Pragmatic Criteria in the Holistic and Analytic Rating of the Disagreement Speech Act of Iranian EFL Learners by Non-native English Speaking Teachers

Minoo Alemi *   **Mohammad Motamedi** **

Abstract

Conveying a strong message within a language stems from not only a linguistically appropriate utterance but also a pragmatically appropriate discourse. Broadly considering various facets of pragmatics, pragmatic assessment has not been potentially brought into perspective. To address this discourse gap, this study, guided by the principles of mixed-method design, pursued three purposes: to inspect the matches and mismatches, to explore rating variations, and to assess the rater consistency between the holistic and analytic rating methods of disagreement speech acts in L2 by non-native English teachers. As a result, 12 different pragmatic situations for disagreement DCTs accompanied by EFL learners' responses to each situation were rated by 50 non-native English teachers. Initially, they were asked to rate it holistically, incorporating both ratings and providing comments. The content analysis of raters' comments indicated sixteen disagreement criteria. The descriptive statistics also revealed variations across different situations. Moreover, the teachers were asked to rate it analytically based on the assessment rubrics adopted from Ishihara and Cohen (2010). The findings of intra-class correlations implied that respondents were more consistent in analytic rating. Moreover, the results indicated that there was a convergence between the two rating methods suggesting that the raters adopted the same level of leniency and severity in rating. Overall, the results accentuated the significance of pragmalinguistic and sociopragmatic

* Associate Professor, West Tehran Branch, Islamic Azad University - Email: alemi@sharif.ir, Corresponding author

** M.A, Sharif University of Technology -Email: mohammed.motamedi@gmail.com

aspects of language for EFL raters. Finally, the results of the present study place a premium on the importance of pragmatic assessment training as well as cultural awareness.

Keywords: Interlanguage pragmatic assessment, ILP rating criteria, Non-native English speaking raters, Speech act, Disagreement

Although pragmatics has turned into an integral part of teaching L2 programs after the development of Bachman's (1990) communicative competence model, assessment and rating of L2 pragmatics have not been studied well. Pragmatics rating is one of the novel issues in the field of interlanguage pragmatics, and its primary focus is centered on a second language learner's knowledge regarding the use of pragmatic rules and practices of the target language; therefore, to speak, it clarifies how to do things with words within the target language (Kasper, 1998). Generally speaking, as pragmatics is intimately tied with people's culture, pragmatic understanding is somewhat different between native and non-native speakers. Discerning the pragmatics of a target language and understanding what native speakers express is crucially essential as any kind of misunderstandings will take place in communication as long as a learner fails to grasp the appropriate meaning or pragmatic function of a speaker's utterance. Therefore, further studies need to be conducted to look for the criteria that non-native teachers use in assessing the appropriacy of speech acts produced in L2.

In comparison with other aspects of language, pragmatic assessment still needs further research (Rose & Kasper, 2001). To further validate this point, Alemi (2012) indicates that nonnative EFL teachers, especially Iranians who do not have access to authentic L2 examples, have serious difficulty in L2 pragmatics and pragmatic assessment. This might be attributed to the lack of valid methods for testing pragmatic interlanguage knowledge or even to the fact that developing a measure of pragmatic competence in an EFL context is instead a complex task (Jianda, 2006). In this regard, a significant number of studies have been conducted recently concerning the rubrics of pragmatic assessment within L2 classes

(Ishihara, 2013; Ishihara & Cohen, 2010). In line with this argument, it is worth mentioning that only a few studies (Alemi, Eslami, & Rezanejad, 2014; Alemi & Tajeddin, 2013; Liu, 2007; Liu & Xie, 2014; Roever, 2008; Taguchi, 2006, 2011; Tajeddin & Alemi, 2014; Walters, 2007; Youn, 2007) have addressed significant pragmatic issues, such as raters bias, behavior, and criteria applied to assess pragmatic competence. These issues place a premium on teaching and assessing pragmatics in language. The growing tendency towards assessing pragmatic ability leads us to implement potential approaches to find the criteria or framework NNES raters use to rate the pragmatic competence of EFL learners. This involves the investigation of the variations between native and nonnative raters, intercultural proficiency as well as rater bias. In particular, the present study sets out to investigate the predominant criteria that nonnative instructors apply across two pragmatic rating methods, which are holistic and analytic. On the whole, this study could illuminate and cover deficiencies such as the absence of pragmatic strategies or criteria for raters and even students and shortage of knowledge concerning cross-cultural differences in individual speech acts.

Literature Review

ILP Assessment

Testing is considered as one of the ways to assess and measure students' abilities. Although measurement lends itself to several forms, such as rating, ranking, and tests, quantification is associated with the numbers, letter grades. The assessment has different types, including formative and summative (Hughes, 1989; Bachman, 1990). As pragmatics is intertwined with people's cultures, pragmatic understanding is rather different between native and non-native speakers. In comparison with other aspects of language, pragmatic assessment is still in its infancy (Rose & Kasper, 2001). This might be attributed to the lack of valid methods for testing pragmatic interlanguage knowledge or even to the fact that developing a measure of pragmatic competence in an EFL context is rather a complex task (Jianda, 2006). Besides, few studies (Taguchi, 2011; Alemi

& Tajeddin, 2013; Tajeddin & Alemi, 2014) have addressed this issue. Pragmatic assessment, a key yet challenging issue, is fairly difficult to conduct due to some reasons. For example, considering the speech act of apologizing, several questions can be posed which are noteworthy, such as how to pinpoint a standard for appropriateness and whether it is what natives actually do or are bound to do and whether natives may necessarily understand it well (Ishihara & Cohen, 2010). Conclusively, pragmatic assessment has failed to be taken into account properly. A framework has been proposed by Hudson, Detmer, and Brown (1995) regarding assessing cross-cultural pragmatics. In line with this argument, this framework revolves around three variables that were adopted from Brown and Levinson's (1987) theory. Briefly, it consists of the speaker's power over the hearer, the distance between interlocutors as well as the obligation of the speaker to conduct speech acts. Scollon and Scollon (2001) were the dominant figures who remarked that three types of politeness systems are embedded in various contexts. These systems are called the "deference politeness system," "the solidarity politeness system," and "the hierarchical system," which consecutively refer to "distance," "imposition," and "power," mentioned by Brown and Lenvison (1987).

Furthermore, Brown (2001) has provided six main methods to assess L2 learners' interlanguage pragmatic assessment, namely the Written Discourse Completion Tasks (WDCT), Multiple-Choice Discourse Completion Tasks (MDCT), Oral Discourse Completion Tasks (ODCT), Discourse Role Play Talks (DRPT), Discourse Self-Assessment Talks (DSAT), and Role-Play self-assessments (RPSA). These methods can be utilized to elicit learners' underlying knowledge of pragmatics. Finally, it is suggested that a combination of both native and non-native speakers is the best way to assess and develop interlanguage pragmatics items (Jianda, 2006).

Cohen (2008) implied that there are more questions about pragmatic assessment for instructional objectives than there are answers. Among the tests employed to assess pragmatic knowledge, discourse completion task (DCT), and, more specifically, written discourse completion test (WDCT)

is one the most well-known open-ended tasks which are utilized for the elicitation of pragmatic knowledge (Yamashita, 2008). DCT is a type of task in which learners have to complete it based on their pragmatic knowledge. Speech acts are elicited through DCTs or role-plays and are then evaluated based on a rating scale provided for native speaker raters (Taguchi, 2011).

In the case of apology, Tajeddin and Alemi (2014) employed English native raters to assess pragmatic competence, and they found that the key elements that native raters consider while assessing L2 pragmatics were the expression of apology, situation explanation, repair offer, promise for the future, and politeness. It was also concluded that despite considering politeness, raters' biases regarding pragmatic rating have different degrees, which is in line with the result of Taguchi (2011). In a latter study, Alemi and Tajeddin (2014) conducted related research on the speech act of refusal, and they succeeded to achieve rating criteria of native and non-native teachers for assessing learners' strategies of refusal. The results of their study indicated that non-native teachers were more lenient and divergent in their ratings, which emphasize the complexity and the gap which exists among native and non-native teachers. Additionally, Alemi and Khanlarzadeh (2015) conducted a study to investigate the pragmatic rating criteria which both native and nonnative English speaking teachers apply on the speech acts of request and complaint. Similarly, the results of their study placed a premium on both pragmalinguistic and sociopragmatic factors, which the raters take into account while rating EFL learner's pragmatic competence.

Holistic and Analytic Rating

The performances of L2 learners are usually measured through two primary rating scales, namely holistic, which is also called global as well as for analytic or componential (Bachman & Palmer, 1996). As a rule of thumb, the holistic rating scale is done in a unitary way based on the overall assessment, while analytic scoring takes into account some subscales to assess different traits of a performance (Carr, 2000). To be

more precise, in analytic scoring, the rater assigns grades to indicate how well the students answer each of the facets of the question or task that is supposed to be addressed in a model answer. In line with this argument, analytic scoring reasonably yields a more objective judgment than holistic methods since it is less vulnerable to irrelevant factors and biases such as handwriting quality or a rater being dramatically influenced by the final part of the students' answers. In general, scoring regarding the analytic method takes a longer time than with a holistic one, comparatively speaking. Therefore, there might be a kind of trade-off among reader agreement, score reliability, and other factors (Klein, Stecher, Shavelson, McCaffrey, Ormseth, Bell, & Othman, 1998). A large number of studies have been conducted to address the differences between these two rating scales by many scholars (Bachman, Lynch, & Mason, 1995; Douglas & Smith, 1997).

Furthermore, it is noteworthy that the holistic method has been favored over the analytic method due to its more comfortable applicable rating scale. It is believed that the simplicity of the scoring process is less time-consuming than complicated ones. In line with this argument, the convenient and quick analysis of holistic methods encourages raters to apply it more frequently. However, teachers might fail to consider the details of learners' errors since the method is a one-shot rating scale (Carr, 2000).

Review of Studies on Speech Acts

In the classification of speech acts by Austin (1975), disagreement is considered a subcategory of commissive speech act due to the speaker's commitment to his or her ideology. In line with this argument, Stalpers (1995) holds that disagreement is viewed as when participants do not share a particular comfort zone regarding a specific subject, which in a way stems from a difference in opinion. In a study by Behnam and Niroomand (2011), it was concluded that although learners might possess a tremendous pragmatic competence, they may not be able to use disagreement strategies due to their lack of linguistic competence. Further,

Samar, Abaszadeh, and Pourmohamadi (2013) found that different factors incorporating interlocutor's power, relationships, background, and, most importantly, context affect the kind of disagreement that occurs. Additionally, Kreutel (2007) investigated the speech acts of disagreement on ESL and native American speakers by using a discourse completion test and figured that language proficiency does not necessarily guarantee pragmatic competence. Walkinshaw (2007) also studied the disagreement speech act of Japanese learners of English and implied that Japanese learners were linguistically able to disagree with native speakers, but they were not willing to do so with power-unequal interlocutors.

As the literature suggests, there seems to be no study about ILP rating concerning the speech act of disagreement. That is to say, most of the previous studies have explored the learners' behavior regarding the speech act of disagreement. Thus, rating analysis of disagreement, which is deemed as a face-threatening speech act, can yield profound results that can be applied to second language teaching and testing. As EFL learners do not have sufficient exposure to L2 input, ILP rating research can be useful. Since L2 pragmatic knowledge is integrated into the native speakers' cultural and social perspectives, they are presupposed in native speakers' minds; therefore, pragmatic understanding for nonnatives as well as L2 learners may not be entirely discernible (Widdowson, 2007).

On the whole, a thorough set of criteria is demanding for both native and nonnative teachers to rate pragmatic ability more efficiently, especially when it comes to the disagreement speech act, which is deemed as one of the most widespread speech acts in our daily life. As a result, providing a yardstick is a contributing factor within the pragmatic assessment. Hence, the reviewed literature offers a justification for conducting this study. In order to carry out this research, the following questions were posed:

1. What criteria may be used by non-native English speaking raters for rating the speech act of disagreement produced by EFL learners?

2. What variations may exist in the ratings of non-native English speaking raters with the speech act of disagreement produced by EFL learners?
3. Is there any significant relationship between the analytic and holistic ratings of non-native English speaking raters concerning the speech act of disagreement produced by EFL learners?

Method

Participants

The principal objective of the present study was to manifest the criteria adopted by non-native English speaking teachers while rating EFL learners' disagreement productions holistically and analytically. The participants composed of 50 non-native English speaking teachers (male and female) as well as 12 Iranian EFL learners chosen based on convenient sampling. In order to have a wide variety of teaching experience, the raters were classified into three levels of 1-5 years, 6-11 years, and above 11 years following their experience. In particular, 22 of the raters held Ph.D., and the rest were M.A. students of Applied Linguistics studying at various universities.

Moreover, the 12 EFL learners for this study were upper intermediate pupils studying at a prestigious English language institute in Tehran. In other words, they had passed elementary and intermediate levels of the language institute. Hence, they were precisely evaluated by the researcher in terms of their knowledge. Thus, they were expected to be capable of using the correct speech acts in the given situations.

Instruments

A written discourse completion test (WDCT) was used for the current study. In particular, items 1 and 2 were taken from Tajeddin and Alemi (2015), while the researcher designed the rest of the items. The DCT was primarily comprised of six situations with various degrees of power, distance, and imposition (Brown & Levinson, 1987). Additionally, the WDCT was judged by two language experts to add to its validity. Carrying out the research, two kinds of WDCTS were employed: In the first one,

the raters had to evaluate based on the holistic scale, while on the second one, they had to rate EFL learners' disagreement productions based on the analytic scale. To be clear, in the holistic phase, the raters were asked to read the learners' output in each situation and rate its appropriateness by the rating scale (1= poor, 2= fair, 3= proficient, 4= native-like). Also, they were supposed to write their applied rating criteria for assessing the pragmatic performance of EFL learners. In the latter phase, the raters were asked to rate the appropriateness of the same responses in situations by a different method. As a result, a table consisting of assessment rubrics developed by Ishihara and Cohen (2010) was provided for the raters to rate them on the scale.

Data Collection Procedure

A mixed-method design was applied to serve the purpose of this study. In particular, the design of this study was sequential exploratory. "A sequential exploratory mixed-method design is a set of procedures that researchers use to collect and analyze qualitative data to explore a topic in a first phase, plan a second phase based on the qualitative findings, and then collect and analyze quantitative data in the second phase to help extend or generalize the qualitative findings" (Creswell & Clark, 2011, p. 397). The justification for pursuing a mixed-method approach is that both quantitative and qualitative research methods by themselves cannot be adequate for demonstrating and explaining the complexity of language studies (Creswell & Clark, 2011). To collect the pertinent data, firstly, the WDCTs were developed and distributed among EFL learners. Afterward, the soft versions of answered WDCTs were prepared and sent to nonnative teachers through email. In the first phase, they were asked to assess the outputs holistically. The raters assessed the responses on the Likert scale and expressed the criteria they adopted while rating. In the second phase, they were asked to rate them analytically based on the rubrics.

Data Analysis

To answer the first research question, which was about discovering the criteria for assessing EFL learners' pragmatic production and the criteria frequency, qualitative and quantitative analyses were employed to derive the criteria patterns in non-native raters. More precisely, the qualitative part of these questions was carried out by using thematic content analysis. Firstly, all comments were summarized, categorized, and analyzed. Such a categorization process resulted in the identification of the dominant patterns discerned by the teachers. Then, the themes were coded and analyzed using frequency count. Subsequently, bar graphs were provided for a better and more accurate interpretation of the results.

Furthermore, to deal with the second research question, which was concerned with variations of ratings in non-native raters, the quantitative part of data analysis, which is descriptive analysis, was employed, which included the calculation of mean and standard deviation of the rating scores for the total DCTs and each situation. Finally, to answer the third research question, which was about the relationship between the analytic and holistic ratings of raters, the inferential analysis was employed, which included the calculation of the inter-rater reliability of raters' ratings. It was computed using intra-class correlation.

Results

Rating Criteria

The first research question explored the criteria in non-native English speaking teachers' criteria in the holistic rating of EFL learners' production of disagreement. The findings of this research question conducted through content analysis suggested several criteria that nonnative raters deemed while assessing EFL learners' holistic disagreements. To elaborate more, the leading criteria are provided in the next phase.

(1) *Grammatical Structure***:** Broadly speaking, this criterion managed to receive the highest degree of attention compared to other criteria. As derived from the comments, nonnative raters were deeply concerned with the linguistic accuracy of the utterances produced by learners. In the same

fashion that language teachers were predominantly focused on the structure of language in old language teaching methods, enormous favor was adduced to this issue among the nonnative raters in assessing the disagreement of learners. The following example illustrates this point.

Example: Although he was trying to express disagreement, there are some grammatical errors and verb-tense-related errors in his speech.

(2) *Appropriate Use of the Speech Act*: The second widely used criterion by the raters was the appropriate use of the disagreement by EFL learners. Several steps are to be followed to produce a correct and appropriate speech act, and a disagreeing expression is to be utilized, which is congruent with the face of the hearer. Besides, the topic is to be appropriately developed while stating disagreement. Here is a sample of a rater's comment in this respect:

Example: The learner, in this case, failed to express his disagreement. It does not sound like a disagreement, and more like he is asked to do something, he is not able to.

(3) *Politeness:* Politeness is a crucial issue within cross-cultural pragmatics owing to the fact it is contingent upon factors such as level of imposition, social distance, and power within a context. A typical disagreement is not approved by any culture or society, even though its understanding may be different. Therefore, observing the face of the interlocutor and using a face-saving and polite language is very important, as was suggested by the raters as well.

Example: The problem is with "it is your chance." S/he could have said, "that happens, scarcely." It would have been politer.

(4) *Directness*: The crux of this criterion lies within the fact that being too direct to someone, especially one with a higher status, can pose a threat to the person's self-image. Accordingly, the speaker is supposed to adjust the level of the directness of his/her speech to the specific context and the values which exist in it. As Brown and Levinson (1987) point out,

utterances could be very direct or indirect. Additionally, this criterion is intimately tied with culture, and its severity varies cross-culturally. Besides, it is intertwined with politeness and formality. An example of employing this criterion is given below:

Example: The person has talked directly about the issue, and he can say it indirectly because the listener is a manager and superior to this person, it is not appropriate to talk like that.

(5) *Justification and Reasoning*: This criterion led the researcher to implicate that disagreements remarked by the learners are to be replete with adequate clues and reasons so as the interlocutor can comprehend the rationale behind the claim. In other words, opposing an idea without justification is like claiming something without evidence. In line with this argument, the raters favored disagreements, which were well-reasoned and followed by supporting details and strategies. The following comment elucidates the point:

Example: The learner appropriately mentioned the reasons why those classes can be suitable for different people. In a nutshell, he referred to excellent points about the problem.

(6) *Vocabulary*: The lexical aspect of our discourse plays a vital role in the appropriateness of our utterances. Moreover, the importance of vocabulary has been long acknowledged in old teaching methods and has been one of the most regarded aspects of learners' communicative competence. Thus, the raters thought of this criterion significant as can be seen in the following comment:

Example: Semantically a little weak in terms of not choosing appropriate vocabulary

(7) *Social Status*: Considering social status, age, and gender of the speakers are crucially crucial that failing to observe these issues is likely to get one in trouble. As an example, a worker needs to be respectful and cautious of the discourse that he/she uses when disagreeing with a higher status like boss or manager. The following example connotes this criterion:

Example: It might not be pragmatically appropriate as the boss has a higher status than you.

(8) *Formality*: This criterion is associated with the speaker's style while expressing disagreement. The type of style we adopt when talking with our close friends is far different from the one we use while talking to strangers. To elaborate more, our speaking style varies from situation to situation from formal to informal. The following remark addresses the significance of formality to the raters.

Example: I do not think we speak like this. It is too formal for a friendly conversation.

(9) *Vagueness or Clarity*: Utterances produced need to stand clear and unambiguous to convey the meaning. Thus, the message is communicated well as long as it is to the point and understandable. The following example is a sample of rater's comment about this criterion:

Example: The utterance sounds a little vague and seems difficult to understand.

(10) *Appropriate Length of Production*: There is almost a norm or so-called 'standard' which the members of any society follow to express disagreement. Culturally, one community may approve of a more prolonged disagreement, whereas the other may prefer a shorter one. To strike a balance, the kind of disagreement made by learners is to have an

appropriate length. The following example illustrates how this criterion matters to the teachers.

Example: The answer is too brief, and the topic development is not good at all.

(11) *Pragmatic Tone*: The tone is a highly important factor in pinpointing the sincerity of our talk. Moreover, having a mild or aggressive tone is intimately tied with the politeness of our speech to our interlocutor. The mentioned comment by the respondent admits this point.

Example: Its tone could have been less severe, and it should have been more lenient.

(12) *Use of Hedging*: The way we disagree with someone sets some boundaries in that we cannot cross the line. To be more exact, if we apply a hedged disagreement like a hedged request, we respect the person's real face and gradually lessen the rudeness and enhance the politeness of our talk. Therefore, raters thought that some of the learners need to take into account more hedges when they oppose someone. The instance below stresses the issue.

Example: The utterance lacks hedging. I think it needs more softeners.

(13) *Social Distance*: It is seen that speakers observe the type of social distance that exists among them. Thus, we speak differently to different people based on the power distance, the intimacy, or the social distance that lies among them. Thus, this level of social distance arises some norms which we adjust to. The mentioned remark by a teacher addresses this point.

Example: The answer is excellent because of the relative social distance between the speakers.

(14) *L1 Transfer*: EFL learners are quite adept at avoiding structures that seem difficult for them. To hide this spot, they quickly refer to their native language, which pragmatically undermines the disagreement expressed. To elucidate, the following comment by a respondent justifies the claim. Example: It seems that the leaner's native language has affected the way he/she has disagreed.

(15) *Knowledge of Conjunctions*: Utterances that lack conjunctions do not seem natural and smooth. To be more precise, the coherence and cohesion of disagreement are likely to be weakened by the shortage of conjunctions. In consequence, the raters paid attention to this criterion and pointed the finger of their blame as to why some of the learners have failed to use conjunctions in their speech.

Example: It is not a convincing answer since there is not a logical connector among the sentences.

(16) *Organization*: Among all the mentioned criteria, the organization received the lowest degree of attention by the learners. Of course, our utterances must have organization or shape to seem appropriate. Therefore, learners need to know when to apply a concluding remark or an introductory remark and how to develop the disagreement to make it understandable.

Example: The utterance has an inappropriate beginning and lacks organization.

The criteria illustrated above had different diversity across the situations. Table 1 represents the frequency of NNESRs in each situation.

Table 1.

Frequency and Percentages of Holistic Disagreement Criteria among Nonnative Raters

Situation	GS	AS	Pol	Dir	JR	VOC	SS	For	VC	ALP	PT	UH	SD	LT	Knc	Org	Total
1	23	6	7	6	9	7	2	5	3	2	1	2	2	5	0	3	83
2	5	21	8	7	5	2	5	2	2	4	0	1	0	0	0	0	62
3	15	14	11	8	5	4	0	2	7	4	1	2	3	0	0	0	76
4	7	8	11	9	5	2	16	4	3	5	3	1	0	0	4	1	79
5	25	6	14	4	4	8	1	6	2	0	1	2	2	1	1	0	77
6	7	8	2	9	8	6	4	3	4	5	3	1	1	1	0	0	62
Total	82	63	53	43	36	29	28	22	21	20	9	9	8	7	5	4	439
Percentage	18.67	14.35	12.07	9.79	8.2	6.6	6.37	5.01	4.78	4.55	2.05	2.05	1.82	1.59	1.1	0.9	100

Note: GS: Grammatical Structure; AS: Appropriate Use of the Speech Act; Pol: Politeness; Dir: Directness; JR: Justification and Reasoning; Voc: Vocabulary; SS: Social Status; For Formality; VC: Vagueness or Clarity; ALP: Appropriate Length of Production; PT: Pragmatic Tone; UH: Use of Hedging; SD: Social Distance; LT: L1 Transfer; Knc: Knowledge of Conjunctions; Org: Organization

As illustrated in Table 1, the order of preferences of nonnative raters' criteria was: GS (18.67%), AS (14.35%), Pol (12.07%), Dir (9.79%), JR (8.2%), Voc (6.6%), SS (6.37%), For (5.01%), VC (4.78%), ALP (4.55%), PT (2.05%), UH (2.05%), SD (1.82%), L1T (1.59%), Knc (1.1%), and Org (0.9%). The results of this study indicated that nonnative raters frequently applied a variety of criteria to make an accurate assessment of students' performance in terms of relatively stable criteria that remained applicable from situation to situation and from individual to individual. On the whole, the results across situations indicated situation-sensitive divergence and convergence in rating. As pointed above, GS and AS were taken into account more often by teachers, while Knc and Org were not preferred by most of the raters. Accordingly, GS turned out to be the most dominant criterion among nonnative raters, and Org appeared to be the last one. All in all, if we are to stress the five frequently used criteria among nonnative raters, it is plausible to state that GS, AS, Pol, Dir, and JR were interestingly the leading criteria among the respondents. In order to present

a more realistic picture of the raters' use of the mentioned criteria, Figure 1 is given below.
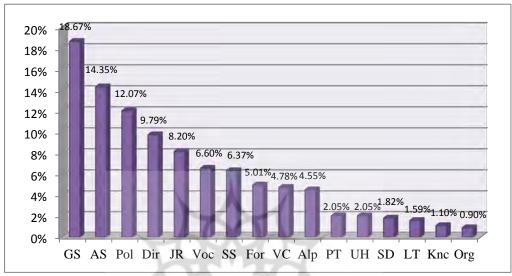

Figure 1.

Disagreement Holistic Criteria among Nonnative Raters

Since the use of criteria within each situation is unsettled and varies concerning the situation, the samples of nonnative raters' comments for each disagreement situation are provided to give a more vivid description and to analyze each situation precisely.

Additionally, the descriptive statistics on disagreement for nonnative raters regarding total disagreement-holistic DCTs are displayed in Table2.

Table 2.

Descriptive Statistics for Disagreement-Holistic Group

Speech Act	N	Minimum	Maximum	Mean	Std. Deviation
D1	50	1.00	4.00	1.70	.81
D2	50	1.00	4.00	1.30	.74
D3	50	1.00	4.00	3.24	.89
D4	50	1.00	4.00	1.74	.94
D5	50	1.00	4.00	2.02	.89
D6	50	1.00	4.00	2.82	1.00

				2.14	.48

Total DH

As Table 2 shows, the mean (M) rating of the 50 nonnative raters for total DCTS was 2.14. It reveals that their overall evaluation of disagreement in the six situations fell within the "fair" point on the scale. Although standard deviation (SD) for the entire situation was comparatively low, the distance between minimum score (1) and maximum score (4) provides a rough account of divergence or dispersion in rating disagreements.

In a similar vein, the descriptive statistics for the total disagreement-analytic DCTs and the six situations thereof for nonnative raters are presented in Table 3. A score on each situation ranged from 1 (poor) to 4 (native-like). As Table 3 indicates, the mean (M) rating of the 50 nonnative raters for total DCTs was 2.20. It illustrates that their overall evaluation of disagreements in the six situations fell within the "fair" point on the scale. However, the standard deviation (SD) for the whole situation was comparatively low.

Table 3.
Descriptive Statistics for Disagreement-Analytic Group

Speech Act	N	Minimum	Maximum	Mean	Std. Deviation
DA1	50	1.00	3.75	2.07	.58
DA2	50	1.00	3.33	1.79	.60
DA3	50	1.75	4.00	3.03	.63
DA4	50	1.00	3.75	1.90	.68
DA5	50	1.33	3.75	2.19	.55
DA6	50	1.00	4.00	2.25	.71
Total DA				2.20	.40

Furthermore, the intraclass correlation coefficient was run to assess the consistency of rating within the disagreement-holistic group, respectively, as shown in Table 4. As can be seen in Table 4, the intraclass correlation was statistically significant, displaying that disagreement-holistic group was consistent in the rating of the speech act of

disagreement, ICC = .53, F = 2.13, df = (49,245), p =.00 (see the descriptive statistics related to this group in Table 2).

Table 4.

Intraclass Correlation Coefficient: Disagreement-Holistic

| | Intraclass Correlation | 95% Confidence Interval | | F Test | | | |
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Average Measures	.53	.29	.70	2.13	49	245	.00

In the same fashion, the results of Table 5 highlighted that the disagreement-analytic group was consistent in the rating of the speech act of disagreement, ICC = .70, F = 3.37, df = (49,245), p =.00 (see the descriptive statistics related to this group in Table 3).

Table 5.

Intraclass Correlation Coefficient: Disagreement-Analytic

| | Intraclass Correlation | 95% Confidence Interval | | F Test | | | |
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Average Measures	.70	.55	.81	3.37	49	245	.00

Finally, the Pearson Product Moment Correlation Coefficient was computed as the estimate of inter-rater reliability of non-native English speaking teachers' holistic and analytic rating. This was computed between the mean of analytic rating components of disagreement scale and the mean of holistic ratings of each student. As shown in Table 6, the Correlation Coefficient was .66, which was statistically significant, implying that there was a convergence between two ratings.

Table 6.

Interrater Reliability of Raters' Ratings for Disagreement

		DAH	DAA
DAH	**Pearson Correlation**	1	.662**
	Sig. (2-tailed)		.000
	N	50	50
DAA	**Pearson Correlation**	.662**	1
	Sig. (2-tailed)	.000	
	N	50	50

** Correlation is significant at the 0.01 level (2-tailed)

Discussion

Exploring the nonnative rating of the EFL learners' pragmatic competence, finding variations, and criteria adopted in the assessment are essential topics that have been rather meager within the literature. EFL learners' pragmatic knowledge is to be examined as their linguistic knowledge is to be. Therefore, a comprehensive set of criteria is required to rate this ability. More precisely, these sets of criteria for each specific speech act help raters to test the learners' interlanguage pragmatic knowledge more consistently and efficiently.

This study probed the possibility of different inclinations to the interlanguage pragmatics construct by non-native English raters concerning disagreement. The findings indicated sixteen criteria that non-native raters employ while assessing the EFL learners' disagreement production. Moreover, the results suggested that the criterion "grammatical structure" was of utmost importance to the raters. This might be because syntactic appropriateness is one of the foremost issues that is given attention within the context of Iran.

In line with the previous arguments, "appropriate use of the speech act" favored the most attention among nonnative raters. Raters considered the sociopragmatic appropriateness of the disagreement as one of the leading factors when learners express disagreement. Regarding the matches and mismatches between the holistic and analytic methods, there

were a few mismatches, which were the use of discourse markers, epistemic stance markers as well as the level of imposition. Moreover, some new criteria were found which were missing in the analytic scale such as justification and reasoning, vagueness or clarity, L1 transfer, knowledge of conjunctions as well as the use of hedging. On the whole, it seemed that raters failed to note some critical criteria like the use of discourse markers and epistemic stance markers and level of imposition. The criterion "justification and reasoning" has been similarly expressed in other speech acts with other titles such as "explanation" by Alemi and Tajeddin (2013) in the speech act of apology by native teachers, Tajeddin, Alemi, and Razzaghi (2014) in the case of apology and understanding impoliteness by both native English speakers and EFL learners, Fraser (1981), Olshtain and Cohen (1983), Holmes (1990), Sydorenko, Maynard, and Guntly (2015), and Alemi and Khanlarzadeh (2015). To further prove the point, "use of hedging" in Paramasivam's (2007) study on Malay participants potentially indicated that disagreement is mostly expressed indirectly with a high degree of mitigation, and it does not tend to be explicit and straightforward. Moreover, Kreutel's (2007) study on ESL and native American speakers also shed light on the fact that native speakers also use mitigating devices such as hedging even more frequently than nonnative speakers. The results are also in line with studies conducted by Pearson (1986) and Beebe and Takahashi (1989), who remarked that native speakers tend to utilize mitigation devices when expressing disagreement.

Blum-Kulka and Olshtain's (1984) research on request and apology addressed factors such as directness. Furthermore, criteria such as politeness, linguistic appropriacy have been remarked in several studies (i.e., Eslami, Jafari, & Mehregan, 2012; Jalilifar, 2009; Murphy & Neu, 1996). Correctly, in the case of disagreement, significant criteria such as politeness, social status, directness have been regarded in literature. For example, Locastro (1986), Brown and Levinson (1987), Behnam and Niroomand (2011), Locher (2004), Samar, Abaszadeh and Pourmohamadi (2013), Pattrawut (2014) have profoundly addressed these criteria in detail.

To provide more examples, Alemi's (2012) study on native and nonnative raters' evaluation of refusal and apology addressed criteria such as "politeness," "reasoning and explanation," which were presented in the current study as well. Besides, Alemi and Khanlarzadeh's (2015) study on the speech acts of request and complaint came up across some criteria among which "politeness," "softeners," (use of hedging), "formality," "logical reasoning and explanation," "linguistic appropriateness," "directness," were observed in the current investigation.

Overall, nonnative teachers adopted criteria from pragmalinguistic, sociopragmatic, and metapragmatic aspects of language. Nevertheless, pragmalinguistic features drew the most significant attention among them, which is correspondingly approved by the outputs of Alcon-Soler and Martinez-Flor (2008), who implied that EFL raters are principally concerned with pragmalinguistic knowledge. However, it stands in contrast with the results of other ILP studies (i.e., Alemi & Tajeddin, 2013; Alemi, Eslami, & Rezanejad, 2014) in which raters are mostly concerned with sociopragmatic features. Consequently, the significance of both pragmalinguistic, as well as sociopragmatic aspects of language, are evident for nonnative raters.

The findings of this research also suggested that nonnative raters employed a justifiable level of convergence between the two rating methods. However, nonnative raters applied a wide range of criteria with significant variations and frequencies, which did not hold constant across all situations. Therefore, the nonnative raters' preferred criteria were not entirely the same in every single situation. In the same fashion, Youn (2007) also claimed that each rater tends to propose a unique pattern that might be contingent upon the test type and speech act. Additionally, this feature is not necessarily limited to nonnative teachers. Instead, as Taguchi (2011) also claims, native raters also do not form a uniform category, and they may vary about their perceptions and understanding of appropriateness, politeness since they come from different cultures, which may have quite different community norms for communicative events. To further justify the argument, Alemi and Tajeddin (2013) also remarked that

rating is regarded as a product-oriented task, although the employed criteria analyses is a process-oriented approach.

On the whole, it seemed as if nonnative raters were not entirely acquainted with pragmatic evaluation and its underlying basics. More importantly, the demand for their training has been pointed in the previous studies (Alemi, 2012; Knoch, 2007; Kasper, 1997; Tajeddin, Alemi, & Pashmforoosh, 2011; Weigle, 1994). To further clarify the point, nonnative raters can optimize their rating through training programs that make them aware of preliminaries regarding pragmatic assessment. The necessity to train EFL raters has also been admitted by Eslami and Eslami-Rasekh (2008).

Concerning the consistency within the holistic and analytic rating methods, the respondents demonstrated higher consistency in analytic rating. Additionally, the findings of intraclass correlation also suggested that they were far more consistent in analytic rating. Finally, there was a more significant convergence among raters. However, the obtained degree again was not close to benchmarks attained by native raters in other studies. In other words, the variation in the use of rating criteria across methods is the matter of divergence existing in evaluating the appropriateness of L2 disagreement production in every single situation, so to speak, pragmatic training and holding workshop is necessary for bringing the raters closer together in their ratings (Knoch, 2007).

Conclusion

This study presented sixteen different criteria employed by NNERs which were: "grammatical structure", "appropriate use of the speech act", "politeness", "directness", "justification and reasoning", "vocabulary", "social status", "formality", "vagueness or clarity", "appropriate length of production", "pragmatic tone", "use of hedging", "social distance", "L1 transfer", "knowledge of conjunctions" and "organization". The frequency through which they applied these criteria enjoyed a wide range of heterogeneity. Besides, there were some mismatches between the two rating methods, which were: discourse markers, epistemic stance markers,

as well as the level of imposition. More importantly, raters also proposed some new criteria such as justification and reasoning, vagueness or clarity, L1 transfer, knowledge of conjunctions as well as the use of hedging. The dominant criteria for the raters were "grammatical structures 18.67% and "appropriate use of the speech act 14.35%".

The non-native raters revealed a higher level of consistency in the analytic phase in comparison with the holistic rating. This might be due to the point that the respondents were not entirely aware of pragmatic assessment and appropriacy.

Even though there was convergence among the ratings, there were significant discrepancies in both rating methods. This level of divergence indicates the necessity of cultural education and pragmatic training for language raters as well as EFL teachers. More importantly, pragmatic training and assessment are to be embedded in the curriculum. Moreover, workshops are supposed to be held regarding pragmatic assessment to encourage a universal understanding and interpretation among nonnative raters. As a result, employing this training, rater inconsistency is diminished. As the standards of rating become universal, it adds to the fairness, reliability, and validity of our language tests. Crucially, changes that are evident in rating across the disagreement situations indicate that pragmatic performance in specific situations is somewhat more difficult to assess since variables such as power, imposition, and distance significantly affect its appropriateness, and it varies cross-culturally.

Last but not least, the other conclusion that can be drawn from this study is that nonnative raters, especially the in-service ones, require assessment rubrics as the one utilized in the study in order to yield a fair score to the pragmatic production of EFL learners. Rubrics can act as a yardstick and help them to reset their standards.

Regarding the results of the current study, several significant pedagogical implications can be offered. This study substantially contributes to the body of research on EFL teachers' pragmatic competence, and it opens a new window to the worldviews of teachers, teacher educators, materials developers, or anyone who has a role in the

process of L2 language learning. Moreover, it accentuates the fact that pragmatic assessment is as critical as a linguistic assessment. Overall, the findings also contribute to policymakers and educators to integrate necessary information to embed cultural and pragmatic dimension to learning materials.

References

Alcon Soler, E, & Martinez Flor, A. (Eds.), (2008). *Investigating pragmatics in foreign language learning, teaching, and testing.* Clevedon: Multilingual Matters.

Alemi, M. (2012). *Patterns in interlanguage pragmatic rating: Effects of rater training, intercultural Proficiency, and self-assessment.* Unpublished Ph.D. dissertation, Allameh Tabataba'i University, Tehran, Iran.

Alemi, M., Eslami, Z. R. & Rezanejad, A. (2014). Rating EFL learners' interlanguage pragmatic competence by non-native English speaking teachers. *Procedia-Social and Behavioral Sciences, 98*, 171-174.

Alemi, M., Khanlarzadeh, N. (2015). *Rating EFL Learners' Production of Request and Complaint by Native and Nonnative English Speaking Teachers.* (Unpublished master's thesis). The Sharif University of Technology, Tehran, Iran.

Alemi, M., & Tajeddin, Z. (2013). Pragmatic rating of L2 refusal: Criteria of native and non-native English teachers. *TESL Canada Journal, 30*, 63–81.

Alemi, M., & Tajeddin, Z. (2014). Pragmatic rating of L2 refusal: Criteria of native and non- native English teachers. *TESL Canada Journal, 30*, 63-81.

Austin, J. L. (1975). *How to do things with words.* Oxford: Oxford University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing.* New York: Oxford University Press.

Bachman, L. F., Lynch, B. K., & Mason, M. (1995). Investigating variability on tasks and rater judgment in a performance test of foreign language speaking. *Language Testing, 12* (2), 238-257.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice.* Oxford: Oxford University Press.

Beebe, L. M., & Takahashi, T. (1989). Sociolinguistic variation in face-threatening speech acts. New York: Springer.

Chastisement and disagreement. In M. R. Eisenstein (Ed.), *The dynamic interlanguage: Empirical studies in second language variation* (pp. 199-218). New York: Plenum Press.

Behnam, B., & Niroomand, M. (2011). An investigation of Iranian EFL learners' use of politeness strategies and power relations in disagreement across different proficiency levels. *English Language Teaching, 4*(4), 204-220. http://dx.doi.org/10.5539/elt.v4n4p204

Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, *5*(3), 196-213.

Brown, J. D. (2001). Pragmatics tests: Different purposes, Different tests. In K. R. Rose & G. Kasper (Eds.), *Pragmatics in language teaching*. New York: Cambridge University Press.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage.* Cambridge: Cambridge University Press.

Carr, N. T. (2000). A Comparison of the Effects of Analytic and Holistic Rating Scale Types in the Context of Composition Tests. *Issues in Applied Linguistics, 11*(2), 207-229.

Cohen, A. D. (2008). Teaching and assessing L2 pragmatics: What can we expect from learners? *Language Teaching*, *41*, 213-235.

Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research* (2nd Ed.). California: Sage.

Douglas, D., & Smith, J. (1997). *Theoretical underpinnings of the Test of Spoken English Revision Project* (TOEFL Monograph Series Report No. 9). Princeton, New Jersey: Educational Testing Service.

Eslami, A., Jafari, D. S., & Mehregan, M. (2012). How do you react to the breakdown after it happens? Do you complain about it?: A

contrastive study on the complaint behavior in American English and Persian. *Procedia-Social and Behavioral Sciences*, *47*, 34-40.

Eslami, Z.R., & Eslami-Rasekh, A. (2008). Enhancing the pragmatic competence of non-native English-speaking teacher candidates (NNESTCs) in an EFL context. In E. Alcón Soler & A. Martínez-Flor (Eds.), *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing* (pp. 178-197). Great Britain: Cromwell Press.

Fraser, B. (1981). On apologizing. In F. Coulmas (Ed.), *Conversational Routine: Explorations in standardized communication situations and pre-patterned speech* (pp. 259–271). New York: Mouton.

Holmes, J. (1990). Apologies in New Zealand English. *Language in Society, 19*(2), 155–199.

Hudson, T., Detmer, E., & Brown, J. D. (1995). Developing prototypic measures of cross-cultural pragmatics. (Tech. Rep. No. 7). Honolulu: the University of Hawai'i at Manoa, Second Language Teaching and Curriculum Center.

Hughes, A. (1989). *Testing for language teachers.* Cambridge Cambridge University Press.

Ishihara, N., & Cohen, A. D. (2010). *Teaching and learning pragmatics: Where language and culture meet.* Harlow, UK: Pearson Education.

Ishihara, N. (2013). Teacher-based assessment of L2 Japanese pragmatics: Classroom applications. In S. Ross & G. Kasper (Eds.), *Assessing second language pragmatics* (pp. 124–148). Basingstoke, UK: Palgrave Macmillan.

Jalilifar, A. (2009). Request strategies: a cross-sectional study of Iranian EFL learners and Australian native speakers. *English Language Teaching*, *2*(1), 46-61.

Jianda, L. (2006b). Assessing EFL Learners' Interlanguage Pragmatic Knowledge: Implications for Testers and Learners. *Reflections on English Language Teaching 5* (1), 1-22.

Kasper, G. (1997). The role of pragmatics in language teaching education. In K. Bardovi-Harlig and B. Hartford (Eds.) *Beyond Methods:*

Components of Second Language Teacher Education (pp. 113-136). New York: McGraw-Hill.

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.

Klein, S. P., Stecher, B. M., Shavelson, R. J., McCaffrey, D., Ormseth, T., Bell, R. M., & Othman, A. R. (1998). Analytic versus holistic scoring of science performance tasks. *Applied Measurement in Education*, *11*(2), 121-137.

Knoch, U. (2007) Diagnostic writing assessment. The development and validation of a writing scale (Unpublished doctoral dissertation). The University of Auckland, New Zealand. Retrieved from https://researchspace.auckland.ac.nz

Kreutel, K. (2007). I'm Not Agree with You. ESL Learners' Expressions of Disagreement. *TESL-EJ*, *11*(3), 1-35.

Liu, J. (2007). Developing a pragmatic test for Chinese EFL learners. *Language Testing, 24*(3), 391-415.

Liu, J., & Xie, L. (2014). Examining rater effects in a WDCT pragmatics test. *Iranian Journal of Language Testing*, *4*(1), 50–65.

Locastro, V. (1986). *Yes, I agree with you, but...: Agreement and disagreement in Japanese and American English*. Paper presented at the Japan Association Language Teachers' International Conference on Language Teaching and Learning, Hamamatsu, Japan.

Locher, M. (2004). *Power and Politeness in action: Disagreements in oral communication*. Berlin: Mouton de Gruyter.

Murphy, B., & Neu, J. (1996). My grade is too low: The speech act set of complaining. In S. M. Gass, & J. Neu (Eds.), *Speech acts across cultures: Challenges to communication in a second language* (pp. 191-216). Berlin: Mouton de Gruyter.

Olshtain, S., & Cohen, A. D. (1983). Apology: A speech act set. In N. Wolfson & E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 18–35). Rowley, MA: Newbury House.

Paramasivam, S. (2007). Managing disagreement while managing not to disagree: Polite disagreement in negotiation discourse. *Journal of Intercultural Communication Research*, *36*(2), 91-116.

Pattrawut, C. (2014). A Cross-Cultural Pragmatic Study: Politeness Strategies and Realizations of the Strategies Used to Perform Student-Lecturer Multiple Disagreements by Native Speakers of Thai and English. *Journal of Educational and Social Research*, *4*(1), 147-158.

Pearson, E. (1986). Agreement/disagreement: An example of the results of discourse analysis applied to the oral English classroom. *International Review of Applied Linguistics, 74*(1), 47-61.

Roever, C. (2008). Rater, item, and candidate effects in discourse completion tests: A FACETS approach. In E. Alcon Soler & A. Martinez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching, and testing* (pp. 249–266). Clevedon, UK: Multilingual Matters.

Rose, K. R. & Kasper, G. (Eds.). (2001). *Pragmatics in language teaching*. New York: Cambridge University Press.

Samar, R. G., Abaszadeh, A., & Pourmohamadi, F. (2013). Investigating Disagreement through a Context-Specific Approach: A Case of Iranian L2 Speaker. *Applied Research on the English Language*, *3*(1), 87100.

Scollon, R., & Scollon, S.W. (2001). 27 Discourse and Intercultural Communication. In D. Schiffrin, D. Tannen, & H. Hamilton (Eds.), *The handbook of discourse analysis*, (pp. 538-547). Oxford: Blackwell.

Stalpers, J. (1995). The expression of disagreement. In K. Ehrlich & J. Wagner (Eds.), *The discourse of business negotiation* (pp. 275–289). Berlin: Mouton de Gruyter.

Sydorenko, T., Maynard, C., & Guntly, E. (2015). Rater Behaviour When Judging Language Learners' Pragmatic Appropriateness in Extended Discourse. *TESL Canada Journal*, *32*(1), 19-41.

Taguchi, N. (2006). Analysis of appropriateness in a speech act of request in L2 English. *Pragmatics*, *16*(4), 513-533.

Taguchi, N. (2011). Rater variation in the assessment of speech acts. *Pragmatics, 21*(3), 453–471.

Tajeddin, Z., Alemi, M., & Pashmforoosh, R. (2011). Non-native teachers' rating criteria for L2 speaking: Does a rater training program make a difference. *Teaching English Language and Literature Society of Iran (TELLSI), 5*(1), 125-153.

Tajeddin, Z., & Alemi, M. (2014). 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.

Tajeddin, Z., Alemi, M., & Razzaghi, S. (2014). Cross-cultural Perceptions of Impoliteness by Native English Speakers and EFL Learners: The Case of Apology Speech Act. *Journal of Intercultural Communication Research*, *43*(4), 304-326.

Tajeddin, Z., & Alemi, M. (2015). *Functional Communication in English*. Tehran, Jungle Press.

Walkinshaw, I. (2007). Power and disagreement: Insights into Japanese learners of English. *RELC Journal, 38*(3)*,* 278–301

Walters, F. S. (2007). A conversation-analytic hermeneutic rating protocol to assess L2 oral pragmatic competence. *Language Testing, 24*(2), 155-183. doi: 10.1177/0265532207076362

Weigle, S. C. (1994b). Effects of training on raters of ESL compositions. *Language Testing, 11*(2), 197–223.

Widdowson, H. G. (2007). *Discourse analysis*, Oxford: Oxford University Press.

Yamashita, S. (2008). Investigating interlanguage pragmatic ability: What are we testing? In E. AlcónSoler & A. Martínez-Flor (Eds.), *Investigating Pragmatics in Foreign Language Learning, Teaching and Testing* (pp. 201–223). Bristol, England: Multilingual Matters.

Youn, S. J. (2007). Rater bias in assessing the pragmatics of KFL learners using facets analysis. *Second Language Studies, 26*(1), 85–163.

Appendix A
Speech Act Questionnaire (Disagreement.Holistic)

Professor's background:
- University Degree:
- Major:
- Gender: Male   Female
- Years of English teaching experience: 1-5   6-10   11+

- Nationality:
- Native Language:

Dear Professor, in the following situations, an English language learner (EFL learner) was supposed to express disagreements. Please read the learner's answer in each situation and rate its appropriateness according to the following rating scale. Then provide your criteria and reasons for the selection of a particular point (1, 2, 3, 4) on the scale.

Disagreement

1. You are in a coffee shop. You are talking to two people from Polytechnic, Danial and Sarah. Danial is a friend, but you have never met Sarah before. Your friend Danial says that he wants to take a writing class because he is not very good at writing essays. Sarah laughs at Danial and says:
"Only stupid people need to do those classes." (She turns to you)
"Don't you think so?" You have heard that the class is very good. You say (To Sarah):

Answer: *Mrs Sarah can we know each other? I have respect for you but please talk politely .And I disagree with your idea because if person aware of own weak point ,it's better try and try for solving the problem with reading the book ,article ,go to the specialized classes and get familiar with top professors for finding the writing skills and learn new rule and new experience.*

1. Poor 2.Fair 3.Proficient   4. Native-like

Criteria:

2. You are in a car with the head teacher of the English school. He is driving to the airport to meet another student. You don't know

the head teacher very well, because he is new at his job. He doesn't know Christchurch very well. As you drive up to a traffic light, he says, "I think we should turn right here." You know where the airport is. You know that he should turn left, not right. If he turns right, you will be five minutes late meeting the student.

Answer: *I`m sorry I can`t. I don`t have enough time to go to class.*

1. Poor 2.Fair 3.Proficient   4. Native-like

Criteria:

---

3. You are an eager follower of football. One of your colleagues invites you to a picnic. Meanwhile, he says that football is a waste of time. What would you say to disagree?

Answer: *Well, I think it's the matter of personal taste and everybody is free to choose what to like and what not to like. Football is my hobby and I prefer to spend my time on it.*

1. Poor 2.Fair 3.Proficient   4. Native-like

Criteria:

---

4. You have an excellent command of English. Your manager invites you for dinner. Now he is discussing a point with you about languages. He tells you that English is the most difficult language to learn. What would you say to disagree?

Answer: *Our minds make the difficulty of the things you should change your attitude.*

1. Poor 2.Fair 3.Proficient   4. Native-like

Criteria:

---

5. You are expecting for the train to arrive in the train station. The train always arrives at 7.30 am and you have been going to work with the train for 15 years, but it is the first time that it has delay.

Meanwhile, a well-dressed elderly lady points out to you and says the train always arrives late. What would you say to disagree?

Answer: *Oh lady, it's not true .I going to work with train for 15 years , and it's first time that it has delay I think it's your chance.*

1. Poor 2.Fair 3.Proficient   4. Native-like

Criteria:

---

6. You own a barber shop. You are running low on budget and one of your workers tells you that "I think we need to raise the prices in order not to be low on budget". However, you know that by raising the prices, you will lose most of your clients. What would you say to disagree?

Answer: *My clients are important to me and I wanna keep them.*

1. Poor   2.Fair   3.Proficient   4. Native-like

Criteria:

Appendix B
Speech Act Questionnaire (Disagreement.Analytic)

Professor's background:
- University Degree:
- Major:
- Gender: Male   Female
- Years of English teaching experience: 1-5   6-10   11+

- Nationality:
- Native Language:

Dear Professor, in the following situations, an English language learner (EFL learner) was supposed to express disagreements. Please read the learner's answer in each situation and rate its appropriateness according to the rating scale. The response sheet is to be found after the questions.

Disagreement
1. You are in a coffee shop. You are talking to two people from Polytechnic, Danial and Sarah. Danial is a friend, but you have

never met Sarah before. Your friend Danial says that he wants to take a writing class because he is not very good at writing essays. Sarah laughs at Danial and says:
"Only stupid people need to do those classes." (She turns to you)
"Don't you think so?" You have heard that the class is very good. You say (To Sarah):

Answer: *Mrs Sarah can we know each other? I have respect for you but please talk politely .And I disagree with your idea because if person aware of own weak point ,it's better try and try for solving the problem with reading the book ,article ,go to the specialized classes and get familiar with top professors for finding the writing skills and learn new rule and new experience.*

2. You are in a car with the head teacher of the English school. He is driving to the airport to meet another student. You don't know the head teacher very well, because he is new at his job. He doesn't know Christchurch very well. As you drive up to a traffic light, he says, "I think we should turn right here." You know where the airport is. You know that he should turn left, not right. If he turns right, you will be five minutes late meeting the student.

Answer: *I`m sorry I can`t. I don`t have enough time to go to class.*

3. You are an eager follower of football. One of your colleagues invites you to a picnic. Meanwhile, he says that football is a waste of time. What would you say to disagree?

Answer: *Well, I think it's the matter of personal taste and everybody is free to choose what to like and what not to like. Football is my hobby and I prefer to spend my time on it.*

4. You have an excellent command of English. Your manager invites you for dinner. Now he is discussing a point with you about languages. He tells you that English is the most difficult language to learn. What would you say to disagree?

Answer: *Our minds make the difficulty of the things you should change your attitude.*

5. You are expecting for the train to arrive in the train station. The train always arrives at 7.30 am and you have been going to work with the train for 15 years, but it is the first time that it has delay. Meanwhile, a well-dressed elderly lady points out to you and says the train always arrives late. What would you say to disagree?

Answer: *Oh lady, it's not true .I going to work with train for 15 years , and it's first time that it has delay I think it's your chance.*

6. You own a barber shop. You are running low on budget and one of your workers tells you that "I think we need to raise the prices in order not to be low on budget". However, you know that by raising the prices, you will lose most of your clients. What would you say to disagree?

Answer: *My clients are important to me and I wanna keep them*

Response sheet (Disagreement situations)

Evaluation: 1. Poor   2.Fair 3.Proficient   4. Native-like

	Situation 1	Situation 2	Situation 3	Situation 4	Situation 5	Situation 6
Linguistic aspect (pragmalinguistic ability):						
Vocabulary	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Grammatical structures	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Appropriate length of production	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Pragmatic tone(how sincere the speaker is)	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Organization(introduction, body, conclusion)	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Use of discourse markers (well, umm, by the way, etc.)	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Epistemic stance markers (I think, maybe, seem, suppose, of course, etc.)	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Cultural aspect (sociopragmatic ability):						
The level of directness, formality, politeness	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
The choice of use of appropriate speech act	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Social aspect (metapragmatic ability):						
Considering social status, age, gender	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Considering degree of social distance with the hearer	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4
Considering the level of imposition	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4	1 2 3 4