

Test-Takers' Perceptions of Paired Speaking Tests and the Role of Interlocutor Variables in Pairing¹

Hassan Soodmand Afshar²

Received: 25/04/2019

Accepted: 27/01/2020

Abstract

As a mixed-method study, this study investigated EFL test-takers' perceptions of paired speaking tests and the role of interlocutor variables (i.e., peer interlocutor's acquaintanceship, gender, and proficiency level) in pairing in a Cambridge English Language Assessment Open Center. The participants ($N = 148$), having experienced sitting either live paired speaking tests of Cambridge ESOL (group P, $n = 53$) or only IELTS solo interview, (group S, $n = 48$) or both (group P-S, $n = 47$), completed a validated Likert-scale questionnaire. Also, 63 of the participants attended, on a voluntary basis, a semistructured interview. Findings indicated group P test-takers had significantly more positive perceptions of paired/group speaking tests than their counterparts in group P-S, who, in turn, had significantly more positive perceptions of pairing than group S participants. Interview results revealed that the majority of the participants preferred friend, same-proficiency-level, and same-gender peer interlocutors. The (culture-specific) reasons for the findings are discussed in detail.

Keywords: Paired/Group Speaking Tests; Perception; Solo Interview; Proficiency; Gender; Acquaintanceship

1. Introduction

Paired/group speaking test format came in vogue, following the emergence of the communicative approach to foreign language education in the 1970s (Borger, 2019; McNamara & Roever, 2006), which stressed the adoption of pair/group work in classroom contexts. Proceeding from this paradigm shift in L2 education, the paired speaking format was introduced to Cambridge English for the Speakers of Other Languages (ESOL) Key English Test (KET) in 1993, revised Preliminary English Test (PET) in 1995, and revised First Certificate in English (FCE) in 1996

¹Please cite this paper as follows:

Soodmand Afshar, H. (2020). Test-takers' perceptions of paired speaking tests and the role of interlocutor variables in pairing. *Journal of Research in Applied Linguistics*, 11(1), 89-123.

²Department of English Language, Faculty of Humanities, Bu-Ali Sina University, Hamedan, Iran; soodmand@basu.ac.ir

(Taylor, 2000). Grounded in the construct of interactional competence, the present study, thus, investigated Iranian EFL test-takers' perceptions of paired/group speaking tests and the role interlocutor variables (i.e., peer interlocutor's gender, proficiency level, and acquaintanceship) might play in pairing.

2. Literature Review

2.1. Theoretical Framework

As Roever and Kasper (2018, pp. 331-332) hold, whereas the solo interview speaking tests reflect "a primarily psycholinguistic individualist perspective," paired/group speaking tests reflect a mainly "sociolinguistic-interactional perspective" which is heavily rooted in "sociolinguistic models of language use" and pays due attention to social and situational contexts of language use.

Similarly, looking at the construct from a sociocognitive perspective, Galaczi and Taylor (2018, p. 221) view spoken interaction as both a cognitive trait, which places emphasis on "knowledge and processing dimension of language use," and a social interactional trait, which highlights "the social interactional nature of speaking."

As Galaczi (2014) maintains, successful interaction ability in an L2 has questioned the concept of communicative language ability (Canale & Swain, 1980) or what Roever and Kasper (2018) call the psycholinguistic individualist perspective. It is now believed that the ability to communicate in an L2 does not merely lie with an individual; rather, it is coconstructed socially in a joint endeavor; an ability called interactional competence (IC) which was, first, introduced by Kramsch (1986) who argued that communication is constructed mutually and reciprocally by the interlocutors during the act of communication. Similarly, Jacoby and Ochs (1995, p. 171) define IC as "a range of interactional processes, including collaboration, cooperation and coordination." Within the same lines, Young (2008, p. 101) maintains, "interactional competence is not the knowledge or possession of an individual person, but is coconstructed by all participants in a discursive practice." It can, thus, be argued that at the crux of IC lies the fact that language notions and events are constructed reciprocally and socially during the process of communication occurring among interactants.

Paired speaking tests are believed to tap test-takers' IC (Kramsch, 1986). That is, in paired/group orals, communication is established socially and "interaction is constructed jointly by all the members together, rather than individually" (Ahmadi & Montasseri, 2019, p. 4). Therefore, as Galaczi (2010) holds, because paired/group oral tests include and assess wider range of tasks and interactional skills, such as topic initiation, topic termination, turn-taking, interactive listening, and interactional

management, they could elicit more language evidence from test-takers and are, thus, more valid.

The mismatch existing between such modern interactionally oriented trends in L2 language teaching which support paired/group interaction and the oral language tests that deal with individuals has been criticized by some scholars in the field (e.g., Kramsch, 1986). These scholars argue that if the interaction in real-life situations as well as under communicatively-oriented classroom pedagogy occurs in pairs and groups, this should naturally be echoed in oral language assessment, a reasoning which seems to justify the rationale behind the adoption of paired/group speaking tests in language testing contexts. Therefore, the significance of and the focus on interactional language ability has resulted in a pervasive use of paired/group class activities in language teaching and paired/group speaking tests in language assessment (Galaczi, 2014).

Despite all the merits mentioned above for IC and interactionally-oriented oral language assessment, the notion has been criticized on several grounds: Firstly, although different scholars have adopted different methodologies to tap into test-takers' IC in paired/group speaking tests, there is still "debate among language testing scholars on how to deal with it" (Lam, 2018, p. 380). Secondly, in spite of the fact that IC has had a huge impact on both spoken language assessment and language teaching, it has also caused tension and confusion as to whether to award individual or shared scores to test-takers' performance in a paired/ group speaking test session (Song & Lee, 2016).

2.2. Advantages and Disadvantages of Pairing

Paired speaking tests are believed to give candidates a sense of relaxation and psychological ease (Wallis, 1995) and produce richer sample of spoken language due to the generation of more varied patterns of exchange and more and greater range of language functions (Taylor, 2001). They are said to be more reflective of real-life daily speech situations in which the speaker and listener coconstruct the communication (Heaton, 1988), add to test fairness and reduce the disadvantage of scores being dependent on only one person (Saville & Hargreaves, 1999), and combine the advantages of both holistic and analytic scoring methods (Weir & Taylor, 2011). The last advantage mentioned here might, however, be the sole feature of Cambridge ESOL Main Suite examinations and not the characteristic of all paired format tests.

Paired tests are thought to result in the generation of more authentic and symmetrical discourse due to the existence of equal power relations among peer interlocutors (Ahmadi & Sadeghi, 2016). The construct underlying paired testing is believed to be much broader in nature and includes, in addition to communicative

competence, Kramsch's (1986) IC, as mentioned earlier. Moreover, paired tests are believed to be received positively by test-takers (Ahmadi & Sadeghi, 2016; Fulcher, 1996), generate more positive washback effect (Bonk & Ockey, 2003), and save assessment time and workload (Davis, 2009).

Despite the advantages voiced for paired speaking tests mentioned above, they have been criticized on several grounds: The use of various prompts in paired tests might significantly affect the discourse pattern produced (Leaper & Brawn, 2019; Leaper & Riazi, 2014), the scoring reliability of paired/group orals might be under question (Leaper & Brawn, 2019; van Moere, 2006), and shy test-takers might be disadvantaged (Bonk & van Moere, 2006). In addition, it is not yet clear whether some of the advantages mentioned above, in reality, affect candidates' performance positively. For example, as Foot (1999) maintains, test-takers' higher relaxation in a paired test might not necessarily lead to a better performance. Similarly, if test-takers produce more varied exchange patterns and more linguistic functions in a paired test, they might not necessarily score as high.

However, it should be noted here that most disadvantages and concerns voiced against paired format speaking tests are associated with how the pairings take place or what has been called in the literature as the *interlocutor effect*. That is, the performance of candidates might be affected negatively not because of their linguistic incompetence, but simply due to CIV (Messick, 1995) originating from such social relationships and social factors as race, social class, profession, and age (Foot, 1999), gender, (Norton, 2005), personality (Berry, 2007), test partner unfamiliarity (O'Sullivan, 2002), and linguistic heterogeneity of the candidates in pairs (Saville & Hargreaves, 1999).

2.3. Various Research Perspectives on Paired Speaking Tests

Paired/group format speaking tests have been researched from different perspectives comprising test-takers' performance in and/or perception of pairing (e.g., Ahmadi & Sadeghi, 2016; Brooks, 2009; Együd & Glover, 2001), such interlocutor variables as proficiency level, gender, and pairmate acquaintanceship (e.g., Bonk & van Moere, 2004; Davis, 2009; Joo, 2019; O'Sullivan, 2002), personality factors like extroversion/introversion, shyness, and talkativeness (e.g., Berry, 2007; Bonk & van Moere, 2004; Joo, 2019; Ockey, 2009), various task types (i.e., planned/unplanned) and prompts (e.g., Lam, 2019; Leaper & Riazi, 2014; Nitta & Nakatsuhara, 2014), the nature of discourse and the features of interaction (e.g., Brooks, 2009; Galaczi, 2008; Nakatsuhara, 2011) and, finally, the construct behind pairing (i.e., IC itself; e.g., Ahmadi & Montasseri, 2019; Lam, 2019). To address the objectives of the present study, the present researcher focused on the first two perspectives and reviewed the literature existing in the field in this regard.

2.3.1. Test-takers' performance in and perception of paired speaking tests

Several studies have been conducted in the field, investigating the performance of test-takers in the paired speaking tests. In a rather recent study, Ahmadi and Sadeghi (2016) assessed Iranian EFL learners' speaking performance in the three groups of monologue ($n = 6$), interview ($n = 8$) and group oral ($n = 9$). The participants in the monologue group talked about the given topic for nearly 5 min, without interacting with an interlocutor of any kind. In the interview group, the interaction occurred between an interviewer and an interviewee and lasted about 7 to 8 min. In the group oral, the participants discussed a topic, introduced by the interviewer in groups of three or more. The results indicated that the test format type significantly differentiated the groups, that is, overall, the test-takers in the group oral showed significantly better performance than their counterparts in the monologue group, who, in turn, outperformed their counterparts in the interview group. However, their results indicated that regarding the discourse produced, the group oral performed better than the other two groups concerning the accuracy of the speech produced, whereas, in terms of complexity, the most complex language was generated by the monologue test-takers followed by the groups oral and interview, respectively.

In another study, Brooks (2009), comparing the performance of eight pairs of international test-takers who sat both paired and individual test formats with comparable speaking prompts, found that, overall, the test-takers in the paired format group achieved higher scores. Brook's further qualitative analysis of the data indicated the interaction patterns of the participants when taking paired format were more complex, and their performance was also linguistically more demanding, thus leading to more interaction and negotiation of meaning.

Test perception, in general, and paired/group oral test perception, in particular, have also received research attention in several studies. For one, Palumbo and Steele-Johnson (2014), comparing test scores (i.e., test performance) and test perceptions of 246 Blacks and Whites found that Blacks were outperformed on a cognitive ability test by their White counterparts because they had misperceptions or negative perceptions of the test taken which confirmed similar previous results in this respect. This might clearly indicate the paramount role test perception can play in test performance. Likewise, Sadeghi, Azad Mousavi, and Javidi (2017), investigating the association between 48 upper-intermediate and advanced EFL learners' self-perceived communicative competence and their task-free and task-based self-assessment of speaking, found a high association between the participants' self-perceived communicative competence (i.e., their perception of the construct) and their self-assessment of speaking. They also found a difference between the learners' task-free and task-based self-assessment. The findings might, thus, stress the

dependability of the perception of the given construct, test, and the like, and its compliance with the real existence of the construct.

Similarly, Együd and Glover (2001), administering a questionnaire to 14 secondary school students in Hungary seeking their perception of paired/individual format speaking tests, found that the participants almost unanimously reported that they preferred paired format over the solo interview, that pairing gave them opportunities to produce better output and that pairing supported good teaching.

Also, Humphry-Baker (2000, as cited in Taylor, 2001), administering a questionnaire to 130 candidates who had already sat various Cambridge ESOL tests (e.g., PET, FCE, Certificate of Proficiency in English [CPE]), investigated their perceptions of the test taken. Candidates mostly agreed or strongly agreed with the majority of the twelve items in the questionnaire, including *I like paired test* and *The test gave me good opportunities to speak*, although they, sometimes, disagreed with such statements as *I had enough time to speak* or *I performed well in the test*, which might caution us against concluding that a positive perception of paired testing necessarily leads to better test performance.

2.3.2. *Interlocutor variables and pairing*

The effect of various interlocutor variables (so-called by Vidacović & Galaczi, 2013), such as proficiency level, pairmate acquaintanceship, gender, personality factors (e.g., extroversion/introversion, assertiveness, shyness, etc.), age, race, L1 background, and the like, on pairing has been investigated in the literature of the field. However, of particular interest to the present study was the impact of the first three variables which are discussed below:

2.3.2.1. *Proficiency level and pairing.*

The impact of peer interlocutor's proficiency level on paired testing has received empirical attention in the literature. Iwashita (1998), investigating the effect of proficiency level on test scores and discourse of twenty adult Japanese learners, found that although being paired with a linguistically stronger partner led to more talk in general, it did not significantly affect the scores gained. Similarly, Nakatsuhara (2006) investigated 24 (17 female and 7 male) undergraduate/graduate students at Essex University with a wide range of L1 backgrounds. Dividing the participants into two groups of intermediate and advanced based on their proficiency level, Nakatsuhara paired each subject randomly with two different interlocutors: Once with one from their own proficiency level and once with a partner from different proficiency level and used the two-way collaborative task in the Certificate of Advanced English (CAE) speaking test. Using independent samples *t* test,

Nakatsuhara found no significant difference between the two groups of linguistically homogeneous pairs.

Davis (2009) investigated the impact of peer interlocutors' varying proficiency level on their partners' paired speaking test performance in a sample of twenty first-year students at a university in China. Dividing the participants into groups of high- and low-English proficiency and pairing them once with a peer interlocutor of similar proficiency and once with a partner of varying (i.e., either higher or lower) proficiency level, Davis found that peer interlocutor proficiency had no considerable impact on Rasch analysis ability measures, and that the difference in the proficiency level of the test-takers did not preclude use of paired format speaking.

Norton (2005) analyzed the data gained from transcriptions of Cambridge speaking tests taken by 27 Japanese and European candidates of FCE (10), CAE (10), and CPE (7) which were all audio- or video-recorded. However, against the general trend of the results of the studies cited above, Norton found that the difference in proficiency level of the test-takers could more benefit the lower-proficiency level test-takers who might have been exposed to "better quality language" and, thus, were "able to incorporate some of their partner's expressions into their own speech" (p. 291).

Thus, although the bulk of the studies mentioned above found no significant impact of the test-takers' varying proficiency level on their partners' language performance, the findings of the studies in this respect are, sometimes, contradictory, calling for the conduct of more research in the field to shed more light on the issue.

2.3.2.2. Acquaintanceship and pairing.

The impact of peer interlocutor acquaintanceship on test performance has also been the focus of several investigations. Norton (2005), for example, found that the candidates in one of the pairs in CAE exam who were friends, produced more talk (3116 words) compared to the mean number of words per CAE interview in that session (which was 2552.75), and were among three of four high-scoring Japanese females paired with a friend. Similarly, another CAE candidate, at the same session and center, paired with a stranger appeared nervous and reticent and, as a result, produced less talk. Although this candidate's main problem might have been her limited linguistic ability, her reticence and unwillingness to produce language might, at least partially, be attributable to being paired with a stranger.

However, it is worth noting here that the findings of Norton (2005) might not fully verify the positive impact of peer interlocutor acquaintanceship on one's performance, firstly, because of the small sample size (as Norton herself contends). Secondly, other factors like proficiency level of the candidates, personality traits like

extroversion/introversion, self-confidence, and so on might have been involved in Norton's participants' producing the highest number of words in that specific session.

Another study conducted on the relationship between peer interlocutor acquaintanceship and test performance in paired format speaking tests is that done by O'Sullivan (2002) who found that learner familiarity impacted upon performance, and that the candidates had a better performance and achieved higher scores when paired with a friend. However, contradicting the findings of the studies mentioned above, Chambers, Galaczi, and Gilbert (2012), adopting a mixed-method approach (i.e., using both questionnaires and interviews), found candidate acquaintanceship did not significantly differentiate the Swiss participants of the study.

2.3.2.3. Gender and pairing.

The role of gender in oral language assessment has also been investigated from various points of view including raters' as well as test-takers' perspectives. For one, Bonk and van Moere (2004), investigating the role of gender and shyness in group oral test performance of 1055 Japanese university students, found that whereas there was a relationship between the test-takers' scores and shyness, gender had no impact on the test scores.

However, O'Sullivan (2002) as mentioned earlier, investigating the impact of peer interlocutor acquaintanceship and gender on test-takers' performance, found both factors affected the participants' test performance.

It is worth mentioning here that, as Brown and McNamara (2004), and Galaczi (2010) maintain, no simple linear, clear-cut, neatly-correlated relationship can be postulated between these interlocutor variables and test performance in paired format speaking tests. Additionally, variability in speaking assessment (e.g., being paired with a person of different gender, proficiency level, and with an acquaintance/stranger, etc.) is part and parcel of real-life IC, which cannot be easily eliminated—a point also made by Swain (2001, as cited in Fox, 2004) and supported by Brown and McNamara (2004).

2.4. Significance of the Study and Statement of the Problem

Although attention to paired/group testing has recently grown due to the emphasis exerted upon communicative approaches to language teaching (Galaczi, 2010; Lam, 2019; Sandlund, Sundqvist, & Nyroos, 2016), paired format speaking tests (e.g., those of high-stakes examinations of Cambridge ESOL) have not received due research attention in the literature until recently (May, 2009). Moreover, relatively little research has been conducted on candidates' attitudes towards or perceptions of paired format speaking tests, in general (Taylor, 2001; Galaczi, 2010). Furthermore, the previously-conducted similar studies were typically carried out with

a restricted number of participants who mainly sat low-stakes, local speaking tests, compared to the participants of the present study who had all experienced sitting such live high-stakes English proficiency examinations as PET, FCE, IELTS, and the like in a real-life (i.e., live) testing situation, which enabled them to give more realistic opinions and evaluation of the test(s) taken. Also, the present study adopted a triangulated mixed-method design (i.e., use of both an interview and a validated structured questionnaire for data collection purposes), which could naturally enhance the generalizability of the findings.

Most importantly, the findings of research on paired/group speaking tests have been shown to be largely context-dependent and culture-specific (Chambers, Galaczi & Gilbert, 2012; Ockey, 2009; O'Sullivan, 2002) and, in effect, no such a study was found to have been conducted in the EFL teaching/testing context of Iran, with participants having experienced taking high-stakes live paired/group tests. Moreover, the rationale behind investigating test-takers' perceptions of the role of interlocutor variables (e.g., language proficiency level, gender, and pairmate acquaintanceship) in pairing in the current study was, firstly, personal. That is, as a certified Cambridge ESOL speaking examiner, I, sometimes, observed confusion, anxiety, embarrassment, and so on when the test-takers were paired with a partner from the opposite sex, different proficiency level, a stranger, and so on. Secondly, as the extensive review of the related literature showed, the previous research findings in this respect were mostly mixed and, sometimes, contradictory which rendered the conduct of the present study inevitable to fill the research gap felt.

Also, it should be noted that in the EFL context of Iran, almost all speaking examinations, both at foreign language institutes and at the university level with students majoring in English language, are conducted in the form of solo one-to-one interviews. These, in fact, are the two microcontexts in the macrocontext of EFL in Iran where the speaking skill, among other language skills, is focused on. The speaking skill is, however, valued at public (i.e., elementary and secondary) schools almost not at all. Neither is it paid attention to at the university level with non-English-major students both in general purposes English (GPE) courses and in English for academic purposes (EAP) courses. The major skills and activities worked on in the latter case include reading, vocabulary, grammar, and translation. Thus, if it is found in the current study that the participants have a more favorable attitude towards paired/group speaking tests, why not incorporating them in foreign language speaking assessments in the country and other similar EFL contexts? Thus, bearing these points in mind and to address the purposes of the study, the following research questions were formulated.

1. Do EFL test-takers' perceptions of paired speaking tests (e.g., that of PET, FCE, etc.) and solo interviews (e.g., that of IELTS) significantly differ?

2. Are such interlocutor variables as language proficiency level, gender, and pairmate acquaintanceship perceived to affect test performance in paired speaking tests?

3. Methodology

3.1. Design of the Study

The present study consisted of two phases: the pilot study and the study. The pilot study was mainly quantitative and adopted a 7-item Likert-scale questionnaire (followed by two open-ended questions) to elicit the required data. The study adopted a mixed method approach (i.e., both questionnaire survey and interview) for data collection and analysis purposes.

3.2. The Pilot Study

Twenty-one participants (12 female and 9 male), all of whom had the experience of sitting the paired format speaking test of Cambridge FCE and/or PET in a Cambridge English Language Assessment Center, took part in the pilot study. Nine of these participants had also experienced sitting the solo interview of IELTS. The age of the participants ranged from 18 to 34, with the mean age being 23.7. All the participants completed a 7-item 5-point Likert-scale questionnaire, ranging from 1 (*Completely Disagree*) to 5 (*Completely Agree*), followed by two open-ended questions. The Likert items of the questionnaire were mainly constructed based on the vast review of the related literature, which were, then, pilot-tested with a group of eight similar test-takers in the center and viewed by two experts in the field, based on whose views necessary adjustments were made. Using Cronbach's alpha internal consistency estimation, the reliability of the pilot study questionnaire was calculated to be 0.85. The pilot study questionnaire can be found in Appendix A.

The questionnaire constructed was, then, administered to all the 21 participants of the pilot study, who all completed and returned it to the researcher.

A Mann-Whitney *U* test was carried out to see whether there were any significant differences between the responses of the two groups in the pilot study. The results indicated the two groups were statistically significantly different in their perceptions of paired speaking tests (Mann Whitney $U = 25.5$, $p = 0.025 < 0.05$). That is, those participants who had experienced sitting both solo interview of IELTS and paired-format Cambridge PET and/or FCE speaking tests had significantly more positive perceptions of the paired speaking tests than the IELTS solo interview.

3.3. The Study

3.3.1. Context and participants of the study

The initial population of the study included 317 students in an English language institute, which is also certified as an authorized Cambridge English Language Assessment Center. The participants had already taken either the paired speaking tests of live PET and/or FCE in the center (group P, $n = 112$), solo interview of live IELTS in IELTS test centers in Iran or in the neighboring countries (group S, $n = 107$), or both (group P-S, $n = 98$) between 2010 and 2017, with the majority having taken the tests from 2012 to 2014. From among these, 196 participants (group P, $n = 72$; group P-S, $n = 61$ and group S, $n = 63$) were selected based on stratified random sampling to ensure the distribution of an almost equal number of participants in each of the three groups.

All the 196 participants were asked to complete an 11-item 5-point Likert-scale questionnaire, out of whom, finally, 148 participants, (group P, $n = 53$; group P-S, $n = 47$; group S, $n = 48$), who voluntarily completed the questionnaire and whose informed consent was also obtained, took part in the study. None of the participants was incorporated in more than one group.

It is also worth noting here that the participants can be regarded as being generally homogeneous concerning their language proficiency. That is, they can, in general, be considered at level B2 of CEFR (Common European Framework of Reference) of Council of Europe because the participants in groups P and P-S had been placed by the center in the CAE1 preparation course because the center authorities were ascertained the majority (nearly 75%) had already passed Cambridge ESOL FCE examination (with grade C or higher), either in the center or other Cambridge ESOL open centers in the country. The remaining participants in groups P and P-S had alternatively been placed in the course (i.e., CAE1) because they had also already taken and passed Cambridge PET (at level B1) with Grade A or “Pass with Distinction,” which is equal to FCE pass score of Grade C (<http://www.cambridgeenglish.org/images/177867-the-methodology-behind-the-cambridge-english-scale.pdf>). The researcher also made sure that the participants in group S had an IELTS total band score of 5 to 7 (this latter score was only 1 case, of course, and the average was 5.75) before enrolling in IELTS preparation courses in the center. According to Cambridge English Language Assessment the IELTS, a total band score of 5.5-6 is equal to level B1 (i.e., FCE) of Cambridge ESOL (<http://www.cambridgeenglish.org/images/28894-cae-comparing-scores.pdf>). It could, thus, be argued that the participants in groups P, P-S, and S were all roughly at level B2 of CEFR and were, thus, fairly homogenous regarding their language proficiency.

With regard to the homogeneity of the participants concerning personality factors, pairmate acquaintanceship, gender, and so on, it could be argued that because the Cambridge ESOL Examinations candidates (including the participants in groups P and P-S) are normally paired randomly for speaking tests in the center in which the present study was conducted (and most probably in all other Cambridge centers), their personality features, pairmate acquaintanceship, gender, and so on might also have been distributed randomly within the pairs and might have affected them almost equally—a reasoning also supported by Leaper and Riazi (2014), as applied to their study.

Also, the participants were all adult learners whose age ranged from 17 to 35, with the mean age being 24.3. Moreover, the participants were all from the same cultural background (i.e., Iran), with the majority either being college/university students or graduates. Therefore, the participants could, to a large extent, be considered uniform in terms of such background variables as language proficiency, education level, age, cultural background, and the like. The characteristics of the participants are summarized in Table 1:

Table 1. *Characteristics of Participants of the Study*

Groups	N	Sex		Percent
		M	F	
P	53	24	29	35.81
P-S	47	20	27	31.76
S	48	21	27	32.43
Total	148	65	83	100

3.3.2. Instruments

3.3.2.1. Structured questionnaire.

The following steps were taken in constructing the structured questionnaire:

1. The (pilot study) questionnaire constructed was given to three colleagues of the researcher, holding a Ph.D. in applied linguistics to receive their expert views. The necessary adjustments suggested, including rephrasing some of the items, were made. One of the colleagues, further, commented the pilot study questionnaire might mostly favor group P and suggested adding some items to obviate this possible drawback. Thus, taking this expert's view into account and drawing on the common patterns of the responses of the pilot study participants to the two open-ended questions of (the pilot study) questionnaire on the advantages of solo interviews, items # 10 and 11 were also built into the questionnaire. Also, items # 7 and 8 were added which along with item # 4 would investigate the test-takers' perceptions of the impact of such interlocutor variables as peer

interlocutor's gender, proficiency level, and pairmate acquaintanceship on their performance in paired speaking tests to provide a more in-depth investigation into the second research question.

2. To further ensure the validity of the questionnaire constructed, Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, Bartlett's test of Sphericity, and principal component factor analysis with Varimax Rotation were conducted. The KMO measure (.83), being greater than .70, showed enough items were predicted by each factor. Also, the results of Bartlett's test ($p = .000$) indicated that there was a high enough correlation between the variables to provide a reasonable basis for factor analysis.
3. The principal component factor analysis with Varimax Rotation was conducted to assess the underlying structure for the 11 items of the questionnaire, the results of which indicated that the items clustered into two groups or factors defined by high loadings: That is, the first factor, which seemed to index "pluses of pairing," loaded strongly on all the 11 questions. Questions # 10 and 11, however, which seemed to index "minuses of pairing," had stronger cross-loadings on the second factor. Also, after rotation, the first and second factors accounted for 38.13% and 12.15% of the total variance, respectively. The items and factor loadings for the rotated factors are displayed in Appendix B, with loadings below .40 being omitted to improve clarity.
4. Using Cronbach's alpha internal consistency estimation, the reliability of the questionnaire was also calculated to be 0.82, which shows that the items form a scale that has reasonable and relatively high internal consistency. A copy of the study questionnaire is available in Appendix C.

3.3.2.2. *Interview.*

To address the purposes of the second research question more deeply and to do multiple-level analysis (Dörnyei, 2007) on the data obtained from the questionnaire, the present researcher asked the participants to also sit a semistructured interview. However, out of the whole 148 participants, only 63 (nearly 42.5%; $nP = 20$, $nP-S = 24$, $nS = 19$) volunteered to attend the interview.

Three out of the four items of the semistructured interview were constructed to specifically and intentionally deal with the participants' perceptions of the impact of various ways of pairing on test performance in paired speaking tests. The last question of the interview dealt with the test-takers' preference for paired format

speaking tests, in general, and their reasons for doing or not doing so, in particular. On average, each interview lasted for about 15 min. With the permission of the interviewees, the interviews were all audio-recorded, transcribed, and subjected to content analysis. A sample of interview questions is enclosed in Appendix D.

4. Results

4.1. Results of Structured Questionnaire Comparing Test-Takers' Perceptions of Paired Testing

The descriptive statistics of the items of the questionnaire for all the three groups of the study are presented in Table 2:

Table 2. *Descriptive Statistics for Individual Items of Questionnaire*

Item	Group	N	Mean	SD	Percent
1	P	53	4.37	0.72	87.4
	P-S	47	3.87	0.75	77.4
	S	48	3.21	0.93	64.2
2	P	53	3.86	0.93	77.2
	P-S	47	4.02	0.94	80.4
	S	48	2.81	0.94	56.2
3	P	53	4.69	0.46	93.8
	P-S	47	4.22	0.59	84.4
	S	48	3.11	1.04	62.2
4	P	53	4.25	0.86	85
	P-S	47	4.02	0.84	80.4
	S	48	3.15	0.88	63
5	P	53	4.37	0.77	87.4
	P-S	47	4.18	0.68	83.6
	S	48	3.13	1.01	62.6
6	P	53	4.39	0.69	87.8
	P-S	47	4.18	0.61	83.6
	S	48	3.28	0.90	65.6
7	P	53	3.82	1.05	76.4
	P-S	47	3.58	1.17	71.6
	S	48	2.91	0.92	58.2
8	P	53	4.18	0.95	83.6
	P-S	47	3.91	0.73	78.2
	S	48	3.45	0.97	69
9	P	53	4.49	0.67	89.8
	P-S	47	3.96	0.95	79.2
	S	48	3.53	0.83	70.6
10	P	53	2.24	0.76	44.8
	P-S	47	2.18	0.80	43.6
	S	48	3.21	1.04	64.2

11	P	53	2.25	0.89	45
	P-S	47	2.38	0.88	47.6
	S	48	3.04	1.06	60.8
All	P	53	3.93	0.32	78.6
	P-S	47	3.68	0.27	73.6
	S	48	3.17	0.39	63.4

To investigate whether the sample in the study was normally distributed, a Kolmogorov-Smirnov normality test was run, the results of which indicated the sample was normally distributed ($z = 0.91$, $p = 0.40 > 0.05$), thus allowing parametric statistics to be applied for data analysis.

To test for the equality of variances in the three groups, Levene's test of homogeneity of variances was conducted, the results of which indicated the principle of the equality of variances was not violated, $F(2, 145) = 2.37$, $p = 0.11 > 0.05$.

Having checked for the normality of the population and having become assured of the homogeneity of the variances in the three groups, a one-way ANOVA was run to investigate whether the three groups were statistically significantly different in their perceptions of paired tests/solo interviews. The results are summarized in Table 3:

Table 3. *Results of One-way ANOVA Investigating Difference Among the Three Groups of P, P-S, and S in Their Perceptions of Paring*

	Sum of Squares	df	Mean source	F	Sig.	η^2
Between Groups	15.06	2	7.35	66.59	0.001	0.47
Within Groups	16.39	145	0.11			
Total	31.46	147				

Note. $F(2, 145) = 66.59$, $p \cong 0.001 < 0.05$

As Table 3 shows, significant differences were found among the three groups in their perceptions of paired tests/solo interviews, $F(2, 145) = 66.59$, $p = 0.001 < .05$. The eta-squared effect size (η^2), being 0.47, shows that the magnitude of the difference among the three groups was high. According to Cohen (1988), the eta-squared effect sizes of .37 and above (up to .45) are considered large, showing the fact that the statistically significant difference observed among various groups is noticeable and is not, thus, negligible. However, to see where this difference lay, a Scheffé post-hoc analysis was conducted, the results of which are presented in Table 4:

Table 4. *Results of Scheffé Post-Hoc Comparison*

Groups	Mean Difference	Std. Error	Sig.	95% confidence interval	
				Lower Bound	Upper Bound
P*P-S	0.24*	0.06	0.002	-.08	.41
P*S	0.76*	0.06	0.001	.59	.92
P-S*S	0.51*	0.06	0.001	.34	.68

Note. The mean difference is significant at the .05 level.

As indicated in Table 4, there were statistically significant differences between groups P and S and between groups P-S and S, as well as between groups P and P-S in their perceptions of the paired tests.

4.1. Interview Results

As mentioned earlier, to do multiple-level analysis on the questionnaire data in order to scrutinize test-takers' perceptions of the impact of interlocutor variables on candidates' test performance and to delve more deeply into the issue (i.e., to address the second research question of the study), a 4-question semistructured interview was conducted with 63 participants.

Two complementary trends were followed in analyzing the interview results: First, because all the four questions of the interview dealt, in reality, with the dichotomies of some sorts (e.g., either the participants preferred paired tests, solo interviews, or had no idea, either they liked same-sex peer interlocutors, different-sex ones, or had no idea, etc.), I, thus, prepared a template or code manual, following the motivated looking approach (Sack, 1984, as cited in Lazaraton, 2002), which is in line with the template organizing style of Grabtree and Miller's (1999, as cited in Dörnyei, 2007), acting on the hunch of the second research question and bearing in mind the questions of the structured questionnaire dealing with the test-takers' attitudes towards the impact of interlocutor variables on their test performance. Second, to extract the participants' reasons and comments about why they believed in what they remarked, the interview transcriptions were subjected to content analysis, and the common patterns and recurring themes of the responses were extracted, which were, then, quantized (Dörnyei, 2007), frequency analyzed, and tabulated.

The first interview question enquired about the participants' perceptions of whether the varying linguistic ability of one's peer interlocutor (i.e., pairing a weaker candidate with a stronger one) could affect his or her performance in paired tests.

Potentially speaking, five different templates or conditions can be postulated here:

1. Having a linguistically stronger peer interlocutor would affect you (i.e., the weaker one) positively.
2. Having a linguistically stronger peer interlocutor would affect you (i.e., the weaker one) negatively.
3. Having a linguistically weaker peer interlocutor would affect you (i.e., the stronger one) positively.
4. Having a linguistically weaker peer interlocutor would affect you (i.e., the stronger one) negatively.
5. Being paired with a candidate of varying linguistic ability has no effect (i.e., either positive or negative) on your performance.

The results of the participants' responses to this question are summarized in Table 5:

Table 5. *Participants' Response Patterns to Impact of Peer Interlocutor's Varying Proficiency Level on One's Test Performance*

Condition	Frequency	Percent
1. Stronger pairmate impacts the weaker one positively.	17	26.98
2. Stronger pairmate impacts the weaker one negatively.	15	23.80
3. Weaker pairmate impacts the stronger one positively.	5	7.93
4. Weaker pairmate impacts the stronger one negatively.	6	9.52
5. Stronger/weaker pairmate has no impact on you.	20	31.74
Total	63	100

As evident from Table 5, rather mixed results were found here. Evidence was found for every one of the five conditions above: That is, nearly 27% of the participants interviewed (i.e., 17 out of 63) agreed that pairing a stronger candidate with a weaker one would more benefit the latter and the stronger candidate could stimulate the weaker one to speak. Nearly 24% were in favor of condition 2; about 8% felt having a linguistically weaker peer interlocutor would affect the stronger one positively (i.e., condition 3 above), and 9.5% were in line with condition 4. However, nearly 32% perceived their peer interlocutors' varying linguistic proficiency level did not affect their test performance.

The second question of the interview sought the participants' preferences for having their peer interlocutors, either from their friends and classmates or from strangers. The results for this question are summarized in Table 6:

Table 6. *Participants' Response Patterns to Impact of Peer Interlocutor Acquaintanceship on One's Test Performance*

Template	Frequency	Percent
1. Friend peer interlocutor is better.	41	65.08
2. Stranger peer interlocutor is better.	8	12.70
3. Friend stranger peer interlocutor has no impact on one's performance.	14	22.22
Total	63	100.0

As Table 6 indicates, the majority of the participants (i.e., 65%) were supportive and appreciative of being paired with an acquaintance. Nearly, 13% were for having a stranger peer interlocutor and another 22% (14 out of 63) perceived it made no difference. The main reasons for the preferences of those interviewees preferring friend/classmate peer interlocutors ($n = 41$) are presented in Table 7. It should be mentioned that some participants gave more than one reason:

Table 7. *Participants' Reasons for Preferring Friend/Familiar Peer Interlocutors*

Template	Frequency	Percent
1. They would help you feel less stressful.	31	75.60
2. They will help increase your self-confidence.	21	51.21
3. With them, the conversation goes ahead more naturally.	15	36.58
4. With them, you become more motivated to speak.	10	24.39
Other reasons	6	14.63

The third question of the interview was concerned with whether the participants preferred to have the same-sex peer interlocutor or one from the opposite sex, the results of which are summarized in Table 8:

Table 8. *Participants' Responses on Gender and Pairing*

Template	Frequency	Percentage
1. Having the same-sex peer interlocutor is better.	38	60.32
2. Having opposite-sex peer interlocutor is better.	10	15.87
3. Gender makes no difference.	15	23.81
Total	63 (F = 36 M = 27)	100

Note. F = Female; M = Male

As shown in Table 8, 60% of the participants interviewed expressed their preference for the same-sex peer interlocutor. Fifteen participants (nearly 24%) reported it made no difference to them. Ten other participants (nearly 16%) reported they preferred to have their peer interlocutors from the opposite sex.

The last question of the interview asked whether the participants liked paired-format and their reasons for that. Table 9 presents interview participants' attitudes towards test pairing:

Table 9. *Participants' Preferences for Paired Speaking Tests/Solo Interviews*

Items	Frequency	Percent
1. Paired format tests are, in general, better and fairer.	39	61.90
2. Individual format tests are better.	11	17.46
3. It makes no difference.	13	20.64
Total	63	100

As the results in Table 9 indicate, nearly 62% of the participants interviewed stated that they preferred paired speaking tests over solo interviews. About 21% expressed indifference towards paired speaking tests. Nearly, 17.5% expressed their preference for individual format tests (i.e., solo interviews). The reasons and comments of the participants preferring paired tests ($n = 39$) are also summarized and presented in Table 10. Needless to mention, some participants gave more than one reason:

Table 10. *Participants' Reasons for Their Preference for Paired Tests*

Reason	Frequency	Percent
1. They are more soothing, relaxing, and stress-free.	34	87.18
2. They are more natural and real.	25	64.10
3. They are much fairer because two examiners examine you.	20	51.28
4. You can help your friend, and they can also help you.	14	35.90
5. You mostly talk with your almost same-proficiency level friend/peer interlocutor (rather than an examiner only).	9	23.07
6. Other reasons.	6	15.38

5. Discussion

As the results obtained from the questionnaire corroborated by the findings of the second question of the interview revealed, the participants in groups P and P-S who had both experienced taking paired speaking tests had significantly more favorable perceptions of paired tests than solo interviews in comparison to those in group S who had no experience of taking live paired speaking tests.

The participants' preference for paired speaking tests corroborate the results of Együd and Glover (2001) which revealed that the participants liked pairing because they thought that the interactions in the paired/group test format resulted in the production of more natural English (a point also supported in the present study by the

findings of the interview in Table 10), and that pairing had good washback effect and affected the students' learning positively. Of course, Együd and Glover's participants believed that pairing could be fruitful with all proficiency levels, a finding which stands in contrast with the ideas of Foot (1999) who maintains that pairing might be effective with only higher-proficiency levels and doubts its effectiveness and success with lower-level proficiency candidates.

The results of the current study are also partially corroborated by the findings by Humphry-Baker (2000), who, experimenting with 130 candidates through a questionnaire, found out that the participants liked pairing, and that pairing gave them good opportunity to speak, though they complained of insufficient time to speak during the exam and talked about their lack of satisfaction with their performance.

As indicated in Table 10, the main reason for the participants' preference for pairing, on which a great majority of them agreed, was psychological in nature: The candidates preferred pairing because it was more relaxing and stress-free, a finding which is not surprising and is widely corroborated by the results of other similar studies and seems to be well accepted (Wallis, 1995).

The second reason for nearly 64% of the participants who preferred paired format speaking tests in response to the fourth question of the interview was that it was more natural and real.

The comments of the participants in this regard are echoed by the results of Brooks (2009) who indicated that, in paired speaking tests, the participants elaborated on their partners' ideas, paraphrased, and produced more sentences, prompted, maintained, and concluded the interaction, and, thus, engaged in and produced more complex interaction patterns, indicative of naturally occurring communication. The findings are also supported by the results by Galaczi (2008), which show the selection of paired speaking tests was more indicative of coconstruction of the communication event occurring in real-life situations. In other words, paired tests mainly tap test-takers' IC which seems to underlie the bulk of real-life communication in which interaction is jointly and often symmetrically coconstructed by the interactants in the society.

However, it should be noted that, as Simpson (2006) maintains, asymmetrical interactions are not necessarily invalid because asymmetrical power relations (e.g., manager-clerk, shop keeper-shop assistant, teacher-student, etc.) already exist in real-life situations—an argument also supported by Fulcher (2015) who maintains, friends' chatting (i.e., symmetrical power relations) is not very much common in higher education academic contexts where students need to interact asymmetrically with professors, librarians, welfare officers, and the like.

One of the significant issues in language testing is the important question of the effect language testing has on teaching or washback effect. Because in realistic classroom contexts, L2 students are to speak and interact with each other in the form of pair work and/or group discussions, it would be a good idea for this to be mirrored in language testing, as echoed by May (1996, p. 137) who maintained, “examiners have noted that students who are accustomed to working with partners tend to be better prepared for this [paired] format.” Thus, as Együd and Glover (2001) rightly put it, exam pairing can support good teaching and can make examination situations and teaching/learning contexts fulfill their complementary roles. This line of reasoning is also in keeping with the tenets of dynamic assessment, which hold that teaching and assessment should not be regarded as distinct activities; rather, they should be considered as fully integrated and at the service of one another (Lantolf & Poehner, 2004).

Arguing within the same lines, Kramsch (1986) asserts there exists a discrepancy between pair/group-based classroom interaction in communicative approaches to L2 teaching and the individually-oriented testing of speaking—a critical point which highlights the need for tapping candidates' IC in oral language assessment by adopting paired/ group format speaking tests, as already discussed.

The present study also set out to investigate the participants' perceptions of the impact of such hotly debated interlocutor variables as candidates' proficiency level, gender, and their acquaintanceship with their test partners on their performance in paired speaking tests.

The first interlocutor variable investigated in the present study was proficiency level. As indicated in Table 5, about 27% of the participants interviewed felt pairing a stronger candidate with a weaker one would more benefit the latter, and that the latter relying on the speech produced by the former can form and organize their own ideas and, thus, cover up their linguistic deficiencies—a finding in line with that of Davis' (2009) who also found that the lower-proficiency level participants produced more language when paired with a higher-level partner.

Homa, one of the participants corroborating this position, stated:

- *I generally think [having] more proficient partner would be more helpful because speaking with a less proficient peer might result in communication breakdown and, accordingly, we cannot extend our discussion. But when your partner is stronger than you, he [/she] will give you some ideas to talk, at least.*

About 8% of the participants interviewed, as stated in Table 5, perceived the situation could more benefit the stronger candidate, simply because comparing

themselves with their weaker peer interlocutor, they would feel more self-confident and comfortable (an argument in favor of condition 3, mentioned in Table 5). Of the 63 candidates interviewed, 15 (nearly 24%) pointed out if their peer interlocutor was stronger than them, it would affect their (i.e., the weaker candidate's) performance negatively (condition two) because the weaker one would not feel psychologically at ease, as remarked by Ziba, a female participant:

- *If your partner is stronger, it can make you feel discouraged and lose your self-confidence.*

Keyvan, another participant, said:

- *Although I am a confident person, I was really affected by the performance of my peer. I believe if your peer is considerably stronger, you may feel stressed and anxious.*

In support of this stance, Sirvan, another participant, remarked:

- *My partner was stronger in FCE exam and talked a lot. It had negative effect on me and I got shy not be[ing] able to catch up with him and I felt anxious during the whole exam. I was very stressful.*

As shown in Table 5, nearly 9.5% of the participants interviewed pointed out something different and stated that having a weaker partner might cause the examiner to have a negative impression of you (i.e., the stronger one) and penalize you, too (i.e., condition 4). Supporting this stance, Habib, one of the interview participants, stated:

- *In my own test experience in FCE, this absolutely affected my performance negatively. I think this is unfair. Why should I be punished for the fault of my partner?*

Of course, this cannot be very well justified, taking the fact into consideration that, at least, in the case of Cambridge ESOL speaking examinations, the speaking examiners are trained, coordinated, monitored, and standardized regularly so much so that they would not punish the given candidate for the low-proficiency level of his or her partner. However, overall, paired/group speaking examiners need to take this point of crucial importance into consideration to guard against test scoring unfairness.

Twenty other participants interviewed (nearly 32%) said that being paired unequally had in practice no effect on their performance, as pointed out by Rostam:

- *In PET exam, my partner was a little stronger than me. She had no effect on me. If you are prepared enough for the test and have good concentration, you can ace the test.*

Thus, as the findings of the study indicated, overall, 68% of the participants interviewed believed that being paired with a peer interlocutor of varying linguistic ability would either advantage or disadvantage them unfairly (i.e., conditions 1-4) and, thus, preferred same-proficiency peer interlocutors—a finding also corroborated by the relatively high mean (i.e., 3.84) of the responses of the participants of all the three groups to question # 8 in the questionnaire, revealing their preference for a peer interlocutor at their own proficiency level in the exam. This is also supported by the results of Bennet (2012) who found that although almost half of the unequally matched candidates were satisfied with their performance and believed it affected them positively; this rate was much higher for the equally-matched pairs, 80% of whom expressed satisfaction and support for pairings of equal linguistic ability.

Based on the findings of the study, it is felt that pairing candidates of unequal linguistic abilities might either advantage or disadvantage their peer interlocutors—an issue which could, ultimately, jeopardize test fairness, which is one of the basic qualities of a test among others like reliability, validity, impact, and practicality. Thus, pairing candidates of equal linguistic abilities would be more ideal and worth doing, as the remarks of the test-takers interviewed indicated. Attempts should, hence, be made to do so as far as practicality issues render it possible to enhance test fairness. However, the problem which might arise is that matching candidates for proficiency level, especially when there is a great pool of candidates to be examined, is easier said than done and is not very much practical.

Concerning the participants' perceptions of the impact of peer interlocutor acquaintanceship on one's test performance, the results indicated that nearly two-thirds of the participants interviewed expressed their satisfaction with having their peer interlocutor from among those they knew which is also confirmed by fairly high mean (i.e., 3.80) of the participants' responses to this question in the questionnaire, indicating the point that they most probably agreed to the idea because a friend peer interlocutor would help one feel relaxed, increase one's self-confidence, make conversation more natural, and motivate one more, as already stated in the interview. In support of this, Arman, one of the interview participants, asserted:

- *I was under stress before taking [the paired] FCE speaking. But as soon as I understood my exam partner was one of my language institute classmates moments before the exam, I got very relaxed and happy. You know, I don't feel at ease talking with strangers especially if the first encounter is in the exam session.*

Quite contrary to the general trend of responses to this question in the interview, Sheida, one of the participants against pairing friends, pointed out:

- *Speaking in front of people who[m] I know makes me stressful because if I cannot speak well in the exam, I'll lose my face in front of my friend. So, it's better for me to have a stranger in the exam.*

Nearly, 22% of the participants expressed their indifference toward the issue, as already mentioned in Table 6. Sohrab said:

- *Basically, it doesn't matter whether I know my peer interlocutor or not because my main focus is on the test and I'm confident about my abilities.*

The overall preference of the participants for friend peer interlocutors is consistent with the findings of similar previous studies, especially with those of O'Sullivan's (2002) and Norton's (2005) that empirically found that the candidates, when paired with a peer interlocutor they knew, had a better performance and achieved higher scores.

However, again, the results should be treated with caution here. The friend pairs' better performance in O'Sullivan's (2002) and Norton's (2005), and other similar studies cannot be solely ascribed to peer interlocutor familiarity; such other factors and personality traits as extroversion/introversion, self-confidence, and so on might have played a role.

As shown in Table 9, nearly 62% of the participants interviewed stated they liked paired speaking tests over solo interviews. The reasons for this mainly related to the relaxing and soothing nature of paired tests, their naturalness, fairness, and so on. Zal, one of the participants interviewed, remarked:

- *I prefer paired format because it decreases my stress. Generally, it is a different experience that makes the situation more friendly and informal which I like.*

Khorshid, another participant, stated:

- *I prefer paired tests as you would call. I took FCE test which I think created a friendly atmosphere for me. Listening to other candidate[s] in paired tests can help you handle the situation more easily and manage your information better.*

Mahtab, another participant who preferred paired tests over solo interviews, maintained:

- *I prefer paired format because it helps a lot and lowers the amount of stress or pressure that you might feel.*

Esfandiyar, another participant favoring paired test, said:

- *Paired format test is my choice because this format gives me more time to think and prepare myself.*

Nearly 17.5% of the participants expressed their preference for solo interviews. Arsham, supporting this stance, remarked:

- *I prefer solo interviews definitely because if the other candidate in the [paired] test gets stressed, I will also lose my control and self-confidence.*

Zari, another participant in favor of solo interviews, said:

- *I've experienced both, but I prefer solo. In paired tests, I become scared of making [a] mistake in front of my friends. So, I am more comfortable in solo tests because nobody will judge my performance except the interviewer.*

As for the participants' perceptions of the impact of gender on paired speaking test performance, the results indicated that nearly 16% of those interviewed were for having opposite-sex peer interlocutor in the exam because they felt this would make the exam more challenging and exciting and create a sense of competition in them to struggle for success, as pointed out by one of the participants. Also, 24% of the participants interviewed stated gender made no difference in their performance. Rahim, a male participant supporting this position, stated:

- *To be honest, it doesn't make any difference. When you are confident and able to adapt yourself to different situations, it doesn't matter whether your partner is from the same or opposite sex, is a friend or a stranger, or higher or lower than you [i.e., is at the same or different proficiency level].*

However, 60% of the participants and, interestingly, about three-fourth of the females interviewed preferred the same-sex peer interlocutor—a point also supported by the relatively high mean (i.e., $M = 3.43$) of the responses of the participants to the corresponding question (question # 7) in the questionnaire. One of the main reasons for this seems to be cultural in nature. That is, in most Middle Eastern communities including Iran, some females could still be found who do not yet feel at ease when talking to a stranger male—something which might be more of a problem for (secondary) school students who have experienced only segregated education at school. Echoing this, Azar, one of the female interviewees, a senior secondary school girl, asserted:

- *I could not perform to my potential in FCE because I was paired with a stranger boy. He mostly interrupted me and continued to speak nonstop which got on my nerves and caused me not to gain the score I expected and deserved in speaking. When I daily talk to my [same-sex] friends at my language institute, I speak much better.*

Tina, another female participant, said:

- *I had my peer interlocutor from my own sex and I liked it because I feel shy with the opposite sex and get worried of making mistakes during speaking because it causes me [to] lose my face.*

Siavash, a male participant corroborating this stance, stated:

- *In fact, [having an exam partner from] my own sex is better because I usually get shy and act awkward in front of women which might affect my score.*

Tahmineh, a female participant having experienced taking FCE, remarked:

- *In the context of Iran where segregated education is being practiced, it creates a stressful situation in the exam session when your pairmate is from the opposite sex.*

6. Conclusion and Implications

The findings showed that there was an overall preference for paired speaking tests in all the three groups. It was also found that the participants, overall, preferred the same-proficiency-level, same-sex, and friend peer interlocutors.

The findings might firstly confirm the candidate-friendly nature of paired speaking tests (including those of Cambridge ESOL exams) and may corroborate their effectiveness in tapping communicative abilities of the learners, especially their IC and, thus, encourage use of pairing in local speaking exams in L2 contexts like Iran in which the bulk of language speaking tests, both at language institutes as well as at the university level, with students majoring in L2s, are administered in the form of solo interviews. That is, I firmly believe that paired testing should be one of the alternatives for tapping learners/test-takers' oral language abilities, along with such other formats as solo interviews because, as Roever and Kasper (2018) rightly argue, the incorporation of IC in oral language assessment opens new horizons on the speaking proficiency construct definition and assessment and increases the validity of oral language assessments.

Secondly, because as indicated by the results of the present study and the findings of other similar studies, the preference for "who to be paired with" in paired format speaking tests seems to be culture- and context-specific, due to a multitude of contextual, social, cultural, and so on factors involved. That is, such interlocutor variables as candidates' proficiency level, their gender, their familiarity with their peer interlocutors, and so on should be taken into consideration in pairing candidates in speaking tests by test administrators, as long as these variables can be practically handled and implemented of course; factors of crucial importance which might, otherwise, affect test performance either positively or negatively in, at least, such

certain cultures and contexts as that of the present study and, thus, jeopardize test fairness.

Therefore, test administrators, for instance, can seek test-takers' opinions on their preference for (speaking) test partnership with same/different gender, friend/stranger, and the like pairmates (by, for example, administering them a brief questionnaire) at the time they are enrolled in and/or entered for the given exam. In the case of Cambridge ESOL paired speaking tests, apart from pairing for the same-sex pairmate which, naturally speaking, can be easily attended because most of test-takers are internal candidates of the centers, the center exams managers or supervisors can simply pair test-takers with their classmates (i.e., friend pairmates) if they prefer so, of course, based on their records in the centers, something which can also be applied to other real-life testing situations. In so doing, attempts must, of course, be made to safeguard test-fairness. Test developers might also be recommended to incorporate these in their test administration materials and guidelines (i.e., examination handbooks, instructions to speaking examiners, etc.) and raise test administrators and speaking examiners' awareness of these potential problems and possibilities.

Also, pedagogically speaking, adopting paired/group oral tests, L2 teachers can kill two birds with one stone: They can both test their students' speaking abilities more authentically, and, at the same time, encourage them to boost their speaking abilities more naturally.

Last but not least, as Galaczi (2014) holds, because test-takers are, at the first-place learners, a meticulous illustration of IC by test developers in terms of "both construct definitions and rating scales" (Brooks, 2009, p. 341) is of crucial importance for both teachers and learners because it can provide a framework for developing interactional abilities in L2 teaching class contexts.

That being said, caution should be exercised in interpreting the results of the present study: First, because the findings of research on the impact of such interlocutor variables and personality factors as proficiency level, gender, pairmate acquaintanceship, and the like on paired/group speaking test performance are, sometimes, mixed, context-dependent, and culture-specific (O'Sullivan, 2002), still more studies in various cultures and testing contexts need to be carried out before we can claim with certainty the superiority of paired/group speaking tests over their counterparts like solo interviews. Second, as Nakatsuhara (2011) maintains, the number of test-takers sitting a speaking test might cause them to produce different language discourse which might be underlined by different constructs. Put simply, different test formats might assess different constructs. Paired tests, for instance,

mainly tap test-takers' IC, which might be different from individually-based L2 competence.

In conclusion, it could be argued that because paired/group orals have been found (in this study and other similar studies) to be more test-taker-friendly and because the curriculum has already started to concentrate more on peer-to-peer interactive communication in real-life situations, paired tests should be added to the testing programs and classroom-based assessments. That is, as rightly put by Ahmadi and Sadeghi (2016, p. 353), test developers, test users, and L2 teachers should "avoid a sole reliance on interview-formatted tests to measure oral proficiency"; rather, they are suggested to use different speaking test formats to provide test-takers with ample and, of course, more realistic opportunities to perform to their best. In other words, paired format speaking tests should be added, as an alternative, to the assessment programs and classroom-based teaching/testing contexts that lack them and that they should be attached more significance by L2 assessment policymakers, test developers and administrators, and teachers, simply because, as Ockey (2009) maintains, paired/group speaking tests can still provide "a reasonably solid basis on which to estimate students' oral ability" (p. 105), and are, thus, worth more attention, investigation, documentation, and implementation.

Note

For ethical considerations and confidentiality purposes, all the names used throughout the study as the participants' names are pseudonyms and not their real names.

References

- Ahmadi, A., & Montasseri, Z. (2019). Interactional competence in paired vs. group oral tests. *Teaching English Language*, 13(1), 1-26.
- Ahmadi, A., & Sadeghi, E. (2016). Assessing English language learners' oral performance: A comparison of monologue, interview, and group oral test. *Language Assessment Quarterly*, 13(4), 341-358.
- Bennet, R. (2012). Is linguistic ability variation in paired oral language testing problematic? *ELT Journal*, 66(3), 337-346.
- Berry, V. (2007). *Personality differences and oral test performance*. Frankfurt am Main: Peter Lang.
- Bonk, W., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing*, 20(1), 89-110.
- Bonk, W. J., & van Moere, A. (2004). *L2 group oral testing: The influence of shyness/outgoingness, match of interlocutors' proficiency level, and gender on*

individual scores. Paper presented at the annual meeting of the Language Testing Research Colloquium, Temecula, California.

- 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.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency. *Language Testing*, 26(3), 341-366.
- Brown, A., & McNamara, T. (2004). "The devil is in the detail": Researching gender issues in language assessment. *TESOL Quarterly*, 38(3), 524-538.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Chambers, L., Galaczi, E. D., & Gilbert, S. (2012). Test-taker familiarity in speaking tests: Does it make a difference? *Research Notes*, 49, 33-40.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing*, 26(3), 367-396.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Cambridge: Cambridge University Press.
- Együd, G., & Glover, P. (2001). Oral testing in pairs. *ELT Journal*, 55(1), 70-76.
- Fox, J. (2004). Biasing for the best in language testing and learning: An interview with Merrill Swain. *Language Assessment Quarterly*, 1(4), 235-251.
- Fulcher, G. (2015). Assessing second language speaking. *Language Teaching*, 48(2), 198-216.
- Galaczi, E. D. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly*, 5(2), 89-119.
- Galaczi, E. D. (2010). Paired speaking tests: An approach grounded in theory and practice. In J. Mader & Z. Ürkun (Eds.), *Recent approaches to teaching and assessing speaking*. IATEFL TEA SIG Conference Proceedings. Canterbury: IATEFL Publications.
- Galaczi, E. D. (2014). Interactional competence across proficiency levels: How do learners manage interaction in paired speaking test? *Applied Linguistics*, 35(5), 553-574.

- Galaczi, E. D., & Taylor, L. (2018). Interactional competence: Conceptualizations, operationalizations, and outstanding questions. *Language Assessment Quarterly*, 15(3), 219-236.
- Humphry-Baker, A. (2000). *Speaking test: Students' perception and performance*. Unpublished MEd thesis, University of Manchester, Manchester.
- Iwashita, N. (1998). The validity of the paired interview in oral performance assessment. *Melbourne Papers in Language Testing*, 5, 51-65.
- Jacoby, S., & Ochs, E. (1995). Coconstruction: An introduction. *Research on Language and Social Interaction*, 28(3), 171-183.
- Joo, M.-J. (2019). Interaction types of students in English-speaking tests. Focused on the impact of personality. *International Journal of Innovative Technology and Exploring Engineering*, 8(3C), 296-303.
- Kramsch, C. (1986). From language proficiency to interactional competence. *The Modern Language Journal*, 70, 366-372.
- 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.
- Lantolf, J. P., & Poehner, M. E. (2004). Dynamic assessment of L2 development: Bringing the past into the future. *Journal of Applied Linguistics*, 1, 49-72.
- Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge: Cambridge University Press.
- Leaper, D. A., & Brawn J. R. (2019). Detecting development of speaking proficiency with a group oral test: A quantitative analysis. *Language Testing*, 36(2), 181-206.
- Leaper, D. A., & Riazi, M. (2014). The influence of prompt on group oral tasks. *Language Testing*, 31(2), 177-204.
- May, L. A. (2009). Coconstructed interaction in a paired speaking test: The rater's perspective. *Language Testing*, 26(3), 397-422.
- May, P. (1996). *Exam classes*. Oxford: Oxford University Press.
- McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Malden, MA: Blackwell Publishing.
- Nakatsuhara, F. (2006). The impact of proficiency level on conversational styles in paired speaking tests. *Research Notes*, 25, 15-20.

- Nakatsuhara, F. (2011). Effect of test-taker characteristics and the number of participants in group oral tests. *Language Testing*, 28(4), 483-508.
- Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pretask planning effects on paired oral test performance. *Language Testing*, 31(2), 147-175.
- Norton, J. (2005). The paired format in Cambridge speaking testes. *ELT Journal*, 59(4), 287-297.
- Ockey, G. J. (2009). The effects of group members' personalities on a test-taker's L2 group oral discussion test scores. *Language Testing*, 26(2), 161-186.
- O'Sullivan, B. (2002). Learner acquaintanceship and oral-proficiency test pair-task performance. *Language Testing*, 19(3), 277-295.
- Palumbo, M. V., & Steele-Johnson, D. (2014). Do test perceptions influence test performance? Exploring stereotype threat theory. *North American Journal of Psychology*, 16(1), 1-12.
- 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.
- Sadeghi, K., Azad Mousavi, M., & Javidi, S. (2017). Relationship between EFL learners' self-perceived communication competence and their task-based and task-free self-assessment of speaking. *Journal of Research in Applied Linguistics*, 8(2), 31-50.
- Sandlund, E., Sundqvist, P., & Nyroos, L. (2016). Testing L2 talk: A review of empirical studies on second language oral proficiency testing. *Language and Linguistics Compass*, 10(1), 14-29.
- Saville, N., & Hargreaves, P. (1999). Assessing speaking in the revised FCE. *ELT Journal*, 53(1), 42-51.
- Simpson, J. (2006). Differing expectations in the assessment of the speaking skills of ESOL learners. *Linguistics and Education*, 17, 40-55.
- Song, B., & Lee, H. (2016). Interactional competence in a paired speaking test: Toward designing an evidence-based rating scale. *English Language Teaching*, 28(2), 133-152.
- Taylor, L. (2000). Investigating the paired speaking test format. *Research Notes*, 2, 14-15.

- Taylor, L. (2001). The paired speaking test format: Recent studies. *Research Notes*, 6, 15-17.
- van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing*, 23(4), 411-440.
- Vidacović, I., & Galaczi, E.D. (2013). The measurement of speaking ability 1913-2012. In C. J. Weir, I. Vidacović, & E.D. Galaczi (Eds.), *Studies in language testing: Measured constructs* (pp. 257-346). Cambridge: Cambridge University Press.
- Wallis, D. (1995). *Testing in twos—The paired format in oral examinations*. UCLES presentation, TATEFL Annual Conference, York.
- Weir, C., & Taylor, L. (2011). Conclusions and recommendations. In L. Taylor (Ed.), *Examining speaking, research, and practice in assessing second language speaking* (pp. 293-313). Cambridge: Cambridge University Press.
- Young, R. F. (2008). *Language and interaction*. New York: Routledge.

Appendixes

Appendix A Pilot Study Questionnaire

Directions: Please read the following questions carefully and choose the number that best suits your opinion about paired-format speaking test (e.g., FCE and PET) and/or solo interview (e.g., IELTS) you sat. Remember 1 (*Completely Disagree*), 2 (*Disagree*), 3 (*No Idea*), 4 (*Agree*), and 5 (*Completely Agree*):

1. Paired format is a better and more enjoyable experience than when I have to talk only to the examiner during the exam.
1 2 3 4 5
2. Paired format is good because the other candidate can help me and/or I can help him or her, too.
1 2 3 4 5
3. I think paired format is good because it eases the tension and makes you relaxed.
1 2 3 4 5
4. Paired format is good but it depends on who you are examined with (i.e., friend pairmate is better).
1 2 3 4 5

5. Being examined in pairs is more similar to real-life situations and looks authentic.

1 2 3 4 5

6. Paired format has good instructional effect; that is, it echoes the type of pair and group communication activities occurring in classroom contexts.

1 2 3 4 5

7. In general, paired-test format is much better and fairer because you are assessed by two examiners.

1 2 3 4 5

8. What other advantages or disadvantages do you think **paired format** speaking **tests** might have over solo interviews? Please elaborate.

9. What advantages or disadvantages do you think **SOLO INTERVIEWS** have over paired format tests? Please elaborate.

Thank You!

Appendix B

Factor Analysis Results

Factor Loadings for the Rotated Factors

Item	Factor Loadings		Communality
	1	2	
1	0.54		.62
2	0.64		.46
3	0.80		.72
4	0.51		.67
5	0.69		.49
6	0.75		.58
7	0.59		.43
8	0.59		.57
9	0.58		.60
10	-0.52	0.55	.68
11	0.41	0.72	.70
Eigenvalues	4.20	1.33	
% of variance	38.13	12.15	

Note. Loadings < .40 are omitted.

Appendix C

Questionnaire of the Study

Directions: Please read the following questions carefully and choose the number that best suits your opinion about paired-format speaking test (e.g., FCE, CAE, PET, etc.) and/or solo interviews (e.g., IELTS) you sat. Remember 1 (*Completely Disagree*), 2 (*Disagree*), 3 (*No Idea*), 4 (*Agree*), and 5 (*Completely Agree*):

1. Paired format is a better and more enjoyable experience than when I have to talk only to the examiner during the exam.

1 2 3 4 5

2. Paired format is good because the other candidate can help me and I can help them too.

1 2 3 4 5

3. I think paired format is good because it eases the tension and makes you relaxed.

1 2 3 4 5

4. Paired format is good but it depends on who you are examined with (i.e., if you are paired with a friend or an acquaintance in the exam, it is much better than when you are paired with a stranger).

1 2 3 4 5

5. Being examined in pairs is more similar to real-life situations and looks more authentic.

1 2 3 4 5

6. Paired format has good instructional effect; that is, it echoes the type of pair and group communication activities occurring in classroom contexts.

1 2 3 4 5

7. I think if I am paired with somebody from my own sex in the exam, I can perform better.

1 2 3 4 5

8. I prefer to have a peer interlocutor at my own proficiency level.

1 2 3 4 5

9. In general, paired format test is much better and fairer because you are assessed by two examiners, rather than one as in IELTS or TOEFL.

1 2 3 4 5

10. I feel if I perform weakly in the exam session in a paired-format test, I lose my face and self-confidence in front of the other candidate, which might affect my score negatively.

1 2 3 4 5

11. I think in comparison to solo interviews, in paired-format speaking tests, it becomes difficult for me to manage everything on my own and, thus, my performance might be affected negatively.

1 2 3 4 5

Thank You!

Appendix D

Semistructured Interview

Directions: Please answer the following questions carefully and honestly. I promise not to take more than 15 min of your precious time.

6. If your partner was linguistically weaker or stronger in paired speaking tests like PET, FCE, and the like you sat, did it affect your performance negatively or positively? Why?
7. Did you wish to have your peer interlocutor from among your friends and classmates or somebody whom you knew, or from strangers or somebody whom you didn't know at all? Why? How did this affect your performance in the exam you took?!
8. Did you wish to have your peer interlocutor from your own sex or from the opposite sex? Why?
9. In general, which one would you prefer for oral/speaking tests: paired-format, that is, two candidates with two examiners as in FCE, or a solo interview, that is, one candidate and one examiner as in IELTS? Why? Why not?

Thank you very much!