THE RELATIONSHIP BETWEEN TEST-TAKERS' L1, THEIR LISTENING PROFICIENCY AND PERFORMANCE ON PAIRED SPEAKING TESTS

AR-A/2015/2

Suwimol Jaiyote, University of Bedfordshire

ABSTRACT

This report presents a study of the relationship between test-takers' L1, listening proficiency and their performance on paired speaking tests. Forty participants from two different L1 backgrounds (20 Urdu L1 speakers and 20 Thai L1 speakers) participated in the study. They took two paired speaking tests: one with a shared L1 partner, and one with a non-shared L1 partner, as well as a listening test and a monologic speaking test to measure their listening and individual speaking ability. After each paired speaking test, the participants were also interviewed about their test-taking experience. All speaking tests and interviews were video-recorded and transcribed. Raters awarded test-takers' analytical speaking test scores (grammar and vocabulary, discourse management, pronunciation and interactive communication) and they also provided written comments on reasons for their scores. Additionally, raters participated in retrospective interview sessions. A mixed-methods approach was utilised to analyse and triangulate different kinds of data. The data analysed in this study were listening and speaking test scores, the raters' retrospective interviews on their perception of the test-takers' speaking performance and written comments on the reasons for awarding the test-takers' speaking test score, the test-takers' retrospective interviews, and the interactional discourse data in the paired speaking formats. Quantitative analysis, Conversation Analysis (CA) and other qualitative analyses were used to inform the relationship between test-takers' listening proficiency, their L1 and their paired speaking performance.

The results showed that there was no statistically significant difference in the test-takers' paired speaking test scores between the two types of pairing (i.e., shared L1 pairs and non-shared L1 pairs). Test-takers' listening test scores significantly correlated with their speaking test scores in the non-shared L1 pairs, while the test-takers' listening proficiency did not matter in the shared L1 pairs. In the non-shared L1 pairs, the greater the listening proficiency that test-takers had, the higher scores they tended to receive in the grammar and vocabulary and discourse management categories. Similarities and differences in communication patterns related to interactive listening between the shared L1 and the non-shared L1 pairs are presented. Some interactional features that illuminate the differences between the shared and non-shared L1 pairs are also described. As pairing with a shared and non-shared L1 partner has both pros and cons, it was recommended to adopt Swain's (1983) notion of "bias for best" in paired speaking test practice. Additionally, recommendations are made for: the further development of a scale of interactive communication category for the paired speaking tests; fairness in awarding each test-taker's performance in paired work; and paired speaking tests in pedagogical settings.

Author

Suwimol Jaiyote is a PhD student at the Centre for Research in English Language Learning and Assessment (CRELLA), University of Bedfordshire. She holds the degree of Master of Arts (English). She has a bachelor degree in Public Administration. She is a lecturer at Naresuan University, Thailand.

Acknowledgements

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

The increasing use of the paired formats has emphasised the fact that "the view of oral test performance as interactive, so central to much current work, means that it is difficult to consider the impact of test-taker characteristics in isolation from those of interlocutors" (McNamara et al 2002, p. 228). McNamara (1996, p. 86) states that awarding speaking performance in paired formats is complicated because test-taker performance is co-constructed with the candidate's underlying competence and other sources, as presented in Figure 1.

Figure 1: 'Proficiency' and its relations to performance (McNamara, 1996, p. 86)
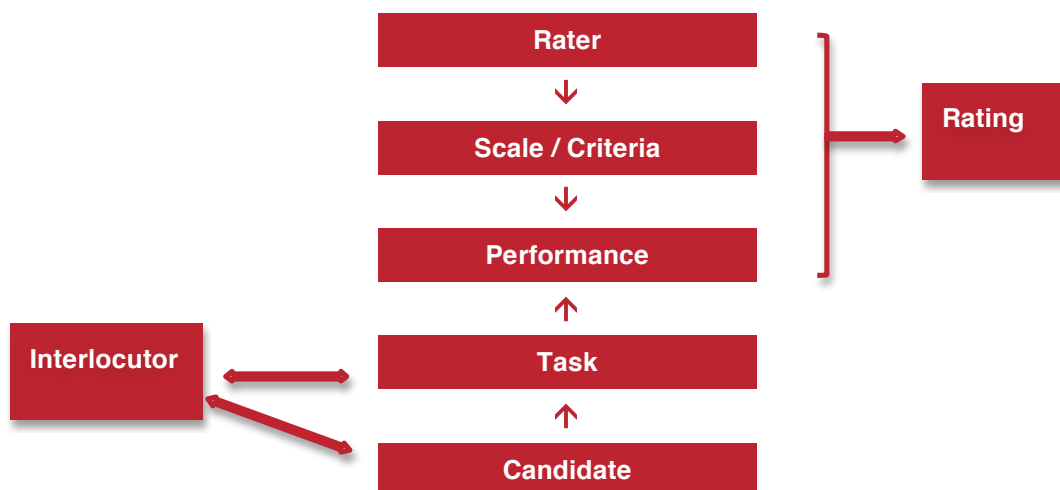
In a testing situation, whom one is paired with is very important (Swain 2001). In the paired formats, Weir (2005, p. 153) states that "an individual's performance is clearly affected by the way the discourse is co-constructed by the person they are interacting with". Therefore, how to pair test-takers should be considered and carefully and appropriately conducted. In neglecting the possible factors which can affect the test-takers' performance, the test-takers might be treated unfairly in the assessment and destroy the validity of the test, as concerns some researchers, e.g., Foot (1999).

Many studies related to co-constructed interactions in assessment of speaking performance in paired or group formats disclose that the following variables could affect the test performance: the test-taker's own and their partner's characteristics, for instance, personality (Berry 1993, 1997, 2007; Ockey 2009, 2011), language proficiency (Iwashita 1998; Norton 2005; Nakatsuhara 2006; Davis 2009), gender and acquaintanceship/familiarity (O'Sullivan 2002; Norton 2005), age (O'Sullivan 2008) and L1 (Lu 2010; Jenkins 1997, 2002). However, there is a lack of studies to examine the correlation of these variables – test-takers' L1 and their listening proficiency – on speaking performance in paired formats in a systematic way. This study aims to fill this gap by investigating the impact of these variables in paired formats and controlling other variables as much as possible (e.g., age range, gender, overall English proficiency, English speaking proficiency and English listening proficiency) which might affect the research findings.

English is widely used around the world and English is viewed as "a lingua franca" (Jenkins 2000) or "an international language" among non-native speakers (NNSs) of English around the world. However, many sociolinguists have talked about "Englishes" or "World English" (Kachru 1992) since in many countries of the "*outer circle*" (e.g., India, and Singapore), English is mainly used in daily life and they have their own standards in English. As there is a growing number of standard varieties of English – not only traditionally recognised standard versions (e.g., standard British and standard American English) as in the past – more research attention has been attracted to the way people in different L2 locations speak with impacts of local language and culture, in terms of "its characteristic accents, its syntactic structures, its lexis, its pragmatic features, and the like" (Jenkins 2006, p. 42).

In 1998, Kachru, who developed the model of the three concentric circles of English, adapted it to the Asian context. The "*inner circle*" is represented by the example of Australia and New Zealand, where people primarily use English as a first language. The "*outer circle*" is exemplified by the case of India and Pakistan, where English functions as an institutionalised language or foreign language, and the "*expanding circle*" is represented by the example of Thailand and China, where English is primarily utilised as a foreign language. Certain characteristics are shared by these three circles; especially that all varieties of English are transplanted and comprise the formal and functional distinctiveness of the varieties of English in Asia (Kachru 1998, p. 93).

As English is used in international contexts, and English language users or learners in different regions speak English with their own accent, there are some concerns on how well NS-NNS and NNS-NNS comprehend each other message and achieve their interactional goal. People in each region possess their own accent and this tends to affect their L2 pronunciation. Even when native speakers hear a new accent for the first time, they may take a little time to get used to it and understand it. This is more problematic for L2 listeners when they converse with a speaker whose accent is unfamiliar for them (Buck 2001). It can cause problems and possibly disrupt the whole process of comprehension followed by communication breakdown.

There are a number of studies providing evidence to support impacts of English spoken by non-shared L1s on different L1 listeners' comprehension in some contexts. Non-native listeners who had the same L1 background with speakers are more able to comprehend the speakers' speech than listeners who had different L1 background with the speakers (Bent and Bradlow 2003; Kachi 2004; Harding 2012). In addition, the listeners who shared L1 background with the speakers seem to get more advantage in comprehending the speaker's message in terms of ability in inferring what the speakers intend to say based on the linguistic and cultural background knowledge that they shared (Kachi 2004). The other aspect related to the influences of L1 on L2 listening comprehension is accent. People from different L1 backgrounds tend to have different accents. Accent is a powerful factor which affects listeners' comprehension. Different accents can cause difficulty in their L2 listening comprehension (Stibbard and Lee 2006; Harding 2012; Ockey and French 2014). A listener who has the same L1 background with a speaker tends to understand the talk more easily. In contrast, a listener who does not share an L1 background with a speaker seems to have some difficulty in understanding the speaker, and the process of listening comprehension is longer than for the listener who shares L1 with the speaker. The stronger the L2 speaker's accent, the less listening comprehension of the L2 listener there is (Ockey and French 2014).

Although sharing L1 background with the speaker seems to be an advantage for listener, it is not always the case. Some studies disclose that listeners do not always gain an intelligibility benefit in L2 spoken by the speakers who share L1 with them (Major et al 2002; Algethami et al 2011). However, since many studies reveal that L1 backgrounds of the speakers and listeners seem to significantly affect L2 listening comprehension, therefore, it is essential to investigate whether non-native test-takers' L1 background affects their L2 interactive listening and their speaking performance in paired speaking test formats.

There has been no study, to the researcher's knowledge, which systematically looks at the effects of L1 on test-takers' listening comprehension in paired speaking tests. Therefore, this study fills the gap by examining both test-takers' speaking-listening proficiencies and L1 factors in the paired formats and identifying interactive listening-related communication patterns in shared L1 pairs and non-shared L1 pairs. It is hoped to gain a better understanding of how L1 and listening proficiency of test-takers are correlated to their paired speaking performance; whether pairing with shared and non-shared L1 partners provide any similarities and differences in communication pattern; and whether shared L1 test-taker pairs comprehend each other better than non-shared L1 test-taker pairs.

Findings of the current study will shed light on paired interaction discourse between shared and non-shared L1. It is also hoped to be beneficial for classroom assessment that consists of international students, and language testing both high- and low-stakes to use paired test formats to cautiously consider matching students/test-takers with a shared or non-shared L1 partner for validity and fairness of the paired tests.

2. RESEARCH QUESTIONS

RQ 1: To what extent is test-takers' performance on paired speaking tests affected by their listening proficiency?

RQ 2: Are there any differences between shared L1 pairs and non-shared L1 pairs in terms of the impact of their listening proficiency on performance in paired oral tests?

RQ 2.1: Are there any differences in speaking test scores when test-takers are paired with shared L1 as compared to (when they are paired with) non-shared L1 partners?

RQ 2.2: What are the similarities and differences in communication patterns between shared L1 pairs and non-shared L1 pairs?

3.   METHOD

3.1   Research design

This study used a mixed method approach in which qualitative and quantitative data were collected in parallel, analysed separately, and then merged. The mixed method approach was utilised in order to gain greater insight into the issue under investigation than using either qualitative or quantitative analysis alone (Dörnyei 2007; Creswell and Plano Clark 2011).

3.2   Participants

The participants for the study consisted of 40 students and two speaking test raters.

Forty pre-sessional English programme students from Urdu (Pakistani language) and Thai (Thai language) L1 backgrounds (10 males and 10 females from each L1 background) participated in the study.

Two female English native speakers, with experience in language assessment, were the speaking test raters and were trained before awarding test-takers' speaking scores.

3.3   Research instruments

The research instruments used in this study consisted of:

- questionnaire
- listening test
- speaking tests (a monologic task and two paired speaking tasks)
- retrospective verbal interview with test-takers and speaking test rater.

3.3.1   Questionnaire

A questionnaire was used to gather information concerning the test-takers' demographic background, their listening-speaking proficiency based on a standardised examination and their familiarity with English spoken by shared and non-shared L1 speakers. The data gathered was used to match test-takers with a partner with similar speaking and listening proficiency levels for the paired speaking tests. Since gender was also controlled in this study, test-takers with the same gender and similar speaking and listening proficiency based on a standardised examination were paired together.
A sample of the questionnaire is illustrated in Appendix 1.

3.3.2   Listening test

The listening test was devised based on questions derived from both the listening test sections of the Cambridge's Preliminary English Test (PET) and the First Certificate in English (FCE) examinations. It consisted of 39 test items, each item being worth one mark.

- 19 test items were taken from PET practice material published by the University of Cambridge ESOL Examinations (Cambridge ESOL 2008). These consisted of multiple choice questions (13 test items) and yes/no questions (6 test items).

- 20 test items were derived from the FCE practice material published by the University of Cambridge ESOL Examinations (Cambridge ESOL 2009). These consisted of multiple choice questions (15 test items) and multiple matching tasks (5 test items).

To ensure that it would give consistent results, the reliability of the listening test needed to be established. Reliability is one feature that is required for the quality of the test. A reliable test is consistent and dependable (Brown 2004, p. 20). Prior to the main study, the listening test was trialled with 30 students (eight male Urdu, seven female Urdu, seven male Thai and eight female Thai L1s) who were in a pre-sessional English language programme at the University of Bedfordshire.

The listening test consisted of dichotomously score items; therefore, Cronbach's alpha was utilised to investigate the reliability of the listening test. The reliability coefficient of the 39 listening test items was .90. According to McNamara (2000, p. 62) a reliability coefficient of 0.90 or better is what we normally look for on a comprehension test.

Although the alpha value seemed acceptable, the discrimination levels of some items were inadequate. The item discrimination is "the extent to which an item differentiates between high- and low- ability test-takers" (Brown 2005, p. 59). Values of item discrimination (discrimination index) level of .20 or higher ($r \geq .20$) were acceptable to measurement. An item with an item discrimination value of less than .20 means it cannot discriminate between strong and weak test-takers. Hence the test items with the values of item discrimination (discrimination index) level lower than .20 ($r < .20$) were discarded from the listening test. Two test items were discarded from the listening test. The reliability of the listening test (37 test items) increased to .91.

3.3.3   Speaking tests

Three speaking test tasks made up the speaking test for the study: a monologic speaking task and two paired speaking tasks (task A and task B). All the tasks (both monologic and paired) were based upon the Part 3 collaborative task used in the Cambridge First Certificate in English (FCE) speaking test. The tasks were taken from the practice material published by the University of Cambridge ESOL Examinations (Cambridge ESOL 2009). The Part 3 collaborative task in the FCE speaking test was used in this study because this task focuses on "sustaining an interaction, exchanging ideas, expressing and justifying opinions, agreeing and/or disagreeing, suggestion, speculating, evaluating, reaching a decision through negotiation, etc." (Cambridge ESOL 2009). In order to achieve these language functions, the task requires test-takers to interact with their partners while shifting between speakers' and listeners' roles. The FCE was deemed appropriate to test the English proficiency of students aiming to enter a university in the UK.

The monologic speaking task was adapted from the FCE Part 3 collaborative task, rather than using monologic tasks such as the FCE Part 2 individual long turn. The adapted collaborative task was considered to be more suitable as a monologic task in this study, because the monologic task had to be as equivalent as possible to the paired speaking tasks in terms of topical and linguistic demands, to enable meaning comparisons between the results of the monologic and paired speaking tasks.

The original FCE Part 3 collaborative task was modified in a way that test-takers were required to respond to two questions on their own for two minutes, instead of discussing them with their partners. The two questions (i.e., *How important are these things for a happy life? Which two are the most important?*) were accompanied by seven pictures. Both questions and all visual stimuli were shown on a big card (1.5 feet x 3 feet). This adaptation allowed raters to assess purely test-takers' monologic speaking proficiency without any interference relating to test-takers' listening proficiency which could result from interaction with a partner. There was only one version of the monologic speaking task and it was used with all test-takers.

The paired speaking tasks were designed based on the original FCE Part 3 collaborative task. Two versions of the paired speaking task were prepared so that one could be used for the shared L1 condition and the other could be used for the non-shared L1 condition. This means that three different test tasks were prepared (one monologic and two paired speaking tasks) with equivalent topical and linguistic features. The two paired speaking tasks consisted of two questions and several pictures. The questions were:

> Task A: *How difficult is it to be successful in these professions?*
> *In which profession is it most difficult to get to the top?*

> Task B: *What are the advantages of having friends?*
> *In which situation are friends most important?*

Each task was illustrated on a 1.5 feet x 3 feet card. It was shown on a table and each pair shared the same card when performing their paired speaking test. The order of the paired speaking task prompts was counter-balanced to control for a potential prompt effect. The paired speaking tests were employed to assess candidates' interactive speaking proficiency potentially involving listening proficiency.

3.3.4 Retrospective verbal interview

The retrospective verbal interviews are retrospective techniques used to examine language learners or test-takers' cognitive processes, and the thoughts or feelings they had while performing a test task or activity (Gass and Mackey 2000). They are carried out immediately after language learners/test-takers have finished the task or activity by utilising audio- or video-recordings of the language learners or test-takers as a stimulated recall tool. In this study, the retrospective verbal interview was employed with both test-takers and raters. All test-takers were asked to take part in a retrospective verbal interview one by one which was conducted immediately after they finished each paired speaking test. A video recording of test-takers' paired speaking performance was used to stimulate their memory while performing the paired speaking test. The interviews aimed to understand better the test-takers' interactive communication manners related to interactive listening, for instance, not responding to a question and producing back-channelling while listening to a partner (checking whether test-takers really comprehended their partner or just wanted the partner to continue speaking).

Additionally, two raters were individually interviewed about their reasons for awarding scores on selected speaking test sessions which helped to justify those scores and to gain greater understanding of raters' thoughts which might not be elaborated on the speaking rating form when awarding the test-takers' speaking scores on each analytical category. All interviews with test-takers and raters were audio-recorded and transcribed orthographically.

3.4   Data collection

The researcher contacted staff from the Language Centre at the University of Bedfordshire to ask for help in accessing information about pre-sessional English students' backgrounds (e.g., L1, country of origin and English proficiency based on the standardised examination). After gaining access to students' background information, the researcher selected participants by using the purposive sampling method. The researcher collected data from 20 Urdu L1 (10 males and 10 females) and 20 Thai L1 (10 males and 10 females) speakers.

Before starting the data collection, test-takers were asked to sign a consent form to demonstrate that they agreed to participate in the study. After that, all test-takers were asked to complete the questionnaire. The test-takers' background information from the questionnaires was used by the researcher to match them with a partner of a similar background (e.g., gender, age range and English speaking proficiency) in the paired speaking tasks.

The researcher coded the test-takers by employing P for Pakistani and T for Thai and using a number after P and T to identify each test-taker, e.g. P01, T01, P02, T02…P20, T20. Test-takers were divided into 10 groups based on their background information from the questionnaire. Each group consisted of four test-takers, who had the same gender, similar age range and similar English speaking and listening proficiencies of which two were Thai L1 speakers and two were Urdu L1 speakers. It should be noted that because of a large number of test-takers (40) in the study, the listening test, and the monologic speaking test were administered on the same day and the two paired speaking tests were administered on over a period of two and a half day.

Test-takers did the listening test following by the monologic test on the same day. Both tests were administered in a quiet classroom. Before the listening test began, all the test-takers were informed about the listening test format and given instructions in both oral and written formats. At this stage, they were allowed to ask questions if anything was unclear. During the listening test, the test-takers were not allowed to speak. The test-takers listened to audio recordings of each listening test part twice. The listening test took approximately 45 minutes.

After all test-takers finished the listening test, test-takers did the monologic test one by one. The monologic test was administered in a closed quiet room with only the researcher and a test-taker present. Other test-takers waited outside the room. Each test-taker was given the instructions for doing the monologic test in oral and written formats before the test started. Firstly, a test-taker was asked to introduce him/herself briefly (approximately one minute) as a warm-up activity and then he/she was given a prompt card for the monologic task and asked to provide individual long turns responding to the task prompt. This activity lasted for two minutes. There was no planning time before speaking. If the test-taker did not finish talking within two minutes, he/she was asked to stop speaking. After completing the task, the test-taker was asked to leave the room and then the next test-taker was invited to the room to do the monologic speaking test.

The paired speaking tests (both tasks) were administered from the next day. As there were 40 test-takers, the researcher collected data over two and a half days. On each of the first two days, 16 test-takers participated; two groups (four shared L1 and four non-shared L1) in the morning and the next two groups (four shared L1 and four non-shared L1) in the afternoon. On the third day, eight test-takers (four shared L1 and four non-shared L1) were tested. Test tasks and order of shared/ non-shared L1 were counter-balanced.

Before each paired speaking test started, test-takers were asked to introduce themselves to their partner (for the paired tasks) briefly (approximately one minute) as a warm-up activity. There was no planning time prior to the paired speaking tasks. All paired speaking test sessions were audio recorded and video recorded.

Following each paired speaking test, each test-taker was individually interviewed by the researcher using the retrospective verbal interview method. All retrospective verbal interviews were audio recorded.

All speaking recordings (40 monologic and 40 paired) of the test-takers' speaking test performance were copied onto DVD and sent to the two raters by post. Special care was taken in the order of the speaking test recordings on the DVDs. The order of all recordings was carefully arranged, to prevent the previous performance of the same test-taker from influencing the raters' judgement of the same test-taker's performance on another task.

The raters awarded test-takers' speaking scores individually by using a public version of the Cambridge FCE speaking criteria (UCLES 2012) as a guideline. It consists of four categories: grammar and vocabulary, discourse management, pronunciation and interactive communication. The band score for each category ranged from 0.0 to 5.0. (See Appendix 2 for speaking criteria). Rater written comments were also collected in order to investigate reasons for their scoring.
After the raters had finished awarding test-takers' speaking-test scores, they were invited to take part in a retrospective verbal interview individually. Eight video recordings were selected for this purpose. The eight sessions were selected in the hope of representing the entire 40 test sessions. Therefore, the selection was based on the researcher's preliminary analysis during the test administration in terms of what seemed to be typical interactional features, as well as the test-takers' genders and proficiency levels. The eight video recordings are displayed in Table 1 below.

Table 1: Video recordings used for the raters' retrospective verbal interview

Test-taker	Gender	English proficiency based on IELTS examination score		Type of pair	
		Speaking	Listening	Shared L1 pair	Non-shared L1 pair
P03	Male	6.0	6.0	P03P04	P03T03
P04	Male	6.0	6.0		
T03	Male	5.5	6.0	T03T04	P04T04
T04	Male	6.0	5.5		
P19	Female	6.0	5.0	P19P20	P19T19
P20	Female	6.0	5.0		
T19	Female	5.5	5.0	T19T20	P20T20
T20	Female	6.0	5.0		

The raters were asked to watch each video recording and to report on their reasons for awarding scores and on their perceptions of the paired performance. All interviews were audio recorded for analysis.

3.5   Data analysis

The SPSS program was used to analyse the demographic data from the questionnaires and scores from the listening test, the monologic speaking test and the two paired speaking tests.

Spearman Correlation was used to examine the strength of the correlations between:

- the listening test scores and the analytical scores of monologic speaking test
- the listening test scores and the analytical scores of paired speaking tests
- the listening test scores and the analytical scores of paired speaking test of shared L1 pairs
- the listening test scores and the analytical scores of paired speaking test of non- shared L1 pairs.

Wilcoxon signed-rank test was used to investigate differences between:

- the analytical scores of monologic and paired speaking tests
- the analytical scores of paired speaking tests of shared and non-shared L1 pairs.

To verify the reliability of the two raters, level of rater consistency was examined after the two raters completed all the ratings. Pearson's correlation was used to test the inter-rater relative reliability in awarding test-takers' speaking test scores in each analytical category. The Pearson correlation coefficients were higher than .80. According to Hinton et al. (2008, p. 364), a reliability coefficient of 0.70 to 0.90 shows high reliability; therefore, the inter-rater reliability when awarding speaking-test scores was satisfactory.

Agreement when awarding scores for the absolute level of performance is called inter-rater absolute agreement. The inter-rater agreement was examined to confirm the consistency of the two raters in terms of awarding the same speaking test scores (monologic and paired speaking tasks) in each analytical category. However, a high percentage for absolute agreement is difficult to achieve in realistic measurements, since "when there are more than four or five rating levels, exact and adjacent agreement may be a more realistic measure to use" (Graham et al 2012, p. 7). The percentage for inter-raters' exact and adjacent agreement when rating the test-takers' monologic and speaking test tasks were 100% in every category. According to Graham et al (2012), an acceptable percentage for exact and adjacent agreement is close to 90%. Hence the percentage for the inter-raters' exact and adjacent agreement when rating the test-takers' speaking test scores in the main study at 100% was clearly acceptable. It suggests that the two raters frequently assigned the same ratings for the test-takers. For more details on the inter-rater relative reliability and the inter-rater absolute agreement, see Appendix 3.

All recordings were transcribed following Conversation Analysis (CA) conventions (Atkinson and Heritage 1984). Appendix 4 shows the transcription symbols. CA was then carried out to explore communication patterns in the paired speaking tests which were related to test-takers' listening abilities and their L1s. The steps taken for the data analysis were as follows.

Firstly, the speaking test scores of the monologic and paired formats in each category were compared using descriptive statistics and non-parametric inferential statistics (Wilcoxon signed-rank test). Then the strength of the correlations between the listening test scores and the speaking test scores in each category awarded on the monologic and the paired formats were compared by using Spearman Correlation test (to answer *RQ 1*).

Secondly, the listening proficiency scores and the speaking proficiency scores awarded on the paired speaking tests between shared L1 and non-shared L1 pairs were compared by using the Spearman Correlation test (to answer *RQ 2.1*). CA was carried out on the paired test data to discover how communication patterns in paired speaking tests occurred (to answer *RQ 2.2*). In addition, CA analysis was carried out to discover any differences in communication patterns between shared L1 and non-shared L1 pairs. The part of conversations in which communication problems occurred was analysed, and this analysis was supplemented by the raters' comments.

Finally, the retrospective interview with individual test-taker and raters was transcribed orthographically and analysed. The raters' written comments on test-takers' speaking performance on each rating category (grammar and vocabulary, discourse management, pronunciation and interactive communication) and retrospective interview data with test-takers and raters were used to support the interpretations of CA, as well as to gain further insights into paired test interactions related to test-takers' L1 backgrounds and listening-related issues.

4. RESULTS

The research findings are presented in this section. The results of the questionnaire are presented first as they were used to gain test-takers' general background. The test-takers' data, related to their L1, gender and English proficiency based on a standardised examination (IELTS), were utilised to match test-takers with a similar background in the paired task between shared L1 pairs and with a similar background except for their L1 in the paired task between non-shared L1 pairs.

The test-takers' age, length of stay in the UK and years of studying English are displayed in this section. There were 20 Urdu and 20 Thai L1 test-takers. Both groups consisted of 10 males and 10 females.

The test-takers' ages ranged from 23 to 34 years old with a mean of 27.20 (SD=2.84). Their ages were relatively similar. All of them were young adults, mostly in their 20s. The length of time they had been in the UK ranged from one month to 18 months (mean=6.60, SD=4.35). Most of them had stayed in the UK for a relatively short time (over half of them had stayed for less than six months). Their length of time studying English varied greatly from one year to 22 years (mean=11.93, SD=5.89). Therefore, it can be said that in this study the first two test-taker variables were relatively consistent, while the last variable differed from one test-taker to another.

A measure of all the test-takers' (20 Urdu and 20 Thai L1) English proficiency based on the IELTS examination was provided by those test-takers who had previously taken the IELTS examination for entry to higher education. Their English proficiency based on the IELTS examination ranged from Band 4.5 to Band 5.5 (mean=5.33, SD=.27). Their *speaking* proficiency ranged from IELTS Band 5.0 to Band 6.0 (mean=5.61, SD=.35) and their *listening* proficiency from IELTS Band 5.0 to Band 6.5 with a mean of 5.28 (SD=.39).

The results for the 20 Urdu and Thai L1 test-takers' opinions on their familiarity with the English spoken by shared and non-shared L1 speakers (i.e., either Urdu or Thai L1 speakers) gained from Likert-scale questions (1=strongly disagree, 2=disagree, 3=neutral opinion, 4=agree and 5=strongly agree) are presented in Table 2.

Table 2: Statistic for Urdu (N=20) and Thai (N=20) test-takers' opinions on their familiarity with the English spoken by Urdu and Thai L1 speakers

		Min	Max	Mean	Median	SD
Urdu	Familiar with the English spoken by **Urdu** L1 speakers	3	5	4.20	4.00	.62
	Familiar with the English spoken by **Thai** L1 speakers	2	5	3.60	4.00	.75
Thai	Familiar with the English spoken by **Urdu** L1 speakers	1	4	2.25	2.00	.79
	Familiar with the English spoken by **Thai** L1 speakers	1	5	3.85	4.00	1.04

Urdu L1 test-takers' opinion on their familiarity with the English spoken by Urdu L1 speakers ranged from 3 (neutral) to 5 (strongly agree), with the mean score of 4.20 (SD=.62). Their opinion on the familiarity with the English spoken by Thai L1 speakers ranged from 2 (disagree) to 5 (strongly agree) and the mean score was 3.60 (SD=.75),

The ratings of the Thai L1 test-takers' familiarity with the English spoken by Urdu L1 speakers ranged from 1 (strongly disagree) to 4 (agree), and the mean score was 2.25 (SD=.79). Their ratings about the English spoken by Thai L1 speakers varied from 1 (strongly disagree) to 5 (strongly agree), with the mean of 3.85 (SD=1.04).

Frequency information for Urdu and Thai L1 test-takers' opinions on their familiarity with the English spoken by Thai L1 speakers is shown in Table 3.

Table 3: Frequency information for Urdu (N=20) and Thai L1 (N=20) test-takers' opinions on their familiarity with the English spoken by Urdu and Thai L1 speakers

		Urdu L1 test-takers' opinion on their familiarity with the English spoken by				Thai L1 test-takers' opinion on their familiarity with the English spoken by			
		Frequency		Percentage		Frequency		Percentage	
		Urdu L1	Thai L1	Urdu L1	Thai L1	Urdu L1	Thai L1	Urdu L1	Thai L1
Valid	**Strongly disagree**	-	-	-	-	3	1	15	5
	Disagree	-	2	-	10	10	-	50	-
	Neutral	2	5	10	25	6	6	30	30
	Agree	12	12	60	60	1	7	5	35
	Strongly agree	6	1	30	5	-	6	-	30
	Total	20	20	100	100	20	20	100	100

Figures 2 and 3 present boxplots for Urdu and Thai L1 test-takers opinion on their familiarity with English spoken by Urdu and Thai L1 speakers.

Figure 2: Boxplot for Urdu (N=20) and Thai L1 (N=20) test-takers' opinions on their familairity with the English spoken by Urdu L1 speakers
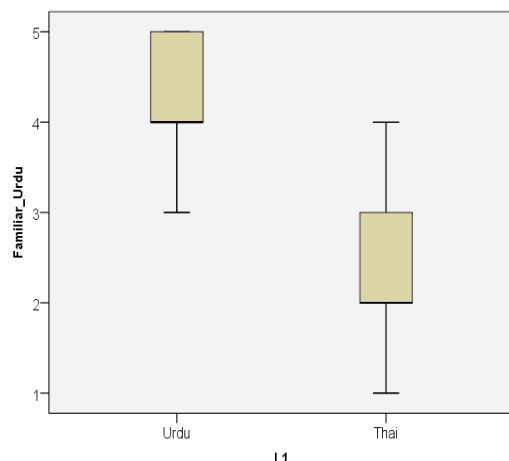
Figure 3: Boxplot for Urdu (N=20) and Thai L1 (N=20) test-takers' opinions on their familiarity with the English spoken by Thai L1 speakers
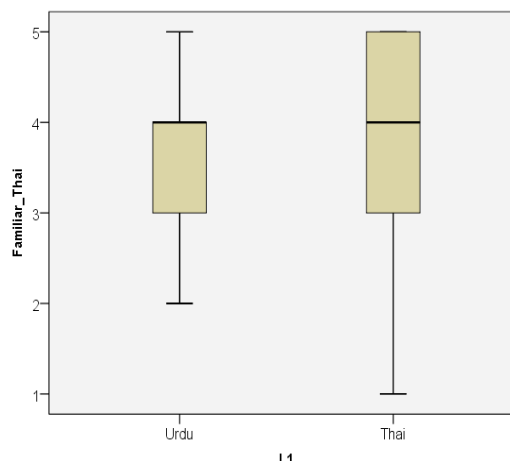
To investigate the differences between Urdu and Thai L1 test-takers' opinions on their familiarity with the English spoken by shared and non-shared L1 speakers (i.e., Urdu and Thai L1 speakers), two non-parametric Wilcoxon signed-rank tests were used. The results are shown in Table 4.

Table 4: Differences between Urdu and Thai L1 test-takers' opinions on their familiarity with the English spoken by Urdu and Thai L1 speakers

Information of	Speakers	Mean	Median	SD	Wilcoxon
Urdu L1 test-takers' opinion on their familairity with the English spoken by:	Urdu L1	4.20	4.00	.62	Z=-2.65 p= .01
	Thai L1	3.60	4.00	.75	
Thai L1 test-takers' opinion on their familiarity with the English spoken by:	Urdu L1	2.25	2.00	.79	Z=-3.67 p=.00
	Thai L1	3.85	4.00	1.04	

Urdu L1 test-takers reported that they were significantly more familiar with the English spoken by Urdu L1 speakers (mean=4.20) than Thai L1 test-takers (mean=2.25). Similarly, Thai L1 test-takers indicated that they were significantly more familiar with the English spoken by Thai L1 speakers (mean=3.85) than Urdu L1 test-takers (mean=3.60), although their ratings were in general lower than the Urdu L1 speakers' ratings across both the categories.

From the statistical results above, one might assume that the participants could understand the English spoken by shared L1 speakers more easily than that spoken by non-shared L1 speakers, in this case, either Urdu or Thai L1 speakers. This is congruent with the findings of Fayer and Krasinski (1987), Bent and Bradlow (2003), Kachi (2004) and Van Engen et al (2010), who reported that the English spoken by shared L1 speakers is more intelligible for a shared L1 listener than non-shared L1 speakers.

4.1   Relationship between listening and speaking test scores in monologic and paired speaking tests (RQ1)

Spearman correlation is used to "correlate data when it is ordinal (one or both variables are not measured on an interval scale), when data is not normally distributed" and it "performs the analysis on the ranks of the scores instead of on the actual data values" (Hinton et al, 2004, p. 300). Since both listening and speaking scores were not normally distributed, Spearman correlation was employed to discover the relationship between:

- listening scores and monologic speaking scores in each analytical category
- listening scores and paired speaking scores in each analytical category.

A summary of the correlations between these variables is presented in Table 5.

Table 5: Correlation between listening and speaking test score (N=40)

	Grammar & vocabulary		Discourse management		Pronunciation		Interactive communication	
	Mono	**Pair**	**Mono**	**Pair**	**Mono**	**Pair**	**Mono**	**Pair**
Spearman's Rho	.19	.32*	.13	.35*	.19	.25	-	.08
Sig. (2-tailed)	.25	.04	.44	.03	.24	.13	-	.63

Mono refers to the monologic speaking test   *Pair refers to the paired speaking tests*

As shown in Table 5, there was no statistically significant correlation between the listening scores and the monologic speaking scores in each analytical category. When considering the relationship between listening scores and paired speaking scores, there were statistically significant correlations between:

- the listening scores and the grammar and vocabulary scores (r_s=.32, p=.04)
- the listening scores and the discourse-management scores (r_s=.35, p=.03) at .05 level.

The strength of the correlations was moderate for both cases. It seems that the better test-takers' listening was, the more correctly and appropriately they used grammar and vocabulary. As grammar and vocabulary elements usually account for a significant amount of the total score variance in skill-specific tests (e.g., Geranpayeh 2007; Shiotsu and Weir 2007), their positive correlation with other skills-based tests (i.e., in this case, listening) may not be surprising.

There was a statistically significant, positive correlation between the listening scores and the discourse management scores in the paired speaking tests. It shows that the better the test-takers' listening was, the more effectively they managed discourse in the paired speaking tests. Intuitively, it makes sense that when the test-takers understand their partner's speech better, they can speak more coherently, relating their output with their partner's utterances. They may not have to ask clarification questions either, which could make their discourse less fluent. Besides, they may not have to hesitate due to comprehension problems, which could result in more fluent discourse.

In addition, to examine whether the test-takers had significantly different scores in each analytical category in the monologic and paired speaking test tasks, a non-parametric Wilcoxon signed-rank test was conducted. The results are shown in Table 6.

Table 6: Statistics for speaking-test scores (N=40)

Category	Task	Mean	Median	SD	Min	Max	Wilcoxon
Grammar & vocabulary (1-5 points)	Mono	3.58	4.00	1.48	.50	5.00	Z = -1.55 p = .12
	Pair	3.36	3.38	1.09	.75	5.00	
Discourse management (1-5 points)	Mono	3.33	4.00	1.43	.50	5.00	Z= -.68 p = .50
	Pair	3.25	3.50	1.05	1.00	4.75	
Pronunciation (1-5 points)	Mono	3.28	3.50	1.39	.00	5.00	Z= -1.08 p = .28
	Pair	3.18	3.13	1.02	1.25	5.00	
Interactive communication (1-5 points)	Mono	-	-	-	-	-	-
	Pair	3.19	3.50	1.06	.75	5.00	

Mono refers to the monologic speaking test   *Pair refers to the paired speaking tests*

There were no statistically significant differences between the test-takers' scores in the monologic or paired speaking tests in each analytical category at the 0.05 level. This means that the test-takers' performances in both types of speaking test did not differ significantly. However, they tended to get slightly better scores in the grammar and vocabulary, discourse management and pronunciation categories in the monologic speaking test than in the paired speaking tests. Although the monologic speaking task may put greater demands on the test-takers than the paired speaking tasks as they have to keep speaking without relying on a paired partner, the monologic task might have given better opportunities for individual test-takers to display what they are capable of in a longer stretch of speech.

4.2 Impact of test-takers' listening proficiency on performance on paired speaking tests for shared and non-shared L1 pairs *(RQ 2.1)*

Since neither listening nor speaking scores were normally distributed, non-parametric Spearman's correlation was employed to examine the correlation between the test-takers' listening scores and speaking scores for the paired speaking tests for shared and non-shared L1 pairs in each analytical category. The results are presented in Table 7.

Table 7: Correlation between listening-and speaking-test scores for shared L1 and non-shared L1 pairs (N=40)

	Grammar & vocabulary		Discourse management		Pronunciation		Interactive communication	
	Shared L1	Non-shared L1	Shared L1	Non-shared L1	Shared L1	Non-shared L1	Shared L1	Non-shared L1
Spearman's rho	.26	.37*	.26	.38*	.22	.22	-.01	.14
Sig. (2-tailed)	.10	.02	.11	.02	.17	.18	.97	.37

*. Correlation is significant at the 0.05 level (2-tailed).

It was found that there was no statistically significant correlation between the test-takers' listening and paired speaking scores for the shared L1 pairs. For the non-shared L1 pairs, there was a statistically significant correlation between:

- test-takers' listening scores and grammar and vocabulary scores (r_s=.37, p=.02)
- test-takers' listening scores and discourse management scores (r_s=.38, p=.02).

They are exactly the same two categories where correlations with listening scores were reported in Section 4.1. Therefore, the correlations shown in Table 7 actually reflected the correlations found in non-shared L1 pairs presented above.

While the strength of the two correlations was only moderate, the positive correlation between the test-takers' listening scores and grammar and vocabulary scores for the non-shared L1 pairs suggests that the test-takers with stronger listening skills tended to score more highly in the grammar and vocabulary category. Furthermore, the positive correlation between the test-takers' listening scores and discourse management scores for the non-shared L1 pairs suggests that the better their listening proficiency, the more effectively they tended to manage discourse. These tendencies were not observed in the shared L1 pairs.

The latter results on discourse management could suggest that when the test-takers who were paired with a non-shared L1 partner had good listening proficiency, they could comprehend the partner and manage to produce more extended stretches of language, provide more relevant and clearer organisation of ideas, and use more cohesive devices and discourse markers. In contrast, the shared L1 pairs might not have required as much English listening proficiency to understand their partner's speech performance, suggesting that even when the test-takers lacked listening skills, they might have understood their shared L1 partners to the extent in which they could manage the discourse as well as those who had better listening skills.

To examine whether the test-takers in the shared and non-shared L1 pairs gained significantly different scores in each analytical category, a non-parametric Wilcoxon signed-rank test was carried out. The results are presented in Table 8.

Table 8: Statistics for the paired speaking scores in the shared and non-shared L1 pairs (N=40)

Category	Mode	Mean	Med	SD	Min	Max	Wilcoxon
Grammar & vocabulary (1-5 points)	shared L1	3.36	3.50	1.12	.50	5.00	Z = -.12
	non-shared L1	3.36	3.50	1.21	.50	5.00	p = .90
Discourse management (1-5 points)	shared L1	3.13	3.50	1.16	.50	4.50	Z = -1.90
	non-shared L1	3.45	3.50	1.12	1.00	5.00	p = .06
Pronunciation (1-5 points)	shared L1	3.19	3.00	1.16	.50	5.00	Z = -.23
	non-shared L1	3.19	3.00	1.10	1.00	5.00	p = .82
Interactive communication (1-5 points)	shared L1	3.23	3.50	1.20	.50	5.00	Z = -.81
	non-shared L1	3.16	3.25	1.28	.00	5.00	p = .86

As illustrated in Table 8, there was no statistically significant difference for each analytical score in the two types of pairing at the 0.05 level.

4.3   Communication patterns related to interactive listening between shared and non-shared L1 pairs *(RQ 2.2)*

A listener's interactive listening can be measured to some extent from a listener's behaviour or response (May 2007; Ducasse and Brown 2009; Ducasse 2010). Hence, this study analysed the test-takers' interactive listening through their behaviour or response in the paired speaking test tasks. In order to support the researcher's interpretation of these communicative events, comments obtained from test-takers' retrospective interviews are used. Further discussions of these observed events will be made in Section 4.3.4 in conjunction with the score results reported in Sections 4.1 and 4.2 and results from raters' retrospective interviews and written comments.

Ducasse (2010) divides interactive listening, which contributes to successful interaction, into two subcategories:

1. *comprehension* that relates to a listener's understanding of a speaker's message and demonstrating it through verbal support to present his/her engagement, encourage the partner to continue speaking or demonstrate comprehension as a listener

2. *supportive listening* that relates to a listener's provision of audible and sometimes non-verbal support to a speaker during a conversation to encourage the speaker to continue speaking and maintain the floor (Ducasse 2010).

The former kind can be evidenced by candidates' comprehension through verbal signs: (1) supplying appropriate vocabulary; and (2) demonstrating comprehension by responding to a partner's message with a relevant contribution. The latter kind can be evidenced by (3) candidates' back-channelling (which may be used with gesturing). According to Ducasse and Brown (2009) and Ducasse (2010), it should be noted that back-channellings do not always indicate the listener's understanding. These three types of performance were used as a guideline to investigate the test-takers' interactive listening in this study.

Some quantification was also attempted in relation to these types of communicative event, and inter-coder reliability related to such quantification is provided wherever appropriate. The quantification was only utilised to support the CA findings. As Schegloff (1993) notes, CA is not fully compatible with quantification, since evidence from a single case could have a significant meaning in CA studies. Quantification is nevertheless useful to provide a more generalisable picture of the interactional features in investigation; however, it is crucial to use this in a meaningful way. It is suggested to use quantification together with single case analysis in order to understand the "environment of relevant possible occurrence" (Schegloff 1993, p. 106). The communication patterns related to interactive listening, which were observed in this study, are divided into two main categories:

1. similarities in communication patterns related to interactive listening between the shared and the non-shared L1 pairs

2. differences in communication patterns related to interactive listening between the shared L1 and the non-shared L1 pairs.

4.3.1 Similarities in communication patterns related to interactive listening between shared and non-shared L1 pairs

Similarities in communication patterns related to interactive listening between the two types of pair were observed in all three categories:

- supplying relevant vocabulary
- demonstrating comprehension
- back-channelling.

An explanation of each communication pattern including supporting excerpts is presented in this section. The supporting excerpt of the non-shared L1 pairs are illustrated first and followed by excerpts of shared L1 pairs: Urdu L1 speaker pairs then Thai L1 speaker pairs.

4.3.1.1 Supplying relevant vocabulary

Supplying appropriate vocabulary occurs when the test-taker provides a word or phrase which his/her partner is searching for. It shows "the partner has been attending and comprehends sufficiently to predict a missing word, which enables the interaction to continue" (Ducasse 2010, p. 76). The CA results of the current study indicated that supplying vocabulary was utilised by the test-takers in both types of pair. The test-taker as a listener provided a word that their partner was searching for to complete their partner's utterance. It was usually triggered when the test-taker as a listener felt that their partner was searching for a word or phrase. By supplying a relevant word or phrase, a listener was able to show his/her engagement and demonstrate comprehension.

Examples of supplying vocabulary in the shared and the non-shared L1 pairs are presented first to show qualitative similarities of such occurrences followed by the quantitative comparison of the number of such occurrences.

Excerpt 1 illustrates a conversation between a non-shared L1 pair. P11 was trying to describe a variety of professions on the prompt card but her talk was very hesitant as shown in a number of filled and unfilled pauses. In line 15, when she attempted to select one profession to focus on further, she could not think of a word to explain it. Her partner (T11) recognised the difficulty from her inhaling and a filled pause of 'er', and assisted P11 by supplying the word "popular" (line 16). P11 accepted T11's assistance by uttering a back-channelling "yeah" and continued presenting her idea.

> **Excerpt 1**
> **Topic: Professions (P11=Pakistani female 11, T11=Thai female 11)**
>
> | | | |
> |---|---|---|
> | L06 | P11: | ah::.hhh (0.5) here ((pointing at a picture)) is a lot of er: |
> | L07 | | mm::it is just like ah:: that is like a model ah:: mm:: er:: |
> | L08 | | (0.3) this one is like a painting |
> | L09 | T11: | uh huh |
> | L10 | P11: | scientist singing |
> | L11 | T11: | yeah |
> | L12 | P11: | ((clearing throat)) yeah so everyone has a lot of |
> | L13 | | advantages and disadvantages in these careers |
> | L14 | T11: | uh huh |
> | L15 | P11: | but in the uk i think this is this one is .hh er::: |
> | → L16 | T11: | popular |
> | L17 | P11: | yeah .hhh this one is (.) more important |

From the retrospective interview with T11, she revealed "My partner couldn't think of a word to explain her idea. So I said 'popular' coz I thought it was the word that she was looking for".

Excerpt 2 shows a conversation between two Urdu L1 test-takers. P20 was talking about good friends and bad friends. When she talked about bad friends, she paused for a while and repeated "they will" twice while searching for a word to describe it. P19 assisted her by saying "destroy" as shown in line 46 and she agreed to use the given word. This helped the conversation continue.

Excerpt 2
Topic: Friends (P19=Pakistani female 19, P20=Pakistani female 20)

L40	P20:	can advise (.) you in good manner and so they can
L41		guide you on (.) right thing ((raising hands)) .hh (0.3)
L42		so (.) you can learn .hh and er: (0.5) er: follow them
L43		(0.3) er: to obtain your objective (.) in right feet .hh
L44		otherwise if you have bad friends (.) they will they
L45		will [er: they=
➡ L46	P19:	[destroy ha ha ha
L47	P20:	=will destroy your life they will they will disappoint
L48		you

In the retrospective interview with P19, she reported "my partner was talking about having bad friends. I thought that having bad friends might give negative effects on our lives, so I said 'destroy'. My partner seemed to agree with me because she used the word I had said." P20 reported "I was trying to find the best word to describe the effect of having bad friends and then my partner helped me by saying that word. You know er it was an exactly word I was looking for."

Supplying a vocabulary was also observed in Thai L1 pairs. Excerpt 3 is a part of a conversation between Thai L1 test-takers. They were discussing the topic of professions. T17 was trying to explain what qualifications might be necessary to become an artist, but she struggled to come up with a relevant word, as indicated by a hand gesture and a short pause followed by a filler, "I mean". Latching with the filler, T18 assisted her by offering the word "talent" as shown in line 54.
T17 accepted the word "talent" by saying "yeah the talent" (line 54). The assistance of T18 helped T17 to continue explaining her idea.

Excerpt 3
Topic: Professions (T17=Thai female 17, T18=Thai female 18)

L52	T17:	ah::: it seems to be er::: easy work but actually it is not
L53		if you don't have the .hh ((moving hand)) (.) i mean=
➡ L54	T18:	=talent yeah↑
L55	T17:	yeah↓ the talent coz you need to make to picture
L56		((moving hands))

When T18 was interviewed after finishing conversing with T17 about why she uttered the word "talent", she said "I was listening to my friend. I think er I thought it's the word she wanted to say, so I spoke it out."

As such, showing interactive listening through supplying relevant vocabulary was found in both types of pairs. It was not only demonstrating the test-takers' comprehension to their partner's talk but also assistance for the conversation to continue smoothly. The frequency for supplying relevant vocabulary of the test-takers in the shared (Urdu-Urdu, Thai-Thai) and non-shared L1 pairs is shown in Table 9.

Table 9: Frequency for supplying relevant vocabulary of test-takers in shared and non-shared L1 pairs

Types of pair	Pairing	Frequency	Percentage
Shared L1	Urdu-Urdu	8	11.77
	Thai-Thai	25	36.76
	Total	**33**	**48.53**
Non-shared L1	Urdu-Thai	35	51.47
Total		68	100.00

As shown in Table 9, the frequency for supplying relevant vocabulary of the test-takers in both types of pair was similar. The shared L1 pairs supplied relevant vocabulary 33 times (48.53%), while the non-shared pairs supplied relevant vocabulary 35 times (51.47%). The number of supplying relevant vocabulary between Urdu and Urdu L1 pairs and Thai and Thai L1 pairs was quite different. While Thai L1 pairs supplied relevant vocabulary 25 times, Urdu L1 pairs supplied relevant vocabulary only eight times.

To confirm the accuracy of the above coding, another coder, who specialises in discourse analytic studies and who received training for the coding scheme of this study, was asked to code all transcripts, to check inter-coder agreement. It was calculated from the number of times the coders agree on coding divided by the total number of coding (Graham et al 2012). The inter-coder agreement rates calculated for coding supplying relevant vocabulary were 85% for the shared L1 pairs and 83% for the non-shared L1 pairs, which indicates the coding reliability was acceptable. The instances initially disagreed between the two coders were discussed until they reached agreement.

A non-parametric Wilcoxon signed-rank test was utilised to investigate the statistical difference in supplying vocabulary of the test-takers in two types of pair.

Table 10: Statistics for supplying relevant vocabulary of test-takers in shared and non-shared L1 pairs (N=40)

Interactive Listening Clue	Mode	Mean	Med	SD	Min	Max	Wilcoxon
Supplying relevant vocabulary	SL	.83	.00	1.15	.00	4.00	$Z = -.56$
	NSL	.88	1.00	1.22	.00	5.00	$p = .58$

SL refers to shared L1 pairs   *NSL refers to non-shared L1 pairs*

As shown in Table 10, there was no difference supplying vocabulary between the test-takers in the shared and non-shared L1 pairs. This means that the test-takers in both types of pair were equally supplying vocabulary when their partner searching for a word or delivering an idea. However, it should be noted that there was a huge difference between Thai L1 pairs and Urdu L1 pairs. This indicates that this communicative pattern is associated more with Thai L1 speakers than Urdu L1 speakers.

4.3.1.2 Demonstrating comprehension

Demonstrating comprehension can be evidenced from the test-takers' comments about a partner's contribution (Ducasse 2010). It can be examined by checking whether it coherently relates to the partner's message and whether the test-taker can respond relevantly to their partner's question. However, short answer (e.g., "yes", "no" or "I agree with you") were not considered as demonstrating comprehension in the current study because the test-takers might be able to provide such short answers without fully comprehending what the partner said, and it is not possible to code these short responses reliably.

An examination of the interactional data suggested that the test-takers in both the shared and non-shared L1 pairs demonstrated comprehension to present the evidence of their interactive listening during a conversation. Excerpt 4 shows a conversation between Thai L1 and Urdu L1 test-takers. They talked about the topic of professions. T03 explained that one profession would require some specific qualifications, one of which was then relevantly followed up by P03, demonstrating P03's comprehension of what T03 had said and added his own idea, as shown in line 57.

Excerpt 4
Topic: Professions (T03=Thai male 03, P03=Pakistani male 03)

L51	T03:	ah:: (0.3) left side the most difficult .hhh they- they-
L52		they need a skill, (.) talent and er some (0.5) more
L53		motivation ((raising hand)) for this uh huh::
L54		businessman, they need money but it's more- more
L55		more creative, more motivation ((looking at a partner
L56		and nodding head))
→ L57	P03:	yeah ((nodding head)) motivation is important because
L58		without motivation, you can't achieve anything ah:
L59		mm:: ((clearing throat))

In retrospective interview, P03 reported "I said 'yeah' and 'nodded my head because I agreed with him about the qualifications of a businessman especially motivation. I think it's important for achievement in every profession."

Next is an example of demonstrating comprehension between a pair of Urdu L1 test-takers.
In Excerpt 5, it was evidenced through responding relevantly to their partner's question and extending the partner's idea. When P12 asked P11 a question "what do you think?" (line 05), P11 responded relevantly by comparing relationship between friends and between family members (line 06).
P12 seemed to agree with P11's idea and she extended and supported P11's idea that people could share everything with friends, while they might not do so with their family members (line 16).

Excerpt 5
Topic: Friends (P11=Pakistani female 11, P12=Pakistani female 12)

L05	P12:	what do you think?
→ L06	P11:	(0.3) ah::: the advantage of having friends (.) er:: is ah::
L07		you know (.) the friendship is a .hhh very (0.5) good
L08		((raising hands)) relation (0.4) rather than (.) ah::
L09		instead of (.) brothers, sisters .hh ah:: (.) you know er: if
L10		you have a friend you discuss your feeling [you::=
L11	P12:	[yeah
L12	P11:	=discuss everything (.) like you don't discuss with your
L13		.hh mm::: (.) very close relatives, sister [or brothers=
L14	P12:	[yeah yeah
L15	P11:	=.hh and er:: like like [a husband ha ha ha
→ L16	P12:	[yes oh yes they-they are really
L17		close because you can share (0.3) ah::: (.) anything .hhh
L29		but even we can't say with our parents
L30		[we can share them everything ((nodding head))
L30	P11:	[ah yeah heh heh heh

Excerpt 6 also shows the demonstrating comprehension between the test-takers in the shared L1 pair. The Thai L1 test-takers were discussing professions. T09 gave examples of a successful football player, when suggesting the importance of talent in order to be successful in football. T10 ratified the topic, supporting T09's idea by referring to the name of the football player mentioned by T09 as shown in line 27. This clearly demonstrated T10's understanding T09's speech.

Excerpt 6
Topic: Professions (T09 = Thai male 09, T10 = Thai male 10)

L24	T09:	they got like so much talent. they got like Ronaldo or
L25		Ronaldinho what- what do you think? .hh they have to
L26		practice a lot or:::
→ L27	T10:	yes actually (.) it's ah:: (0.5) if you talk about Ronaldo
L28		or Messi or some people like this. i think (0.3) .hh you
L29		can say (0.5) they are (0.3) born to be a football player
L30	T09:	ok uh huh

In the retrospective interview with T10, he reported: "I agreed with my partner. The famous football players like Ronaldo and Messi are talented. And I think they were born to be the football player".

As shown in the excerpts above, relevantly responding to or appropriately developing what his/her partner had said demonstrates a listener's interactive listening and comprehension. If clarification requests are made, this illustrates the test-takers' incomprehension. When such requests are used to negotiate meaning until the listener fully understands the speaker's message and the listener is ready to change the role from the listener to the speaker, it shows the test-taker's comprehension. If clarification questions were not asked when necessary, or if the speaker cannot manage to respond to such questions, the unclarity or ambiguity could cause communication breakdowns.

The frequency for demonstrating comprehension of the test-takers in the shared (Urdu-Urdu, Thai-Thai) and non-shared L1 pairs is shown in Table 11.

Table 11: Frequency for demonstrating comprehension of test-takers in shared and non-shared L1 pairs

Types of pair	Pairing	Frequency	Percentage
Shared L1	Urdu-Urdu	56	19.58
	Thai-Thai	78	27.27
	Total	**134**	**46.85**
Non-shared L1	Urdu-Thai	152	53.15
Total		286	100.00

As shown in Table 11, the frequency for demonstrating comprehension of the test-takers in the shared and non-shared L1 pairs was 134 (46.85%) and 152 (53.15%). Urdu and Urdu L1 pairs demonstrated comprehension 56 times (19.58%), while Thai and Thai L1 test-taker pairs demonstrated comprehension 78 times (27.27%) out of the total number of demonstrating comprehension of the test-takers in both types of pair, 286 times.

The inter-coder agreement rates calculated for this coding were 81% for the shared L1 pairs and 80% for the non-shared L1 pairs, which indicates the coding reliability was acceptable.

The quantitative data of demonstrating comprehension in both types of pair was analysed by a non-parametric Wilcoxon signed-rank test.

Table 12: Statistics for demonstrating comprehension of the test-takers in shared L1 and non-shared L1 pairs (N=40)

Interactive listening clue	Mode	Mean	Med	SD	Min	Max	Wilcoxon
Demonstrating comprehension	SL	3.35	3.50	2.01	.00	8.00	Z = -.66
	NSL	3.80	3.00	2.74	.00	11.00	*p* = .51

SL refers to Shared L1 pairs   *NSL refers to Non-shared L1pairs*

As illustrated in Table 12, there is no statistically significant difference in the communicative occurrences that demonstrated comprehension between the test-takers in the shared and non-shared L1 pairs. It shows that the test-takers in both types of pair equally demonstrated their comprehension to their partner's talk.

4.3.1.3 Back-channelling (e.g., uh huh, yeah, yes, mm)

Back-channelling is defined as a speech sound produced by an interactive listener to provide supporting feedback to a speaker while the speaker maintains the floor (Ducasse and Brown 2009; Ducasse 2010). It is used by a listener to let a speaker know that he/she is listening to, and understanding, what the speaker is saying. There are two sorts of back-channelling: verbal (e.g., *yeah*, *ok*, *uh huh*, and *mm*) and non-verbal, i.e., nodding head (Ducasse and Brown 2009; Ducasse 2010; Shelly and Gonzalez 2013). It should be noted that Ducasse (2010) did not include other non-verbal signals in her study, for example, gaze and gesture, laughter, body position and facial expression as part of interactive listening. These signals were categorised as interpersonal non-verbal communication, which was one of the three categories (interpersonal non-verbal communication, interactive listening and interactional management) for achieving interactive communication in pairs. The current study followed Ducasse's study, and did not consider other non-verbal signals, except nodding head, as back-channelling. This was also due to the difficulty of analysing other non-verbal signals reliably.

In this study, there was evidence that the test-takers in both the shared and non-shared L1 pairs used back-channelling during conversing.

It was found that the test-takers in both types of pair used back-channelling to encourage their partner to continue speaking, as shown in Excerpts 7 and 8. Excerpt 7 is an example of back-channelling use between the test-takers in the non-shared L1 pair. Thai L1 and Urdu L1 test-takers were talking about the topic of friends. While P01 was presenting his idea, T01 said "yes" and "yeah yeah", as shown in lines 13 and 18, to illustrate his listening to, and supporting, P01 to continue speaking.

> **Excerpt 7**
> **Topic: Friends (T01=Thai male 01, P01=Pakistani male 01)**
>   L11 P01: even they're helping in an exam (0.3) while we're
>   L12   sitting together (0.5) he can help us [even we don't do=
> → L13 T01:   [yes ha ha ha
>   L14 P01: =it we always take (0.5) er: he's doing the exam he or
>   L15   she .hh we can say that [first to do my exam ha ha ha
>   L16 T01:   [ha ha ha
>   L17 P01: [that's the friendship]
> → L18 T01: [yeah yeah   ] or we can
>   L19   go (0.5) to travelling with friends so:: .hh

In T01's retrospective interview, he was asked why he uttered "yes" and "yeah yeah" when listening to his partner speaking. T01 answered: "I said 'yes' and 'yeah yeah' because I got what my partner said. I always say these words when I'm listening to someone speaking". The data from the interview with T01 confirms that he used back-channelling to show his engagement with and understanding of his partner's speech.

Next is a part of a conversation between two Urdu L1 test-takers. It exhibits how a listener (P02) demonstrated his interactive listening and supported his partner (P01) to speak by producing back-channelling. P02 uttered back-channelling, e.g., "mm" (line 08) and sometimes accompanied it with gesture, e.g., "yeah" with head nodding (line 10). These actions enabled the speaker to continue explaining his idea.

Excerpt 8
Topic: Professions (P01=Pakistani male 01, P02=Pakistani male 02)

L01	P01:	how difficult is to be successful in these professions like
L02		ah playing football .hhh and::: singing, painting, and
L03		dancing and er: the this doctor ((pointing at a picture))
L04		mm: having a doctoring degree .hh like er:: as a sport
L05		professional we will go for sport first (.) like sampling
L06		football how difficult these are to become .h come in this
L07		and have a success in this sport? hh (0.3)
→ L08	P02:	mm::
L09	P01:	like when we don't get a good coach (0.7)
→ L10	P02:	yeah ((nodding head))
L11	P01:	we can't get success in this (0.5) football match we can't
L12		(.) get something new (0.3)

An interesting point was found from Excerpt 8. P02 used back-channelling when his partner paused for a while as shown in lines 8 and 10. In the retrospective interview with P02, he said: "I wanted my friend to keep speaking. That's why I said 'mm' and 'yeah' when he stopped speaking. I understood what he said. While I was listening to him, I was also thinking of how to express my idea". Although P02 seemed not to pay much attention to listening to P01, he did understand what P01 was saying. P02 used back-channelling only when P01 paused speaking to encourage P01 to continue speaking while P01 was thinking of what to say next when he got a turn. This finding is congruent with what Ducasse (2010) found in her study: a listener does not only listen to a speaker, but he or she is thinking of what to say next.

An example of utilising back-channelling in a conversation of a Thai L1 pair is illustrated in Excerpt 9. Thai L1 test-takers were conversing about friends. T07 answered T08's question about the number of his close friends. While T07 told that he had five or six friends, T08 listened and used back-channelling "yeah yeah" (line 90) to show his interactive listening and encouragement T07 to continue speaking.

Excerpt 9
Topic: Friends (T07=Thai male 07, T08=Thai male 08)

L88	T08:	ah:: yeah how- how many close friends do you have?
L89	T07:	i think (0.3) about five or six (0.5) friends
→ L90	T08:	yeah yeah
L91	T07:	i think i have close friends and that (.) both i can
L92		tell him ev- everything

In the retrospective interview with T08, he was asked why he said "yeah yeah" while listening to his partner. He reported: "I wanted to let my friend know that I was listening to him and understood what he had said".

The frequency for back-channelling use of the test-takers in the shared (Urdu-Urdu, Thai-Thai) and non-shared L1 pairs are shown in Table 13.

Table 13: Frequency for back-channelling use of test-takers in shared and non-shared L1 pairs

Types of pair	Pairing	Frequency	Percentage
Shared L1	Urdu-Urdu	118	18.02
	Thai-Thai	140	21.37
	Total	**258**	**39.39**
Non-shared L1	Urdu-Thai	397	60.61
Total		655	100.00

The frequency for back-channelling use of the test-takers in the shared and non-shared L1 was 258 (39.39%) and 397 (60.61%). The test-takers in the non-shared L1 pairs used back-channelling more frequently than the shared L1 pairs. The pairs of Thai L1 test-takers used back-channelling more frequently than Urdu L1 pairs. The Urdu L1 test-taker pairs used back-channelling 118 times (18.02%), whilst the Thai L1 test-taker pairs used it 140 times (21.37%).

The inter-coder reliability rates for this coding in the shared and non-shared L1 pairs were 83% and 84%. Hence, the inter-coder reliability was acceptable.

A non-parametric Wilcoxon signed-rank test was used to investigate the frequency of back-channelling between the test-takers in the shared and non-shared L1 pairs.

Table 14: Statistics for back-channelling of the test-takers in shared L1 and non-shared L1 pairs (N=40)

Interactive listening clue	Mode	Mean	Med	SD	Min	Max	Wilcoxon
Back-channelling	SL	6.45	5.00	4.73	1.00	21.00	Z = -2.86
	NSL	9.93	9.00	6.15	.00	26.00	p = .00

SL refers to shared L1 pairs   *NSL refers to non-shared L1 pairs*

Table 14 shows that there was a statistically significant difference in back-channelling use between the shared and non-shared L1 pairs at .01 level. The test-takers in the non-shared L1 pairs (mean=9.93) used back-channelling significantly more frequently than the shared L1 pairs (mean=6.45). Even though the numbers of back-channelling differed between the two types of pair, the use of back-channelling can still be considered as a similarity between the two pairs, since it was frequently observed in both pairs.

4.3.2 Differences in communication patterns related to interactive listening between shared and non-shared L1 pairs

While both shared and non-shared L1 pairs showed several similarities in communication patterns related to interactive listening, there were also some differences, in terms of (1) attempting to understand a partner completely, (2) understanding an unclear utterance, and (3) misunderstanding because of different cultural background. Examples with the explanation of the differences are presented in the following section.

4.3.2.1 Attempting to understand a partner completely

It is not surprising that some communication breakdowns occur during interactive communication between two non-native speaking (NNS-NNS) test-takers. The communication breakdowns found in this study seem to relate to both the speaker's and listener's limited linguistic ability, as well as the effect of their L1 background.

When a communication breakdown occurs, learners usually attempt to solve it in order to achieve their interactive communication. They use various explicit strategies, such as clarification requests and body language, to problematise and repair the miscommunication to achieve their communication goal. They may also exploit their background knowledge, and personal and social awareness to solve the miscommunication (Hahn and Watts 2011).

However, analysis of communication breakdowns in the shared and non-shared L1 revealed some differences between the two groups. While the test-takers in the shared L1 pairs always tried to solve them by using various strategies, the non-shared L1 pairs did not always do so. The example of how the test-takers in the non-shared L1 pairs responded to communication breakdowns is presented in this section.

Communication breakdown between a shared L1 pair is presented in Excerpt 10. While T05 had been explaining a difficulty of being a doctor, T06 pronounced the word "voice" which confused T05 because he did not expect to hear the word "voice" in a context of a doctor's profession. As seen in line 25, T05 said "yeah?" and looked at his partner to signal that he did not comprehend what his partner had said. After a short pause, he nodded his head and he tried to understand his partner by rightly guessing that T06 was talking about another element of the prompt card, a singer (lines 29-30).

Excerpt 10
Topic: Professions (T05=Thai male 05, T06=Thai male 06)

L20	T05:	when you before before (.) er:: ((moving hand)) while
L21		you study in the in the (.) doctor school they have to (.)
L22		like work (.) hard reads a lot of book (0.3) [ah::
L23	T06:	[voice
L24		((moving hand))
➤ L25	T05:	yeah? ((looking at a partner))
L26	T06:	it's like a voice
L27	T05:	((looking at a partner)) the- (0.3) yeah ((nodding head))
L28	T06:	the voice
➤ L29	T05:	what do mean mean singer↑ ((pointing at a picture)) the
L30		last one right↑
L31	T06:	yeah↓

From the retrospective interview with T05, he said: "I said 'yeah?' because I was wondering how the voice related to a doctor's job. That's why I asked my partner after that". This indicates the communication problem observed in this shared L1 pairs seemed to come from the word which was uttered completely out of the context of the topic they were speaking about. It was not from their difficulties in decoding their partner's utterance. This excerpt presents the attempt of the test-takers in the shared L1 pair to solve the communication problem surprisingly smoothly by making a right guess.

Another example of the test-takers in the shared L1 pair's response to the breakdown is presented in Excerpt 11. The Urdu L1 test-takers (P05, P06) were talking about professions. While P06 was trying to explain, with several pauses, that being a scientist was the most difficult profession to get to the top. He compared it with being a doctor and a singer, and extended his idea on being the singer. P05 summarised what P06 had said and requested confirmation from P06. P06 confirmed P05's understanding by saying "yeah". When P05 continued extending his idea on the singer, P06 realised that P05 misunderstood him. Hence, he repaired by saying "no no no" and clarified his idea (line 40).

Excerpt 11
Topic: Professions (P05=Pakistani male 05, P06=Pakistani male 06)

L29	P06:	so (0.3) this ((pointing at a picture)) the scientist (.)
L30		is much much difficult profession (0.3) that (.) to the top
L31		(0.3) i think so. (0.9) It's much difficult than the doctor (0.3)
L32		than singer (.) singing is (.) most popular in most er (.) most
L33		(.) part of the world but (0.5) scientist
L34	P05:	ok you think that singer singer er singing (0.3) ((pointing
L35		at a picture)) is the most popular in the world
L36	P06:	yeah
L37	P05:	technique there is a lot of competition
L38	P06:	y(h)ah ha ha
L39	P05:	and she need to be work hard
L40	P06:	no no no (0.5) i think scientists ((pointing at a picture))
L41		er:: have to be work hard because it it is er: totally mentally
L42		job (0.4) we have to be mentally presence (.) they don't er:
L43		cerebrate and holidays
L44	P05:	yes of course

In the retrospective interview with P06 about what he was thinking when he said "no no no", he reported: "I thought he understood what I said that being a singer was the most difficult profession to get on the top. But when he talked more, I knew he understood that I thought being a singer was the most difficult profession. That's why I said 'no no no' and explained more about it". This example illustrates that the occurred communication problem did not come from P05's difficulties in comprehension P06's utterance, but it came from his misinterpretation P06's message because the way P06 delivered his idea was rather ambiguous. This example obviously shows that when the communication problem occurred in the shared L1 pairs, they managed to identify the source of the problem easily and reached mutual understanding very smoothly.

Such proactive attempts to understand a partner when the communication breakdown occurred did not seem to happen as frequently as in the non-shared L1 pairs. One of the ways in which the test-takers in the non-shared L1 pairs behaved when faced with communication breakdowns or failures to comprehend their partners were being quiet with gesturing and back-channelling. It was interesting that six out of 25 communication breakdowns (see Table 15 at the end of this section) observed from the non-shared L1 pairs were to do with them not responding to their partner's question. They just kept quiet without trying to rectify the problem proactively. This was not found in the shared L1 pairs. Four out of six instances of being quiet when they were asked a question occurred at the beginning of the discussion. In the retrospective interviews with those test-takers, they reported that they were quiet because they were taking some time to think what the question was and how to react to the question that they did not understand.

Excerpt 12 shows an example of a test-taker in a non-shared L1 pair who did not answer a question because she did not comprehend it; she just stayed quiet and looked at her partner as shown in line 64. This signalled that P14 did not understand T14's question, which led T14 to repeat "close friends" since she thought that her partner did not understand this key word. After that P14 understood the question and was able to answer it.

Excerpt 12
Topic: Friends (T14=Thai female 14, P14=Pakistani female 14)

L61	T14:	yes ((nodding head)) and (2.5) .hhh heh heh heh (1.0)
L62		((looking at pictures)) er: do you have close friend /kloʊs flen/?
L63		((looking at a partner))
→ L64	P14:	(1.3) ((looking at a partner))
L65	T14:	close friends /kloʊs flen/
L66	P14:	yes (.) i have close friend (0.3) but a few here but more
L67		than er:: back home and mm:: .hh you know here
L68		everyone is so busy with their job and study. they don't
L69		(0.3) get er:: enough time to (0.4) to be with you (0.3) so
L70		we have less of time here for friends but (.) er: back
L71		home i have many friends but . hhh i cannot call them all
L72		the time (.) as the time's different and (.) er: i'm
L73		here and we cannot discuss usually as we were there.

In the retrospective interview with P14, she reported that at first she did not know that T14 asked her about "close friend" so she was quiet. P14 might have found it difficult to understand T14's accent in her first hearing due to P14's unfamiliarity with her partner's accent which was from a different L1 background (Harding 2012), even though she reported in the questionnaire that she had neutral opinion on the familiarity with English spoken by Thai L1 speakers. T14's pronunciation of "close friend" indeed had some L1 influence as being pronounced as "/kloʊs flen/". To consider P14's listening proficiency, she got 20 out in the 37 item listening test and Band 5.5 from the IELTS listening part. Her listening proficiency was therefore not great, but it was not likely that she has difficulty in understanding such a frequent phrase. Indeed, P14 knew the meaning of "close friend" and understood it in her second hearing. This result may be explained by Jenkins's (2002) finding that NNS-NNS interaction with below bilingual proficiency level possibly fails to use contextual cues to solve difficulties in listening comprehension which derived from pronunciation error of their partner.

Excerpt 13 is another example of non-shared L1 test-takers' reactions to communication breakdowns. It is a conversation between T03 and P03 under the topic of professions. P03 was giving his opinion about how to become successful in professions and giving examples to support his talk. When P03 finished illustrating his opinion, he signalled T03 to take the speaking floor by looking at him and raising a hand. However, T03 seemed not to understand what P03 was trying to signal to him, and after a long pause (2.5 seconds) T03 was still quiet and looking at a prompt card (see line 64). P03 realised that T03 did not take the floor, so he gave up on letting T03 take the floor and tried to extend his opinion.

Excerpt 13
Topic: Professions (T03=Thai male 03, P03=Pakistani male 03)

L51	P03:	yeah ((nodding head)) motivation is important because
L52		without motivation you can't achieve anything ah::
L53		mm:: ((clearing throat)) (1.5) i think i think er: er:: to
L54		do artist profession is- is the most difficult to get on the
L55		top .hhh so i think (0.5) anyone i- i in in any
L56		professions you (0.3) need to work hard to get (.) at the
L57		level er: success because either you're a doctor or
L58		you're surgeon or you're surgeon or you're
L60		businessman .hhh or you're in any professionals i think
L61		you need to be work hard ((looking at a partner and
L62		raising a hand))
L63		(2.5)
→ L64	T03:	((looking at a prompt card))
L65	P03:	in- in in my in my sense any professions is difficult
L66		(0.3) everyone to get some (.) achievement or call
L67		.hhh like if you're a sportsman [you need to

When T03 was interviewed about why he did not show his own idea when gestured, he revealed: "I was quiet because I was thinking of how to answer the questions in a prompt card". The conversation between the test-takers in this non-shared L1 pair shows when a speaker had different L1 background with a listener, the listener seemed less interested in their partner's speech, and they sometimes even ignored their partner by keeping quiet and thinking of what to say next (Ducasse 2010). As shown in the excerpt, T03 did not even signal that he understood his partner by using back-channelling and this caused a communication breakdown between P03 and T03. A lack of interactive listening such as back-channelling can cause unsuccessful interaction (Galaczi 2004). To consider T03's listening proficiency, he got 18 scores in the 37 item listening test and had IELTS listening Band 5.0. He rated 2 on the 5-point Likert scale questionnaire related to familiarity with English spoken by Urdu L1 speakers which means that he disagreed with it. From the data, it can be assumed that he had limited listening proficiency and he was not familiar with English spoken by his Urdu L1 partner. This might have discouraged him to try to understand his partner.

As discussed in Section 4.3, back-channelling is one sub-category of interactive listening, which is a part of successful speaking-listening interaction. Back-channelling can be used to encourage the speaker to continue speaking, and Ducasse and Brown (2009) named this category of listening "a supportive listening". However, as Ducasse and Brown also noted, back-channelling does not necessarily mean that a listener really understands their partner.

Excerpt 14 shows a test-taker in a non-shared L1 pair who used back-channelling to encourage his partner to speak until his partner finished a turn and then initiated and developed his own topic. In lines 45 and 49, T08 signalled that he did not understand P08's speech through his body language (frowning). Instead of making a clarification request, T08 used back-channelling while listening to his partner and waited until his partner finished his turn. Then T08 started his turn by saying, "ah:: ok so your turn ha ha ha ok tch! in my my turn..." as shown in line 53.

Excerpt 14
Topic: Professions (P08=Pakistani male 08, T08=Thai male 08)

L40	P08:	er:: .hhh for talent er:: yeah er: for to become ah::
L41		successful ah:: artist .hh it-it basically depend upon er::
L42		(.) the mind of painting is er:: personality of person
L43		painting every painter person er: .hh to perform
L44		personality for person who is too shy [to come and=
→ L45	T08:	[mm:: ((frowning))
L46	P08:	=perform in front of many people ah::er:: and especially
L47		in these days so .hh it is very difficult for him or
L48		her to a become successful artist
→ L49	T08:	mm:: ((frowning))
L50	P08:	so and they are so: so many people ah:: er:: mm:: who are
L51		develop talent so they (.) in my opinion they- they can
L52		become successful artist
L53	T08:	ah:: ok so your turn ha ha ha ok tch! in my my turn (.)
L54		this is ah: (0.3).hhh in this pic:tures ((pointing all
L55		pictures)) every people (0.4) who want to successful in-
L56		in (.) in the professions .hhh i think (0.4) it is it is
L57		difficult for every people to- to make a professional (0.3)
L58		for their (.) career

In a retrospective interview with T08, he disclosed that he did not understand everything that his partner had said. This seems to be a reason why he was frowning when listening to his partner. However, he also pretended to understand his partner, as Ducasse (2010) found in her study. Although T08 appeared to understand his partner by displaying back-channelling, he did not extend his partner's idea. Instead, he initiated his turn with a new topic, as he did not understand his partner's speech completely. Considering T08's listening proficiency, he got 19 in the 37 item listening test and had IELTS listening Band 5.5. In the questionnaire, he strongly disagreed to familiarity with English spoken by Urdu L1 speakers. His listening proficiency was not great, but it was not too limited either. Therefore, his failure to understand his partner completely might have been more to do with his unfamiliarity with English spoken by Urdu L1 speakers.

Another example of using back-channelling merely for supporting listening by non-shared L1 pairs without completely understanding their partner's speech is the conversation between T14 and P14 in Excerpt 15. They were talking about friends. T14 was trying to explain the advantages of having friends. She seemed to have difficulty in delivering her idea, as observed from her frequent pauses during her turn and her hand gestures when she could not express her idea or think of appropriate vocabulary. P14 might have noticed that T14 had difficulty in explaining her idea, but P14 did not help her. P14 instead used back-channelling to encourage T14 to keep speaking. It is not likely that P14 comprehended T14 completely as T14's utterance even included the non-English word "*fortuner*" (line 12). Nevertheless, P14 did not attempt to reach mutual understanding, and P14 instead initiated her own topic when T14 gave her a turn.

Excerpt 15
Topic: Friends (P14=Pakistani female 14, T14=Thai female 14)

L07	T14:	mm:: for me my friends (1.0) make me get better
L08		always make me get better ((waving hands))
L09	P14:	ok
L10	T14:	when i have some problems
L11	P14:	uh huh
L12	T14:	(0.5) and (0.8) .tch! when i live here (.) i have fortuner
L13	P14:	uh huh
L14	T14:	(0.5) i (.) i can ((moving hand)) (0.3) practice my
L15		speaking
L16	P14:	uh huh
L17	T14:	and (.) what about you? ((pointing at a partner))
L18	P14:	for me my friend is when i'm sad (0.3) i can call

In the retrospective interview with P14, she said: "I was listening to my partner and thinking of how to formulate my idea in the same time". This shows that P14 did not pay much attention to her partner and did not show her interactive communication apart from superficial back-channelling because she was worrying about how to present her idea. This might be the cause of lack of assistance to her partner by supplying a word or correcting the wrong word use.

Table 15 shows the number of communication breakdowns, attempts to solve the breakdowns, and solved breakdowns in the shared and non-shared L1 pairs.

Table 15: Number of misunderstanding/communication breakdowns, attempts to solve them and solved breakdowns

	Shared L1 pairs		Non-shared L1 pairs	
	Frequency	**Percentage**	**Frequency**	**Percentage**
Number of misunderstandings/ communication breakdowns	12	100.00	25	100.00
Number of attempts to solve the breakdown	12	100.00	19	76.00
Number of breakdowns solved	12	100.00	15	60.00

As shown in Table 15, there were 12 communication breakdowns which occurred in the shared L1 pairs and 25 communication breakdowns in the non-shared L1 pairs. When the breakdowns occurred in the non-shared L1 pairs, there were only 19 attempts (76%) to solve the breakdown. Interestingly, all breakdowns in the shared L1 pairs were solved, while only 15 out of the 25 breakdowns (60%) were solved in the non-shared L1 pairs. It shows that the test-takers in the shared L1 pairs always attempted to solve the occurred breakdowns and they seemed to succeed in solving them easily, while the test-takers in the non-shared L1 pairs seemed not to do so.

The inter-coder agreements in identifying the communication breakdown of the shared and non-shared L1 pairs were 80% and 81%, the attempt to solve the communication breakdown of the shared L1 pairs was 81% and non-shared L1 pairs was 80%, and the solved communication breakdown of the shared L1 pairs was 83% and the non-shared L1 pairs was 85%. The results of the inter-coder reliability in coding the communication breakdown, attempt to solve and solved breakdowns were consistently higher for non-shared L1 pairs, but the agreement rates were acceptable for all cases.

4.3.2.2 Understanding unclear utterances and incorrect word use

In the shared L1 pairs, even unclear utterances and incorrect word use were understood by a partner but this did not happen in the non-shared L1 pairs. In addition to test-takers' limited listening proficiency, misunderstanding and communication breakdown can also occur from unclear utterance, incorrect word use and incorrect grammatical use. It is interesting that these factors did not seem to be very problematic for the test-takers in the shared L1 pairs. In contrast, they seemed to understand easily their partner's utterances even with errors, as shown in Excerpt 16, which is a conversation between Thai L1 test-takers.

In line 69, T07 asked T08, "how did your close-close friends?", mistakenly omitting a main verb. Omitting a main verb is not considered as Thai L1 transfer, as the Thai language does not allow verb omission. Nevertheless, T08 could answer the question correctly. As a result, the conversation continued successfully.

Excerpt 16
Topic: Friends (T07=Thai male 07, T08=Thai male 08)

L66	T08:	for me i know (.) i know many people ((moving
L67		hands)) but- but (.) ah:: the best friends of mine only
L68		a:bout two or three people only i can talk to (0.3)ah[::
L69	T07:	[how
L70		did your close- close friends?
L71	T08:	i study together about the (0.3) more than ten years

When T08 was interviewed, he reported that he did not even notice that the question was grammatically incorrect. He somehow interpreted correctly the partner's intention to say: "how did you meet your close friends?"

Another example is given below in Excerpt 17, which shows a part of the conversation between Thai L1 pairs (T19, T20) about friends. In line 33, T19 intended to say, "a girlfriend and a boyfriend do something (activities) together", but she mistakenly used the verb "make" instead of the verb "do". Nonetheless, T20, even without making a clarification request, interpreted what T19 intended to say, "they do something/some activities together" and then T20 delivered her idea about it. In Thai language, "make" and "do" have the same meaning and it is a common mistake of Thai speakers of English in using these two verbs.

Excerpt 17
Topic: Friends (T19=Thai female 19, T20=Thai female 20)

L28	T20:	that's it and this yeah it's like (0.9) [playing music=
L29	T19:	[er:::
L30	T20:	=[together
L31	T19:	[what do you think about (0.7) er::: (0.3) girlfriend and
L32		(.) boyfriend ah::: ((moving indexes closely)) (0.5)
L33		make something together?
L34	T20:	for the girl they is like maybe share feeling
L35	T19:	mm:::
L36	T20:	or talk something because like girl more talkative .hh than
L37		than [guy so they will more share .hh share feeling than =
L38	T19:	[huh huh huh
L39	T20:	= than guy .hhh but for guy they just (.) ok go together
L40		or play the music but they not [talk yeah talk too much
L41	T19:	[not not mm:: ((nodding
L42		head))

In T20's retrospective interview, T20 reported: "I knew what she meant. In my language, 'make' and 'do' have the same meaning". This showed that T20 could understand T19 easily because they have the same L1 background knowledge and she knew that it was a common mistake of English use of Thai L1 speakers. May (2007) also reported that test-takers from the same L1 background understood each other easily while speaking in an L2 when raters had difficulty in comprehending the test-takers' talk.

By contrast, there was no such instance in non-shared L1 pairs. Test-takers in non-shared L1 pairs could not comprehend unclear utterances or answer unclear questions without explicit negotiation of meaning, as exemplified in Excerpt 18. In this excerpt, T05 asked P05 a question about the advantages of having friends, and he also asked: "Is there anything normal?" (line 2), which was ambiguous and did not relate to the previous question. P05 was quiet for a while and said "Pardon?" with looking at her partner to signal that he did not understand, so T05 clarified what he meant in line 5.

Excerpt 18
Topic: Friends (P05=Pakistani male 05, T05=Thai male 05)

L01	T05:	let's start with the first question that that what are-
L02		what are the [advantages of having friends? =
L03	P05:	[advantages
L04	T05:	= is there anything (.) normal?
L05	P05:	(0.7) pardon? ((looking at a partner))
L06	T05:	it's easy yeah ((moving hands)) (.) advantages of
L06		friends, friend is (0.3) i think it's a good thing, ∘isn't
L07		it? ((scratching head))
L07	P05:	friends in our life is gonna be cheerful (0.3) having
L08		[joyful
L09	T05:	[share (0.3) experience together ((moving hands))

In the retrospective interview with P05, he reported: "I said pardon because I didn't get his question. Luckily he explained it, so I could answer it". His listening proficiency was relatively high, scoring 26 in the 37 item listening test, and he had IELTS listening Band 6.5. He was unable to answer the question, not because of his listening proficiency, but because the question was not related to the prior utterance of the partner. He was confused by the question which was unclear and unrelated to the context. He could not guess the speaker's intention either.

From the above examples, it might be suggested that even unclear utterances can be understood easily in the shared L1 pairs without relying on explicit meaning negotiation. This could be because the shared L1 pairs may know English mistakes common to the same L1 speakers, and because they may utilise the same L1 linguistic background (Bent and Bradlow 2003; Kachi 2004; Hahn and Watts 2011) and cultural background (Kacki 2004) to interpret their partner's speech. Furthermore, the questionnaire result of this study disclosed that the test-takers were more familiar with English spoken by the shared L1 speakers than by the non-shared L1 speakers (see Section 4), which seemed to support the findings here. The familiarity with the English spoken as shared L1 speakers seemed to help enhance mutual understanding even when an utterance is unclear or ambiguous.

4.3.2.3 Misunderstanding because of different cultural background

The test-takers in the non-shared L1 pairs seemed to have difficulty in understanding the non-shared L1 partner because of their different cultural backgrounds. It could even bring a communication problem. For instance, Excerpt 19 illustrates miscommunication between P10 and T10 about how to make new friends. T10 suggested that his partner, P10, should arrange a party. P10 imagined that when arranging the party, it must have alcohol, which was prohibited according to his religious beliefs. From the retrospective interview with P10, he disclosed: "I am Muslim and our religious don't allow us drinking alcohol. I tried to tell my partner about it". However, P10 was trying to explain his reasons only very implicitly referring to his lack of skill in arranging parties and his lack of self-confidence. T10 seemed not to understand his partner's hidden problems with arranging the party and proposed P10 to shift to a new topic (line 69).

Excerpt 19
Topic: Friends (P10=Pakistani male 10, T10=Thai male 10)

L45	P10:	=and er: (.) ha ha i have a .hh very short list of friends
L46		(.) yeah i only have two or three friends ha ha [and the=
L47	T10:	[uh huh
L48	P10:	=mm:: (.) from my [part
L49	T10:	[you- you can do the party a lot man
L50		[yeah if you want to make a lot of friends [ha ha ha ha
L51	P10:	[ha ha ha [yeah ha ha ha
L52		actually problem's that i'm not good at party [ha ha ha
L53	T10:	[oh yes
L54	P10:	er:: i have not having some (0.3) lots of the friends
L55		((moving hands)) [and (.) but er:: (.) in my start (0.3)
L56	T10:	[mm::
L57	P10:	[like ah:: i am here as a new here [er:: one month ago in=
L58	T10:	[uh huh [uh huh
L59	P10:	the uk er::: and er:: i also (0.3) er: get some (.) .hh kind of
L60		(.) the mm::: (.) lack of confident [also having some=
L61	T10:	[uh huh uh huh
L62		((nodding head))
L63	P10:	=like (0.3) to er:: is having problem ((moving hands)) to
L64		make with er: (.) the friends [ev:ery friend and what's=
L65	T10:	[uh huh
L66	P10:	=happing
L67	T10:	right uh huh
L68	P10:	and er::: [.hh i must
L69	T10:	[so let's go to the next [hih hih hih

Without explicit mention of P10's real reason, related to his cultural background, the misunderstanding occurred. This indicates that explicit explanation is the key to successful communication in the non-shared L1 pairs with different cultural backgrounds.

4.3.3 Additional interactional features between test-takers in shared and non-shared L1 pairs

Two additional interactional features were observed in the current study. These were that the test-takers provided their L1 back-channelling while listening and inserted L1 words in their speech.

4.3.3.1 Providing their L1 back-channelling while listening

When the test-takers were listening to their partner, back-channelling was utilised to show their engagement with and comprehension of what their partner had been saying. Sometimes they uttered their L1 back-channelling automatically. This was evidenced only in Thai L1 test-takers but not in Urdu L1 test-takers. Thai back-channellings were found three times for two Thai L1 test-takers. The interesting point is L1 back-channelling occurred only when those Thai L1 test-takers were paired with the non-shared L1 partner. Below is an example of Thai back-channellings pronounced unconsciously by a Thai test-taker.

P08 and T08 were talking about professions. P08 was expressing his idea of how to be successful in an artistic profession. While listening to P08, T08 presented his interactive listening by supplying a word as shown in line 29 and pronouncing Thai back-channelling (/ˈɔː/) as shown in line 34, which has a similar meaning to "I see" in English.

> **Excerpt 20**
> **Topic: Professions (P08=Pakistani male 08, T08=Thai male 08)**
> | L26 | P08: | artist er::: to become successful artist eh: it depends |
> | L27 | | on (0.3) basically depends on .hhh er:: practice |
> | L28 | | [the- the er:: how much the percent ex- er:: = |
> | L29 | T08: | [practice |
> | L30 | P08: | exam ((moving hand)) or .hhh (.) give them more time |
> | L31 | | to practice then er: you(.) you er: get familiar of the |
> | L32 | | work or exam (.) practice makes something perfect |
> | L33 | | [the same thing is ah:: very applicable in the case= |
> | L34 | T08: | [/ˈɔː/ |
> | L35 | P08: | =of artist |

In the retrospective interview with T08 as to why he pronounced Thai back-channelling, he reported: "I didn't know why I said it. I didn't realise that I pronounced Thai word. I intended to listen to my friend and tried to understand him". T08 unconsciously uttered Thai back-channelling perhaps because he was giving much effort to comprehending his partner's speech. Therefore, he was not aware of L1 back-channelling use, which he pronounced unconsciously. In addition, L1 back-channelling use seemed not to cause any communicative problem for his partner who did not share L1 background with him.

4.3.3.2 Inserting L1 word while speaking

Not only the Thai L1 test-takers who took the role of listener uttered their L1 back-channelling, but the Thai L1 test-taker who took the role of speaker also unconsciously inserted their L1 word. This evidence was observed only for one Thai L1 test-taker as illustrated in Excerpt 21. It is a conversation between P19 and T19. They were talking about professions. T19 seemed to have difficulty in showing her idea as shown in many pauses during her speech, while P19 patiently listened to T19 until P19 heard T19 say the word "professor", which did not seem to be relevant to a businessman context. This led P19 to ask for a clarification from T19. In responding to the request, T19 unconsciously pronounced the Thai word "/bæb/" (line 79) which means "to be like" in English.

Excerpt 21
Topic: Professions (P19=Pakistani female 19, T19=Thai female 19)

L66	P19:	[yah I can't say anything about it because
L67		she's doing sport sciences [so maybe it's easy for her
L68	T19:	[yeah yeah if- if someone
L69		ah::: (0.5) ah:: (0.6) get training i think (0.4) get
L70		training a lot ah::: .hh (0.7) maybe everyone can (0.9)
L71		((moving hands)) play or (0.4) can do anything about
L72		exercise or sport (0.3) ah:: it's well (.) and: but for me
L73		i think a businessman ((pointing at a picture)) is .hh
L74		(0.4) ah:: (0.7) a few of people (.) ah: can (1.2) make
L75		((moving hand)) (.) them to profess- professor (.) er::
L76		and (0.3) .tch! (0.6) [er:::
L77	P19:	[what do you mean by
L78		businessman is professor?
L79	T19:	er::: (.) like /bæb/ er::: ((looking at a ceiling)) (1.5)
L80		professional ((looking at a wall)) (0.7) like
L81		((waving hands)) (0.7)
L82	P19:	professional
L83	T19:	yeah yeah professional

When T19 was interviewed after the conversation between her and P19 finished, she said: "I couldn't explain my idea. I was thinking of Thai words before translating it into English. It was very difficult for me to speak. But I was able to understand my partner. She talked a lot". Therefore, the Thai word seemed to be unintentionally pronounced in the mental translation process, probably due to a lack of working memory capacity.

While no general conclusion can be drawn from the use of their L1 by only three test-takers, it is interesting to find such examples of L1 use in the non-shared L1 pairs. Unlike Jenkins' (1997) study where speech accommodation was observed in shared L1 pairs to increase mutual understanding, this study observed the incidences of L1 use only in non-shared L1 pairs.

From the two excerpts that showed L1 back-channelling and a L1 word in the non-shared L1 pairs, it seems that such unconscious usage relates to these Thai test-takers' limited English proficiency, their limited use of English in daily lives and their stress under the testing context. The test-taker in Excerpt 20 used L1 back-channelling when he was trying very hard to comprehend his partner's speech who spoke English with a different accent from him, and the test-taker in Excerpt 21 used an L1 word by mistake when she had difficulty in expressing her idea in English.

In contrast, even if the Urdu L1 test-takers in this study had a similar level of English proficiency, they did not use their L1. This may be explained by the differential use of English in their home countries; in other words, their familiarity with speaking and listening to English in their daily lives. As described earlier in section 1, the Thai L1 test-takers were from the "expanding circle" of Kachru's (2005) Concentric Circle model of the spread of English, and the Urdu L1 test-takers are from a country in the "outer circle" of the Concentric Circle model, as English is used in institutions in Pakistan.

4.3.4 Raters' perceptions towards shared and non-shared L1 pairs' interaction and their listening proficiency

This section reports the raters' perception gained from written comments and the retrospective interview of the shared and the non-shared L1 pairs. There are two main findings regarding the interactional behaviour of the test-takers in both types of pairs:

1. test-takers in the shared L1 pairs seemed more relaxed and more interactive than the non-shared L1 pairs

2. test-takers in the shared L1 pairs seemed to understand each other easily and they helped each other to solve the communication breakdowns, while the test-takers in the non-shared L1 pairs did not try to do so.

4.3.4.1 Non-shared L1 pairs being more relaxed and interactive

The raters noticed that the test-takers in the shared L1 pairs seemed to be more relaxed and more interactive than in the non-shared L1 pairs. Data gained from raters' retrospective interviews are presented below.

In the raters' retrospective interview, Rater 1 reported:

- "test-takers seem to feel more comfortable when interacting with a shared-L1 partner".

- P03 and T03 pair: "P03 speaks too fast and too much sometimes. He dominates the conversation, so the other person doesn't get much chance to speak. I think he tries to encourage his partner to speak, but he speaks a lot more than the other one. T03 responses are very short. He lacks interaction which does not extend the conversation. He could possibly be pretending to understand".

- P04 and T04 pair: "There is not much interaction. P04 just keeps speaking. He doesn't encourage his partner to speak. T04 uses a great variety of language and expands his ideas. He just says what he thinks, but there isn't much interaction".

- P03 and P04 pair: "P03 uses a variety of language to recall something that happened, but there are many grammatical errors. P04 asks and answers questions and sometimes he picks up on his partner's word and develops it which elaborates the flow of the conversation and helps it to be like a genuine conversation".

- P19 and T19 pair: "It's not really an interaction. One dominates the conversation while the other one rarely speaks and there is no conclusion at the end".

Rater 2 reported:

- T03 and T04 pair: "They tend to encourage each other's talk via smiling and back-channelling use, for example, 'mm', 'huhuh', 'yeah'. They take turns to talk. No interruption. T03 shows more of interest in the speech of T04 by smiling, saying 'huhuh', 'yeah' and 'it's true'".

- P19 and T19 pair: "T19 looks more serious, less relaxed and more attentive than when she talked to her Thai interlocutor. She doesn't try to initiate or present her ideas. She allows her partner to dominate the conversation. P19 obviously dominates the interaction and talks over her partner. She doesn't allow her to talk. She doesn't even ask for any opinion from her interlocutor. It isn't interaction".

- P20 and T20 pair: "T20 looks more serious, less relaxed and more attentive than when she talked with her Thai interlocutor. This may be because she feels more comfortable to talk to a partner from the same L1 background than from the different L1 background. T20's speaking role when talking with Pakistani interlocutor is less than talking with Thai interlocutor".

As shown in the data above, both raters perceived that the test-takers in the shared L1 pairs were more comfortable to interact with each other and produced more interactive talk than the non-shared L1 pairs. The test-takers in the non-shared L1 pairs did not extend the partner's ideas but only initiated their own topics, and some test-takers, especially, Thai L1 test-takers, allowed their partner to dominate the conversation, while they just listened to them with back-channelling or nodding responses.

4.3.4.2 Shared L1 pairs' increased mutual understanding, and their attempts to solve communication breakdowns

Both raters reported that the test-takers in the shared L1 pairs seemed to understand each other easily. They tended to encourage each other to talk through interactive listening devices (i.e., back-channelling and supplying vocabulary), and when they faced difficulty in delivering idea, their partner helped to extend it which made the conversation continue. On the other hand, the test-takers in the non-shared L1 pairs did not always try to solve communication breakdowns, and this was also captured in the raters' written comments and retrospective interview data.

In Rater 1's written comments and interview, comments included:

- P05 and P06 pair: "It's a good and lively conversation between both participants. They react to each other questions and comment on each other's statements".

- P08 and T08 pair: "T08 seems not to understand his partner when his partner gives a lengthy talk, because he does not extend or respond his partner's idea. He is frowning while listening to his partner and immediately initiates his own topic when his partner finishes his turn".

- P13 and T13 pair: "There is virtually no interaction between the participants. P13 doesn't ask questions, comments, and links to contribution to what a partner has said. T13 just stops and looks at her partner to indicate it is her turn to speak. No questions and oral reaction apart from three 'mm's".

- P14 and T14 pair: "P14 only speaks, contributes to discussion when prompted. Leaves it to her partner to organise interchange. No questions asked. Doesn't speak even when there is a silence".

- P15 and P16 pair: "This pair holds lively discussion. P15 interacts with a partner and develops her points. Intersperses comments and asks questions. P16 interacts with her partner, questions and comments. It any silence threatens, she fills in with further comments to keep discussion going".

In Rater 2's retrospective interviews and written comments, she reported:

- T19 and T20 pair: "The one who is more talkative tends to help the other one when she is confronted with speaking trouble; she expresses her difficulty by admitting it and asking for help".

- P19 and T19 pair: "T19 has no room to talk and doesn't show any signs of want/need to talk at all. She just listens to her partner and waits for her turn. No interruption ever. No argument given. T19 always smiles and listens to her partner. She doesn't try to interrupt or take a turn from P19. There are a lot of pronunciation problems, for example, final consonant sounds. One (P19) often has extra final –s; the other (T19) leaves it in some words. 'Sport science' which sounds like 'spot scient /spɒt sɑɪns/' will never be understood by any other English speakers. But the interlocutor (P19) doesn't show any sign of misunderstanding because she only initiates her idea without extending to her partner's idea".

- P11 and P12 pair: "P11 uses back-channellings well, initiates and responds negotiating towards an outcome. Both are very cooperative and appear to get along well. P12 also uses back-channellings well, negotiates towards an outcome and links to a partner's contributions".

The above comments from the raters indicate the test-takers in the shared L1 pairs understood each other easily. They were collaborative in co-constructing the conversation. In addition, they helped each other to develop the conversation by asking questions, responding and extending their partner's ideas. Besides, they helped each other solve communication breakdowns and maintained the conversation. For example, in the case of a threatening silence or when their partners were awkward responding to a question, they assisted by clarifying the question (see Excerpt 8), while in the non-shared L1 pairs the test-takers did not talk until one test-taker could not tolerate the silence and attempted to break the silence by repeating vocabulary or presenting his/her own idea (see Excerpts 12 and 13).

5.   CONCLUSIONS

This paper has reported on a study of the relationship between test-takers' L1, their listening proficiency and their performance on paired speaking tests. Findings indicated that in non-shared L1 pairs, the test-takers with higher listening proficiency tended to gain higher scores in grammar and vocabulary, as well as discourse management in the paired speaking tests. In contrast, the test-takers in shared L1 pairs did not have any influence of their listening proficiency on their paired speaking test scores. It was also found that there was no significant difference for paired speaking scores in any of the four categories between shared L1 and non-shared L1 pairs.

Analysis of the speaking test discourse and retrospective verbal interviews, elaborated on these statistical results, found similarities and differences in communication patterns related to interactive listening between shared and non-shared L1 pairs. This shows that the test-takers' L1 and their listening proficiency correlated to their paired speaking in some extents. The test-takers tended to understand their shared L1 partners better than non-shared L1 partners because they shared the knowledge of their L1. In addition, they were not able to understand their non-shared L1 partners completely because of their different cultures. Therefore, when using paired speaking tasks either in classroom assessment or language testing, test-takers' L1 background should be considered in order to prevent bias.

5.1   Implications of the findings and contributions of the present study

The study provides several contributions and implications for language testing research and practice. These contributions and implications are discussed in terms of: the use of mixed methods approach in language testing research (Section 5.1.1); paired speaking testing practice (Section 5.1.2); rating paired speaking tests and fairness (Section 5.1.3); and paired work and paired speaking tests in pedagogical settings (Section 5.1.4).

5.1.2 The use of mixed methods approach in language testing research

The study used a mixed methods approach in order to obtain more comprehensive understanding of the findings. Mixed methods have been widely used in the language testing field because it fixes weak points of each method. Quantification is beneficial to generalise research findings and prevent personal bias and over-interpretation from employing qualitative analysis only. While the usefulness of mixed methods, especially including CA methodology, to research paired- and group-speaking formats has been previously demonstrated (e.g., Nakatsuhara 2009, 2013; van Moere, 2007), the current study has highlighted once again the usefulness of the mixed method approach by systematically and comprehensively analysing and triangulating these different kinds of data. The data analysed in this study were listening and speaking scores, the raters' perception on the test-takers' speaking performance from retrospective interviews and written comments on the reasons for awarding the test-takers' speaking score, the test-takers' retrospective interviews as well as the interactional discourse data in the paired speaking formats. The most crucial contribution of this study is that it has strengthened the interdisciplinary connection between quantitative analysis, CA and other qualitative analyses by emphasising the methodological benefits of CA use to inform the relationship between test-takers' listening proficiency, their L1 and their paired speaking performance.

The study involved transcriptions of 40 paired sessions in total from the test-takers who were from two different L1 backgrounds. The sample size is relatively small, but it is still considered to be sufficient for using quantitative analysis to support the CA results. Although using quantification in CA is arguable regarding its unsuitability for interactional studies, which require a meaningful explanation and understanding in detail of the relevant occurrence's environment (Schegloff 1993), this study has achieved in presenting in-depth discourse data with reasonable generalisation.

Since the previous literature on the impacts of test-taker characteristics on their paired speaking scores has mainly focused on the features of paired speaking discourse (e.g., Berry, 2004; Nakatsuhara 2004; Galaczi, 2004, 2014), research on how non-native speaking test-takers with shared and non-shared L1 interact with the target language in paired speaking tests is rarely found. Therefore, this is systematic and comprehensive research which provides a more insightful perspective on the paired interaction discourse with the focus on the test-takers' L1 backgrounds. The analysis of the paired speaking discourse of the test-takers in the shared and non-shared L1 pairs has confirmed that there is much understanding that can be perceived from the mixed methods approach. For example, the interesting similarities and differences in communication patterns related to the test-takers' interactive listening, as well as the additional interactional features between the test-takers in the shared and non-shared L1 pairs have been found. The most interesting findings were: that the test-takers in the shared L1 pairs seemed to understand their partner's message easily even if there was unclear or incorrect grammar or word use; that the shared L1 pairs tended to be more collaborative when communication breakdowns occurred and they solved these problems successfully while the non-shared L1 pairs did not always attempt to repair communication breakdowns; and that Thai L1 test-takers unconsciously pronounced Thai back-channelling and word when interacting with a non-shared L1 partner. Data from the other resources – the test-takers' scores, raters' perceptions from the written comments and retrospective interviews and the test-takers' retrospective interviews – support the CA results. Without triangulating the data from these resources, it is hard to perceive in-depth understanding of the findings. Therefore, it is highly recommended to utilise the mixed methods approach to study the test-takers' interaction in order to gain comprehensive understanding of the test-takers' interactive listening and interactional discourse and to generalise the results to other, similar contexts.

5.1.2 Paired speaking testing practice

According to McNamara (1996), awarding test-taker speaking performance in the paired formats is co-constructed with the test-taker's underlying competence and other sources. This study provided the empirical data to inform test designer decisions about whether test-takers' L1 and their partners' L1 backgrounds should be included in the construct of paired speaking tests.

There are two types of threat to test validity that test designers should bear in mind, i.e., construct under-representation and construct-irrelevant variance (Messick 1989). One of the underlying aims of this study was to offer some evidence to help understand how to conceptualise the L1 backgrounds of test-takers and their partners against the two types of threat to test validity. This study has illustrated the effect of test-takers' L1 and their listening proficiency on paired interactions, and it highlighted the importance of interactive listening skills in relation to these variables for effective communication in paired speaking tests. Based on the results of this study, it can be suggested that test-takers' L1 variable should be a part of the paired test construct, if the test is to be a good predictor of test-takers' achievement in interactive conversation. While the L1 variable may affect test-takers' communication patterns and advantage and disadvantage different L1 test-takers, it still contributes to creating an interactional environment where test-takers can display their interactive listening skills and interactional competence. It is also important to keep in mind that the L1 background variable is also the variable that is affecting conversation between a shared and non-shared L1 partner in the real world.

Nevertheless, for the paired speaking tests in the setting of test-takers from multi-L1 backgrounds, test designers should consider the implication of forming shared and non-shared L1 pairs, in terms of test fairness. Test-takers may gain some advantages or disadvantages in their paired speaking interaction due to their partners' L1 backgrounds. As shown in this study, test-takers in the same L1 pairs seemed to gain some advantages in terms of more intelligibility, interaction with less stress, more collaborative interaction and more successful repairs of communication problems than the test-takers in the non-shared L1 pairs. On the other hand, test-takers in the shared L1 pairs might gain disadvantages in terms of the careless use of grammar and vocabulary and pronunciation because they seemed to understand each other easily even with unclear utterances, incorrect word or grammatical use, which would be awarded negative scores by raters. In this sense, the test designers should consider which types of pairing would be the best to measure the test construct.

Pairing with a shared and non-shared L1 partner has pros and cons related to different aspects, and therefore, if time and resources allow, utilising both types of pair for all test-takers in the testing context of paired speaking formats is recommended. However, this is likely to be too idealistic in real practice, especially in the high-stakes standardised examination. It is not cost-effective or time-efficient in a large-scale testing context, and it would be very difficult to find a partner from the same L1 background for all test-takers, especially when one's L1 is not a common one. Therefore, adopting Swain's (1983) notion of "bias for best" in paired speaking test practice is recommended in order to "create conditions in testing that allow for best performance" (Fox 2004, p. 244). That is, to allow test-takers to select their own partner for their best interactional performance in pairs.

5.1.3   Rating paired speaking tests and fairness

This section presents implications for rating paired speaking interaction, which are perceived in the present study as follows: (1) further development of a scale of interactive communication category for the paired speaking tests; and (2) fairness in awarding each test-taker's performance in pair.

5.1.3.1 Further development of a scale of interactive communication category for the paired speaking tests

CA results have shown that paired speaking formats have high potential to elicit test-takers' interactional behaviour and are appropriate to assess the test-takers' interactive communication proficiency. In all of the Cambridge English for Speakers of Other Languages (ESOL) Main Suite examinations, the interactive communication category is comprised in the rating scale. Since this study used the speaking task of Part 3 of the FCE, the FCE speaking criteria for the interactive communication category is presented in Table 16.

Table 16: The FCE Speaking criteria for the interactive communication category (University of Cambridge Local Examination Syndicate)

Band	Interactive communication
5	Initiates and responds appropriately, linking contributions to those of other speakers. Maintains and develops the interaction and negotiates towards an outcome.
4	Performance shares features of Bands 3 and 5.
3	Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.
2	Performance shares features of Bands 1 and 3.
1	Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.
0	Performance below Band 1.

The interactive communication scale does not mention interactive listening skills explicitly, although we can assume from the phrase "responds appropriately" that test-takers must have enough listening proficiency and interactive listening skills to respond to their partner's message appropriately. "Linking contributions to those of others" also requires a degree of listening proficiency. Based on the importance of the interactive listening skills for successful paired interaction highlighted in this study, it is recommended that the scale should refer to the interactive listening ability. Inclusion of the interactive listening ability in the interactive communication scale is also suggested by Galaczi (2014). It is also recommended that the scale should clearly specify what types of response can be considered as "appropriate", i.e., whether short response, back-channelling use and non-verbal response are acceptable as providing an appropriate response. Indeed, this study found that to respond with those features did not always mean that the test-takers really understood their partner, and they were used to encourage their partner to continue speaking and superficially demonstrate that they were listening. This evidence was also found in the studies of Ducasse and Brown (2009) and Ducasse (2010).

Additionally, there was evidence that the test-takers in some pairs allowed their partner to dominate a conversation – this type of interaction is called "asymmetric interaction" (Galaczi 2004) or "dominant/passive" (Storch 2002) – making it difficult to award a score in interactive communication (May 2009). To provide a rater training session and clear description of interactive communication might be beneficial for minimising the problem and increasing the fairness in awarding test-takers' interactive communication scores.

Based on the results of this study, the following descriptions for interactive communication can be suggested for inclusion in the criterion.

1. Test-takers are awarded positive scores when:

- they are able to use back-channelling and non-verbal signals effectively and appropriately in presenting that they are understanding and listening to their partner and supporting their partner to talk

- they are able to help their partner when he/she has difficulty in searching for appropriate vocabulary or presenting his/her ideas

- they encourage their partner to talk

- they are able to take the speaking floor when their partner dominates it

- they are able to manage with communication breakdowns by negotiating meaning (e.g., requesting clarification and checking their own understanding).

2. Test-takers are awarded negative scores when:

- they dominate a conversation or allow their partner to dominate the conversation

- they do not attempt to solve a communication breakdown

- they do not assist their partner to continue a conversation when he/she faces difficulty in delivering an idea

- they do not help their partner to find an appropriate word when it is obvious that he/she is searching for it

- they instantly shift a topic or initiate a new topic abruptly, resulting in parallel interaction

- they respond with a short answer or minimal acknowledgement but are not able to extend or develop it.

Furthermore, on producing unintelligible words to raters, for example, L1-influenced words or L1 back-channelling, rater discussion may be recommended before awarding a score (if more than one rater is involved). Rater training about whether or not to give that test-taker a negative score, in particular how it should be treated in the interactive communication scale may also be advisable. L1 back-channelling and L1-influenced words are intelligible to the test-takers in the shared L1 pairs, and the goal of the interactive communication is that the test-takers are able to understand each other's talk, initiate and respond to each other appropriately, maintain and develop the interaction. It is, in this case, the interaction between the test-takers – not the raters – that matters. How to deal with what is unintelligible for raters but intelligible for the paired test-takers, as also illustrated in May (2007), needs to be discussed in light of the concept of World Englishes and the test construct to be measured in each specific test.

To summarise, it is suggested to include clear descriptors as exemplified above in the rating scale of the interactive communication category in order to provide more accurate rating scores.

5.1.3.2 Fairness in rating of an individual's performance in pairs

Even though the paired speaking tests can elicit various interactional features from the test-takers and their construct is in accordance with the interactionalist perspective, there are some points to consider in order to rate the test-takers' speaking performance with fairness. As stated earlier, the characteristics of the test-takers' partner unavoidably affect the test results. Even if these variables can be part of the test construct, how to evaluate test-takers' performance fairly while their performance is co-constructed with their partner's performance is worth consideration. How to rate the test-takers' performance in pairs fairly and correctly and whether to rate them separately have been raised as concerns.

Due to the difficulty in rating test-takers' interactional effectiveness, May (2007) suggests sharing the score for the interactional effectiveness to test-takers' performance in pairs and awarding scores in other categories separately. Nevertheless, Nakatsuhara (2009) argues that to share scores does not always guarantee assessment fairness, since there are some cases where test-takers fail to involve quiet members to cooperate in the interaction even though they attempted very hard to scaffold others' participation. The findings of this study are congruent with Nakatsuhara's findings. There was some evidence of asymmetrical interaction (e.g., Excerpt 13), which showed test-takers trying to stimulate their partner to talk but they did not succeed. Hence, it seems unfair for them to be penalised for the interactive communication. In addition, in this study it was observed that the test-takers, especially in the non-shared L1 pairs, performed parallel interaction (e.g., Excerpt 14) which were commented by the raters as non-interaction because they initiated their own talk without extending their partner's ideas. For this reason, it is suggested to rate the test-takers' performance in pairs separately. To overcome the difficulty of separate scoring, clear and more detailed definitions of the interactive communication scale, as suggested earlier, will be beneficial.

The other possibility in awarding the interactional communication score with fairness is to assess test-takers with various speaking test formats to reflect their real speaking proficiency as much as possible. For example, the Cambridge ESOL is using this approach in its suite of English language examinations in which all test-takers have to do a monologic speaking format, a two-way interaction between test-takers (paired speaking format) and a three-way interaction with the other test-taker and an examiner. Additionally, in the testing contexts of multicultural and L1 backgrounds, following Swain's (1983) "bias for best" technique by allowing test-takers to select their own paired partner is recommended in order to perform their best interaction.

5.1.4   Paired work and paired speaking tests in pedagogical settings

The paired formats could elicit rich language functions from test-takers and the test-takers could provide positive feedback on the paired speaking formats. Therefore, it is recommended to utilise the paired speaking formats in classroom settings to develop the linguistic and interactional competence of language learners. In the classroom contexts where there are only learners from the same L1 background, they rarely use L2 to communicate with each other in their real life. The paired speaking formats can help them produce various language functions and encourage them to use L2 to communicate with each other with confidence. This might help them develop their skills in interactive communication.

In the international classroom settings, using the paired formats is also highly recommended because learners can practise how to interact with non-shared L1 speakers of English, which will be beneficial for them when they have contact with people from various L1 backgrounds in the real world, outside the classroom. Additionally, if possible, teachers should provide them with opportunities to experience paired interaction with shared and non-shared L1 partner, and give feedback on their performance in terms of each analytical category, for example, grammar and vocabulary, discourse management, pronunciation and interactive communication. In doing this, they will help students learn how to communicate with both shared and non-shared L1 partners effectively, and how to improve their linguistic proficiency and their interactional skills. Furthermore, teachers should teach students how to negotiate meaning when facing communication breakdowns to reach an interaction goal, and encourage them to initiate meaning negotiation during paired interaction. It is also important to point out that dominating the conversation and interacting in the parallel interaction pattern (solo vs. solo) will result in unsuccessful communication and low scores in the interactive communication in the testing context. These things can help students interact with people from various L1 backgrounds with confidence and effectiveness.

5.2 Limitations of the study and directions for further study

Although the current study provides various contributions and implications, there are some limitations in terms of selecting participants and generalisation.

Firstly, due to practical constraints, it was difficult to find both L1 groups of participants from the same circle of Kachru's (1998) model of the three concentric circles of English. However, the researcher attempted to control its possible effect on participants' English proficiency levels by selecting participants with similar speaking and listening proficiency levels based on their IELTS band scores. Short interviews with the participants from the outer-circle of English about their use of English in their daily life were also used at the participant selection stage as well. Even though the researcher did her best to control the possible variables that might affect the test-takers' interactional performance, there might be some implicit effects which might have confounded the results of this study. Hence, it is recommended to replicate this study with test-takers in other L1s who are in the same concentric circle of English and have similar speaking and listening proficiency level.

Secondly, the results of this research with test-takers from Urdu and Thai L1 backgrounds might not be generalisable to other L1 contexts because people from different cultural and L1 backgrounds own different interactional patterns (Lu 2010). For the benefit of generalisation, further studies should investigate the effect of test-takers' L1 and their listening proficiency on their speaking performance from other cultural and L1 backgrounds. One of the communication breakdowns which occurred in the non-shared L1 pairs in this study resulted from the test-takers' different cultural backgrounds. Hence, studies on the effect of different cultural backgrounds of test-takers on their interactional communication are required. In such studies, great care should be taken to select the topics used to elicit the interactional competence of test-takers from different cultural and L1 backgrounds, to prevent some test-takers from gaining advantages or disadvantages from the topic used.

Additionally, this study showed some interesting findings related to communication breakdowns that occurred during interaction of the test-takers in the shared and non-shared L1 pairs. The communication breakdowns in the non-shared L1 pairs were more frequent than in the shared L1 pairs. The test-takers in the shared L1 pairs always helped each other to solve the communication breakdown, while the test-takers in the non-shared L1 did not always attempt to do so. Surprisingly, the test-takers in the shared L1 pairs were able to solve all communication breakdowns, while the test-takers in the non-shared L1 pairs could not. Therefore, further studies should focus on communication breakdowns that occur in the NNS-NNS interaction in the shared and non-shared L1 pairs, and how they solve those communication breakdowns. The results of such studies will be highly beneficial to gain a better understanding of NNS-NNS interaction, and to facilitate more effective communication by both shared L1 and non-shared L1 speakers who use English as a communication tool.

In addition, this study has used only the FCE collaborative task to investigate the test-takers' speaking performance in pairs. For further studies, use of different task types such as an information exchange task is recommended, in order to see whether the same results are obtained with different types of task. Studies with a bigger sample size are also recommended to obtain more generalisable results.

REFERENCES

Algethami, G., Ingram, J. and Nguyen, T. (2011). The interlanguage speech intelligibility benefit: The case of Arabic-accented English. In J. Levis and K. LeVelle (Eds) *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference,* Sept 2010 (pp. 30–42), Ames, IA: Iowa State University.

Atkinson, J.M., and Heritage, J. (1984). *Structures of social action.* Cambridge: Cambridge University Press.

Bent, T. and Bradlow, A.R. (2003). The interlanguage speech intelligibility benefit, *Journal of the Acoustical Society of America,* vol 114 (3), pp. 1600–1610.

Berry, V. (1993). Personality characteristics as a potential source of language test bias. In A. Huhta, K. Sajavaara and S. Takala (Eds), *Language testing: New openings.* Jyvaskyla, Finland: Institute for Educational Research, pp. 115–124.

Berry, V. (1997). Ethical considerations when assessing oral proficiency in pairs. In A. Huhta, V. Kohonen, L. Kurki-Suonio and S. Luoma (Eds), *Current developments in language testing*. Jyvaskyla, Finland: Jyvaskyla University Press, pp. 107–123.

Berry, V. (2004). *A study of the interaction between individual learner differences and oral performance test facets*. Unpublished PhD thesis, University of London.

Berry, V. (2007). *Personality differences and oral test performance*. Frankfurt: Peter Lang.

Brown, H.D. (2004). *Language assessment: principle and classroom practice*. New York: Longman.

Brown, J.D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment.* New York: McGraw-Hill.

Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.

Cambridge ESOL. (2008). *Cambridge Preliminary English Test*. Cambridge: Cambridge University Press.

Cambridge ESOL. (2009). *Cambridge First Certificate in English*. Cambridge: Cambridge University Press.

Coolican, H. (2013). *Research Methods and Statistics in Psychology*, 5th ed. New York: Routledge.

Creswell, J.W. and Plano Clark, V.L. (2011). *Designing and conducting mixed methods research*, 2nd ed. California: Sage Publications.

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment, *Language Testing,* vol 26(3), pp 367–396.

Dörnyei, Z. (2007) Creating a motivating classroom environment. In J. Cummins & C. Davison (Eds), *International handbook of English language teaching,* vol 2, pp. 719–731. New York: Springer.

Ducasse, A.M. and Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction, *Language Testing*, vol 26(3), pp. 423–443.

Ducasse, A.M. (2010). *Interaction in paired oral proficiency assessment in Spanish: Rater and candidate input into evidence-based scale development and construct definition*, Frankfurt: Peter Lang.

Fayer, J.M., and Krasinski, E. (1987). Native and non-native judgments of intelligibility and irritation, *Language Learning* vol 37, pp. 313–326.

Foot, M.C. (1999). Relaxing in pairs, *ELT Journal,* vol 53(1), pp. 36–41.

Fox, J. (2004). Biasing for the Best in Language Testing and Learning: An Interview with Merrill Swain, *Language Assessment Quarterly,* vol 1(4), pp. 235–251.

Galaczi, E. (2004). *Peer-peer interaction in a paired speaking test: The case of the First Certificate in English.* Unpublished PhD dissertation, Teachers College, Columbia University, New York.

Galaczi, E.D. (2014). Interactional competence across proficiency levels: How do learners managing interaction in paired speaking tests? *Applied Linguistics,* vol 35(5), pp. 553–574.

Gass, S.M. and Mackey, A. (2000). *Stimulated recall methodology in second language research*. New Jersey: Lawrence Erlbaum.

Geranpayeh, A. (2007). Using structural equation modelling to facilitate the revision of high stakes testing: The case of CAE, *Research Notes* 30, pp. 8–12.

Graham, M. Milanowski, A. and Miller, J. (2012). *Measuring and promoting inter-rater agreement of teacher and principal performance ratings*, retrieved 10 January 2014 from: http://www.files.eric.ed.gov/fulltext/ED532068pdf

Hahn, L. and Watts, P. (2011). (Un)Intelligibility tales. In J. Levis and K. LeVelle (Eds), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University, Ames, IA, pp. 17–29.

Harding, L. (2012). Accent, listening assessment and the potential for a shared-L1 advantage: DIF perspective, *Language Testing,* vol 29(2), pp. 163–180.

Hinton, P.R., Brownlow, C., McMurray, I. and Cozens, B. (2008). *SPSS Explained*. New York: Routledge.

Hutchby, I. and Wooffitt, R. (2005). *Conversation Analysis*. Malden: Blackwell Publishers Inc.

Iwashita, N. (1998). The validity of the paired interview in oral performance assessment, *Melbourne Papers in Language Testing,* vol 5(2), pp. 51–65.

Jenkins, J. (1997). Testing pronunciation in communicative exams. In M. Vaughan-Rees, (Ed), *A special issue of Speak Out! Bringing together the interests of the IATEFL pronunciation and testing signs*, pp. 7–11.

Jenkins, J. (2000). *The Phonology of English as an International Language*. Oxford: Oxford University Press.

Jenkins, J. (2002). A socio-linguistically based, empirically researched pronunciation syllabus for English as an International Language, *Applied Linguistics,* vol 23, pp. 83–103.

Jenkins, J. (2006). The spread of EIL: A testing time for testers, *ELT Journal,* vol 60(1), pp. 42–50.

Kachi, R. (2004). *Factors predicting native and non-native listeners' evaluative reactions to Japanese English*. PhD thesis, Ohio State University

Kachru, B.B. (1992). World Englishes: approaches, issues and resources, *Language Teaching,* vol 25(1), pp. 1–14.

Kachru, B.B. (1998). English as an Asian Language, *Links & Letters* 5, pp. 89–108.

Kachru, B.B. (2005). *Asian Englishes Today: Asian Englishes beyond the Canon*. Hong Kong: Hong Kong University Press.

Liao, S.C., Hunt, E.A. and Chen, W. (2010). Comparison between inter-rater reliability and inter-rater agreement in performance assessment, *Ann Acad Med Singapore,* vol 39, pp. 613–618.

Lu, Y. (2010). 13 Chinese EFL learners' discoursal performance in the Cambridge ESOL FCE Speaking Test. In L. Cheng and A. Curtis (Eds), *English Language Assessment and the Chinese Learner*. New York: Routledge. pp. 175–189.

Major, R.C., Fitzmaurice, S.F., Bunta, F. and Balasubramanian, C. (2002). The effects of non-native accents on listening comprehension: Implications for ESL assessment, *TESOL Quarterly,* vol 36(2), pp. 173–190.

May, L. (2007). *Interaction in a paired speaking test: the rater's perspective*, Unpublished PhD thesis, The University of Melbourne.

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

McNamara, T.F. (1996). *Measuring second language performance*. London: Longman.

McNamara, T.F. (2000). *Language Testing*. Oxford: Oxford University Press.

McNamara, T.F., Hill, K. and May, L. (2002). Discourse and assessment, *Annual Review of Applied Linguistics*, vol 22, pp. 221–242.

Messick, S. (1989). Validity. In R.L. Linn, (Ed), *Educational measurement (3rd edition)*. New York: Macmillan.

Nakatsuhara, F. (2004). *An investigation into conversational style in paired speaking tests*. Unpublished MA thesis, University of Essex.

Nakatsuhara, F. (2006). The impact of proficiency level on conversational styles in paired speaking tests, *Cambridge ESOL Research Notes*, vol 25, pp. 15–20.

Nakatsuhara, F. (2009). *Conversational styles in group oral tests: How is the conversation co-constructed?* Unpublished PhD thesis, University of Essex.

Nakatsuhara, F. (2013). *The co-construction of conversation in group oral tests*. Frankfurt: Peter Lang.

Norton, J. (2005). The paired format in the Cambridge speaking tests, *ELT Journal*, vol 59(4), pp. 287–296.

Ockey, G.J. (2009). The effects of group members' personalities on a test-taker's L2 group oral discussion test scores, *Language Testing*, vol 26, pp. 161–186.

Ockey, G.J. (2011). Self-consciousness and assertiveness as explanatory variables of L2 oral ability: a latent variable approach, *Language Learning*, vol 61 (3), pp. 968–989.

Ockey, G.J. and French, R. (2014). From one to multiple accents on a test of L2 listening comprehension, *Language Testing*, vol 3, pp. 1–24.

O'Sullivan, B. (2002). 'Learner acquaintanceship and oral proficiency test pair-task performance', *Language Testing*, vol 19(3), pp. 277–295.

O'Sullivan, B. (2008). *Modelling Performance in Tests of Spoken Language*. Frankfurt: Peter Lang.

Salkind, N.J. (2011). *Statistics for people who (think they) hate statistics*, 4thed. California: Sage.

Schegloff, E.A. (1993). Reflections on quantification in the study of conversation, *Research on Language and Social Interaction*, vol 26(1), pp. 99–128.

Shelly, L. and Gonzalez, F. (2013). Back Channelling: Function of Back Channelling and L1 Effects on Back Channelling in L2, *Linguistic Portfolios* 2, pp. 98–108.

Shiotsu, T. and Weir, C.J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance, *Language Testing*, vol 24(1), pp 99–128.

Stibbard, R.M. and Lee, J.I. (2006). Evidence against the mismatched inter-language speech intelligibility benefit hypothesis, *Journal of the Acoustical Society of America*, vol 120, pp. 433–442.

Storch, N. (2002). Patterns of interaction in ESL pair work, *Language Learning*, vol 52(1), pp. 119–158.

Swain, M. (1983). Large-scale communicative language testing: A case study, *Language Learning and Communication*, vol 2, pp. 133–142.

Swain, M. (2001). Examining dialogue: Another approach to content specification and to validating inferences drawn from test scores, *Language Testing*, vol 18(3), pp. 275–302.

UCLES. (2012). *First Certificate in English: Handbook for teachers*, available online at http://www.cambridgeenglish.org/cae-handbook-for-teachers.pdf

Van Engen, K.J., Baese-Berk, M., Baker, R.E., Choi, A., Kim, M. and Bradlow, A.R. (2010). The wildcat corpus of native- and foreign-accented English: communicative efficiency across conversational dyads with varying language alignment profiles, *Language and Speech*, vol 53(4), pp. 510–540.

Van Moere, A. (2007). *Group oral test: How does task affect candidate performance and test score?* Unpublished PhD thesis, University of Lancaster.

Weir, C.J. (2005). *Language Testing and Validation*. New York: Palgrave Macmillan.

Appendix 1: Questionnaire

Please fill in <u>all sections</u> of this questionnaire.

1   About Yourself

Name: ....

Student ID number: ....

Email:..

Gender: Male / Female (please circle)   Age:................ years old

Country of origin: ....

First Language (language you speak at home): ....

How long have you been staying in the UK? .... years .... months

2   English Proficiency

Have you taken any English language tests?   **YES/ NO**   (please circle)

If **YES**, please give details:

Test (eg, FCE, IELTS, TOEFL)	Date taken (DD/MM/YY Eg, 01/12/05)	Grade or Score (if known)		
		Overall	**Listening**	**Speaking**

Appendix 2: Speaking test criteria

Band	Grammar and vocabulary	Discussion management	Pronunciation	Interactive communication
5	Shows a good degree of control of a range of simple and some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a wide range of familiar topics.	Produces extended stretches of language with very little hesitation. Contributions are relevant and there is a clear organisation of ideas. Uses a range of cohesive devices and discourse markers.	Is intelligible. Intonation is appropriate. Sentence and word stress is accurately placed. Individual sounds are articulated clearly.	Initiates and responds appropriately, linking contributions to those of other speakers. Maintains and develops the interaction and negotiates towards an outcome.
4	Performance shares features of Bands 3 and 5.			
3	Shows a good degree of control of simple gramatical forms, and attempts some complex grammatical forms. Uses a range of appropriate vocabulary to give and exchange views on a range of familiar topics.	Produces extended stretches of language despite some hesitation. Contributions are relevant and there is very little repetition. Uses a range of cohesive devices.	Is intelligible. Intonation is generally appropriate. Sentence and word stress is generally accurately placed. Individual sounds are generally articulated clearly.	Initiates and responds appropriately. Maintains and develops the interaction and negotiates towards an outcome with very little support.
2	Performance shares features of Bands 1 and 3.			
1	Shows a good degree of control of simple grammatical forms. Uses a range of appropriate vocabulary when talking about everyday situations.	Produces responses which are extended beyond short phrases, despite hesitation. Contributions are mostly relevant, depite some repetition. Uses basic cohesive devices.	Is mostly intelligible , and has some control of phonological features at both utterance and word levels.	Initiates and responds appropriately. Keeps the interaction going with very little prompting and support.
0	Performance below Band 1.			

(UCLES 2012, p. 59)

Appendix 3:
Inter-rater relative reliability and inter-rater absolute agreement on the speaking tests

Inter-rater agreement is defined as the absolute agreement of the scores from different raters, and inter-rater reliability is the relative consistency of scores from different raters across all students (Liao, Hunt and Chen 2010, p. 617). The inter-rater reliability is "the measurement of the consistency between evaluators in the *ordering* or *relative standing* of performance ratings, regardless of the absolute value of each evaluator's rating" (Graham et al 2012, p. 5). Hence to assure scores from a measurement is consistent and fair, it is necessary to measure both the inter-rater reliability and the inter-rater agreement (Graham et al 2012). The test-takers' speaking test scores (the monologic and paired speaking tests) on each analytical category from the two raters were examined for both the inter-rater relative reliability and the inter-rater absolute agreement.

For inter-rater relative reliability, the monologic and paired speaking test scores in each analytical category (grammar and vocabulary, discourse management, pronunciation and interactive communication) from the two raters were tested using Pearson's correlation. Pearson's correlation measures the correlation among the linear variables (Hinton et al 2004, p. 297) and therefore, measures the relative reliability of the two raters.

However, a Pearson coefficient considers only relative order, it does not mean the raters award the same scores. Agreement when awarding scores for the absolute level of performance is called inter-rater absolute agreement. In order to determine the inter-rater absolute agreement, the scores given by the two raters were also examined by calculating the percentage for raters' scoring agreement (Salkind 2011, p. 108).

Table 17 illustrates the results for inter-rater reliability. It should be noted that there were no data for interactive communication in the monologic speaking test, since the category was not applied due to its unsuitability for the monologic speaking task.

Table 17: Inter-rater reliability when scoring the monologic and paired speaking tests

	Grammar & vocabulary		Discourse management		Pronunciation		Interactive communication	
	Mono	Pair	Mono	Pair	Mono	Pair	Mono	Pair
Pearson Correlation	.93**	.92**	.88**	.91**	.89**	.92**	-	.90**
Sig. (2-tailed)	.00	.00	.00	.00	.00	.00	-	.00

**. Correlation is significant at the 0.01 level (2-tailed).
Mono refers to the monologic speaking test   Pair refers to the paired speaking tests

Table 17 clearly shows that there is significant positive correlation between the two raters when scoring the monologic and paired speaking tests in each analytical category at the 0.01 level. It shows that both raters consistently awarded scores for test-takers' monologic and paired speaking tests. The Pearson correlation coefficients for the monologic speaking test of grammar and vocabulary, discourse management and pronunciation were 0.93, 0.88 and 0.89, respectively. The Pearson correlation coefficients for the paired speaking tests of grammar and vocabulary, discourse management, pronunciation and interactive communication were .92, .91, .92 and .90, respectively. According to Hinton et al. (2008, p. 364), a reliability coefficient of 0.70 to 0.90 shows high reliability. Therefore, the inter-rater reliability when awarding speaking-test scores was satisfactory.

The inter-rater agreement was examined to confirm the consistency of the two raters in terms of awarding the same speaking test scores (monologic and paired speaking tasks) in each analytical category. It is shown as a percentage of absolute agreement and was calculated from "the number of times raters agree on a rating divided by the total number of ratings" (Graham et al 2012, p. 7). A high percentage for absolute agreement is difficult to achieve in realistic measurements, since "when there are more than four or five rating levels, exact and adjacent agreement may be a more realistic measure to use" (Graham et al 2012, p. 7). To find the percentage of exact and adjacent agreement it is necessary "to calculate the percentage of times ratings fall within one performance level of one another (e.g., counting as agreement cases in which rater 1 gives a 4 and rater 2 gives a 5)" (Graham et al 2012, p. 7). The results for the percentage of exact agreement, as well as adjacent agreement, when scoring the 40 test-takers monologic speaking and paired speaking test tasks are illustrated in Table 18.

Table 18: Percentage of inter-raters' absolute agreement and exact and adjacent agreement when scoring monologic and paired speaking tests

Category	Task type	Absolute agreement		One point of difference		Exact & adjacent agreement	
		Freq	%	Freq	%	Freq	%
Grammar & vocabulary	**Mono**	26	65.00	14	35.00	40	100.00
	Paired	40	50.00	40	50.00	80	100.00
Discourse management	**Mono**	20	50.00	20	50.00	40	100.00
	Paired	44	55.00	36	45.00	80	100.00
Pronunciation	**Mono**	22	55.00	18	45.00	40	100.00
	Paired	52	65.00	28	35.00	80	100.00
Interactive communication	**Mono**	-	-	-	-	-	-
	Paired	39	48.75	41	51.25	80	100.00

Freq refers to frequency
Mono refers to the monologic speaking test   *Paired refers to the paired speaking test*

The percentages for inter-raters' absolute agreement when scoring the test-takers grammar and vocabulary, discourse management and pronunciation categories of the monologic speaking test task were 65.00, 50.00 and 55.00, respectively. The percentages for inter-raters' absolute agreement when scoring the grammar and vocabulary, discourse management, pronunciation and interactive communication categories of paired speaking were 50.00, 55.00, 65.00 and 48.75, respectively. The percentage for the inter-raters' absolute agreement when awarding the test-takers' monologic and paired speaking test scores was low, therefore, the percentage of inter-raters' exact and adjacent agreement was calculated.

The percentage for inter-raters' exact and adjacent agreement when rating the test-takers' monologic speaking test was 100.00 in the grammar and vocabulary, discourse management and pronunciation categories. The percentages for the inter-raters' exact and adjacent agreement when rating the test-takers paired speaking-test tasks were also 100.00 in every category. According to Graham et al (2012), an acceptable percentage for exact and adjacent agreement is close to 90%. Hence the percentage for the inter-raters' exact and adjacent agreement when rating the test-takers' speaking test scores in the main study at 100% was clearly acceptable. It suggests that the two raters frequently assigned the same ratings for the test-takers.

Appendix 4: Transcription symbols

(0.5)	Number of a time gap in tenths of a second
(.)	A pause which is less than two-tenths of a second
=	A latch between utterances
[	An onset of overlapping talk
.hh	A speaker's in-breath
hh	A speaker's out-breath
(( ))	A non-verbal activity
-	A sharp cut-off of the prior word or sound
:	Stretch sound or letter
( )	An unclear fragment on the tape
(guess)	The transcriber's best guess at an unclear utterance
.	A stopping fall in tone
,	A continuing intonation
?	A rising intonation
Underlining	An emphasised word or sound
° °	The talk is quieter than the surrounding talk
> <	The talk is quicker than the surrounding talk
< >	The talk is slower than the surrounding talk
Hah, huh, heh	Laughing

(based on Atkinson & Heritage, 1984)

British Council
Assessment Research
Awards and Grants

If you're involved or work in research into assessment, then the British Council Assessment Research Awards and Grants might interest you.

These awards recognise achievement and innovation within the field of language assessment and form part of the British Council's extensive support of research activities across the world.

THE RELATIONSHIP BETWEEN TEST-TAKERS' L1, THEIR LISTENING PROFICIENCY AND PERFORMANCE ON PAIRED SPEAKING TESTS

Suwimol Jaiyote