' characteristics and relationships (NES & NNES)
- (3) Strategy use (NES & NNES)
- (4) Authenticity (NES & NNES)
- (5) Sincerity (NES & NNES)

- (6) Fluency (NES & NNES)
- (7) Linguistic accuracy (NES & NNES)
- (8) Cultural issues (NES)
- (9) Affective considerations (NES)

(1) Politeness: Being polite and using polite sentences when complimenting someone was the first compliment criterion. Both the native and non-native participants of this study claimed that politeness is an important factor to consider. For example, some responses from the EFL learners were rated very low because the raters thought that they were not polite enough. Two examples of this criterion derived from NES and NNES rating comments are given below.

NES comment: It's always appropriate to say you like someone's clothing, but asking where she bought the dress could be made more polite by using phrases like "Would you mind telling me....." or "Do you mind if I ask....."

NNES comment: It was a very good compliment. It is polite, generous and sincere.

(2) Interlocutors' characteristics and relationships: The second criterion mentioned by the raters was a care for the characteristics of the interlocutors when complimenting someone. The teachers considered the age, gender, social status, and level of formality of the interlocutors very important factors to consider. The example comments are presented below:

NES comment: I think this answer is ok because it's a friend, but if the individual answering is a male and the friend is a female it can be a little uncomfortable.

NNES comment: This quite informal answer was appropriate, provided the two friends were close enough. If they weren't that intimate, the very last sentence shouldn't have been expressed.

Strategy use: This strives for a tactful utterance regarding the steps and moves of a compliment, having a sense of creativity when complimenting, considering relevance issue when complimenting, and a cautious use of explicit or implicit types of compliment. This

criterion was mostly realized in the form of some suggestions or guidelines for having a better type of compliment.

NES comment: *Put the compliment first – about the article. Then you can say something about the writer. The object of the compliment needs to come first. Also it's important to establish relevance among the sentences.*

NNES comment: *Appropriate, the guest showed his/her happiness of being there and also being with the friend, actually indirectly, I think.*

Authenticity: The fourth criterion important to the raters in this study was authenticity of EFL learners' pragmatic productions. This criterion pertains to such issues as the naturalness of utterances, what native speakers naturally say in the proposed situations, and the normalness issue. The following examples are samples of what the raters thought:

NES comment: *Americans wouldn't use "My lovely grandma" – they would just use Grandma as a name. But the compliment on the bag is fine.*

NNES comment: *Do native speakers really speak like this? I don't think so.*

Sincerity: The fifth criterion mentioned by the native and non-native raters was sincerity. The raters emphasized on the importance being candid and sincere when complimenting, without any sign of flattery and sycophancy.

NES comment: *I think the response reflects a genuine interest in the dress as the compliment seems to like it very much and wants to get a similar one.*

NNES comment: *If you are really interested in what he said, then this would be a great statement leading up to a future discussion.*

Fluency: The sixth criterion involves fluency of utterances which is related to issues such as order of sentences, length of speech, appropriate phrase use, being well-stated and appropriate use of idioms. Two examples for this criterion are presented below:

NES comment: *It's too descriptive for someone you don't know, too complicated, we don't talk like. The speaker could use some better idiomatic expressions than this one.*

NNES comment: *It is rather direct and informal and the amount of language used is short.*

Linguistic accuracy: The raters thought that one of points to be considered when complimenting is paying attention to the basic structural rules of English. That is to say, a perfect compliment must be appropriate with regard to issues such as grammar, lexicon, and structure. That way the utterance will be well-stated.

NES comment: *Several of the small words, the prepositions and articles are wrong. They should be added, changed, or deleted. The first sentence would still be a little strange, but would be better with the correct word choices.*

NNES comment: *I rated this as "1" because there are enough grammar and word choice mistakes to impair meaning somewhat.*

Cultural issues: The eighth criterion, exclusive to the native raters, was a care for the contextual and cultural issues when complimenting. It is concerned with issues such as considering the situation and context, being careful on what the interlocutor said next or before, paying attention to context of talk, and being watchful of cultural issues.

NES comment: *This compliment is fine to me, although I think that the way you speak to grandmothers and other elderly family members may differ from culture to culture. I'm relatively informal with my grandmother for example, but that may not be the case with everyone.*

Affective considerations: This ninth criterion was also mentioned only by the native raters. They thought that the interlocutors should care for feelings of the other person and avoid being negative while talking. They thought that neither producer of compliment nor responder to a compliment should belittle the other side. They should care for each other's feelings, be kind, and use

appropriate words and expressions. An example of comments from NES raters related to this criterion is presented below:

***NES comment:** I'm sure making an A on an English test was a huge accomplishment, and this doesn't really praise him, it sounds more like it would bring him down.*

After extracting the criteria by raters in rating the speech acts of compliment, in the second phase of data analysis, quantitative section, the frequency and percentage of those criteria were calculated. In the following section, a detailed analysis of the criteria related to this speech act and their frequency and percentage will be discussed. The specific criterion frequency for each of the seven situations is presented in Table 2.

As the table depicts, the majority of the observed criteria were *strategy use* (NES = 28.72%, NNES = 16.66%), affective considerations (NES = 13.12%, NNES = 0%), *politeness* (NES = 12.63%, NNES = 22.45%), interlocutors' characteristics and relationships (NES = 11.14%, NNES = 19.05%), *linguistic accuracy* (NES = 11.14%, NNES = 9.53%), and *sincerity* (NES = 8.17%, NNES = 8.85). In addition, the least frequent criteria mentioned by the raters were *authenticity* (NES = 7.68%, NNES = 13.27%), *fluency* (NES = 4.21%, NNES = 9.87%), *cultural issues* (NES = 2.73%, NNES = 0%).

Table 2: Frequency of compliment criteria in different situations among NES and NNES raters

Situation		LA	Aut	Sinc	Pol	Flu	IR	CI	AC	SU	Total
Situation 1	native	F 9	4	9	13	2	1	2	8	14	64
	P	14.06%	6.25%	14.06	20.31%	3.12	1.56%	3.12%	12.5%	21.87%	
non-native	F	5	10	6	15	1	14	0	0	14	65
	P	7.69%	15.38	9.23	23.07%	1.53	21.53%	0	0	21.53%	
Situation 2	native	F 4	9	2	34	2	11	1	6	9	78
	P	5.12%	11.53	2.56	43.58%	2.56	14.10%	1.28%	7.69%	11.53%	
non-native	F	2	6	2	11	1	10	0	0	6	38
	P	5.26%	15.78	5.26	28.94%	2.63	26.31	0	0	15.78%	
Situation 3	native	F 4	2	2	3	3	3	4.65%	11.62	19	43
	P	9.30%	4.65%	4.65	6.97%	6.97	6.97%	4.65%	11.62	44.18%	
non-native	F	4	8	2	7	2	5	0	0	4	32
	P	10%	20%	5%	17.5%	5%	12.5%	0	0	10%	
Situation 4	native	F 10	5	5	0	7	12	2	10	11	62
	P	16.12%	8.06%	8.06	0	11.29	19.35	3.22%	16.12	17.74%	
non-native	F	9	6	3	8	11	6	0	0	3	47
	P	18.75%	12.5%	6.25	16.66%	22.91	12.5%	0	0	6.25%	
Situation 5	native	F 15	7	3	0	0	7	2	3	16	53
	P	28.30%	13.20	5.66	0	0	13.20%	3.77%	5.66%	30.18%	
non-native	F	1	4	9	15	5	1	0	0	4	48
	P	2.08%	8.33%	18.75	31.25%	10.41	2.08%	0	0	8.33%	
Situation 6	native	F 2	1	3	1	0	0	0	9	35	51
	P	3.92%	1.96%	5.88	1.96%	0	0	0	17.64	68.62%	
non-native	F	3	3	3	4	6	3	0	0	10	32
	P	9.09%	9.09%	9.09	12.12%	18.18	9.09%	0	0	30.30%	
Situation 7	native	F 1	3	9	0	3	11	2	12	12	53
	P	1.88%	5.66%	16.98	0	5.66	20.74%	3.77%	22.64	22.64%	
non-native	F	4	2	1	6	3	8	0	0	8	32
	P	10.81%	5.40%	2.70	16.21%	8.10	21.62%	0	0	21.62%	
Total native		45	31	33	51	17	45	11	53	116	404
Total non-native		28	39	26	66	29	56	0	0	49	294
Percentage native		11.14	7.68%	8.17%	12.63%	4.21	11.14%	2.73%	13.12	<u>28.72%</u>	
Percentage non-native		9.53%	13.27	8.85%	<u>22.45%</u>	9.87	19.05%	0	0	16.66%	

NES-NNES Compliment Ratings

Our second research question in this study was:

- (2) Is there any significant difference between native and non-native English speaking raters in rating the speech act of compliment produced by EFL learners?

Research question two was raised to investigate the difference between NES and NNES teachers in rating the speech act of compliment produced by EFL learners. To address the research question, descriptive statistics, chi square and t-test were calculated. The descriptive statistics for the total seven situations for native and non-native raters are presented in Table 3. The score given on each situation ranged from 1 (highly inappropriate) to 5 (most appropriate). As Table 3 shows, the mean (M) rating of the 60 native raters for total DCTs was 3.35. It shows that their overall evaluation of compliments

in the seven situations fell within the “somehow appropriate” point on the scale. Although standard deviation (SD) for the total situations was fairly low, the distance between minimum score (1) and maximum score (5) provides a rough account of divergence or dispersion in rating of each specific compliment situations.

Table 3: Descriptive statistics of ratings by NES and NNES raters for compliment

	Group	N	Min.	Max.	Mean	Std. Deviation
Situation1	NN	60	2	5	3.63	.84
	N	60	2	5	4.02	.77
Situation2	NN	60	1	5	3.38	1.13
	N	60	1	5	3.02	1.03
Situation3	NN	60	1	5	3.57	1.01
	N	60	1	5	4.03	.92
Situation4	NN	60	1	5	2.33	1.08
	N	60	1	5	3.10	1.21
Situation5	NN	60	1	5	3.67	.93
	N	60	1	5	3.55	.96
Situation6	NN	60	1	5	2.77	1.33
	N	60	1	4	1.78	.82
Situation7	NN	60	1	5	3.30	1.06
	N	60	1	5	4.00	1.07
Total	NN	60	2.29	4.00	3.23	.40
	N	60	2.43	4.29	3.35	.45

* N=Native, NN= Non-native

The table shows that the mean rating among the native and non-native raters was 3.35 and 3.23 respectively. In addition, the highest rating among non-native raters was assigned to Situation 5, with $M = 3.67$. Among the native raters, the highest rating was assigned to situation 3 with $M = 4$. Moreover, the lowest SD related to situation 1 for both native (.77) and non-native raters (.84). In addition, the lowest rating was applied to Situation 4 which was 2.33 among the non-native raters and to situation 6 with mean rating of 1.78 among native raters. The results in this section show that in almost all situations, some degree of disagreement among the raters could be observed.

Besides, an analysis of chi-square was run to probe any significant relationship between natives' and non-natives' criteria regarding the speech act of compliment. The results of chi-square ($\chi^2(1) = 41.567, P < .05$) indicated that the difference observed in Table 4 was statistically significant. Thus the null-hypothesis as there is not any significant difference between natives' and non-natives' criteria regarding the speech act of compliment was rejected.

Table 4: Chi-square results for the speech act of compliment by nationality

Chi-Square	41.567 ^a
Df	1
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 296.5.

Moreover, in order to find out whether there is any significant difference between the ratings of the native and non-native raters, an independent samples t-test was run between the ratings of the two groups for the speech act of compliment. The results are presented below in Table 5.

Table 5: Independent samples t-test for compliment ratings between native and non-native raters

	T	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Total compliment	-1.5	118	.12	-.12	.079

Table 5 depicts the results of the independent t-test. The results in this section ($t(118)=-1.53, p=.12$) illustrate that there was not any significant difference between the native and non-native raters in the rating of the EFL learners' pragmatic productions regarding the speech act of compliment.

DISCUSSION

The main goal of this research study was exploring the criteria that native and non-native English speaking raters consider during rating EFL learners' pragmatic productions regarding the speech act of compliment. The results indicated that non-native raters overall considered seven macro criteria during their ratings. They included: "linguistic accuracy", "fluency", "politeness", "authenticity", "sincerity", "strategy use", and "interlocutors' characteristics and relationships". The results of the study further showed that native raters were to some extent more watchful of some more details and listed nine macro criteria. The two additional criteria mentioned by the native raters included: "cultural issues", "affective considerations".

The results of this study concerning the criteria favored by the raters were in line with some other few research studies done. For example, Taguchi (2011) also found similar results among native raters concerning the two speech acts of request and opinion. In his study, the raters also emphasized the importance of issues such as politeness, content, directness, strategies used, and clarity.

The findings were also compatible with Alemi (2012) and the criteria she found among the raters for the speech act of refusal. Exact matches include the criteria of politeness, relevancy, a care for cultural matters, and honesty. Moreover, the criteria were also comparable with Tajeddin and Alemi (2013) who also discovered similar criteria regarding the speech act of apology among native raters.

This indicates whereas some rating criteria seem to be common among different speech acts, there are also some criteria which are specific to some certain speech acts. It seems that some criteria such as the criterion "politeness" are universal among the raters. That is to say, they can be applied in rating of every speech act. On the other hand, the lack of consistency of different studies on pragmatic rating on different speech acts illustrates that not all speech acts attract the same rating criteria by the raters. This further reinforces the need for doing separate studies on different speech acts in order to discover the criteria that the raters may adhere to when rating pragmatic productions.

Moreover, it was observed that there was a significant difference in the frequency of compliment rating criteria between the native and non-native raters. This low correlation may have been originated from not having a one to one correlation between the cultures of the two groups. The sociopragmatic norms of the two groups were not in fact much similar in complimenting. It seems that the native group was much more lenient in complimenting than the non-native group. This lack of congruency in the two groups' sociocultural norms may have influenced the raters' ratings. That is to say, nationality had a significant effect on the frequency of the criteria selected by the raters. This again may have originated from the specific viewpoints attributed to the two groups of native and non-native raters. They were from different backgrounds and had different ideas regarding the appropriateness of the responses. Compliment is one of the speech acts which are exposed to a lot of discrepancy and divergence among people of different cultures and backgrounds.

Along with all consistencies or inconsistencies among the native and non-native raters, the researchers are not about to conclude that in all cases, we need to use the benefit of this consistency in native ratings and forget all about the non-native ratings. But this is for sure to say that both native and non-native raters are in need of some training programs and instructions in order to be more competent and skilled in ratings (Taguchi, 2011). In this research study we examined what non-native and native raters considered while rating compliment productions. But it is not to be concluded that what native speakers considered normal, can always be regarded as the main pattern and model. Both groups seem to be in need of pragmatic instructions in order to become more aware and attentive of major and minor points to be considered during ratings. Still, this need seems to be more crucial for non-native raters and they need more attention regarding the instruction issue. This instruction constitutes such matters as the language itself and grammatical issues and also a concern for teaching of pragmatic issues.

Research has shown that non-native teachers do not feel much confident about their language proficiency and seem to have a weaker pragmatic competence (Eslami-Rasekh, 2005; Pasternak & Bailey, 2004). Therefore, it can be argued that non-native teachers are not only in need of more training and education in general language ability, but also need more instruction for developing their pragmatic competence. And of course, Iranian non-native teachers are not an exception.

This need for raising pragmatic awareness of teachers seems to be even more emphasized in EFL contexts, for example in a country like Iran. Many factors may help one to raise his/her pragmatic awareness, among which being exposed to intercultural communications seems to be one of the most important ones. Just as EFL learners need to interact with native English speakers in order to develop their language ability, the same is true for teachers. Unfortunately in Iran the EFL teachers do not have much access to native English speakers and more specifically native ESL teachers, to share knowledge with. That is to say, Iranian language teachers do not receive much implicit instruction on pragmatics. The fact is that they mostly are in need of explicit education on pragmatics.

CONCLUSION AND IMPLICATIONS

The researchers' main objective in this study was exploring the criteria that native and non-native raters considered during rating EFL learners' pragmatic productions regarding the speech act of compliment. The results indicated that overall the two groups considered nine macro criteria. They included: "linguistic accuracy", "authenticity", "sincerity", "politeness", "fluency", "interlocutors' characteristics and relationships", "cultural issues", "affective considerations", and "strategy use".

As mentioned before in the results section, some inconsistencies could be observed among the raters in their ratings and criteria. This inconsistency can further lead to rater inconsistency in the assessment of pragmatic productions by learners. The issue seems to be more critical in EFL contexts and also developing countries where the topic of pragmatics is still new and needs more attention regarding its education and assessment.

Although many different researchers emphasized the role and importance of development of language proficiency for non-native language teachers (Mahboob, 2004; Medgyes, 1994; Pasternak & Bailey, 2004; Samimy & Brutt-Griffler, 1999), not much attention is still paid to issues such as pragmatic awareness of non-native teachers. Based on Bachman's (1990) model of language competence, in order to communicate effectively, one needs an appropriate amount of knowledge on both grammatical issues and pragmatic competence.

The results of this study can have important implications for material designers and teacher trainers who are in charge of developing material for teacher education courses and instructing teachers on being a better teacher and cognizant of minor and major issues in pedagogy. This is further proved by referring to scholars such as Biesenback-Lucas (2003) and Rose (1997) who claimed that ESL teacher education programs do not focus much on pragmatic issues and are mostly negligent on these concerns. In fact, there are only a handful of sources which have dealt with the importance attached to the issue of pragmatic competence in teacher education programs (e.g. Bardovi-Harlig and Hartford, 1997; Eslami-Rasekh, 2005; Rose, 1997).

Bio-data

Minoo Alemi is Assistant Professor of Applied Linguistics at Islamic Azad University, Tehran-west Branch. She is the founder of Robot-Assisted Language Learning (RALL) and the co-founder of Social Robotics in Iran, which she achieved as a post-doctoral researcher at the Center of Excellence in Design, Robotics, and Automation (CEDRA) of Sharif University of Technology. Her areas of interest include discourse analysis, interlanguage pragmatics, and materials development. She has been a presenter in many (inter)national conferences and the author of many papers published in journals like *The Asia-Pacific Education Researcher* and *TESL Canada Journal*.

Atefeh Rezanejad received her Master's degree in TEFL from Sharif University of Technology in 2013 and currently she is pursuing her doctoral degree at Allameh Tabataba'i University, Tehran, Iran. She has been teaching English for over 10 years and her main areas of interest include intercultural communicative language teaching, ethics, academic writing, sociolinguistics, discourse analysis, and textbook evaluation.

References

- Alemi, M. (2012). *Patterns and variations in native and non-native interlanguage pragmatic rating: Effects of intercultural identity, self-assessment, and rater training*. Unpublished Ph.D. dissertation, Allameh Tabataba'i University, Tehran, Iran.
- Alemi, M., Eslami, R. Z., & Rezanejad, A. (2014a). Rating EFL learners' interlanguage pragmatic competence by non - native English speaking teachers. *Procedia - Social and Behavioral Sciences*, 98, 171-174.
- Alemi, M., Eslami, R. Z., & Rezanejad, A. (2014b). 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 (JTLS)*, 6(3), 21-49.
- Ang-Aw, H. T., & Goh, C. C. M. (2011). Understanding discrepancies in rater judgment on national-level oral examination tasks. *RELC Journal*, 42(1), 31-51.
- Bachman, F. L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bardovi-Harlig, K. & Hartford, B. S. (Eds) (1997). *Beyond methods: Components of second language teacher education*. New York: McGraw Hill.
- Biesenback-Lucas, S. (2003). Preparing students for the pragmatics of e-mail interaction in academia: A new/forgotten dimension in teacher education. *Teacher education interest section newsletter*, 18(2), 3-14.

- Blum-Kulka, S., & Kasper, G. (1993). Interlanguage pragmatics: An introduction. In G. Kasper and S. Blum-Kulka (Eds.), *Interlanguage pragmatics* (pp. 3-17). New York: Oxford University Press.
- Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing*, 20(1), 1-25.
- Brown, A. (2005). *Interviewer variability in oral proficiency interviews*. Frankfurt am Main: Peter Lang.
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.
- Du, Y., Wright, B. D., & Brown, W. L. (1996). Differential facet functioning detection in direct writing assessment. Paper presented at the Annual Meeting of the American Educational Research Association, New York.
- Ducasse, A. M. (2009). Assessing paired orals: Raters' orientation to interaction. *Language testing* 26(3), 423-443.
- Eckes, T. (2005). Examining rater effects in TestDaF writing and speaking performance assessments: A many-facet Rasch analysis. *Language Assessment Quarterly*, 2(3), 197-221.
- Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing*, 25(2), 155-185.
- Eslami, R., Z. (2005). Raising the pragmatic awareness of language learners. *ELT journal*, 59(3), 199-208.
- Galloway, V. B. (1980). Perceptions of the communicative efforts of American students of Spanish. *Modern Language Journal*, 64(4), 428-433.
- Johnson, J. S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, 26(4), 485-505.
- Kasper, G. (1989). Interactive procedures in interlanguage discourse. In W. Oleksy (Ed.), *Contrastive pragmatics* (pp. 189-229). Amsterdam: John Benjamins.
- Lee, C. (2012). 'Cute your hair' – 'Na'. An exploratory study of Hawaii Creole English compliments and their responses. University of Hawaii Working Papers in English as a Second Language 9.
- Lim, G. S. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing*, 28(4), 543-560.
- Mahboob, A. (2004). Native or non-native: What do students enrolled in an intensive English program think? In L. D. Kamhi-Stein (ed.), *Learning and Teaching from Experience: Perspectives on Non-native English-speaking Professionals* (pp. 121-149). Ann Arbor, MI: The University of Michigan Press.

- May, L. (2007). *Interaction in a paired speaking test: The rater's perspective*. Ph.D. dissertation: The University of Melbourne.
- May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3), 397-421.
- Medgyes, P. (1994). *The Non-Native Teacher*. London: MacMillan.
- Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Orr, M. (2002). The FCE Speaking test: using rater reports to help interpret test scores. *System*, 30(2), 143-154.
- Pasternak, M. & Bailey, K. M. (2004). Preparing non-native and native English-speaking teachers: Issues of professionalism and proficiency. In L.D. Kamhi-Stein (Ed.), *Learning and Teaching from Experience: Perspectives on non-native English-speaking Professionals* (pp. 155 - 176). Ann Arbor, MI: The University of Michigan Press.
- Plough, I. C., Briggs, S. L., & Van Bonn, S. (2010). A multi-method analysis of evaluation criteria used to assess the speaking proficiency of graduate student instructors. *Language Testing*, 27(2), 235-260.
- Roever, C. (2007). DIF in the Assessment of Second Language Pragmatics. *Language Assessment Quarterly*, 4(2), 165-189.
- Roch, S. G. (2007). Why convene rater teams: an investigation of the benefits of anticipated discussion, consensus, and rater motivation. *Organizational Behavior and Human Decision Processes*, 104(1), 14-29.
- Rose, K. R. (1997). Pragmatics in the classroom: Theoretical concerns and practical possibilities. *Pragmatics and language learning*, 8(2), 267-292.
- Samimy, K. & Brutt-Griffler, J. (1999). To be a native or non-native speaker: Perceptions of non-native speaking students in a graduate TESOL program. In G. Braine (Ed.), *Non-native Educators in English Language Teaching* (pp. 127-144). Mahwah, NJ: Erlbaum.
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.
- Son, M. (2010). *Examining rater bias: An evaluation of possible factors influencing elicited imitation ratings*. MA project, Brigham Young University.
- Taguchi, N. (2011). Rater variation in the assessment of speech acts. *Pragmatics* 21(3), 453-471.
- Tajeddin, Z., & Alemi, M. (2012). Pragmatic rater training: Does it affect rating accuracy and consistency? *Iranian Journal of Language Testing*, 4(1), 66-83.
- 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

- Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing*, 24(2), 155-183.
- Wigglesworth, G. (1994). Patterns of rater behavior in the assessment of an oral interaction test. *Australian Review of Applied Linguistics*, 17(2), 77-103.
- Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing*, 28(1), 31-50.

Appendix

Speech act rating questionnaire

Teacher's background:

- Name (optional):
- University Degree:
- Major:
- Gender: Male Female
- Years of English teaching experience: 1-6 7-11

Dear EFL/ESL Teacher: *In the following situations, an English language learner was supposed to compliment someone. Please read the learner's answer in each situation and rate its appropriateness according to the following rating scale. Then please kindly provide your criteria and reasons for the selection of a particular point (1, 2, 3, 4, or 5) on the scale.*

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

- You have just read a new, outstanding article on an interesting topic by one of your professors. When you meet your professor how would you compliment him on that article?

Answer: I just read your new interesting article and I'm so curious to know more about it.

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

- You go to your friend's house to help her study math. She is dressed in her best clothes because she has just got home from a party. How would you compliment her on her dress?

Answer: Hey, I like your dress! It's really nice. Can you take me where you bought it?

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

3. You are visiting Turkey for Norouz holiday (It's the New Year holiday in Iran), staying at a friend's house. What would you say to compliment the beautiful views?

Answer: *You are so lucky, because you can see these views every day.*

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

4. You're in a swimming pool. Suddenly you notice that a person swimming next to you is swimming professionally. How would you compliment him/her?

Answer: *When you swim, you have move through water like a fish with beautiful synchronization for your hands and legs. I wish I could swim as perfect and as professionally as you do.*

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

5. Your grandmother has just returned home from shopping. You notice that she has just bought a new handbag. How would you compliment your grandmother?

Answer: *Oh my lovely grandma, It's really beautiful.*

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

6. You are a member of the scientific board. You notice that Mr. Ahmadi, one of your students, is always dressed professionally. How would you compliment him?

Answer: *You are the best student that I've ever had.*

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

7. Your younger brother tells you that he got an “A” in his last English test. How would you compliment him?

Answer: *I know this test has taken a lot of hard work and effort, but I think you deserved the results and I am proud of you. You earned it and I am sure you will have a bright career ahead of you.*

1. Highly inappropriate 2. Inappropriate 3. Somewhat appropriate 4. Appropriate 5. Most appropriate

Criteria:

