

The Effect of Self-Assessment and Conference on EFL Students' Production of Speech Acts and Politeness Markers: Alternatives on the Horizon?

Parviz Birjandi

Professor, Islamic Azad University,
Science and Research Branch, Tehran
birjandi@srbiau.ac.ir

Maryam Pezeshki

Ph.D. Candidate, Allameh Tabataba'i University
pezeshkimaryam@gmail.com

Abstract

Alternative assessment approaches received considerable attention soon after a discontent with traditional, one-shot testing. These approaches, however, have been used only to improve learners' linguistic ability despite communicative models of language which pointed that knowledge of language also involves pragmatic ability (Bachman, 1990; Bachman & Palmer, 1996). The present study tries to explore the effect of two alternative assessment approaches (self-assessment and conference) on students' production of four speech acts (apology, complaint, request, and refusal) plus politeness markers. A sample of 64 EFL students participated in this study. Metapragmatic explanations were given to three groups of university students studying English as a foreign language. The two treatment groups received instruction through self-assessment and conference and a comparison group was exposed to the conventional one-shot testing. The results of the One-Way ANOVA conducted after the treatment revealed an outperformance of alternative assessment approaches to the conventional one in the production of the four speech acts and politeness markers in WDCTs. A qualitative analysis of students' self-ratings in the self-assessment group revealed that they focused more on linguistic criteria during the initial sessions and on pragmatic ones by the end of the treatment. It can be concluded that alternative assessment approaches are beneficial to students' production of speech acts and politeness markers. In the end, the applicability and the positive effect of alternative approaches in pragmatic assessment are emphasized.

Keywords: Alternative Assessment, Interlanguage Pragmatics (ILP), Pragmatic Assessment, Conference, Self-assessment, Metapragmatic Awareness, Written Discourse Completion Test (WDCT)

Received: March 2012; Accepted: December 2012

1. Introduction

1.1. Alternative Assessment

The inception of alternative assessment in the US was made along with a discontent with the overuse of multiple-choice tests which revealed low levels of validity compared with real-world performance and produced a negative washback effect on teaching and learning (Palm, 2008). Educators were dissatisfied with the tests they used in the classrooms since they could not distinguish the process of students' learning and plan further instruction (Tzagari, 2004). The common belief was that traditional testing procedures could not represent the target construct and they could not reveal the real progress made (Lachat & Spruce, 1998). It was argued that such tests were robust in delineating the quantity and output of learning, but they could not guarantee the real world performance with the same vigor (Lombardi, 2008). The partial representation of the construct under study and remoteness from the real-world performance in target language situations in traditional tests provoked scholars to put forward the idea of alternative assessment approaches.

Although different scholars have provided diverse definitions for alternative assessment (Tzagari, 2004), the following definition seems to capture the essence of a set of assessment procedures which are gathered under the title of alternative assessment:

Assessment procedures which are less formal than traditional testing, which are gathered over a period of time rather than being taken at one point in time, which are usually formative rather than summative in function, are often low-stakes in terms of consequences, and are claimed to have beneficial washback effects (Alderson & Banerjee, 2001, p. 228).

The Effect of Self-Assessment and Conference on EFL Students'...

Different assessment methods were proposed in the alternative move, such as portfolios, journal-writing/diaries, think-aloud, peer-assessment, conference, and self-assessment. The last two of these approaches (conference and self-assessment) are the focus of the present study.

Conferences are discussions or conversations between teachers and the whole class, or groups of students and individuals, which occur after a piece of work is completed or while the work is in progress (Genesee & Upshur, 1996). Conferences can help teachers to provide feedback to students through “meta-linguistic explanations” either with individual students or with small groups (Bitchener & Knoch, 2009). They involve students meeting teachers to discuss a piece of work or the process of learning (Brown & Hudson, 1998). Genesee and Upshur (1996) believe conferences are advantageous because they help in “understanding the processes, strategies, and approaches students use” and aid the teacher in understanding the difficulties students have, their use of what is taught to them, processes students go through to do some tasks, and their interests in activities they are undertaking (p. 109). Further benefits enumerated by Genesee and Upshur (1996) for the use of conferences are presented below: They help students:

- to be self-reflective,
- to assume responsibility for their own learning,
- to collaboratively set individual learning goals,
- to assume ownership of learning,
- to recognize and enjoy their accomplishments,
- to communicate orally in one-to-one conversations with their teachers about schoolwork in ways that are important to them (p. 109).

In spite of all these advantages conferences are believed to be “relatively time-consuming, difficult and subjective to grade, and typically not scored or

rated at all” (Brown & Hudson, 1998, p. 664). Nevertheless, conferences can be used in EFL contexts when the teacher has conversations with individual students to help them with difficulties they have in performing tasks and exercises. It is generally used to provide students with feedback on their work. Although it has been generally used in writing, it is not much common in speaking (McNamara, 2001).

Self-assessment is the procedure in which students try to assess their own language knowledge. It helps students to evaluate their performance and reflect on it, which can help them to promote awareness of learning and exercise some control on their own learning and to motivate them (Butler & Lee, 2010; Oscarson & Apelgren, 2011). It also informs teachers about the affective state of each student which is in line with humanistic approaches to learning and learner-centered pedagogy (Butler & Lee, 2010; Little, 2005). This causes the teachers to look at each student as a whole person, considering the cognitive dimension together with their affective aspect. Through exercising self-assessment both teachers and students feel a sense of “shared responsibility” towards assessment (Little, 2005, p. 322).

Self-assessment has been proved to be effective because of promoting learners’ “self-regulatory learning and autonomy” (Butler & Lee, 2010, p.6). Furthermore, as Butler and Lee (2010) state “through self-assessment, students can become aware of goals and expectations, monitor their learning processes and progress, and evaluate their own state of understanding” (p. 8).

Many studies have been conducted on self-assessment to date (Bachman & Palmer, 1998; Blanch & Merino, 1989; Butler & Lee, 2010; Janssen van-Dieten, 1989; Matsuno, 2009; Oscarson, 1989; Ross, 1998). However, studies conducted on the validity of self-assessment (Bachman & Palmer, 1989; Blanche & Merino, 1989; Ross, 1998) have produced mixed results originated from

The Effect of Self-Assessment and Conference on EFL Students'...

individual differences among students (such as age, proficiency level, and anxiety), the skill which was being assessed, and the formulation of questions or items (Butler & Lee, 2010). Ross (1998) carried out a meta-analysis of studies done in the literature with a focus on the validity of self-assessment. He believes studies on self-assessment concerned themselves too much with the two traditional concepts of validity and reliability, ignoring the beneficial effects of this alternative type of assessment (such as motivating students and teachers).

Nevertheless, all studies carried out with conference and self-assessment had their focus on the linguistic aspect of language competence and were oblivious to the pragmatic aspect which is intricately intertwined with it.

1.2. Interlanguage Pragmatics (ILP)

Pragmatics is “the study of language from the point of view of users, especially of the choices they make, the constraints they encounter in using language in social interaction and the effects their use of language has on other participants in the act of communication” (Crystal, 1997, p. 301, cited in Kasper & Rose, 2001). Pragmatic competence is “concerned with the ability to bridge the gap between sentence meaning and speaker meaning in order to interpret the indirectly expressed communicative intention” (Jung, 2002, p. 4). It is divided into two components: Pragmalinguistic and Sociopragmatic knowledge.

Pragmalinguistics is the consideration of forms and means of expressing intentions; for example, the knowledge of how to use questions or modals to indirectly request for something relates to pragmalinguistics (Roever, 2006). It is defined as the “intersection of pragmatics and linguistic forms” (Kasper & Roever, 2005). Sociopragmatics, on the other hand, involves a concern for power, distance, imposition or some other conventions such as “mutual rights,

obligations and conventional courses of action” that occur in a community while using the language (Roever, 2006). It is everything about what we need to do to whom and when (Kasper & Roever, 2005).

Models of communicative competence proposed by Canale (1983 cited in Tada, 2005) and Bachman and Palmer (1996, a revision of Bachman’s 1990 model) included pragmatics as one component of communicative language ability which is inter-related with linguistic ability. Similar to the linguistic aspect of communicative competence, pragmatic competence consists of so many subcomponents, such as speech acts, politeness, back-channeling, dialect and language variation, discourse markers, indirectness, metapragmatics, phatic expressions (pre-sequence, prosodic, register), and turn-taking (Yamashita, 2008). However, an inclusion of all these components in one test of pragmatics is a cumbersome and laborious task which makes it nearly impossible to undertake. Of all the subcomponents, speech acts are the ones being widely used as they lend themselves better to testing.

Austin (1962 cited in Tada, 2005) first proposed his model of speech acts by maintaining the felicity conditions; the model was later supplanted by Searl’s (1969, cited in Tada, 2005) idea of performatives. He believed people do not use the language just to make statements about facts; they use language to perform acts (Walters, 2004). Many types of speech acts have been proposed, but those mostly relevant to the study of interlanguage pragmatics include requests, refusals, apologies, complaints, complements, and suggestions.

Another component of pragmatics which has not received much attention regarding both its instruction and assessment is politeness. It is believed that non-native speakers use fewer politeness markers and strategies in their conversations than native speakers (Dufon, 2008); therefore, the instruction of these markers and strategies may contribute to L2 learners’ use of these

The Effect of Self-Assessment and Conference on EFL Students'...

strategies. House and Kasper (1984, cited in Watts, 2003) proposed an elaborate taxonomy of politeness markers used by native speakers in various situations. The strategies include, but are not limited to, “politeness markers” (such as the word *please* to show respect or ask for help), “play-downs” (such as the use of past tense or interrogatives to reduce the face-threatening act), “hedgers” (such as *kind of* or *sort of* to let the addressee choose his/her purpose without any imposition), and “committers” (like *I think* which are used to reduce the amount of commitment to a statement by the speaker).

Pragmatics has not received enough attention with regard to assessment. Although communicative competence involves both pragmatic and linguistic abilities, not all tests or assessment procedures include pragmatics as a component. Few tests try to assess the pragmatic knowledge of testees, not including among them TOEFL and IELTS with holistically scored speaking sections which give no possibility of assessing pragmatic knowledge independently and on its own right (Grabowski, 2007). Even the alternative assessment procedures used so far have focused only on the linguistic aspect of language.

2. Purpose of the Study

Different alternative procedures used in the studies conducted so far, have ignored the pragmatic side of language competence, focusing exclusively on the linguistic side (e.g., Little, 2005; Matsuno, 2009; Oscarson & Apelgren, 2011). Furthermore, various tests of pragmatics have been used in the literature in a one-shot testing of the students' comprehension and production and have not paid much attention to the *process* of learning pragmatics.

Taking a pragmatic point of view, the present study tried to compare the effect of different alternative assessment procedures (conference and self-

assessment) with a focus on pragmatics and to examine whether using any of the different alternative assessment procedures (conference and self-assessment) had any significant effects on students' pragmatic production as opposed to the one-shot traditional testing. In other words, if the assessment of pragmatics was done through the process-oriented approach of alternative assessment, would it help to improve students' production of speech acts and politeness markers? Furthermore, the present study aimed to examine self-ratings of students in the self-assessment group to see what criteria they used in the assessment of their pragmatic production. Therefore, the following research questions were addressed in this study:

1. Is there any significant difference between alternative and non-alternative assessment procedures as far as students' written production of speech acts (refusal, complaint, request, and apology) and politeness markers are concerned?
2. What pragmatic criteria do students apply in self-assessment of their speech act production?

3. Method

3.1. Participants

The participants of the present study were 64 freshman university students majoring in English Translation and English Literature at two universities in Tehran. These participants ranged in age from 18 to 24 years old. The study was conducted in the Speaking and Listening course in the first semester at the universities. All the participants had studied English at different institutes or high school for at least seven years before being admitted to university. Since grammar was not the focus of the present study, participants were homogenized only based on their pragmatic production ability. Therefore, they

The Effect of Self-Assessment and Conference on EFL Students'...

were corrected whenever their grammatical mistakes hindered communication or if they asked the instructor for correct forms.

Treatment was carried out in three different groups: The *self-assessment* group consisted of 20 students and there were 24 students in the *conference* group and 20 participants in the one-shot testing, *comparison* group.

Since a randomization of the sample was not possible, three intact classes at the university with the same level of pragmatic production proficiency (as revealed by the use of native speaker strategies in the pretest) were chosen. However, assignment of the three groups to the two experimental (self-assessment and conference) or comparison groups was done randomly.

3.2. Instrumentation

3.2.1. Written Discourse Completion Tasks (WDCTs)

Since the focus of the present study was on the *production* of four speech acts including refusal, complaint, request and apology, WDCTs were given to students both as pretest and posttest. Based on the results of the pretest WDCTs students were also homogenized regarding their pragmatic level. The test battery consisted of 20 WDCTs on the four speech acts where five scenarios were allocated to each speech act. The items on the test also varied with regard to levels of imposition, power and distance. Care was taken to include a mixture of the three factors for each speech act. If the first task was high imposition, high distance, and different power relations, the second included low imposition, low distance and the same power relations. Students' responses to the 20 scenarios were rated on a Likert scale from 1 to 5 based on the native speaker strategies criteria (gained from the corroborated results of previous studies in the website of Center for Advances in Research on Language Acquisition <http://www.carla.umn.edu/speechacts/>) and politeness

markers (from House & Kasper's (1984, cited in Watts, 2003) taxonomy of politeness markers).

WDCT validation. In order to prepare the scenarios and validate the pretest and posttest tasks, four steps, as suggested by Liu (2006), were taken. Another group of 144 EFL university students and four native speakers participated in the validation stage of the study. The first step involved an *exemplar generation* which was used to obtain topics of the scenarios for each of the speech acts (request, refusal, complaint, and apology) from students themselves (Liu, 2006). Students were first given a brief training and an example of the specific speech act and were then required to write down five most recently occurring examples of the speech act in their daily life, either in English or in Persian. Altogether, 64 EFL university students participated in this stage of WDCT preparation. Most students wrote about four or five scenarios that occurred in their daily life. In general, about 210 situations were generated by the students, most of which were new. Only about 35 out of 210 scenarios recurred for two, three or four times with similar themes. Therefore, about 160 situations were selected and revised regarding their grammatical or spelling mistakes, however, original meanings were kept intact. The aim was to create a test with an acceptable level of face validity to students.

The second stage involved *likelihood investigation* (Liu, 2006) in which students were required to identify the likelihood of any of the situations happening in their daily life. Another group of 60 university students participated in this phase of the study. They were asked to identify the likelihood of the scenarios using a five-point Likert scale. A score of five meant the scenario happened recurrently in their daily life and the probability of its recurrence was high. After the students rated the scenarios, the average of the ratings was calculated for each scenario. Forty scenarios were selected and care

The Effect of Self-Assessment and Conference on EFL Students'...

was taken to involve those with combinations of the three contextual factors (power, distance, imposition) to set them for the next stage of WDCT production, i.e., *metapragmatic assessment* (Liu, 2006).

In this stage, 10 students were given instructions on the three contextual factors (power, distance, and imposition), their significance in communication, and some example situations in which the three factors occurred with various strengths. After that, they were asked to identify these factors for each of the scenarios.

The last stage involved a *pilot study* of the scenarios in which 10 EFL students and four native speakers were requested to provide their appropriate answers to each of the scenarios to check if any problems would appear in reacting to the scenarios or comprehending them. Furthermore, since the scenarios were written based on the students' input, two native speakers were asked to double-check if the scenarios sounded authentic, native-like English.

3.2.2. DCTs on the Four Speech Acts (Refusal, Complaint, Request and Apology)

Each group received eight sessions of instruction on the four speech acts (two sessions for each). For each speech act there were two DCTs which involved students in different levels of imposition, power and distance during each session. Scenarios which were taken from students in the validation phase of the WDCT preparation and had been excluded from the pretest were used as classroom tasks after students received metapragmatic explanations.

3.2.3. Pragmatic Criteria Questionnaire

An open-ended questionnaire was used in order to investigate the criteria students resorted to during the rating of their production of each of the speech acts in the self-assessment group. Therefore, after students received explicit instruction on each of the speech acts and provided answers to the related tasks, they rated their own answers on a scale of 1 to 5 (from extremely inappropriate to extremely appropriate). Students in the self-assessment group were then required to write down what criteria they used in order to rate their production of the speech acts.

3.2.4. Politeness Markers Taxonomy

House and Kasper's (1984, cited in Watts, 2003) taxonomy of politeness markers and strategies was used to teach politeness markers to students after they were taught the native speaker speech act strategies.

3.3. Procedure

Three intact groups of students participated in this study, two of whom received instruction through alternative assessment procedures (conference and self-assessment), whereas in the other group (the comparison/control group) the conventional one-shot testing was used. The study continued for 10 weeks at the university and all the three classes were taught by the same instructor. Two sessions were allocated to the pretest and the posttest and eight more sessions were allotted to the treatment (two sessions for each of the four speech acts and the relevant politeness markers).

All the students took the WDCT pretest in order to check their level of pragmatic knowledge and also to check their homogeneity. One week after the

The Effect of Self-Assessment and Conference on EFL Students'...

pretest was administered, the instruction started. Two sessions were allocated to each speech act. In the first session the strategies used by native speakers were taught. Students were provided with metapragmatic explanations on the native-speaker strategies used for each speech act. These strategies were taken from the corroborated results of previous studies in the website of Center for Advances in Research on Language Acquisition <http://www.carla.umn.edu/speechacts/>. For example, Native American speakers were found to use positive opinion (For example, I'd love to), gratitude, and future acceptance in refusing an invitation:

• *I'd like to, Sam, but I have lots of work to do for the weekend ... thank you for inviting me, though. Hope you enjoy the party. Maybe next time!*

In the second session, politeness markers and strategies relevant to each speech act were taught. For example, indirect strategies were said to be more polite in making requests. Instead of saying “*open the door*”, students were taught to use a more indirect polite structure like, “*would you mind opening the door?*” using “consultative devices” in the taxonomy. Meanwhile, students received instruction on the three contextual factors in speech acts: power, distance, and imposition. They were taught about the variations among these factors in different situations and with different people. For example, they were told that they should consider power relations when talking to their boss versus when talking to a friend. Since pragmatic and linguistic abilities are interrelated, students in the three groups were given feedback on their grammatical mistakes whenever necessary. After the preliminary instruction, each group went through its specific treatment, which is explained below, and then they sat for the posttest.

Conference group. Students received metapragmatic explanations for the strategies used by native speakers, the relevant politeness markers and the

three contextual factors (power, distance, and imposition) for each speech act. To help minimize an overload of information given to students, politeness strategies in the taxonomy were taught step by step over the sessions so that students could have enough time to learn and acquire them. After each speech act strategy was explained, the students were given a scenario and were asked to provide an appropriate response based on the strategies that were taught to them. Then, they were divided into six groups of four students each. The instructor had discussions over the students' responses with them in each of the six groups, either with the students one by one or with the group, and talked about the ways they could revise their answers to make them more appropriate to the situation considering the variables in the scenario. However, the preference was given to students' self-correction and peer-correction in which they were asked to give their ideas about their own answer and that of their friends. The instructor waited to hear the students' comments on their responses and the ways they thought their responses could be improved. Genesee and Upshur's (1996) guidelines for conducting conferences in the classroom were used with relevant changes to suit the assessment of pragmatics. Examples of questions the instructor asked the students were:

- what do you like about your response to this situation?
- how do you think you can improve this response using the strategies I gave you?
- which strategies did you use in your response? Would you show them to me?

The treatment in this group continued with the same procedure as students received one scenario each session and then manipulated all strategies and markers for each of the speech acts.

The Effect of Self-Assessment and Conference on EFL Students'...

Self-assessment group. Students in this group also received explicit instruction and metapragmatic awareness on strategies for each of the speech acts. Over a period of two sessions, all the strategies for a specific speech act, the relevant politeness markers and the three contextual factors were taught and students practiced using the strategies in the scenarios given to them.

After the students received instruction on the strategies, a scenario was given to them and then they were asked to write an appropriate answer using those strategies. Then the instructor asked the students to rate their own performance on a scale of 1 to 5 from 'very inappropriate' to 'completely appropriate'. Since an uninformed self-assessment technique was used in this study with the aim of not providing the learners with criteria on assessment, an open-ended questionnaire was used in this group which required the students to write down the criteria they used to rate their own production of the speech acts. The purpose was to decipher what criteria were more significant for the learners themselves. After the students finished their ratings, the instructor gave them an appropriate response to the situation so that they could compare their own responses with it. The treatment continued with the same procedure until all the strategies, politeness markers, and contextual factors for the four speech acts were taught to students.

Comparison group. After taking the pretest, students in this group were given instructions on the strategies for each speech act. Two sessions were devoted to the strategies of each speech act, politeness markers and contextual factors. After the instruction in each session, students were given a scenario and were asked to write down their responses to each. Their responses were scored by the instructor and given back to them the next session. After the eight sessions of instruction, students in this group sat for the posttest.

4. Results

The results of the analyses of the students' pretest and posttest WDCTs and a qualitative analysis of the criteria students used in the self-assessment group to rate their pragmatic production are presented in this section.

4.1. Is There Any Significant Difference between Alternative and Non-alternative Assessment Procedures as far as Students' Written Production of Speech Acts and Politeness Markers Are Concerned?

To see whether the two experimental and one comparison groups differed in their performance on WDCTs, a pair of One-way ANOVA procedures was conducted, one before the treatment to make sure of the homogeneity of the sample and to see if they were similar in their performance before the treatments, and another after the treatment. Table 1 presents the descriptive statistics of the pretest scores of the groups.

Table 1. Descriptive Statistics for the Pretest WDCT Scores of Self-assessment, Conference, and Comparison Groups

Group	Mean	SD	Skewness	Kurtosis
Conference	37.2	9.8	.17	.23
Self-assessment	41.1	10.8	.25	.33
Comparison	38.3	14.2	.20	1.15

The means of the conference, self-assessment and comparison groups were 37.2, 41.1, and 38.3 and the standard deviations were 9.8, 10.8, and 14.2, respectively, which revealed that students in the three groups performed almost equally before the treatment. None of the amounts for skewedness and kurtosis for the three groups exceeded -2 or +2, which revealed no skewed or peaked distributions (Bachman, 2004). To make sure of the normality of the

The Effect of Self-Assessment and Conference on EFL Students'...

distributions, however, a Kolmogorov-Smirnov test was also carried out. None of the amounts for the K-S test exceeded the critical value, which revealed the three groups had all normal distributions before the treatment. Because the number of students in the three groups was not the same, a Levene's test of equality of variances was conducted to ensure that the three groups had equal variances.

The observed value for the Levene's statistic was less than that of the critical value. Therefore, the three groups were said to have equal variances. To ensure that the three groups were homogeneous before the treatment and did not differ significantly from one another, a One-way ANOVA was conducted.

Table 2. One-Way ANOVA for the Pretest Scores of Conference, Self-assessment, and Comparison Groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	169.2	2	84.6	.62	.54
Within Groups	8296.8	61	136.0		
Total	8466.0	63			

The One-Way ANOVA showed that the differences in scores between the comparison group (N=20, M=38.3, SD=14.2), the self-assessment group (N=20, M=41.1, SD=10.8), and the conference group (N=24, M=37.2, SD=9.8) were not statistically significant. Therefore, the three groups were not significantly different from one another before the treatment. After the treatment, another One-way ANOVA was conducted with the posttest WDCT scores of the students.

Table 3. Descriptive Statistics for the Posttest WDCT Scores of Self-assessment, Conference, and Comparison Groups

Group	Mean	SD	Skewness	Kurtosis
Conference	79.9	13.8	1.8	3.1
Self-assessment	79.3	9.2	1.3	1.8
Comparison	40.5	12.5	.0	1.4

The mean scores for the conference, self-assessment and comparison groups were 79.9, 79.3, and 40.5, and the standard deviations were 13.8, 9.2, and 12.5, respectively. The two experimental groups had similar mean scores, while the comparison group was different. The values for skewness and kurtosis did not exceed -2 or $+2$, except for the conference group with a kurtosis of 3.1 which shows the distribution might not be normal. Therefore, to investigate whether the scores of the three groups had normal distributions in the posttest, a Kolmogorov-Smirnov Test was conducted.

The values observed with the K-S test were not significant for any of the three groups, which revealed the groups had normal distributions after the treatment. To make certain of the equality of variances of the three groups, a Levene's Test was conducted.

Results of the Levene's test were not significant. Therefore, the groups were shown to have equal variances after the treatment. A One-way ANOVA was conducted to see whether the differences between the means of the groups were significant (Table 4).

Table 4. One-way ANOVA with the Posttest WDCT Scores of the Self-assessment, Conference, and Comparison Groups

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	21063.3	2	10531.6	71.8	.00
Within Groups	8944.2	61	146.6		
Total	30007.4	63			

As can be seen in table 4, the differences among the scores in the three groups were statistically significant, $F(2, 61) = 71.8, p < 0.05$. That is, there was a significant difference between the groups after the treatment. Therefore, the null hypothesis of no difference between the groups was rejected. To see where

The Effect of Self-Assessment and Conference on EFL Students'...

the difference lied, a *post hoc* Scheffe test was conducted. The results are shown in Table 5.

Table 5. Scheffe Test for the Posttest ANOVA results of Conference, Self-assessment, and Comparison Groups

(I) Groups	(J) Groups	Mean Difference (I-J)	Sig.
Conference	Self-assessment	.5	.99
	Comparison	39.4*	.00
	Self-assessment	38.8*	.00
Comparison			

Only the differences between the self-assessment group and the comparison group and those of the conference group and the comparison group were significant. That is, both groups receiving treatment using alternative assessment procedures outperformed the one-shot assessment comparison group in their WDCT performance.

4.2. What Pragmatic Criteria Do Students Apply in Self-Assessment of Their Speech Act Production?

In order to answer the second research question, a qualitative analysis of the students' answers to items in the open-ended questionnaire, regarding the criteria they used for the ratings of their responses to each scenario, was conducted. An open coding approach was used in which the students' criteria for rating were analyzed and were grouped under the same theme and then those themes were named by the researchers. Based on the analysis, five groups of rating criteria were identified as no specific criteria, grammatical criteria, politeness criteria, contextual factors criteria (by referring to the level of

imposition, power relationships and distance), and pragmatic strategies criteria.

No specific criteria. No criteria were used during the initial sessions. Students produced outputs responding to each scenario with what they believed was the best way to perform in that situation. However, they did not present any rationales. For example, in asking a friend for help in order to borrow a book from the library, one student gave the following response:

“Mina, can you tell me how I can borrow a book from university’s library? I need a book for my lecture.”

She gave a low rating of 2 out of 5 to her response and simply wrote:

“I think it is a good method for request.”

Students in this group also tended to give a lower-than expected score to their responses to each scenario even if their answers were suitable for the situation based on the speech act and politeness strategies they were taught. It seemed they tended not to put enough confidence in their responses. For example, a student in this group gave 3 out of 5 to her response which was:

“Excuse me, can you tell me how I can borrow a book from the library?, otherwise I may borrow the book from other people.”

This student believed *“maybe my answer isn’t true”*, although she used the three strategies for performing a request. One student even mentioned that she used the three parts of a request but she was not sure whether her production was true or not. Another student rated her response 3 because she believed she might have made a mistake in her output: *“Zahra, would you please give me your book you borrowed from library?”*

“I think that maybe my answer is wrong.”

Grammatical criteria. About 30 percent of students talked about the grammatical mistakes they had in their responses and rated them merely on the

The Effect of Self-Assessment and Conference on EFL Students'...

basis of grammatical criteria even if they were aware of using the speech act and politeness strategies. For example, one of the students rated his response 4 “*because I am not sure not to have a grammatical mistake in my request, but I used all of the 3 structures of a request*”. Another student rated her sentence 3 out of 5 “*because I think part three in my statement is false and I didn't use a good clause in my statement*”.

The grammatical criteria, however, were used only for the first two or three sessions of the treatment in this group. Gradually, students changed their rating criteria toward pragmatic ones.

Politeness criteria. Nearly half of the students prioritized politeness as their criteria for rating. They believed their responses were suitable because they were polite to the person they were talking to. One student, for example, wrote “*because I suppose I don't know her/him very well, I don't speak with him/her like a friend*”. Another student said “*because the person is someone we don't know, we have to be careful and polite*”.

Even toward the end of treatment using politeness criteria abounded in students' reasons for ratings, giving reasons like “*because my complaint is polite*” or “*the request for non-recurrence is a little less polite*” or “*we should say in a polite way, I think, with the use of questions or transfer of the blame from the person to the object*” for their ratings.

Contextual factors criteria. The second session of each speech act instruction revolved around the contextual factors and politeness markers relevant to each speech act. About thirty percent of students used only these factors for their ratings after they were taught the relevant factors. These students did not mention the native speaker strategies in their comments at all and only resorted to the contextual factors. For example, students wrote sentences like: “*I think I consider approximately the three types of contextual*

factors.” or “The imposition is high and the social distance is about high. The power relationship is high because I speak with my professor so I have to be more polite.” or “The power relationship is high and so is social distance. So, I think it’s a good request.” or “Because the dorm manager is more powerful than us and we should speak to her politely. She has a high social distance from us”.

Students emphasized on power relationship, distance, imposition, and politeness more than they did on native speaker strategies. Most of the ratings had these factors as their criteria. Only some students talked about the native speaker strategies.

Pragmatic strategies criteria. Towards the end of treatment sessions, more than 70 percent of students used the pragmatic strategies criteria as the only criteria for giving a rating. For example, one of them wrote *“I just came by to see if I could talk about my problem. I think maybe the problem is with the radiator. I appreciate it if you would check the radiator. I hate to talk about it but I have to.”* in making a complaint to the dorm manager for the rooms being cold. Then she wrote:

“5, I used explanation of purpose, complaint, request for solution and request for non-recurrence.”

All of these criteria were the native speaker strategies for making a complaint. Other students said they *“used the rules”* or *“used the strategies”* in their ratings. Even after using pragmatic criteria, however, some students were still worried about their sentences being grammatically correct or having spelling mistakes. For example, one student wrote: *“I should talk about last night, although I don’t like to. I wish we could solve this problem for me if you have time. Last night heaters had problems, I think, and they couldn’t warm my room and it was really cold. Is it possible to repair them?”* Then, she wrote:

The Effect of Self-Assessment and Conference on EFL Students'...

“4 out of 5. I think I used forewarning and I used ‘we’, questioning, depersonalization of the problem and mitigators. If I don’t have spelling and meaning wrong, I think it’s a good score.”

By the end of treatment sessions, students stopped underrating their answers. They even sometimes gave a rating of 5 to themselves because they used all the strategies. For example, in one WDCT students were given a situation in which the teacher asked them to work in groups, but for the reasons they expressed, they did not like working in groups and so they wanted to refuse the teacher’s request. One student wrote: *“Is it possible not to work in groups? Because I’m not sure I can use others’ ideas on my project. I prefer my project to be the best because I know it is going to be successful.”* And then he rated his response 5, *“because I can convince my teacher and promise to do the best”*. He seemed completely confident about the answer because he could manage the situation. Some other students started giving higher ratings to their answers because they thought their answers were *‘complete’*, meaning they used all strategies.

5. Discussion

As is apparent from the results of ANOVA with the WDCT scores of students in the posttest, both groups taught through alternative assessment approaches (self-assessment and conference) outperformed the comparison group which was exposed to a conventional, one-shot method of assessment in their performance on the WDCTs. The results of the present study reveal that the two alternative assessment procedures were advantageous to students as far as their pragmatic production was concerned. The benefits provided by these procedures in contributing to students’ learning have been enumerated and discussed by many researchers (e.g., Brown & Hudson, 1998; Butler, & Lee,

2010; Dłaska & Krekeler, 2008; Genesee & Upshur, 1996; Lombardi, 2008; Oscarson, 1989; Tsagari, 2004). Considering, to our knowledge, that no study has tried to practice alternative assessment procedures in the instruction of pragmatics, the results of the present study reveal that these procedures are beneficial to students not only with their linguistic knowledge (as was shown in previous studies), but also with their pragmatic knowledge. Although pragmatic knowledge is not limited to knowledge of speech acts and politeness, the results reveal that after the treatment students were able to use both speech act strategies used by native speakers and politeness markers.

Furthermore, the results of this study confirm those of the other researchers working on self-assessment and speaking of the beneficial effects of this method of assessment (Bachman & Palmer, 1998; Blanch & Merino, 1989; Butler & Lee, 2010; Janssen van-Dieten, 1989; Matsuno, 2009; Oscarson, 1989; Ross, 1998). While previous researchers worked on linguistic ability, the present study had its focus on the pragmatic one.

However, findings of the qualitative analysis showed that students used the grammatical criteria for rating in initial treatment sessions. The reason could be they did not have any experience of pragmatics being taught to them and what was important for them all through their courses at school was producing grammatically correct sentences. At first, students were really obsessed with their sentence grammar. They asked questions about the grammatical correctness of their responses. Since grammar and pragmatics are interrelated, their questions were answered by the instructor. However, the instructor tried to focus their attention on pragmatics and emphasize that language was not limited to grammar. After the instructor explained speech act strategies and contextual factors relevant to different situations, which lent themselves to various politeness markers, students tended to focus more on the

The Effect of Self-Assessment and Conference on EFL Students'...

sociopragmatic features and the relationships with the addressee in the scenario, the imposition of that situation and how much they needed to express themselves in a polite manner. By changing their criteria in self-ratings, students understood the significance of other criteria than grammatical ones. They explicitly stated their pragmatic reasons for ratings they gave to their answers during the last sessions of treatment.

A factor observed by the instructor during teaching in the self-assessment class was the feeling of inconvenience students had in the class. Students in the study had not experienced self-assessment before in other classes. After they wrote their answers to each situation, they had to encounter the task of rating their own performance which was really cumbersome and difficult for them as they did not feel confident about their responses. The reason for their lack of confidence could be they had always been given a score by teachers before. Although the instructor wrote a suitable response to each scenario, considering the strategies, contextual factors, and politeness after students finished their answers, they felt really insecure when they understood the teacher was not going to play any part in the ratings. They tended to give themselves a lower-than-expected rating during the first sessions although they used all the strategies they were taught. To the end of the treatment sessions, however, and after students compared their responses with those of the instructor every session, they tended to realize how they could use the criteria for rating. Therefore, their ratings were higher in the last sessions.

6. Conclusion

To sum up, the two alternative approaches (self-assessment and conference) used in the present study were found to be advantageous to students as far as their production of the speech acts and politeness markers were concerned.

Students in the conference and self-assessment groups were able to use the native speaker strategies and politeness markers in their productions of the WDCTs. Therefore, the present study might be a proof to the benefits of alternative assessment approaches (self-assessment and conference) in improving students' pragmatic production of speech acts and politeness markers. Alternative assessment approaches may, therefore, help students produce speech act strategies and politeness markers in a process-oriented approach to pragmatic production.

However, as mentioned earlier in the present article, pragmatics is not limited to speech acts and politeness. Researchers interested in this area may prefer to work on other aspects of pragmatics, such as implicatures, discourse markers, phatic expressions, and turn-taking and examine the effects of alternative assessment approaches on these aspects. Moreover, other alternative approaches like portfolios, journal writing, or peer-assessment could be used to see what effects these procedures can have on the pragmatic performance of students. Students' production of pragmatics was focused on in the present study. Studies with a focus on the comprehension of pragmatics can be conducted using MC tests of pragmatics. Researchers in the present study prepared and validated a test of pragmatics. Tests can also be made which consist of both pragmatic and linguistic components.

The present study was only an attempt to help interested teachers who feel a need for a change in their assessment methods. As was revealed in the study, the two alternative assessment procedures were beneficial to students' learning of the speech acts and politeness markers. However, more research needs to be carried out to consolidate the results of the present study and to apply these approaches in the classrooms for pragmatic instruction. Meanwhile, teachers can work on the pragmatic knowledge of their students at the same time they

The Effect of Self-Assessment and Conference on EFL Students'...

try to improve their linguistic knowledge since language is not limited to linguistic ability.

One of the limitations of the present study could be the use of native speaker norms for instruction and assessment while English is being considered as an international language. It was believed that the use of native speaker norm made rating students' responses more straightforward.

References

- Alderson, J. C., & Banerjee, J. (2001). Language testing and assessment (Part 1). *Language Teaching*, 34(4), 213-236.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. (2004). *Statistical analyses for language assessment*. Cambridge, UK: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6(1), 14-29.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bitchener, J., & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37, 322-329.
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3), 313-340.
- Brown, J. D., & Hudson, Th. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-674.
- Butler, Y. G., & Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1), 5-31.

- Dufon, M. A. (2008). Language socialization theory and the acquisition of pragmatics in the foreign language classroom. In E. Alcon-Soler & A. Martinez-Flor (Eds.), *Investigating pragmatics in foreign language learning, teaching and testing* (pp. 25-44). Multilingual Matters: USA.
- Genesee, F., & Upshur, J. A. (1996). *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Grabowski, K. C. (2007). Reconsidering the measurement of pragmatic knowledge using a reciprocal written task format. *Teachers College, Columbia University Working Papers in TESOL & Applied Linguistics*, 7, 1-48.
- Janssen van-Dietten, A. M. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6(1), 30-46.
- Jung, J. Y. (2002). Issues in acquisitional pragmatics. *Working Papers in TESOL & Applied Linguistics*, 2, 1-13.
- Kasper, G., & Roever, C. (2005) Pragmatics in second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 317-334). Mahwah, NJ: Lawrence Erlbaum.
- Kasper, G., & Rose, K. (2001). Pragmatics in language teaching. In G. Kasper & K. Rose (Eds.), *Pragmatics in language teaching* (pp. 1-9). Cambridge, UK: Cambridge University Press.
- Lachat, M. A., & Spruce, M. (1998). *Assessment reform, equity, and English language learners: An annotated bibliography*. New York: The Education Alliance.
- Little, D. (2005). The common European framework and the European language portfolio: Involving learners and their judgments in the assessment process. *Language Testing*, 22(3), 321-336.
- Liu, J. (2006). Assessing EFL learners' interlanguage pragmatic knowledge: Implications for testers and teachers. *Reflections on English language teaching*, 5, 1-22.

The Effect of Self-Assessment and Conference on EFL Students'...

- Lombardi, M. M. (2008). *Making the grade: The role of assessment in authentic learning*. Colorado, US: EDUCAUSE.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100.
- McNamara, T. (2001). Language assessment as social practice: Challenges for research. *Language Testing*, 18(3), 333-349.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6(1), 1-13.
- Oscarson, M., & Apelgren, B. M. (2011). Mapping language teachers' conceptions of student assessment procedures in relation to grading: A two-stage empirical inquiry. *System*, 39, 2-16.
- Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research and Evaluation*, 13(4), 1-11.
- Roever, C. (2006). Validation of a web-based test of ESL pragmalinguistics. *Language Testing*, 23(2), 229-256.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1), 1-20.
- Tada, M. (2005). *Assessment of EFL pragmatic production and perception using video prompts* (Unpublished doctoral dissertation). Temple University, Philadelphia, PA.
- Tsagari, D. (2004). Is the life beyond language testing? An introduction to alternative language assessment. *CRILE working papers*, 58, 2-23.
- Walters, F. S. (2004). *An application of conversation analysis to the development of a test of second-language pragmatic competence* (Unpublished doctoral dissertation). Illinois University, Urbana, Illinois.
- Watts, R. J. (2003). *Key topics in sociolinguistics: Politeness*. Cambridge: Cambridge University Press.

Yamashita, S. (2008). Investigating interlanguage pragmatic ability: What are we testing? In E. A. Soler & A. Martinez-Flor (Eds.), *Investigating pragmatics in foreign language learning teaching and testing* (pp. 201-223). Bristol: Multilingual Matters.

