

The Journal of Teaching Language Skills (JTLS)

6 (3), Fall 2014, Ser. 76/4

ISSN: 2008-8191. pp. 21-49

**Iranian Non-native English Speaking Teachers' Rating
Criteria Regarding the Speech Act of Compliment:
An Investigation of Teachers' Variables**

M. Alemi

Assistant Professor, Islamic Azad
University, Tehran-west Branch,
Humanities Faculty
email: alemi@sharif.ir

Z. Eslami Rasekh

Associate Professor, TEFL
Texas A & M University
email: zeslami@tamu.edu

A. Rezanejad *

Ph.D student in TEFL
Allameh Tabataba'i University, Tehran, Iran
email: rezanejad_a85@yahoo.com

Abstract

Among topics in the field of pragmatics, some seem to be in a more rigorous need of investigation. Pragmatic assessment and specifically the issue of pragmatic rating are among issues which deserve more thorough consideration. The purpose of this study was to examine rater criteria and its consistency and variability in the assessment of Iranian EFL learners' production of compliments based on the teachers' gender and teaching experience of the Iranian non-native English speaking raters (INNESRs). The data for this study were collected through WDCTs rating questionnaire from sixty Iranian EFL teachers and were later analyzed through descriptive statistics, t-tests and Chi-squares. The results of the study showed that Iranian EFL teachers consider seven macro criteria when rating EFL learners' pragmatic productions regarding the speech act of compliment. The criteria include "politeness" (26.37%), "interlocutors' characteristics and relationships" (22.83%), "variety and range" (19.68%), "socio-pragmatic appropriateness" (14.17%), "sincerity" (10.23%), "complexity" (9.84%), and "linguistic appropriacy" (8.66%). The results of the t-test and Chi-squares further showed that whereas there was no significant difference in the teachers' ratings based on their gender and teaching experience, the difference was significant in the frequency of rating criteria

Received: 2/14/2014 Accepted: 10/15/2014

*Corresponding author

provided by the raters. To conclude, the results of the study reinforce the need for rater training regarding the assessment of pragmatic productions based on pragma-linguistic and socio-pragmatic norms.

Keywords: inter-language pragmatics, rating, speech act, EFL teachers, compliment

1. Introduction

Most models of communicative competence (e.g. Bachman, 1990; Bachman & Palmer, 2010; Canale, 1983; Canale & Swain, 1980) consider pragmatics as one of the main components of language ability. Moreover, nowadays it is recognized that learning a second language is something much more than a mere learning of grammatical structures and vocabulary items. It is believed that in order to be a successful L2 learner, the learners also need to have a good command of functional and sociolinguistic functions of language. Therefore pragmatic competence, the ability to perform language functions in the relevant social context, is an integral component of language learning.

Many researchers (Blum-Kulka, 1982; Cohen & Olshtain, 1981; House, 1982; Kasper, 1981; Thomas, 1983; Wolfson, 1981) ascertain that language learners who have an acceptable knowledge of grammar and vocabulary may fail to communicate effectively. Widdowson (1978) believes that second language learners' pragmatic failure is due to cross-linguistic differences in realization of rules related to various speech acts. He states that there is an equal tendency on the side of second language (L2) learners to transfer rules of use (related to contextual appropriacy) and rules of usage (related to grammatical accuracy).

Pragmatic competence is said to be an integral part of the overall language competence. Second language learners would not gain much success in their second language learning career without noticing this aspect of communicative competence. As eliminating the pragmatic facet from the process of second language education would cause deficiencies in learning, the same difficulties may arise from omitting pragmatic dimension from the assessment aspect of language learning. Many Iranian language instructors have faced situations in which language students with high command of grammatical points and vocabulary could not make a simple compliment, request politely for something, or respond to the compliment of a professor in English.

That is to say, pragmatic assessment makes an important contribution to the realm of language education and more specifically pragmatic instruction. Different scholars (e.g. Beebe & Takahashi, 1989; Kasper, 1998; Thomas, 1983) have investigated the issue of pragmatic assessment from different perspectives and had differing views regarding the definition and usability of the notion of pragmatic assessment. For example, according to Thomas (1983), for L2 learners' productions to be pragmatically well-formed, two types of judgment are employed: pragmalinguistic assessment (concerned with linguistic issues) and sociopragmatic judgment (concerned with social distance, level of power, and degree of imposition). Nevertheless, a new area of studies related to pragmatic assessment is the issue of rating pragmatic productions. Not many research studies have investigated the concept up till now. In response to this lack of research, this paper is an attempt to investigate pragmatic assessment and more specifically pragmatic rating among Iranian non-native English speaking raters.

2. Literature Review

2.1 Background on interlanguage pragmatic assessment

The issues of research on pragmatics and intercultural pragmatics have attracted attention for many decades and gained momentum recently. However, the point is that the concept of pragmatic assessment and testing has a more recent history. Roever (2007, p. 165) also agrees that assessment of second language pragmatics is "a relatively recent part of L2 testing, and not many tests exist".

The concept of pragmatic assessment was first introduced by Oller (1979). It mainly focused on contextual relevance of L2 learners' language use. 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)

Researchers concerned with and interested in developing and validating pragmatic tests in L2 pragmatics have used six main types of tests introduced by James D. Brown (2001) as their instruments to measure and test learners' pragmatic proficiency. They are named as: (1) the Written

Discourse Completion Tasks (WDCT), (2) Multiple-Choice Discourse Completion Tasks (MDCT), (3) Oral Discourse Completion Tasks (ODCT), (4) Discourse Role Play Tasks (DRPT), (5) Discourse Self-Assessment Tasks (DSAT), and (6) Role-Play self-assessments (RPSA). In addition, tests of L2 pragmatics either focus on sociopragmatics or pragmalinguistics aspects of language assessment. It is argued (Roever, 2007, p. 166) that both “Sociopragmatics and pragmalinguistics are intertwined in real-world language use, and users need both to function in communication”.

In the domain of sociopragmatics and its testing, some major research studies can be mentioned. For example, Cohen and Olshtain (1981) made an effort to investigate a role play test to measure L2 learners’ performance regarding the sociocultural appropriateness of their speaking ability and deciding on whether a rating scale can effectively be used to assess learners’ sociocultural competence. Their study was mainly concerned with the speech act of apology, measuring students’ ability to act appropriately in the specific context by selecting appropriate forms of language. In their study they were able to categorize culturally and stylistically inappropriate L2 productions regarding the speech act of apology, but not successful in developing a rating scale to measure sociocultural competence.

As Roever (2007) claimed, the largest research study on testing sociopragmatics is done by Hudson, Brown, and Detmer’s (1995) validation study. They developed six types of tests: written discourse completion tasks, multiple choice discourse completion tasks, oral discourse completion tasks, self-assessments, role play discourse tasks and role play self-assessments. Their tests, whose participants were ESL students in the United States, were meant to be an attempt to systematically develop tests of pragmatic knowledge on the speech acts request, apology, and refusal. After gathering the scores on all these tests and reporting the descriptive statistics, reliability and validity of the measures, they concluded that only the MDCT type of measurement was not much successful.

Another study with similar results to Hudson, Brown, and Detmer’s (1995) was the one done by Yamashita (1996). In fact, the same instrument used by Hudson, Brown, and Detmer’s (1995) was adapted to Japanese students. She also tried to investigate the effectiveness of these pragmatic tests. The participants of her study were mainly Japanese students who studied English as a second language and were selected from four different universities in Japan. A specific feature of her study was that she didn’t use the English version of the tests, but translated them into Japanese. The result

of her study also showed that all the measures were appropriate for the purpose of measuring language students' pragmatic ability except the MDCT format one. Some years later, Yoshitake (1997), also applied some of the tests in the EFL context of Japan with Japanese-speaking learners of English. She used the DCTs, MDCTs, ODCTs and role plays with EFL university students in Tokyo in order to figure out their effectiveness.

In the domain of testing pragmalinguistic issues of language, an all-inclusive instrument to be mentioned is Roever's (2005, 2006) web-based test of ESL pragmatics. The advantage of his test battery is that it is not designed for any specific L1 group and tests learners' knowledge of three aspects of pragmalinguistics: implicature, routines, and speech acts. "The overall test construct assumes that these components of pragmatic knowledge are to some degree related because they are affected by similar developmental factors (most notably exposure and L2 proficiency), but they differ in the degree to which these developmental factors influence them" (Roever 2007, p. 167).

Still, the point is that, as Jianda (2006) claims, the realm of language testing does not adequately consider the issue of pragmatic testing and this topic still needs more investigations. He continues, we don't have enough number of tests to measure learners' pragmatic proficiency. Meanwhile, Bachman (1990) asserts that pragmatic knowledge can't be separated from the knowledge on language proficiency. Jianda (2006) states that the reason for not having enough pragmatic tests is that developing a measure of pragmatic competence in an EFL context is very difficult.

Moreover, a new area of research in the area of pragmatic assessment is related to the issue of rating and rater criteria. Recently some research studies have been concerned with emphasizing the importance of coming up with unified and comprehensive rating criteria in pragmatic assessment. For example, Alemi (2012) investigated the criteria that native English teachers and non-native Iranian English teachers consider when rating EFL learners' pragmatic productions regarding the speech act of apology and refusal. In her study, she has discovered that teacher raters used five macro criteria to rate the speech act of apology (*expression of apology, explanation/reasoning, politeness, repair offer, promise for future*) and eleven 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*), rate the speech act of refusal.

In the same line, Tajeddin and Alemi (2013) focused on the criteria that native English raters considered while rating EFL learners' pragmatic competence. They mainly focused on the speech act of apology. Besides discovering the criteria of the raters, the researchers also emphasized the existence of any bias among the raters. To fulfill this purpose, 51 educated native English teachers, from the U.S., the U.K, Australia, New Zealand, and Canada took part in their study. They rated six diverse pragmatic situations for an apology discourse completion task (DCT) which were accompanied by an L2 learner's response to each situation. The raters were asked to rate the appropriateness of the productions as well as producing their comments regarding the answers. The analysis of the raters' justifications revealed five macro criteria frequently applied in their rating. They included: *expression of apology, situation explanation, repair offer, promise for future, and politeness*. FACETS procedure was also utilized to trace the rater bias. Results depicted 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, as there were many variations in their ratings.

2.2 Background on the speech act of compliment

According to Hobbs (2003, p. 249), "A compliment is a speech act which explicitly or implicitly bestows credit upon the addressee for some possession, skill, characteristic, or the like, that is positively evaluated by the speaker and addressee." As Wolfson (1983, p. 89) pointed out, compliments "grease the social wheels" and accordingly function as "social lubricants".

Furthermore, Holmes (1986) states that complimenting as a speech act, requires multifaceted skills in sociolinguistics. She also notes that speech act of complimenting has "a darker side" (Holmes, 1995, p. 119). That is to say, a compliment may be interpreted the opposite and regarded as an offensive utterance. Moreover, if the complimentee feels that he must return a compliment back to the complimenter, the compliment may be regarded as a face-threatening act (Holmes, 1986).

One of the pioneering studies in this regard is the one done by Wolfson and Manes (1980) on English. In their study, they were concerned with identifying specific lexical, syntactical, and functional features of the speech acts of compliment and compliment response. The results of their study showed that the structure of the speech act of compliment tends to follow

some formulas and patterns and furthermore people use little adjectives when complimenting.

The speech act of compliment has been studied in diverse contexts and by different scholars using different methods. Previously, the focus was on the realization of this speech act in English and mainly in the United States, recently studies on compliments and other speech acts have expanded and included different contexts and cultures. For instance, Sharifian (2005) investigated the discrepancies in productions related to speech act of compliment in Persian and English in an academic context. He was mainly concerned with an exploration of the concept of *shekastenafsi* (translated as *modesty*) in relation to compliments. Whereas Persian speakers considered downgrading the compliment received as a sign of modesty, this act was not acceptable by English speakers. Sharifian also found that in Persian, people feel that they need to return the compliment to be polite and so that the other interlocutor will also feel fine.

Another study in this area is by Daikuhara (1986). He explored the use of the two speech acts compliment and compliment response among Japanese and English people of the same status. Whereas, some of his findings were similar to the ones by Wolfson and Manes (1980), he also found some differing results. For example, it was observed that people of the same status use compliments to convey solidarity and in many cases they only pay compliment to receive more information on the topic. But it seems that the specific finding of this study was related to the notion of transfer. Daikuhara observed that many Japanese speakers use expressions such as "no, that's not true" which was actually a transfer from Japanese. Another finding of this study was that Japanese speakers never used the syntactic pattern "I like/love + N" which may be very common in English.

Yu (2005) investigated Chinese and American complimenting behavior. He analyzed the complimenting behavior from different perspectives, such as strategies, functions, topic, and interlocutors' relationships. Regarding the strategy use, both groups used direct and indirect compliments, but Chinese learners used more indirect ones. Regarding the function, it was observed that Americans use compliments mainly as a conversation opener. However Chinese people use it as a real praise. In addition, while Americans complimented more on the topics such as appearance and possessions, Chinese people complemented more on functions and performance. Finally, compliments were more widespread among people of equal status in Chinese culture.

Golato (2003) and Huth (2006) also extended this line of research. They examined the speech act of compliment through conversation analysis (CA). Their participants were American English students of German. They found that American English speakers tended to say expressions such as “thank you” in response to a compliment. However German speakers tended to agree with the compliment being paid.

Purpose of the study

The literature on pragmatic assessment and the studies done particularly on the domain of speech acts indicated that there are two major gaps in the literature. First of all, pragmatic rating is still a young area of study and in need of more in-depth and profound investigations. In addition, the second niche in the literature goes back to the domain of speech acts and, more specifically, compliment speech act. The speech act of compliment is one of the most pragmatically controversial speech acts which is subject to great variations in different cultural contexts. The point is that most of the previous studies have investigated the learners’ behavior toward the speech act of complimenting. An area which still deserves more thorough exploration is EFL teachers’ behavior regarding the speech act of compliment and mainly their rating criteria in assessing EFL learners’ pragmatic productions. In response to this need in the literature, this study aims at finding the criteria and patterns that Iranian EFL teachers hold to when rating EFL learners’ pragmatic productions regarding the speech act of compliment. The following research questions are addressed in this study:

1. What criteria underpin INNESRs’ (Iranian Non-native English Speaking Rater) rating of the EFL learners’ compliment productions?
2. Is there any significant difference in INNESRs’ (Iranian Non-native English Speaking Rater) ratings and rating criteria based on their gender?
3. Is there any significant difference in INNESRs’ (Iranian Non-native English Speaking Rater) ratings and rating criteria based on their teaching experience?

3. Method

3.1 Participants

The main participants in this study consisted of sixty Iranian non-native English speaking teachers, 34 females and 26 males, with different teaching experiences from different language centers in Iran. They were asked to rate EFL learners’ pragmatic competence (based on the collected pragmatic

productions regarding the compliment speech act). One main criterion for choosing the raters was their academic level. In order to have a homogeneous group of raters, all raters had an MA degree. Also all of the teachers majored in TEFL (Teaching English as a Foreign Language) and had some years of teaching experience. Their teaching experience was further categorized into 2 levels of 1-6 years, and 7-11 years, due to the spread of teachers in each group.

3.2 Instrument

The instrument that the researchers used to collect data from the raters was a WDCT (Written Discourse Completion Test) prepared by the researchers themselves. It was based on the answers obtained from the EFL learners. The survey constituted seven situations which covered different occasions in which one would compliment someone. In the selection of the situations to be included in the survey, the three variables by Brown and Levinson (1987) played an important role. They include:

- (1) Relative power: the differentiation between the listener and speaker because of such matters as rank, degree, professional status.
- (2) Social distance: the social distance between the listener and speaker as a result of familiarity or shared solidarity due to group membership.
- (3) Degree of imposition: the degree of imposition imposed by the language used within the cultural context.

In order to have a representative sample of different scenarios, the researchers tried to have a balanced mixture of situations covering all these different factors. The efforts finally resulted in a survey of seven situations with different occasions on the speech act of compliment.

Furthermore, the main theme of the situations was around the EFL learners' family, social, and academic life. The seven situations on the speech act of compliment included: *complimenting a professor on his published paper, complimenting a close friend on her clothes, complimenting a beautiful view, complimenting a stranger who swims professionally, complimenting one's grandmother on her new bag, complimenting a student, and complimenting one's brother on his success.*

In order to come up with our final survey to be handed over to Iranian EFL teachers, the DCT situations were given to Iranian EFL learners. They were supposed to read the situation and write down what would they exactly say in that situation. Later in the selection of learners' answers to include in the survey, the same elements and factors discussed above were taken into

account. Lastly, the researchers ended this stage with a survey which included seven situations, each accompanied by one answer from Iranian EFL learners. This was the main survey of the study which was filled out by Iranian EFL teachers. Iranian teachers were asked to read the EFL learners' response to each of the situations and rate them on a Likert scale from "Highly inappropriate" to "Most appropriate" (see the appendix for the complete form of the questionnaires). Then they were asked to write their criteria and reasons for their ratings. Table 1 is a sample of the survey.

Table 1. A sample situation from the speech act rating questionnaire

Situation: 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?

EFL learner 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:

In addition, the DCTs were preceded by demographic questions about name (optional), university degree (BA, MA, PhD), gender (male or female), and teaching experience (1-6 years, 7-11 years) which was placed at the beginning of the questionnaire.

3.4 Data collection

As mentioned before, the data for this research was collected in two major phases. In phase one, the DCTs were distributed among EFL learners. In each situation the learners were asked to read a hypothetical situation and respond. The students were asked to write down exactly the compliment they would give in that situation.

In the second phase, non-native Iranian EFL teachers rated the selected responses on a 5-point Likert scale from "highly inappropriate" to "most appropriate". Following that, they pointed out their criteria for their rating. In order not to influence the teachers rating criteria, no specific criteria were provided to the teachers. What's more, the data was collected either by sending the DCTs to the participants through email or administering it in person.

3.5 Data analysis

The study was aimed at finding the patterns and criteria non-native teachers use in their rating of speech acts produced by Iranian upper-intermediate and advanced EFL students through content analysis. Furthermore, the quantitative part of this study draws on descriptive and inferential statistics, including calculating the frequency and percentage and going through measures such as t-tests, correlations, and Chi-squares.

4. Results

The first research question in this study was:

1. *What criteria underpin INNESRs' (Iranian Non-native English Speaking Rater) rating of the EFL learners' compliment productions?*

The examination of what INNESRs produced as their reasoning and justifications for appropriateness of the EFL learners' compliment productions revealed seven macro criteria. The macro criteria emerged after investigation of what they had mentioned regarding the EFL learners' productions in seven compliment situations (each accompanied with one answer from learners) presented to them. We decided on the final seven macro criteria after many discussions among the researchers and also consulting with a panel of expert who were knowledgeable on pragmatics.

The list of the macro criteria are given below, followed by some examples:

- (1) Politeness
- (2) Interlocutors' characteristics and relationships
- (3) Variety and range
- (4) Sociopragmatic appropriateness
- (5) Sincerity
- (6) Complexity
- (7) Linguistic appropriacy

Politeness: The first compliment criterion was politeness of the utterances. The raters considered this an important issue to consider when complimenting someone, especially someone older or of a higher status. To be polite one needs to use respectful language based on the context of situation and the interlocutors. An example comment from the data related to the rating and politeness as an issue is given below:

Very polite for two strangers who haven't met before.

Interlocutors' characteristics and relationships: The second criterion was interlocutors' characteristics and relationships. When rating compliment responses of the learners, the raters considered the age, gender, social status, and level of formality of the interlocutors very important factors to be cautious of. An example of this criterion by one of the EFL teachers is presented below:

The language is formal and is not suitable for a friendly talk. It seems that it is a conversation between two strangers.

Variety and range: The third criterion involves variety and range. This criterion embraced issues related to having a tactful, thoughtful, and considerate type of production. It consisted issues such as following the steps and moves of a compliment, having a sense of creativity when complimenting, and a cautious use of explicit or implicit types of compliment in relation to the situation. An example for this criterion is presented below:

I found this compliment very cliché. It didn't seem natural at all. The learner could act in a much better way.

Sociopragmatic appropriateness: The fourth criterion pointed out by the raters was the sociopragmatic appropriateness of the compliments, whether they are authentic, native like, and natural or not. Most of the INNESRs thought that for a compliment production to be appropriate, it should resemble what native English speakers in real situations. An example of raters' comment regarding this criterion is mentioned below.

It's not a natural and authentic way to compliment someone. I've never heard an American say something like this.

Sincerity: The fifth criterion was sincerity. The raters noted how important it is to be honest and sincere when complimenting, without any sign of flattery and sycophancy. Related example is presented below:

The expression "my lovely grandma" makes it an insincere and fake compliment.

Complexity: The sixth criterion important to Iranian EFL teachers was complexity of EFL learners' pragmatic productions. Length of speech produced while complimenting and an appropriate use of idiomatic expressions by students were mentioned as important criteria in the assessment of compliments. The following example illustrates this point:

Too formal and short for a conversation.

Linguistic accuracy: Appropriateness in relation to such issues as grammar, lexical choice, structure, etc. was the last criterion used by raters.

Iranian EFL teachers thought that for a compliment to be well-stated, it is imperative to follow basic structural rules of English. The following example illustrates this point:

In my idea a better adjective instead of "good" could be used by the student.

The specific criterion frequency for each of the seven situations is presented in Table 2. As Table 2 shows the criterion “*politeness*” was the most frequent criterion considered by the raters. Totally, 22.48% of the raters mentioned politeness as a factor for their ratings. 19.46% of the participants considered “*interlocutors’ characteristics and relationship*” as a criterion for rating. The criterion of “*variety and range*” was regarded by 16.77% of the respondents. The criteria “*sociopragmatic appropriateness*” and “*sincerity*” were mentioned by 12.08% and 8.72% of the raters respectively. The least frequent criteria stated as reasons for the rating by the teachers were “*complexity*” and “*linguistic accuracy*” which were stated by 8.38% and 7.38% of raters correspondingly.

Table 2. Frequency of compliment criteria in different situations among INNESRs

Situation	Linguistic accuracy		Complexity		Politeness		Sociopragmatic appropriateness		Sincerity		Variety & range		Interlocutors' characteristics & relationship		Total
	F	P	F	P	F	P	F	P	F	P	F	P	F	P	
Situation 1	5	7.14%	1	1.42%	15	21.42%	10	14.28%	6	8.57%	14	20%	14	20%	70
Situation 2	2	5.26%	1	2.63%	11	28.94%	6	15.78%	2	5.26%	6	15.78%	10	26.31%	38
Situation 3	4	12.5%	2	6.25%	7	21.87%	8	25%	2	6.25%	4	12.5%	5	15.62%	32
Situation 4	9	19.56%	11	23.91%	8	17.39%	6	13.04%	3	6.52%	3	6.52%	6	13.04%	46
Situation 5	1	2.08%	5	10.41%	15	31.25%	4	8.33%	9	18.75%	4	8.33%	10	20.83%	48
Situation 6	3	9.37%	6	18.75%	4	12.5%	3	9.37%	3	9.37%	10	31.25%	3	9.37%	32
Situation 7	4	12.5%	3	9.37%	6	18.75%	2	6.25%	1	3.12%	8	25%	8	25%	32
Total	22		25		67		36		26		50		58		298
Percentage	7.38%		8.38%		22.48%		12.08%		8.72%		16.77%		19.46%		

In order to be more specific concerning the results related to each specific situation, a summary of the findings is presented in Table 3. It provides information on the number of the situation, its mean rating, the most frequent criterion, and the raters' main comment in that situation.

Table 3. A summary of important information regarding each situation

Situations	Mean rating	Leading criterion	Teachers' main comments
1	3.63 / appropriate	Politeness (mentioned by 21.42% of teachers)	Most of the raters mentioned the complexity of the situation, that is to say complimenting a person of a higher status by someone in a lower social level, which makes it more difficult for the EFL learners.
2	3.38 / appropriate	Politeness (mentioned by 28.94% of the teachers)	Many of the teachers stated that the last section of the answer implies a high degree of imposition. They also mentioned the effect of cultural norms on the acceptability of this response.
3	3.57 / appropriate	Sociopragmatic appropriateness (mentioned by 25% of the teachers)	They doubted whether this sentence would be used by the native speakers in real life. In many cases, even in the other scenarios, this go togetherness with what is natural in the native speaking community was an important criterion for the raters.
4	2.33 / inappropriate	Complexity (mentioned by 23.91% of teachers)	Almost all raters considered it an unusual and strange type of complimenting someone on swimming which is furthermore not authentic and natural.
5	3.67 / appropriate	Politeness (mentioned by 31.25% of teachers)	In this scenario the learners are expected to compliment their grandmother, an older person. Based on Iranian culture, people are highly expected and recommended to respect the elderly.
6	2.77 / somewhat appropriate	Variety and range (mentioned by 31.25% of the teachers)	Is there any logical relationship between being well-dressed and being a good student?
7	3.30 / somewhat appropriate	variety and range" and "interlocutors' characteristics and relationship" (equally mentioned by 25% of teachers)	What most agreed on was that the language does not fit the relationship between the speakers. That is to say, based on the expected informal relationship between two brothers, this type of compliment is too much formal.

Moreover, in order to shed light on the divergence and convergence of INNESRs' ratings, the descriptive statistics for the total seven WDCT situations of the speech act of compliment was calculated. The results are summarized in Table 4.

Table 4. Descriptive statistics for compliment rating

	N	Min	Max	Mean	Std. Deviation
Situation1	60	2	5	3.63	.843
Situation2	60	1	5	3.38	1.136
Situation3	60	1	5	3.57	1.015
Situation4	60	1	5	2.33	1.084
Situation5	60	1	5	3.67	.933
Situation6	60	1	5	2.77	1.332
Situation7	60	1	5	3.30	1.062

As Table 4 depicts the mean rating of the raters for the total DCTs was 3.23. This means that on average, the raters considered the EFL learners' pragmatic productions as somewhat appropriate. Nevertheless, the ratings are much divergent in each of the seven compliment situations. The ratings in each situation vary from 1 (highly inappropriate) to 5 (most appropriate). That is to say, the minimum rating was 1 and the maximum was 5. The exception to this finding is only situation one which has a minimum rating of 2. What can be discovered from this variation in the individual and total ratings is that great variation and divergence existed in the Iranian EFL teachers' compliment ratings. Table 5 further highlights the frequency and percentage by which each of the options on the survey (highly inappropriate, inappropriate, somewhat appropriate, appropriate, and most appropriate) were selected by the INNESRs.

Table 5. Selection rate of options on the survey by INNESRs

	Situation 1		Situation 2		Situation 3		Situation 4		Situation 5		Situation 6		Situation 7	
	F	P	F	P	F	P	F	P	F	P	F	P	P	
Highly inappropriate (1)	0	0	3	5%	3	5%	15	25%	1	1.7%	13	21.7%	3	5%
Inappropriate (2)	6	10%	12	20%	4	6.7%	20	33.3%	5	8.3%	14	23.3%	10	16.7%
Somewhat appropriate (3)	8	30%	14	23.3%	19	31.7%	18	30%	18	30%	15	25%	21	35%
Appropriate (4)	28	46.7%	21	35%	24	40%	4	6.7%	25	41.7%	10	16.7%	18	30%
Most appropriate (5)	8	13.3%	10	16.7%	10	16.7%	3	5%	11	18.3%	8	13.3%	8	13.3%

Furthermore, another point worth mentioning is the standard deviation in different situations and the total standard deviation which is a sign of the dispersions in the ratings. As Table 4 manifests, the total standard deviation was .40 and the situation with the highest standard deviation was situation six with a standard deviation of 1.33. As depicted, situations two, three, four, six, and seven had a standard deviation of more than one, which is considered a high variation. This standard deviation can be regarded as a rough account of the discrepancies between the raters.

In order to figure out whether the results of ratings among the Iranian EFL teachers in their rating of the compliment productions of EFL learners were statistically significant, inferential analysis was employed which included the calculation of inter-rater reliability of raters' ratings. As this study entailed presence of multiple raters, the researchers utilized the intra-class correlation procedure to compute inter-rater reliability.

The inter-rater reliability index for the ratings by the teachers was calculated (Table 6) and was equal to .03 ($p > .05$). This inter-rater reliability index indicates that there was not any significant correlation in the ratings of the Iranian EFL teachers which further proves the lack of systematic consistency and stability among the raters.

Table 6. Intraclass correlation coefficient between the raters for the speech act of compliment

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.005	-.042	.075	1.032	59	354	.419
Average Measures	.031	-.396	.364	1.032	59	354	.419

Teachers' Gender

The second research question in this study was:

2. *Is there any significant difference in INNESRs' (Iranian Non-native English Speaking Rater) ratings and rating criteria based on their gender?*

Another aim of the study was to investigate whether the teachers' ratings and more specifically their criteria differ based on their gender or not. Table 7 summarizes the results concerning the mean rating for different situations

based on the gender of the raters. The results show that, whereas in situations one, two, and four, the overall mean rating was higher among male teachers (3.69 vs. 3.59, 3.62 vs. 3.21, & 2.35 vs. 2.32 respectively) than female teachers, in the other four situations female teachers had a higher rating than male ones.

Table 7. Descriptive statistics for male and female teachers' ratings

	Gender	N	Mean	Std. Deviation	Std. Error Mean
Situation1	Male	26	3.69	.788	.155
	Female	34	3.59	.892	.153
Situation2	Male	26	3.62	1.134	.222
	Female	34	3.21	1.122	.192
Situation3	Male	26	3.50	.860	.169
	Female	34	3.62	1.129	.194
Situation4	Male	26	2.35	1.413	.277
	Female	34	2.32	.768	.132
Situation5	Male	26	3.65	.892	.175
	Female	34	3.68	.976	.167
Situation6	Male	26	2.62	1.299	.255
	Female	34	2.88	1.365	.234
Situation7	Male	26	3.19	1.201	.235
	Female	34	3.38	.954	.164
Total	Male	26	3.2308	.41419	.08123
	Female	34	3.2395	.41122	.07052

To find out whether this difference in ratings was significant or not, an independent t-test was run. Table 8 depicts the results of the independent t-test. The results in this section ($t(58) = .081$, $p = 0.936$) illustrate that there was not any significant difference between the female and male raters in the rating of the EFL learners' pragmatic productions regarding the speech act of compliment.

Table 8. Independent samples t-test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	Df	Sig. (2- tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Total compliment	.187	.667	-.081	58	.936	-.00873	.10747	-.22385	.20639

Nevertheless, in order to come up with more accurate results concerning this difference and shed light on the differences between raters based on their gender, we also decided to consider their criteria during the ratings. Before doing any statistical procedure, first of all we were interested in knowing about the frequency of the seven macro criteria among male and female INNESRs. Table 8 summarizes the results regarding this concern.

As Table 9 depicts, the criterion “linguistic accuracy” was stated by 4.02% of the female raters and 5.70% of male raters. 5.36% of female raters and 4.69% of male raters went for the criterion “complexity”. With a noticeable difference between genders, the criterion “politeness” was selected by 13.08% and 7.04% of female and male raters respectively. The criterion “sociopragmatic appropriateness” however didn’t face much variation among male and female raters. It was opted by 7.04% and 6.04% of female and male raters correspondingly. Moreover, whereas 6.37% of the raters commented on the criterion “sincerity”, only 2.68% of the male raters thought that the pragmatic productions need to be sincere. The variation among the raters was even much greater regarding the criterion of “variety and range”. It was in fact nearly eight fold. It was stated by 32.65% of female and 4.69% of male raters respectively. And finally, the last criterion “interlocutors’ characteristics and relationships” was stated by 12.08% of female and 6.71% of male non-native English speaking raters.

Table 9. Frequency of the different criteria based on the gender of the raters

Gender	Linguistic accuracy		Complexity		Politeness		Sociopragmatic appropriateness		Sincerity		Variety & range		Interlocutors' characteristics & relationship	
	F	P	F	P	F	P	F	P	F	P	F	P	F	P
Female	12	4.02%	16	5.36%	39	13.08%	21	7.04%	19	6.37%	32	32.65%	36	12.08%
Male	17	5.70%	14	4.69%	21	7.04%	18	6.04%	8	2.68%	14	4.69%	20	6.71%

In the next stage, in order to probe any significant difference between male and female raters in the criteria they adhered to while rating EFL learners' productions regarding the speech act of compliment, an analysis of chi-square was run.

As displayed in Table 10, the female INNESRs (N = 132, Residual = 23) used more criteria during their ratings than the male INNESRs. In fact, the positive index of residual among female raters indicated that the frequency of their use of criteria was beyond expectation. In addition, the residual index for the male raters (Residual = -23) was negative, i.e. they had used criteria in ratings, but less than what was expected.

Table 10. Frequencies, expected and residual values; speech act of compliment by gender

	Observed N	Expected N	Residual
Male	86	109.0	-23.0
Female	132	109.0	23.0
Total	218		

The results of chi-square ($\chi^2(1) = 9.70$, $P < .05$) further indicated that the difference observed in Table 11 was statistically significant. Thus the null-hypothesis as there is not any significant difference in Iranian EFL teachers' criteria in rating the learners' production of compliments based on the teachers' gender is rejected.

Table 11. Chi-square speech act of compliment by gender

	Gender
Chi-Square	9.706 ^a
Df	1
Asymp. Sig.	.002

a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 109.

Teaching Experience

The third research question in this study was:

3. *Is there any significant difference in INNESRs' (Iranian Non-native English Speaking Rater) ratings and rating criteria based on their teaching experience?*

The other aim of the study was to discover whether the rating criteria of Iranian non-native English speaking (INNESR) raters differ based on their teaching experience. The data on the teaching experience of the raters were divided into two main categories. The two categories include:

a.(Group 1): 1-6 years of teaching experience (43.3% of the raters),

b.(Group 2): 7-11 years of teaching experience (56.7% of the teachers).

Frequency of the different criteria selected by the INNESR based on their teaching experience is presented in Table 12. As can be seen, in all cases and regarding all criteria, the frequency of the mentioned criteria were more among INNESR of group 2 (those with teaching experience of 7-11 years) compared to group 1 (those with teaching experience of 1-6 years). As depicted, among the participants of both groups, the criterion "politeness" was the most frequent one, selected by 7.38% of respondents in group 1 and 12.41% of raters in group 2.

Table 12. Frequency of the different criteria selected by the INNESR based on their teaching experience

Teaching Experience	Linguistic accuracy		Complexity		Politeness		Sociopragmatic appropriateness		Sincerity		Variety & range		Interlocutors' characteristics & relationship	
	F	P	F	P	F	P	F	P	F	P	F	P	F	P
1-6 years	10	3.35%	9	3.02%	22	7.38%	17	5.70%	10	3.35%	21	7.04%	21	7.04%
7-11 years	20	6.7%	20	6.7%	37	12.41%	23	7.71%	17	5.69%	26	8.72%	34	11.4%

In order to figure out whether there is any significant difference in the rating scores of the INNESRs of group 1 and group 2, an independent samples t-test was run. The results are depicted in Table 13. The results ($t(58) = .009$, $p = 0.993$) demonstrate that there was not any significant difference between the rating scores of group 1 (those with 1-6 years of teaching experience) and group 2 (those with 7-11 years of teaching experience).

Table 13. Independent samples t-test results

	Levene's Test for Equality of Variances		t-test for Equality of Means						
	F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper
Total compliment	.743	.392	.009	58	.993	.00097	.10747	-.21416	.21610

To be more accurate and exact regarding the results presented, in this section we would investigate the existence of any significant difference among the raters of the two groups based on the frequency of the criteria they adhered to. To do so, an analysis of Chi-square was run to explore any significant relationship between the INNESRs of the different teaching experiences.

As displayed in Table 14, the INNESRs who had a teaching experience of 7 to 11 years ($N = 132$, Residual = 23), used more criteria during their ratings than the INNESRs with 1 to 6 years of teaching experience. That is to say, the positive index of residual indicates that the frequency of the criteria stated by those in group 2 was beyond expectation. The residual index for the INNESRs with 1 to 6 years of teaching experience (Residual = -23) was negative, i.e. while rating EFL learners pragmatic productions regarding the speech act of compliment, they have mentioned some criteria, but that was less than what was expected.

Table 14. Frequencies, expected and residual values

	Observed N	Expected N	Residual
1-6 Years	86	109.0	-23.0
7-11 Years	132	109.0	23.0
Total	218		

Furthermore, the results of Chi-square ($\chi^2(1) = 9.70$, $p < .05$) indicated that the difference observed in Table 15 was statistically significant. As a result, the null-hypothesis as there is not any significant difference in INNESRs criteria in rating the learners' productions of compliments based on the teachers' teaching experience was rejected.

Table 15. Chi-square speech act of compliment by teaching experience

	Experience
Chi-Square	9.706 ^a
Df	1
Asymp. Sig.	.002
a. 0 cells (0.0%) have expected frequencies less than 5. The minimum expected cell frequency is 109.	

5. Discussion

The purpose of this study was to explore the criteria INNESRs consider when rating EFL learners' pragmatic productions, with a specific focus on the speech act of compliment. Also the consistency of the raters' criteria and their variability based on the teachers' gender and teaching experience were of interest. Regarding the criteria that INNESRs considered when rating EFL learners' compliment productions, the results of the study illustrated that the raters considered seven major criteria in their ratings. They included: *Politeness, Interlocutors' characteristics and relationships, Variety and range, Sociopragmatic appropriateness, Sincerity, Complexity, and Linguistic accuracy.*

The results also showed that the criterion "politeness" was the most prominent one. In many situations, the raters mentioned that the response by EFL learners is either considered polite or rude. It seems that this factor was the first thing that came to the teachers' mind. This finding is in line with similar study by Alemi (2012) done on the speech acts of apology and refusal. In her study, Alemi also found that Iranian raters considered politeness of the productions more important than anything else.

Besides not being very lenient in specifying the criteria and adhering to only one type of comment (insisting on politeness issue) on the side of some raters, another interesting point worth mentioning was that in some cases in this study, the teachers didn't provide any criteria when rating EFL learners' productions and only rated the productions based on the Likert scale and assigned a number from 1 to 5 to the productions (they were of course eliminated from the study in order not to damage the reliability of the findings of the study). This may originate from two reasons. Either they were not patient enough to explain what they thought or they didn't have any specific reason in mind. The first justification seems to be very weak,

since all of the teachers were chosen based on their willingness to contribute and did not have any time limitation. In some cases the process of data collection from some teachers even took near a month. That is to say they all were eager to contribute in this research study.

It seems that the other justification would be more acceptable. Observing many blank spaces in the collected surveys may have been driven from the lack of knowledge and academic awareness of the teachers. This is also reinforced by referring to such studies as (Eslami-Rasekh, 2005; Pasternak & Bailey, 2004) which showed that non- native English-speaking teacher candidates are not really sure of themselves about their English language proficiency level and it seems that their pragmatic competence is weaker than their organizational competence. This claim is further supported by keeping in mind that in many cases the teachers themselves admitted that they *feel* one response is appropriate or not well-formed, but didn't have any specific reason or justification for that in mind. In addition to having no criteria in many cases, the results of this study also showed that Iranian raters didn't have consistency in their rating. The results of intra-class correlations among Iranian EFL teachers proved that the raters were not homogeneous in their ratings and criteria. That is to say there wasn't consistency in their ratings. This can further cause a lack of rater consistency in ratings of pragmatic productions and also distort the concept of test fairness.

This fact in itself highlights the need for paying an additional attention to issues associated with the pedagogical aspects of pragmatic competence. According to Eslami Rasekh & Eslami Rasekh (2008), specific facets of L2 pragmatics (pragmalinguistic and sociopragmatic) do not develop satisfactorily without instruction. Fortunately, good news is that pragmatic awareness can be acquired with having a pedagogical focus on pragmatic competence (Bardovi-Harlig & Dornyei, 1998; Eslami rasekh & Eslami Rasekh, 2008; Niezgodá & Rover, 2001). Bardovi-Harlig and Dornyei (1998) and Niezgodá and Rover (2001) argued that pragmatic awareness can be acquired with a focus on pragmatic competence in educational curriculums. Unfortunately, the point is that neither EFL learners nor teachers receive decent education regarding the pragmatic issues of language and the educational system is in need of fundamental revisions. However, there are only a handful of sources which have addressed the issue of pragmatics in ESL teacher education programs. They include Bardovi-Harlig and Hartford (1997), Eslami-Rasekh (2005) and Rose (1997).

In addition to lack of pragmatic training for the Iranian EFL teachers, another reason for their lack of ability in rating effectively with coherent criteria can be attributable to the fact that not only Iranian EFL teachers, but also all Iranians on average do not have much interaction with native English speakers. That is to say Iranians primarily learn English academically and in classes in language institutes or universities with non-native Iranian teachers. Only a very few number of them have regular contact with native English speakers and are familiar with their cultural issues or have lived in a native English speaking country. Therefore, Iranian teachers didn't have any clear picture of sociopragmatically appropriate talk in English. Interestingly, in some cases their norm for choosing a response as appropriate or not appropriate was what they have seen in English original movies.

In our second research question, we were interested in knowing whether there exists any significant difference between male and female INNESRs in terms of their ratings and criteria. The results indicated that whereas there was not any significant difference between the two groups regarding the rating numbers they assigned to EFL learners' productions, the difference was significant based on the criteria that male and female raters considered during their ratings. The results showed that the female raters were more considerate and watchful in their ratings and provided more criteria. That is to say, female raters tended to be more exact in their justifications and provided more criteria. They tried not to be only limited to a selection of numbers while rating.

On the other hand, male raters were less willing to provide criteria, reasons, and justifications for their ratings. This might in fact have originated from the distinct biological and mental characteristics of male and female raters. It seems that females are more accurate in their everyday life too. Generally, women tend to be more careful and try to do everything with great care. This might have affected their ratings too. They didn't consider a mere provision of rating numbers as an accurate rating. But also tried to explain exactly why they thought so.

Moreover, another purpose of this study was observing the effect of teaching experience on INNESRs ratings. The results showed that whereas there wasn't any significant difference in the rater's ratings, a significant difference was observed in the criteria they applied. It was observed that teachers with more teaching experience used more criteria than those with less teaching experience. This finding in itself seems to be logical. Those

teachers who had a teaching experience of fewer than six years were not professional enough. They didn't have much idea on the EFL learners' productions. On the other hand, teachers with more teaching experience had a more thoughtful consideration of the pragmatic productions. They tried to clearly comment on the productions and provide accurate criteria for their ratings.

6. Conclusion and Implications

The purpose of this study was discovering the criteria Iranian EFL teachers would consider when rating the pragmatic productions of EFL learners. Whereas the raters considered seven macro criteria, the results indicated that the raters didn't have consistency in their ratings. As discussed before Iranian EFL teachers are not or rarely trained regarding pragmalinguistic and sociopragmatic issues of language teaching. This is of course true regarding the EFL learners. As pragmatic issues are much culture-bound, people from different cultures and societies need enough instruction in pragmatics to be competent. Pragmatic instruction and training are needed not only for the EFL learners, but also more importantly for EFL teachers.

The results of this study showed that Iranian EFL teachers were not consistent in their criteria and ratings. That is to say, they didn't follow any specific procedure and standard rule in their ratings. This reinforces the need for a systematic and comprehensive training educational program for Iranian EFL teachers. A unified instruction on pragmalinguistic and sociopragmatic issues of language can help the teachers to be consistent in their ratings to a great extent.

The findings also have important implications for teacher training course instructors and material developers. It seems that an important area missing in the content of language teacher education programs is an emphasis and more attention to pragmatic competence. This itself has two interrelated aspects. On the one hand, the teachers themselves need to be taught about pragmalinguistics and sociopragmatic aspects of language. Pragmatics is one aspect of language which needs more attention to be paid to, especially in the EFL context of Iran. Lack of knowledge and awareness of pragmatics was one issue present in this study among the teachers. On the other hand, EFL learners are the next group who need more instruction regarding the pragmalinguistics and sociopragmatic side of language. Instructors in teacher training courses should remind teachers to be mindful of teaching these aspects of language to the students.

References

- 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.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
- Bardovi-Harlig, K., & Dornyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *TESOL Quarterly*, 32(2), 233-262.
- Bardovi-Harlig, K., & Hartford, B. S. (1997). *Beyond methods: Components of second language teacher education*. The McGraw-Hill Second Language Professional Series: Directions in second language learning. New York: McGraw Hill.
- Beebe, L. M., & Takahashi, T. (1989). Sociolinguistic variation in face-threatening speech acts: Chastisement and disagreement. In M. Eisenstein (Ed.). *The Dynamic Interlanguage: Empirical Studies in Second Language Variation* (pp. 199-218). New York: Plenum.
- Blum-Kulka, S. (1982). Learning to say what you mean in a second language: a study of the speech act performance of Hebrew second language learners. *Applied Linguistics*, 3(1), 29-59.
- Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language*. Cambridge: Cambridge University.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In R. Schmidt, & J. C. Richards (Eds.). *Language and communication* (pp. 2-27). London: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47.
- Cohen, A. D., & Olshtain, E. (1981). Developing a measure of sociocultural competence: The case of apology. *Language Learning*, 31(1), 113-134.
- Coates, J. (2004). *Women, men, and language: A sociolinguistic account of gender differences in language*. Longman/Pearson, Harlow, Essex.

- Daikuhara, M. (1986). A study of compliments from a cross-cultural perspective: Japanese vs. American English. *Working Papers in Educational Linguistics*, 2(2), 103-134.
- Eslami-Rasekh, Z. (2005). Raising the pragmatic awareness of language learners. *ELT Journal*, 59(2), 199-208.
- 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.
- Golato, A. (2003). Studying compliment responses: A comparison of DCTs and recordings of naturally occurring talk. *Applied Linguistics*, 24(1), 90-121.
- Hobbs, P. (2003). The medium is the message: Politeness strategies in men's and women's voice mail messages. *Journal of Pragmatics*, 35(2), 243-262.
- Holmes, J. (1995). *Women, men and politeness*. London: Longman.
- Holmes, J. (1986). Compliments and compliment responses in new Zealand English. *Anthropological Linguistics*, 28(4), 485-508.
- Hudson, T., Brown, J. D., & Detmer, E. (1995). *Developing prototypic measures of cross-cultural pragmatics* (Vol. 7). Natl Foreign Lg Resource Ctr.
- Huth, T. (2006). Negotiation structure and culture: L2 learners' realization of L2 compliment-response sequences in talk-in-interaction. *Journal of Pragmatics*, 38(12), 2025-2050.
- Jianda, L. (2006). Assessing EFL Learners' Interlanguage Pragmatic Knowledge: Implications for Testers and Learners. *Reflections on English Language Teaching*, 5(1), 1-22.
- Kasper, G. (1981). *Pragmatische Aspekte in der Interimsprache*. Tübingen: Gunther Narr.
- Kasper, G. (1998). Interlanguage pragmatics. In H. Byrnes (Ed.). *Learning Foreign and Second Languages: Perspectives in Research and Scholarship* (pp. 183-208). New York: The Modern Language Association of America.
- Oller, J. W. (1979). *Language tests at school: A pragmatic approach*. London: Longman.
- Manes, J., & Wolfson, N. (1981). The compliment formula. In C. Florian (Ed.). *Conversational routine: Explorations in standardized*

- communication situations and pre-patterned speech* (pp. 115-132). The Hague: Mouton.
- Niezdoda, K., & Rover, C. (2001). Pragmatic and grammatical awareness: A function of the learning environment? In K. R. Rose, & G. Kasper (Eds.). *Pragmatics in language teaching* (pp. 63-79). Cambridge: Cambridge University Press.
- 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.
- 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.
- Roever, C. (2007). DIF in the Assessment of Second Language Pragmatics. *Language Assessment Quarterly*, 4(2), 165-189.
- Rose, K. R. (1997). Pragmatics in the classroom: Theoretical concerns and practical possibilities. In L. F. Bouton (Ed.). *Pragmatics and language learning* (pp. 267-295). Urbana, IL: University of Illinois at Urbana-Champaign.
- Sharifian, F. (2005). The Persian cultural schema of shekastehnafsi: A study in compliment responses in Persian and Anglo Australian speakers. *Pragmatics and Cognition*, 13(2), 338-361.
- 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*. DOI 10.1007/s40299-013-0118-5
- Thomas, J. (1983). Cross-cultural pragmatic failure. *Applied Linguistics*, 4(2), 91-112.
- Widdowson, H. (1978). *Teaching language as communication*. London: Oxford University Press.
- Wolfson, N. (1981). Compliments in Cross-Cultural Perspective. *TESOL Quarterly*, 15(2), 117-124.
- Wolfson, N. (1983). An empirically based analysis of complimenting in American English. In N. Wolfson and E. Judd (Eds.). *Sociolinguistics and language acquisition (Issues in second language research)*, (pp. 82-95). United States: Newbury House Publication.

- Wolfson, N., & Manes, J. (1980). The compliment as a social strategy. *Papers in Linguistics: International Journal of Human Communication*, 13(3), 391-410
- Yamashita, S. O. (1996). *Six measures of JSL pragmatics*. Honolulu: Second Language Teaching & Curriculum Center of University of Hawaii at Manoa.
- Yoshitake, S. S. (1997). *Measuring interlanguage pragmatic competence of Japanese students of English as a foreign language: A multi-test framework evaluation*. Unpublished doctoral dissertation, Columbia Pacific University, Novata, CA.
- Yu, M. (2005). Sociolinguistic competence in the complimenting act of native Chinese and American English speakers: a mirror of cultural value. *Language and Speech*, 48(1), 91-119.