Diagnostic Assessment of Interactional Competence in Paired Speaking Tests: Investigating Rating Accuracy of Iranian EFL Learners

Masoome Azmoode Sis Abad[1], Gholam-Reza Kiany[2*], Gholam-Reza Abassian[3]

[1]P Ph.D. Candidate, Department of English, Islamic Azad University, Science and Research Branch, Tehran, Iran

[2*]Associate Professor, Department of English, Tarbiat Modarres University, Tehran, Iran

[3]Associate Professor, Department of English, Imam Ali University, Tehran, Iran

Abstract

Assessment of interactional competence (IC) as a multi-componential construct poses several challenges for both teachers and administrators. Having a diagnostic assessment perspective through stakeholders' involvement, this mixed methods study attempts to provide information about students' strengths and weaknesses in IC. The paper first explored the distinctive effects of diagnostic self- and peer-assessments on the development of Iranian EFL learners' IC; it then examined the accuracy of the learners' diagnostic assessment of IC in paired speaking tests. The learners' perception toward the application of diagnostic assessment was also investigated qualitatively. To this end, 60 students majoring in English translation at Islamic Azad University participated in this study. Taking the instructor's ratings as the criterion, over the course of 12 weeks, the accuracy of the learners' diagnostic self- and peer- assessments was investigated. Data analysis, using T-test and MANOVA, confirmed that while the two groups of diagnostic self- and peer- assessments had considerable improvement in IC, there was not any significant difference between the two groups' gain. In addition, no statistically significant difference was found between the accuracy of diagnostic self-, peer – assessment and those of instructor- assessment throughout the course. Furthermore, the results showed that the participants held favorable perception toward the application of diagnostic assessment.

Keywords: Assessment stakeholders; Diagnostic assessment; Interactional competence; Paired speaking tests

INTRODUCTION

IC as an increasingly influential theoretical construct is the focus of research inquiry into the social dimensions of second language teaching, learning, and assessment. IC was first coined by Kramsch (1986) who criticized the proficiency tests for stressing the static content structure of language at the expense of the dynamic aspect. Since then, speaking assessment researchers have tried to offer definition of IC either through delineating features important to raters or features distinguishing different proficiency levels (Lam, 2019). Recently a macro definition of IC was proposed by Galaczi and Taylor (2018) who define this construct as the speakers' capability to co-construct interaction in a meaningful and purposeful way, taking into account sociocultural and pragmatic aspects of the speech situation and event. In fact one certain element that seems to be central to any account of IC is its non-monologic nature which involves the co-construction of meaning through discursive practices (Young, 2000).

*Corresponding Author's Email:
kiany_gh@modares.ac.ir

Although the number of studies on IC has been expanded in recent years, this construct has not been well-theorized yet, hence its theoretical aspect as well as practical implementation need to be more developed to inform teaching, learning and assessment of IC in a learner-friendly and more comprehensive way (May, Nakatsuhara, Lam, & Galaczi, 2019). For the assessment of IC, for instance, there exist a number of challenges which have made it a critical issue. One important gap made by the proponents of IC is that earlier models of L2 competence overlooked a salient feature of interactional ability; namely, the microskills which let interlocutors engage in interactions (Dings, 2014). The importance of IC microskills, for both learning and assessment of this construct, is reflected in recent definition and operationalization of IC as well (Galaczi & Taylor, 2018). In practice also the interactional abilities of EFL learners are highly depended on their IC microskills' mastery level, thus in instructional settings particular attention is needed to detect the nature and causes of the pupils' weaknesses in IC microskills. Perhaps one of the best ways to strengthen the assessment practices of IC and capture its micro-level features is through diagnostic assessment because it provides detailed and fine grained information on strengths and weaknesses of learners in interaction. In fact, it assesses what the learners already know and the nature of their learning weaknesses, which if undiagnosed may limit their learning outcomes (Terwase & Obbadare-Akpata, 2018). It is worth noting that such diagnostic information can be of great help to both learners and instructors because once learning problematic areas are identified, instructors can make essential instructional adjustments and plan for subsequent remedial learning.

Despite the importance of diagnostic assessment, especially in the domain of second and foreign language (SFL) testing and assessment, still this field is underdeveloped and almost poorly theorized; consequently there are several potential problems for the implementation. Firstly, in language learning contexts the number of real diagnostic tests of foreign language proficiency is scarce. This condition is more critical for productive skills since the existing diagnostic tests (like DIALANG) use a computer-scored system; therefore due to the nature of such skills, it is virtually impossible to assign pre-programmed, computer- based scores for productive skills (Alderson, 2005). Secondly, when students take tests it usually takes time to receive diagnostic reports, thus the diagnostic feedback lacks immediate relevance. Therefore, designing purpose-built diagnostic tests and checklists and conducting ongoing, formative diagnostic assessment a long with stakeholders' involvements can assist the implementation of diagnostic assessment in instructional contexts. Obviously engaging learners in the act of diagnosis would empower them to gain a clear picture of their own strengths and weaknesses in a skill, so that they can learn more effectively. This is mainly in accordance with one principle of diagnostic assessment which lays stress upon benefitting from diverse stakeholders' views in diagnostic decisions for it can offer a better insight into particular learning difficulties (Alderson, Brunfaut & Harding, 2014). But in EFL contexts, not adequate research attention has been directed toward the diagnostic assessment of language skills in general and to the stakeholders' involvement in diagnosis in particular.

Exploring the past research in the area of speaking assessment shows that while there is a growing interest in examining different aspects of IC (e.g. May, 2009, 2011; Nakatsuhara, 2013; Roever & Kasper, 2018), diagnostic assessment of learners' strengths and weaknesses in IC micro-level has been almost overlooked. The lack of research in bridging theoretical and descriptive discussions on IC could be due to the researchers' ignorance of the practical usefulness of this construct. In the context of English language teaching in Iran also, little or almost no research has been done with the purpose of conducting diagnostic assessment of IC through learner involvement. Undoubtedly, such an in-depth treatment of IC can contribute to language teaching and learning through raising instructors 'and students' awareness of its micro-level features, and to the assessment practices

through developing a wider recognition of IC (May et al. 2019).

REVIEW OF LITERATURE
IC Assessment
The present interest in IC originates from early criticisms of the Chomskyian dichotomy between language competence and performance (Chomsky, 1965). Based on the subsequent research in this area, later the concept of communicative competence has been developed by the researchers (e.g., Canale & Swain, 1980). However in the 1990s, the premises of communicative competence were criticized for representing a static and monocentric perspective of language competence (Skogmyr Marian & Balaman, 2018). Young (2013), for instance, problematize the exclusive focus on a single individual's contribution to interaction. Hence a new line of research has been developed advocating the idea that actions, activities and abilities are "jointly" constructed by "all" participants in interaction. Earlier, this constructivist view of competence and interaction had been addressed by kramsch (1986) as "interactional competence".

The research in L2 learning and assessment has extended substantial but quite different insights concerning IC features. In this regard, Young (2000) identified four characteristics for IC; first it is related to language used in particular discursive practices rather than to the language user's ability, independent of context. Second, IC is identified by the co-construction of discursive practices by *all* participants in interaction .Third the IC theory defines a set of interactional resources which speakers use in particular ways to co-construct meaning. And fourth refers to the significance of identifying resources brought to the interaction. Basically the assumptions that communication is co-constructed and context-dependent are central in IC conceptualization (Young, 2000) but these two features pose certain challenges for the assessment and test interpretation of IC (Borger, 2019), thus they have been much debated so far laying a dual perspective: while some are in favor of awarding shared scores to speakers' contribution in acknowledgement of the co-constructed nature

of IC (e.g., May, 2009, 2011) others have stressed the significance of disentangling individual contributions in spoken interaction (e.g., Nakatsuhara, 2013).

Another salient issue associated to IC assessment is the test formats. As Galaczi and Taylor (2018) noted, in recent years for the assessment of oral skills, paired and group speaking test formats were favored over individual test format. This is partly because of the shift toward a more communicative approach in language teaching and learning and partly due to the limitations of individual oral test formats. May (2011) refers to the advantage of paired speaking test formats asserting that through co-construction of discourse the interlocutors have the chance to show a broad spectrum of interactional competencies with a partner rather than with an examiner. Therefore, the paired/group test format was perceived as a viable alternative to the individual oral test format which provides more equal conversational rights and responsibilities in interaction (Galaczi & Taylor, 2018).

Application of Diagnostic Assessment to IC Assessment
Although in the last few years, the need for conducting research on IC was noticed by the researchers (eg., Lam, 2019; Ross, 2018; Tecedor, 2016; Youn, 2019) still its practical operationalization has not been much acknowledged in language learning contexts. It is noteworthy that in recent years particular research attention has been given to the interactional micro-skills in IC research. Among the researchers, Nakatsuhara et al. (2018) define five broad domains of interactional skills namely, initiating new idea, keeping the discussion go over several turns, negotiating for an outcome and using body language properly. Such interactional micro-skills which were drawn from the existing theoretical and empirical IC research would have certain implications for the learning and assessment of this construct. Even though, exploring the previous research reveals that studies on IC assessment with a focus on its microskills (eg., May et al. 2019; Nakatsuhara et al, 2018) are

less than sufficient. To fill such research gap in the literature, the IC micro-level features can be assessed diagnostically to provide detailed and fine grained information about the quality of learners' interaction.

In particular, through diagnostic assessment the learners' areas of strengths and weaknesses in language skills are identified in order to help improve learning outcome (Terwase & Obbadare-Akpata, 2018; Lee, 2015; Lee & Sawaki, 2009). Providing diagnostic feedback is an important element in performing diagnostic language assessment because once the strengths and weaknesses are identified in a skill, the related diagnostic feedback can help learners and instructors take necessary actions for eliminating the identified weaknesses (Lee, 2015). Indeed, the kind of feedback offered to assessment stakeholders, especially to students and instructors, has become a dominant issue in diagnostic assessment discussions (Poehner, Zhang & Lu, 2014). Similarly, in the area of IC, providing learners with necessary feedback on the interactional skills is very crucial too, (May et al. 2019).

Nevertheless, until recently the domain of language testing and assessment lacked a comprehensive theory of diagnostic assessment (Alderson, 2005); however later Alderson et al. (2014) proposed a tentative theoretical framework comprising a set of principles for implementing diagnostic assessment. Reviewing the literature on speaking assessment reveals that so far most of the previous studies have mainly dealt with diagnostic assessment of speaking ability in general (Kazemi & Tavassoli, 2020; Tozcu, 2016) which despite their significance, these inquiries fail to diagnose and assess the learners' strengths and weaknesses in interactional microskills. Furthermore, in the design of past diagnostic studies, the viewpoints of assessment stakeholders were rarely taken into consideration. Despite the importance of learner involvement in the process of diagnostic assessment (Harding *et al.* 2015), this area is almost under-researched. But the need for students' involvement in language learning and assessment have been widely acknowledged in higher education, asserting that the alternative means of assessments

namely, self-and peer- assessment foster students' autonomy and facilitate self-directed learning. In this regard self- and peer -assessment of oral performance have long been examined empirically (Cheng & Warren, 2005; Han & Fan, 2019; Lee & Chang, 2005; Lu, 2018) providing a large body of research which casts more light on its various potential aspects. However, little or almost no research attention has been made to the diagnostic assessment of IC through learners' involvement. To this end, the following questions were raised:

RQ1. Is there any statistically significant difference between Iranian EFL learners' diagnostic self-and peer assessment groups in the development of IC in paired speaking tests?

RQ2. Is there any statistically significant difference between the accuracy of self-, peer- and instructor ratings in diagnostic assessment of Iranian EFL learners' IC in paired speaking tests?

RQ3. How do Iranian EFL learners perceive diagnostic assessment of IC?

METHOD
Participants
Sixty Iranian EFL learners aged between 18 and 23 took part in this study. The participants were 16 male and 44 female students majoring in English translation at Islamic Azad University, Shahr Qods Branch. The participants were selected based on their performance on the Oxford Placement test (OPT) in two intact classes. In fact, randomizing the subjects was not practically feasible for the administration norms of the university. The participants were also arbitrarily assigned into two groups of diagnostic self- and peer-assessment, each containing 30 students.

Instruments
English Language Proficiency Test
The Oxford Proficiency Test (OPT) version 1.1 UCLES (2001) was administered to check the homogeneity of the participants. Based on the results, the participants' general English proficiency was approximately at intermediate level.

IC diagnostic checklist

The IC checklist developed by Nakatsuhara et al. (2018) was used for the purpose of diagnostic assessment. The checklist contains a number of IC micro-level features for tapping into the participants' ability in initiating new idea, keeping the discussion go over several turns, negotiating for an outcome and using body language appropriately. The checklist components each consist of a number of sub components. The adaptation was made for the performance level, and it holistically assessed through a four-point Likert scale ranging from 1 to 4 (1 = Poor, 2 = Fair, 3 = Good, and 4 = Excellent). The checklist was also adapted to provide the essential description of each ability level in order to be easy -referent for diagnostic assessment purposes.

IC pretest and posttest

The IC pretest and posttest were used to check the subjects' level of IC performance prior and after the instructional treatment. Thus, in both groups the participants were assigned to form paired groups and talk about a given topic for about ten minutes. The performance of each pair was audio-video recorded simultaneously for instructor and rater's scoring and judgment.

Open-ended questionnaire

The participants' perception of implementing diagnostic assessment was determined via an open-ended questionnaire. The first draft included a number of questions addressing the learners' overall perception of the course including the benefits, drawbacks and difficulties that they experienced in performing diagnostic assessment as well as their perceived effectiveness of the course. The questionnaire was evaluated in terms of content relevance and representativeness of the items by two expert teachers who made the necessary revisions and assured the appropriateness of the questionnaire. The proofed draft was ultimately used for eliciting the intended responses from the students.

Semi-structured focus group interview

To find an in-depth account of the learners' perception and cross-validate the obtained qualitative data a semi-structured focus group interview was conducted with 20 students who were selected randomly in five groups. During the interview a number of questions, similar to open-ended questions, were asked to obtain the students' attitudes about the intended course. The interview of the learners was audio-recorded for subsequent analyses.

Procedures

Since the present study was a mixed-methods research with an embedded design, both quantitative and qualitative approaches were applied. In this study, data collection took 15 weeks in regularly scheduled class period. At the outset of the study, the OPT was administered to all the participants in order to ensure homogeneity of the students in terms of English language proficiency. Thus, 60 students whose scores were one standard deviation above or below the population mean were selected. Following that, the participants were pretested on IC in paired speaking tests, so that they were asked to talk about a given topic for about ten minutes. Before the instructional treatment the participants received training to use the IC checklist for the purpose of diagnostic self- and peer- assessment. To this end, the instructor introduced the diagnostic assessment principles and criteria of IC assessment to all the subjects. Furthermore, she discussed the potential benefits of self and peer assessment and introduced the related techniques. Subsequently the participants received training to do diagnostic self- and peer- assessment of IC. In so doing, they were provided with the IC checklist based on which the components as well as the descriptors of each performance level were elaborated on; so that the IC micro-level features, including the participants' ability in initiating new idea, keeping the discussion go over several turns, negotiating for an outcome and using body language appropriately, were discussed thoroughly in both groups. Moreover, the instructor modeled rating the interaction of some students based on the diagnostic assessment checklist; she also asked the participants to rate some sample performance as well. Then, she randomly selected some rated diagnostic checklists and

displayed them on a video projector to analyze and comment on the students' ratings in order to establish the criteria of IC diagnostic assessment. Being familiar with the checklist, the students in the self-assessment group practiced diagnostic assessment of their own performance; and in the other group, the learners assessed their peers' IC performance diagnostically.

After receiving two training sessions, during the treatment process, every session the participants were provided with a topic, about which they were required to talk in pairs for about ten minutes. Before that, the instructor asked some questions to be discussed in the class about a given topic; so that the students could collect all their ideas on that specific topic. For maximum interaction the participants were encouraged to share different points of view while observing the criteria of IC. After brainstorming the students discussed the topic in pairs. Then, based on the IC diagnostic checklist the students were required to diagnose and assess their own IC strengths and weaknesses in self-assessment group and that of their peers' in peer assessment group. To check the subjects' developmental growth as well as the accuracy of their diagnostic assessment ratings, every session the performance of each pair was audio-video recorded for subse-

quent instructor- diagnostic assessment. Thus, the participants received the teacher diagnostic assessment as well as related diagnostic feedback on every performance in a weekly basis; thus they could ponder upon any possible evaluative mismatches between their own assessment and that of the instructor for further improvement. To provide opportunities for remedial learning, every session the instruction was tailored according to the overall diagnosis of the learners' strengths and weaknesses in IC microskills. In fact based on the learners' IC performance, the instructor pointed to the most challenging IC microskills and helped them overcome their weaknesses. Therefore, the participants received remedial assistance in the problematic areas throughout the course.

After the fulfillment of the instructional treatments, the participants took the IC posttest in both groups. To achieve more reliable results, an external rater, who was an experienced university instructor was asked to rate the performance of the participants at the pretest and posttest. Having received the twelve week-long intervention, the participants filled out the diagnostic assessment open ended questionnaire in both classes. Finally, the instructor randomly invited five groups of students to take part in the focus group interview.

Table 1
Inter-Rater Reliability Coefficients for IC Pretest and Posttest

Pearson Correlations		**Pre-Rater2**	**Post-Rater2**
Pre-Rater1	Pearson Correlation	.787**	
	Sig. (2-tailed)	.000	
	N	60	
Post-Rater1	Pearson Correlation		.833**
	Sig. (2-tailed)		.000
	N		60

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

RSULTS
Quantitative data analysis
Before testing the null hypotheses, the assumption for normality of scores in both groups were checked. Based on Kolmogorov-Smirnov and Shapiro-Wilk tests results ($p < .05$) the scores were normally distributed. Furthermore, the results of Levene's test (sig. = .30) confirmed that the assumption of equality of variances was not violated.

Testing the first null hypothesis
The first research question intended to probe learners' IC development through diagnostic self and peer –assessment in paired speaking tests. The following null hypothesis was formulated accordingly:

H01: There is not any statistically significant difference between Iranian EFL learners' diagnostic self- and peer- assessment groups in the development of IC in paired speaking tests.

In order to address this research hypothesis, the IC post-test scores were used to run an independent samples t-test. The results of the independent samples t-test indicates (Table 2) that there was no significant difference between the diagnostic self-assessment group (M= 17.36, SD= 1.70) and the diagnostic peer-assessment group (M=17.41, SD= 1.56; t (29) = -.135, p=.893, two-tailed) in terms of their performance in the IC post-test. The effect size of the differences in the means was also very small (eta squared = .0003).

Table 2
Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Post Speak	Equal variances assumed	1.091	.301	-.135	58	.893	-.05733	.42327	-.90460	.78993
	Equal variances not assumed			-.135	57.598	.893	-.05733	.42327	-.90473	.79006

Testing the second null hypothesis

To probe the second research question, any probable significant difference between the accuracy of the diagnostic self-, peer- and instructor- assessments of IC was explored during 12 instructional sessions. So the following null hypothesis was formulated:

H02: There is not any statistically significant difference between the accuracy of self-, peer- and instructor-ratings in diagnostic assessment of Iranian EFL learners' IC in paired speaking tests.

To understand the difference, a multivariate analysis of variance (MANOVA) was run on the cumulative data of all sessions in all IC microskills. The data consisted of 1800 different numbers for each of the three groups making a total of 5400 marks given to students' performance by themselves, their peers, and the instructor during 12 sessions in 5 different IC microskills. Table 3 shows descriptive statistics for the data.

Table 3
Descriptive Statistics

Groups	Mean	Std. Deviation	N
Teacher score	3.4339	.52919	1800
Self score	3.5390	.51361	1800
Peer score	3.2819	.55601	1800

Sample size, equality of covariance matrices, and equality of error variance were checked to make sure that no assumption has been violated. Based on Tabachnick and Fidell (2013, p. 253), a sample size of at least 20 in each cell should ensure 'robustness'. The number of cases in each cell is provided as part of the MANOVA output. In this case, there are many more than the required number of cases per cell. As Box's Test of Equality of Covariance Matrices indicates (Table 4) the data does not violate the assumption of homogeneity of variance-covariance matrices since the obtained significance value is larger than .001 (Sig=.291).

Table 4
Box's Test of Equality of Covariance Matrices

Box's M	41.226
F	1.116
df1	36
df2	24661.533
Sig.	.291

Levene's test of equality of error variances was also used to make sure that assumption of equality of variances for variables is not violated. As Table 5 shows, none of the variables indicated significant values; thus the assumption was not violated.

Table 5
Levene's Test of Equality of Error Variances

		Levene Statistic	df1	df2	Sig.
Self score	Based on Mean	1.838	12	1784	.083
	Based on Median	1.217	12	1784	.265
	Based on Median and with adjusted df	1.217	12	1203.276	.265
	Based on trimmed mean	1.740	12	1784	.053
Peer score	Based on Mean	.945	12	1784	.500
	Based on Median	.960	12	1784	.485
	Based on Median and with adjusted df	.960	12	1754.726	.485
	Based on trimmed mean	.935	12	1784	.510

Table 6
Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Group	Pillai's Trace	.019	1.120	30.000	3568.000	.298	.009
	Wilks' Lambda	.981	1.120[b]	30.000	3566.000	.298	.009
	Hotelling's Trace	.019	1.120	30.000	3564.000	.299	.009
	Roy's Largest Root	.012	1.457[c]	15.000	1784.000	.113	.012

The results of the MANOVA test are indicated in Table 6 ($F= 1.12$; Wilks' Lambda = .981, $p= .298$, partial eta squared = .009), according which there is no significant difference between the scorings of the two groups and those of instructor- assessments.

Furthermore, test of between-subject effect in Table 6 shows that there is no significant difference in the two groups compared with the instructors' scores since the obtained significance values are less than .025. As discussed by Tabachnick & Fidell (2016), to interpret the result of Between-subject effects, the alpha level should be adjusted to avoid Type I error. Therefore, alpha level is divided by the number of dependent variables, which is called Bonferroni adjustment. Here, there are two dependent variables; therefore we divide .05 by two which makes .025. The obtained values, as shown in Table 7 are .59 and .14 for diagnostic self-scores and peer-scores respectively. Thus, there is no significant difference between the two groups in comparison with instructor's scores.

Table 7
Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Group	Self score	3.465	15	.231	.875	.593	.007
	Peer score	6.393	15	.426	1.383	.147	.011

In addition, to make sure that the overall picture we obtained from the data was in line with each one of the 12 sessions, the results obtained from the last session was analyzed. The results of MANOVA for the last session's data, as indicated in Table 8, shows that there was no significant difference between diagnostic self-, peer- and instructor-assessment scores in five subskills, too. As the table 7 indicates, the obtained significance values for the subskills are .68, 1.13, 1.26, .83, and 1.38 which are all above .025, thus statistically insignificant.

Table 8
Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Initiatenewidea12	Initiate new idea 12	.719	5	.144	.688	.635	.061
	Idea development 12	1.463	5	.293	1.136	.353	.097
	Collaboration 12	.850	5	.170	1.267	.292	.107
	Negotiate for an outcom12	1.193	5	.239	.831	.533	.073
	Body language use 12	1.951	5	.390	1.385	.245	.116

Qualitative data analysis

The third research question was concerned with the participants' perception of their involvement in the diagnostic assessment of IC. The qualitative data collection was conducted through a semi-structured focus group interview and an open-ended questionnaire. To analyze the data, first the most frequent responses given to the open-ended questionnaire were categorized then the interview responses were transcribed, coded and finally analyzed using content analysis. The content analysis of the learners' responses revealed three general themes: Outcomes 2) Benefits 2) Drawbacks

Outcome

The first dimension demonstrates the learners' perceived outcomes in the process of diagnostic assessment. Two themes appeared under this dimension: improvement of IC features and finding a diagnostic perspective in spoken interaction. The majority of the students (88%) reported an improvement in IC at the end of the course. According to the responses, the learners' interactional abilities improved considerably through the application of diagnostic assessment. In addition, many students (76%) believed that they found a diagnostic vision in their interaction and they learned how to identify their strengths and weaknesses in IC by the end of the course. Some extracted responses are:

"I have less difficulty in conversation, now." I don't ignore the important features of interaction anymore."; "I can speak more effectively in discussion." said one student,

"Before I wasn't aware of my weaknesses in interaction, but now I can almost identify them." said another student.

Benefits

This dimension indicates the participants' perceived benefits in the process of diagnostic assessment. This dimension was classified into three themes: Criteria awareness, Reflection and Finding motivation. 86% of the participants found that diagnostic assessment assisted them to consider various IC criteria. One student commented: "I learned to focus on the quality of spoken interaction. I paid less attention to the IC features, but now I try to consider the important elements of interaction." Furthermore, many students (75%) asserted that performing diagnostic assessment encouraged them to reflect on interactional micro-skills more critically to recognize their strengths and weaknesses. Moreover, some students (42%) were inspired by the course outcomes; they were motivated to employ the criteria of diagnostic assessment to progress in other language skills, as well. A few excerpts are: "Throughout the course diagnostic assessment helped me think about my weaknesses in interaction." "That was so interesting when I learned some important aspects of interaction." "I really enjoyed when I could evaluate my performance based on IC checklist".

Drawbacks

This dimension indicates difficulties or drawbacks that the students perceived through performing diagnostic assessment. This dimension was

classified into three themes: Difficulty, Imprecision and Being time-consuming .While the criteria of diagnostic assessment were well clarified to the participants, majority of them (90%) found it difficult to perform it, particularly at the beginning of the course. In response to the question if they had any problems in applying diagnostic assessment one student commented: "At first it was almost difficult to diagnose the problematic areas. Sometimes rating my peer's performance was a bit challenging."

One of the specific problems associated with diagnostic assessment through learner involvement was that, some students (43%) felt that they were inaccurate in this process; specifically, they expressed some degree of uncertainty and imprecision in performing diagnostic assessment. They thought they might award themselves or their peers with higher or lower expected scores. In addition, some of the students (51%) asserted that completing the diagnostic checklists was so time-consuming. Some extracted responses are: "At first I was not sure enough about the exact performance level in interaction" and "Once my classmate underestimated my performance." "It was time-consuming to fill the diagnostic checklists, particularly at the beginning of the course."

DISCUSSION

Taking a diagnostic approach, this study accounted the stakeholders' involvement in the assessment of IC and proved that while the learners in both diagnostic self- and peer-assessment groups had considerable improvement in the development of IC, no significant difference was found between the two groups' gain in promoting their interactional skills. There could be two plausible explanations for this result. The first one is based on the fact that the application of diagnostic assessment via self- and peer assessment along with instructor- assessment throughout the course helped the learners find awareness regarding their strengths and weaknesses in IC micro-skills and eventually their performance in paired speaking tests were improved. The second explanation, which was also supported qualitatively in the present research, is that when

the learners were informed about the IC features and the required diagnostic assessment criteria, they found better insights and awareness concerning the interactional micro-skills, thus they have achieved a developmental growth in IC throughout the course. Previous research (Kazemi & Tavassoli, 2020; Tozcu, 2016) also confirmed the effectiveness of diagnostic assessment on improving learners' speaking ability. Tozcu (2016), for instance, explored the role of diagnostic assessment in learners' oral proficiency for narrating past events. The research findings revealed that learners who took part in the diagnostic assessment interview and received an individualized learning plan had noticeable improvement in basic sentence structures.

Concerning the assessment design, in this study individual scores were given to each candidate since it has been suggested that in the assessment of IC joint scores rather than individual scores would be unfair (Nakatsuhara, 2013). This argument, which stems from the co-constructed nature of IC, is one the most debated issues in IC research and in spite of unanimous consensus in this regard, still researchers are challenged on how to deal with it (Lam, 2018).

Furthermore taking the instructors' ratings as the criterion, the rating accuracy of the learners' diagnostic assessment was investigated during the whole instructional sessions. The results of the MANOVA on the cumulative data of 12 sessions in all micro-skills as well as the analysis of the results from the last session showed no significant difference between the scorings of the two groups and those of the instructor's. It can be concluded that involving learners in the process of diagnostic assessment empower them with the necessary insights regarding the nature and causes of their weaknesses in IC micro-skills so that their interactional skills a long with the accuracy of their assessment improved substantially over time. The results of the current research also show that diagnostic peer-assessment was closer to the instructor- assessment. This finding is quite compatible with a previous study done by Hirai and Yokouchi (2019) who explored the capabilities of EFL students' diagnostic assessment for speaking

test. Focusing on the quality of peer-assessment, this study revealed that students' peer assessments were as accurate as the instructor's on non-linguistic features but far less so in the assessment of linguistic aspects. Among the very few studies which examined the role of learners' involvement in the assessment of IC, May et al. (2019) report on a project that was targeted to identify IC features in order to develop a checklist and feedback materials for the assessment of IC micro-skills. Similar to the current research, in May et al. study a diagnostic view of IC was implied since the researchers believe that if feedback is formulated and delivered efficiently, it can aid students to understand their strengths and weaknesses in interaction which ultimately would help them improve their interactional abilities.

While the previous research on IC assessment disregards both diagnostic assessment of learners' interaction and the role of stakeholders' involvement in the assessment of IC, in the design of many studies which investigated the learners 'oral performance through self- and peer- assessment (Chen, 2008; *De Grez,* Valcke, & Roozen, 2012; Han & Riazi, 2017; Ma & Winke, 2019) some diagnostic purposes are implied. For instance, in Chen's (2008) study, it was proved that self-assessment helped the learners recognize their strengths and weaknesses so that they eventually managed to analyze and eliminate their weaknesses. Moreover, concerning the rating accuracy of self- and peer- oral assessment, reliability and validity issues have been largely discussed in the related literature (eg; Han, 2018; Ma & Winke, 2019; Salehi & Sayyar Masoule, 2017); however some delicate points need to be taken into consideration. The first is that, measurement accuracy itself is not the only relevant issue in learners' assessment (Boud & Falchikov, 1989) unless the main purpose is to apply learners' given scores in formal assessment; second, the inaccuracy or imprecision in students' assessment do not necessarily invalidate or diminish the pedagogical and educational benefits of such assessment practices (Han & Riazi, 2017).

The results also indicated that the participants had a positive perception about the application of diagnostic assessment. The findings highlight the evidence for employing diagnostic assessment in EFL settings specifically by offering students supportive condition to apply the criteria of diagnostic assessment through learner involvement. Being in the same vein, Jang (2005) in her mixed methods study asserted that, it is of paramount importance to engage teachers and students in evaluating the effects of diagnostic approach and use because the ultimate goal of diagnosis would be 'change' in actions and perceptions which mainly leads to improvement in targeted areas. The results of this study are quite in line with Jang, Dunlop, Park and van der Boom's (2015) findings, showing that the inclusion of self-assessment in diagnostic feedback promotes learners' critical reflection and subsequent learning planning. Their findings also pinpoint the fact that students' perceptions regarding their own learning orientations and ability highly affect the way they processed diagnostic feedback.

The results that students had a strong desire to be involved in the assessment of their own or peer's performance appear to validate Oscarson's (1989) assertion that learner participation in assessment enriches the assessment practices and this could also reinforce their learning autonomy and motivation. Although in this study the learners perceived some degree of difficulty and imprecision in performing diagnostic assessment they were still willing to be involved in this process, interestingly this adds more weight to the obtained results. One plausible reason as to why students' viewed this form of self- and peer- assessment as valuable could be associated with the fact that the subjects gained a diagnostic vision throughout the course, recognizing the significance of IC micro-skills in their interaction. If self- and peer- assessment is regarded as learning oriented assessment which entail both short- and long-term outcomes (Thomas, Martin & Pleasants, 2011), it is conceivable that the students have perceived wider learning implications of alternative assessment as well.

CONCLUSION
Having a diagnostic assessment perspective, this paper showed how students could rea-

sonably assess their own and their peers' IC performance and demonstrated developmental growth in this construct. The findings of this research showed that providing EFL students with the opportunity to perform diagnostic self- and peer- assessment of IC is an effective way to improve their interactional skills and minimize evaluative mismatches in their ratings. It further revealed areas in the learners' performance that called for improvement and increased their awareness regarding their strengths and weaknesses in IC. Moreover, students who participated in this study demonstrated willingness to engage in diagnostic self- and peer- assessment. It is worth noting that, while diagnostic assessment procedures were rarely practiced in Iranian EFL contexts, the learners performed quite well in this process, thus the findings of the present research can offer evidence for implementing diagnostic assessment of IC in EFL contexts which can help teachers and learners locate problematic areas in learners' interaction for providing appropriate remedial instruction.

The findings of the present research entail theoretical as well as practical implications. At the theoretical level, as far as the assessment of IC is concerned, there is scope to advance the theoretical definition of IC (Galaczi & Taylor, 2018). Therefore, with the specific focus on diagnostic assessment of IC, as an under-researched area, this research might be regarded as a pioneering step in applying diagnostic assessment through learner involvement. At the practical level, a crucial implication for instructors is to train students to engage in the ongoing process of diagnostic assessment which can ultimately help to recognize their strengths and weaknesses, monitor their learning and reinforce their autonomy over time. Another implication is for main EFL stakeholders such as policy makers, syllabus designers and materials developers to implement diagnostic assessment procedures in order to refine the language curricula and strengthen their educational practices.

Furthermore, major limitations of this research ought to be acknowledged. The sampling method (intact classes) which was used with Iranian English language learners in

the present study obviously bears some implications for generalizability. Furthermore, the variable of students' proficiency level was not considered in this research; therefore, the effect of proficiency level on the accuracy of students' diagnostic self- and peer- assessment requires more empirical exploration. The IC studies reviewed in this research also suggest a number of directions for further inquiries; the issue of individual versus joint scores, for instance, is a possible research area in the context of the paired speaking tests. Notably, the inherent complexity in the assessment of candidates' interaction is an area that warrants more research (Borger, 2019); therefore more inquiries can be done focusing on different variables in other learning contexts with larger sample size.

Finally, it is hoped that more research endeavor in the area of speaking diagnostic assessment will contribute to a better recognition of IC and create useful insights to the teaching, learning and assessment of this construct.

References

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency: The Interface Between Learning and Assessment* (illustrated, reprint ed.). Bloomsbury Academic.

Alderson, J. C., Brunfaut, T., & Harding, L. (2014). Towards a Theory of Diagnosis in Second and Foreign Language Assessment: Insights from Professional Practice Across Diverse Fields. *Applied Linguistics*, *36*(2), 236-260.

Borger, L. (2019). Assessing Interactional Skills in A Paired Speaking Test: Raters' Interpretation of The Construct. *Apples - Journal of Applied Language Studies*, *13*(1), 151-174.

Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, *I*(1), 1-47.

Chen, Y. M. (2008). Learning to Self-Assess Oral Performance In English: A Longitudinal Case Study. *Language Teaching Research*, *12*(2), 235-262.

Cheng, W., & Warren, M. (2005). Peer Assessment of Language Proficiency. *Language Testing*, *22*(1), 93-121.

Chomsky, N. (1965). *Aspects of the Theory of Syntax* (Vol. 11). MIT press.

De Grez, L., Valcke, M., & Roozen, I. (2012). How Effective Are Self- And Peer Assessment of Oral Presentation Skills Compared with Teachers' Assessments? *Active Learning in Higher Education*, *13*(2), 129-142.

Dings, A. (2014). Interactional Competence and the Development of Alignment Activity. *The Modern Language Journal*, *98*(3), 742-756.

Falchikov, N., & Boud, D. (1989). Student Self-Assessment in Higher Education: A Meta-Analysis. *Review of Educational Research*, *59*(4), 395-430.

Galaczi, E., & Taylor, L. (2018). Interactional Competence: Conceptualisations, Operationalisations, and Outstanding Questions. *Language Assessment Quarterly*, *15*(3), 219-236.

Han, C. (2018). A Longitudinal Quantitative Investigation Into The Concurrent Validity of Self and Peer Assessment Applied To English-Chinese Bi-Directional Interpretation in An Undergraduate Interpreting Course. *Studies in Educational Evaluation*, *58*, 187-196.

Han, C., & Fan, Q. (2020). Using Self-Assessment As A Formative Assessment Tool in An English-Chinese Interpreting Course: Student Views and Perceptions of Its Utility. *Perspectives*, *28*(1), 109-125.

Han, C., & Riazi, M. (2018). The Accuracy of Student Self-Assessments of English-Chinese Bidirectional Interpretation: A Longitudinal Quantitative Study. *Assessment & Evaluation in Higher Education*, *43*(3), 386-398.

Hirai, A., & Yokouchi, Y. (2019). An Investigation of EFL Learners' Diagnostic Assessment Capabilities for a Classroom-Based Speaking Test. *ARELE: Annual Review of English Language Education in Japan*, *30*, 209-224.

Jang, E. E. (2005). *A Validity Narrative: Effects of Reading Skills Diagnosis on Teaching and Learning in The Context of NG TOEFL*. University of Illinois at Urbana-Champaign.

Jang, E. E., Dunlop, M., Park, G., & van der Boom, E. H. (2015). How Do Young Students With Different Profiles of Reading Skill Mastery, Perceived Ability, and Goal Orientation Respond to Holistic Diagnostic Feedback? *Language Testing*, *32*(3), 359-383.

Kazemi, N., & Tavassoli, K. (2020). The Comparative Effect of Dynamic vs. Diagnostic Assessment on EFL Learners' Speaking ability. *Research in English Language Pedagogy*, *8*(2), 223-241.

Kramsch, C. (1986). From Language Proficiency to Interactional Competence. *The Modern Language Journal*, *70*(4), 366-372.

Lam, D. M. K. (2018). What Counts As "Responding"? Contingency on Previous Speaker Contribution As A Feature Of Interactional Competence. *Language Testing*, *35*(3), 377-401.

Lam, D. M. K. (2019). Interactional Competence with and without Extended Planning Time in a Group Oral Assessment. *Language Assessment Quarterly*, *16*(1), 1-20.

Lee, S.-K., & Chang, S.-H. (2005). Lerner Involvement in Self- and Peer-assessment of Task-based Oral Performance. *Second Language Research*, *41*, 711-735.

Lee, Y.-W. (2015). Diagnosing Diagnostic Language Assessment. *Language Testing*, *32*(3), 299-316.

Lee, Y.-W., & Sawaki, Y. (2009). Application of Three Cognitive Diagnosis Models to ESL Reading and Listening Assessments. *Language Assessment Quarterly*, *6*(3), 239-263.

Lu, L. (2018). An Analysis of Peer-Assessment in Chinese as a Second Language Classroom Presentation. *Chinese Language Teaching Methodology and Technology*, *1*(3), 18.

Ma, W., & Winke, P. (2019). Self-Assessment: How Reliable Is It in Assessing Oral Proficiency Over Time? *Foreign Language Annals*, *52*(1), 66-86.

May, L. (2009). Co-constructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, *26*(3), 397-421.

May, L. (2011). Interactional Competence in a Paired Speaking Test: Features Salient to Raters. *Language Assessment Quarterly*, *8*(2), 127-145.

May, L., Nakatsuhara, F., Lam, D., & Galaczi, E. (2020). Developing tools for learning oriented assessment of interactional competence: Bridging theory and practice. *Language Testing*, *37*(2), 165-188.

Nakatsuhara, F. (2014). *The Co-construction of Conversation in Group Oral Tests*. Peter Lang Verlag. https://www.peterlang.com/document/1044236

Nakatsuhara, F., May, L., Lam, D., & Galaczi, E. (2018). *Learning Oriented Feedback in The Development and Assessment of Interactional Competence (Research Notes, Issue 70)*. Cambridge Assessment English.

Oscarson, M. (1989). Self-Assessment of Language Proficiency: Rationale and Applications. *Language Testing*, *6*(1), 1-13.

Poehner, M. E., Zhang, J., & Lu, X. (2015). Computerized Dynamic Assessment (C-Da): Diagnosing L2 Development According to Learner Responsiveness to Mediation. *Language Testing*, *32*(3), 337-357.

Roever, C., & Kasper, G. (2018). Speaking in Turns and Sequences: Interactional Competence as A Target Construct in Testing Speaking. *Language Testing*, *35*(3), 331-355.

Ross, S. (2018). Listener Response as A Facet of Interactional Competence. *Language Testing*, *35*(3), 357-375.

Salehi, M., & Masoule, Z. S. (2017). An Investigation of The Reliability And Validity of Peer, Self-, and Teacher Assessment. *Southern African Linguistics and Applied Language Studies*, *35*(1), 1-15.

Skogmyr Marian, K., & Balaman, U. (2018). Second Language Interactional Competence and Its Development: An Overview of Conversation Analytic Research on Interactional Change Over Time. *Language and Linguistics Compass*, *12*(8), e12285.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using Multivariate Statistics* (Vol. 5). pearson Boston, MA.

Tecedor, M. (2016). Beginning Learners' Development of Interactional Competence: Alignment Activity. *Foreign Language Annals*, *49*(1), 23-41.

Terwase, T. N., & Oluwatoyin, C. (2014). Diagnostic Assessment: A Tool for Quality Control in Education. *Educ Res Rev*, *1*(1), 17-24.

Thomas, G., Martin, D., & Pleasants, K. (2011). Using Self-and Peer-assessment to Enhance Students' Future-learning in Higher Education. *Journal of University Teaching & Learning Practice*, *8*(1), 52-69.

Tozcu, A. (2016). The Effectiveness of Diagnostic Assessment on the Development of Turkish Language Learners' Narrative Skills as an Oral Proficiency Interview (OPI) Task. *Journal of the National Council of Less Commonly Taught Languages*, *19*, 61-96.

Youn, S. J. (2020). Managing Proposal Sequences in Role-Play Assessment: Validity Evidence of Interactional Competence Across Levels. *Language Testing*, *37*(1), 76-106.

Young, R. F. (2000). *Interactional Competence: Challenges for Validity* Annual Meeting of the American Association of Applied Linguistics, Vancouver, Canada.

Young, R. F. (2013). Learning to Talk The Talk and Walk The Walk: Interactional Competence in Academic Spoken English. *Ibérica*, *25*, 15-38.

Biodata

Masoome Azmoode Sis Abad is a Ph.D. candidate in TEFL at Islamic Azad University, Science and Research Branch, Tehran. She has been teaching English since 2004 and she has taught various TEFL courses at Azad University and teaching training centers within the last decade. Currently, she is a university lecturer at IAU. Her main areas of research interest include diagnostic language assessment and second language teaching and learning issues.
Email: *Azmoode2014@gmail.com*

Dr. Gholam-Reza Kiany is an Associate Professor at Tarbiat Modares University. He received his Ph.D. in language and linguistics from the University of Essex, UK. His main areas of research include testing and related issues, research methodologies, and program evaluation. He is the author of many books and scholarly articles.
Email: *kiany_gh@modares.ac.ir*

Dr. Gholam Reza Abbasian is an Associate Professor of Applied Linguistics at Imam Ali University and Islamic Azad University, South Tehran Branch His area of interest is language assessment, applied linguistics, and translation studies. He has published numerous articles in national and international journals.
Email: *gabbasian@gmail.com*