Fumiyo Nakatsuhara

The Co-construction
of Conversation
in Group Oral Tests

1 Introduction

This book explores the nature of co-constructed interaction in group oral tests by investigating the impact of a test-taker's own and his/her group members' extraversion and oral proficiency levels on conversational style in groups. The impact is examined across three task types and two group sizes, in order to build a comprehensive picture of the impact mediated by different task qualities and conditions.

This study is largely motivated by the lack of language testing research on group oral test discourse, although group oral formats have now been widely utilised. While an increasing number of studies have recently demonstrated the influence of different test-taker characteristics on the scores awarded in paired and group oral tests, and explored discourse features in paired oral tests (e.g. Berry, 2007; Galaczi, 2004; May, 2007; O'Sullivan, 2008; Ockey, 2006), little is known about how test-takers with different characteristics interact in group oral tests. Therefore, in order to gain a better understanding of multi-person discourse in group oral tests (with more than two test-takers), this study examines group test discourse data using both quantitative and qualitative approaches, while focusing on test-takers' extraversion and oral proficiency variables. Additionally, this book also attempts to contribute to the discussion about whether and to what extent variations caused by test-taker characteristics should be considered as variables within the construct of group oral tests.

1.1 Rationale for the Study

Over the past few decades, oral language testing has been increasingly influenced by the introduction of various formats. Paired and group formats, where non-native candidates are paired or grouped to interact with one another while being assessed, have become particularly popular tools. Whilst assessments of oral production are sometimes not adopted, due to practical constraints, paired and especially group formats are much more time- and cost-effective than other types of oral assessments, such as an examiner-candidate interview format. Paired and group formats are also capable of eliciting richer language functions from test-takers than interview tests, thus providing great potential to assess communication ability (e.g. Brooks, 2008; Ffrench, 2003). Furthermore, paired and group formats are likely to match more closely the types of tasks and conversations that students may encounter in the communicative language classroom, and in the real world. Therefore, paired and group tests tend to be introduced into educational settings in the hope of promoting a positive washback effect on classroom teaching and learning (e.g. Csépes, 2009; Masubuchi, 2003).

In fact, the pairing of test-takers is currently standard practice in the Cambridge ESOL Main Suite examinations. Group oral testing has been introduced into a number of high- and low-stakes tests, especially in educational settings. To understand how these formats are used, the next section introduces a few examples of the paired and group oral tests currently being practised.

1.1.1   Current Practice in Speaking Tests in Pairs and in Groups

Paired testing is now common throughout the Cambridge ESOL Main Suite examinations.[1] Since these Cambridge tests are widely used all over the world, the introduction of paired formats was a milestone in the area of testing speaking. As one example of paired oral tests, this section will first describe the First Certificate in English (FCE). The FCE represents Cambridge Level Three (out of five) and is approximately equivalent to B2 of the Council of Europe Common European Framework of Reference for Languages (CEFR) (UCLES, 2007: 4). The speaking component of the FCE test lasts approximately 14 minutes, and the speaking test, comprising four parts, is conducted with two candidates (see Table 1-1).

In Part 1, the interlocutor interviews candidates individually, asking them questions about themselves, such as their present circumstances, past experience and future plans. In Part 2, with visual stimuli, candidates separately produce an approximately one-minute turn, which gives information and expresses opinions via a comparison-contrast description. Part 3, also based on visual materials, involves a collaborative task by candidates and requires negotiation and appropriate turn-taking. The topic is then expanded on in Part 4 in a discussion that the interlocutor leads with both candidates. Therefore, although paired candidates are present throughout the whole testing process, only Part 3 requires the paired candidates to engage in conversation with minimum intervention from the interlocutor, while the other parts are designed to elicit different discourse patterns. The individual performance of the paired candidates is assessed by both an interlocutor who awards a mark for *global achievement* and an assessor who gives marks on four analytical criteria: *grammar and vocabulary, discourse management, pronunciation* and *interactive communication* (UCLES, 2007: 86).

---

1   During the 1980s, the paired format started as an optional format for the FCE and CPE. It became obligatory for the CAE in 1991, for the KET in 1993, for the PET in 1995 and for the FCE in 1996. The CPE also started to employ a paired format as a compulsory format in 2002 (Ffrench, 2003).
2   These group oral tests listed here will be discussed in more detail in Chapter 2.
3   This is different from the current English educational practice in Japan. Obligatory English education has started for the 5th and 6th graders at primary schools since 2011

Table 1-1: Structure of the FCE speaking test (UCLES, 2007: 75)

Part	Task type and format	Focus
Part 1 (3 minutes)	A conversation between the interlocutor and each candidate (spoken questions).	General interactional and social language.
Part 2 (A 1-minute 'long turn' for each candidate, plus a 20-second response from the second candidate)	An individual 'long turn' for each candidate, with a brief response from the second candidate. In turn, the candidates are given a pair of photographs to talk about.	Organising a larger unit of discourse; comparing, describing, expressing opinions.
Part 3 (3 minutes)	A two-way conversation between the candidates. The candidates are given spoken instructions with written and visual stimuli, which are used in a decision-making task.	Sustaining interaction; exchanging ideas, expressing and justifying opinions, agreeing and/or disagreeing, suggesting, speculating, evaluating, reaching a decision through negotiation, etc.
Part 4 (4 minutes)	A discussion on topics related to the collaborative task (spoken questions).	Expressing and justifying opinions, agreeing and/or disagreeing.

Whilst paired oral formats are widely used in such international, standardised tests, group oral formats seem to have been mostly used in local educational settings, which are usually resource-limited. The use of group oral tests in educational settings has been reported in various parts of the world, including Finland (Folland & Robertson, 1976; Liski & Puntanen, 1983), Israel (Reves, 1991; Shohamy et al., 1986), Zambia (Hilsdon, 1991), China (He & Dai, 2006), Hong Kong (Berry, 2007) and Japan (Bonk & Ockey, 2003; Negishi, 2005). Since the present research was carried out in Japanese upper-secondary schools (see 1.2), three group oral tests in Japan and Hong Kong are introduced below, to exemplify some group oral tests in practice.[2]

First, the Hong Kong Advanced/Supplementary Level Examination (A-level/ AS-level) is the matriculation examination conducted by the Hong Kong Examination and Assessment Authority (HKEAA). It is a large-scale, high-stakes test, with a total of 36,608 candidates in 2007 (HKEAA, 2007). Among dozens of subjects, almost every school candidate takes the 'Use of English' examination, which has included a 20-minute oral component since 1994. Candidates are usually examined in groups of four (with a minimum of three) and the test consists of two different formats: (i) an individual presentation based on a reading text, and (ii) group discussion in a university-like setting, such as a small academic seminar. In the first part, candidates are given 10 minutes to read a

---

2   These group oral tests listed here will be discussed in more detail in Chapter 2.

passage of 300 words and make a 1.5 to 2 minute-individual presentation on the basis of the passage. After each candidate has made his/her presentation, the group is given two minutes to prepare for a 10-minute discussion. The topic of the discussion is related to the reading text in the first part. The group discussion part is assessed on the following criteria: *range of vocabulary and structures, fluency, intelligibility* and *participation and meaningful contribution*, while the presentation part is assessed on *intelligibility, accuracy of pronunciation and grammar, fluency* and *technical aspects* (e.g. eye contact) (HKEAA, 2005: 134-135).

The second example of group oral tests is a medium-stakes university in-house test in Japan. The Kanda English Proficiency Test (KEPT) aims to assign students' end-of-year English grades, as well as to assess the effectiveness of the English programme regarding students' English proficiency for curriculum evaluation and development. The test has been administered since 1992 and is administered annually to around 1,700 students (Van Moere, 2007: 126). KEPT includes a speaking component in the form of an oral discussion in groups of four (or, where one student is absent, a group of three). Test-takers are first allowed one minute to read about a topic on a written prompt; then, they are asked to discuss the given topic for a minimum of eight minutes. The performance of each student is marked on five criteria: *pronunciation, fluency, grammar, vocabulary* and *communicative skills/strategies* (Kobayashi & Van Moere, 2004: 3).

Finally, the third example is a regional group oral discussion contest for secondary school students in the Ibaraki prefecture in Japan. The contest, *Interactive English Forum*, has been organised annually since 1999, with the purpose of fostering positive washback in the classroom and making English classes more communication-oriented (Masubuchi, 2003). In the contest, one or two students selected from each secondary school perform three discussion tasks, in a group of three, with students from different schools. All three tasks are five-minute free-discussion tasks on general topics reflecting the content of the educational guidelines (e.g. dreams, hobbies, family and 'my town'), and students are regrouped with new members for each discussion. It is unique in that, in order to create a more collaborative atmosphere, three repeated assessments do not award scores to an individual student, but to a group as a whole. The sum of the three assessment scores that each student attains becomes an individual score to compete in the contest (Ojima, a member of the Board of Education at the Ibaraki prefecture, 2005, personal communication). The ratings are based on three criteria: *intelligibility of expression, cooperativeness* and *appropriateness of expressions* (Masubuchi, 2003: 88-89).

This section has introduced one paired oral test and three group oral tests that are currently in use. While paired formats have also been established as a tool in international standardised tests, group formats seem to be mostly utilised in local educational settings. Furthermore, paired formats in such international tests

are utilised together with other formats, whereas group formats employed in rather resource-limited situations often play an assessment role as a stand-alone method.

1.1.2 Co-constructed Interactions in Oral Skills Performance Assessment

As described above, paired and group oral tests are quite widely used, and the growth of these formats highlights how "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: 228). This raises a growing concern with regard to "whose performance" is being assessed in performance assessments, where candidates' performances are co-constructed as an outcome of interaction (McNamara, 1996: 85-87). Here, it is worth noting that the word *co-construction*, as defined by Jacoby and Ochs (1995: 171), refers to "the joint creation of a form, interpretation, stance, action, activity, identity, institution, skill, ideology, emotion, or other culturally meaningful reality", and "the *co-* prefix in *co-construction* is intended to cover a range of interactional processes, including collaboration, cooperation and coordination. However, co-construction does not necessarily entail affiliative or supportive interactions" (ibid.: 171).

McNamara (1996: 86) illustrates the complexity of oral performance, as shown in Figure 1-1, and suggests that we should be aware that assigned scores are the result of co-constructed performance, influenced not only by the candidate's underlying competence, but also by other sources.

Figure 1-1: 'Proficiency' and its relation to performance (McNamara, 1996: 86)
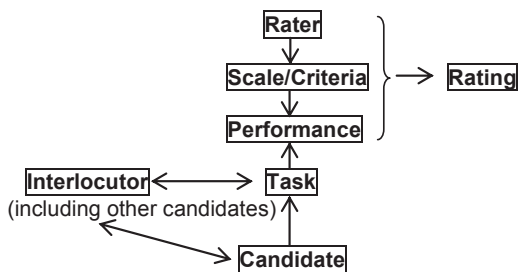
Among other factors, the recent popularity of paired and group oral testing has drawn particular attention to the influence that a test-taker could be subject to from his/her paired or group candidates, and it has been stressed that we need to understand that the pairing and grouping of test-takers should be conducted appropriately (e.g. O'Sullivan, 2008). The lack of research in this area also raises some doubts about the value of these formats, as ignorance of the possible

influence of such factors, if any, might result in potential unfairness to candidates, and thus affect the validity of a test.

Recently, a number of studies have investigated the effects of the variables associated with paired and grouped test-takers (as well as the interlocutor in the interview format) on candidates' performance and the scores awarded. The variables include gender (e.g. O'Loughlin, 2002; O'Sullivan, 2000), personality (e.g. Berry, 2007; Ockey, 2006), acquaintanceship (e.g. Norton, 2005; O'Sullivan, 2002), language proficiency (e.g. Csépes, 2009; Iwashita, 1998; May, 2007) and cultural background (e.g. Young & Halleck, 1998). These studies have shown the possible influence of these variables on paired and group test performances and scores, although the results have often been mixed. Some of these test-taker characteristics have also been examined collectively by multi-variable studies (e.g. Bonk & Van Moere, 2004; O'Sullivan, 2008), and it was found that test-takers' and interlocutors' characteristics could affect performance in a far more complex way, through some factors interacting with other factors (O'Sullivan, 2008). In other words, the evidence is growing to suggest that test-takers' performance (and scores in some circumstances) is likely to be influenced by their paired or group members.

However, in spite of the fact that a number of studies have investigated the impact of test-taker characteristics on scores, how and why these influences are manifested through the paired and group testing discourse has been minimally researched. To date, there have been only a small number of studies providing detailed insights into co-constructed discourse features in dyads and groups related to test-taker characteristics (e.g. Galaczi, 2004; Gan, 2010; Lazaraton & Davies, 2008; May, 2007; Nakatsuhara, 2006). Moreover, to my knowledge, no discourse research into group oral tests has been carried out that takes intra-group test-taker characteristics variation into consideration. Given that the group oral format is now widely used in both high- and low-stakes tests, especially as a stand-alone format, close examination of group oral discourse is urgently needed to understand how test-taker characteristics affect group oral discourse, and how group members co-construct interactions in this format. Furthermore, in order to conduct a close inspection of any test discourse, research should necessarily take task aspects into account, as interactional features could largely depend on the types of task and the conditions of task implementation (Skehan, 1998). Such close examination of the discourse of group oral tests would allow us to rethink what exactly group oral tests are measuring, and it would take our understanding of the group oral test construct one step forward.

In this context, this book addresses three research questions about the impact of the two most relevant test-taker characteristics, extraversion and oral proficiency level, on conversational style in group oral tests (see 2.4.1 for the reasons for selecting these two test-taker characteristics). The first question concerns the general impact of a test-taker's and his/her group members'

extraversion and oral proficiency levels on conversational style in group oral tests, regardless of task types or task implementation conditions. The second and third questions examine how these influences (if any) are articulated through different task types and conditions, respectively. Three different task types (information-gap, ranking and free-discussion tasks) are selected to provide different task qualities, and two group sizes (groups of three and groups of four) are used to provide different task conditions.

1.2 Setting for the Research

The research was conducted in five upper-secondary schools in Japan. These five schools were selected to cover a representative sample of Japanese upper-secondary school students (see 3.2 for more details). The research participants were both male and female students all of whom had Japanese as their first language, and most of them were in either second or third grade (from 16 to 18 years of age). For those who participated in this study, their compulsory formal English education started in lower-secondary school at the age of 12 or 13.[3] All students participating in the research had three 50-minute English classes per week throughout their three years in lower-secondary school, and at the time of data collection they were having six to nine 50-minute (or 65-minute) English classes per week in upper-secondary school (MEXT, 1999a: 64-65, 1999b: 13). The primary goal of English education in upper-secondary schools is to acquire *practical communication ability* (MEXT, 2003), and the target level specified in the course of study seems to be approximately *B1* level in the CEFR (Nakatsuhara, 2007; 2009).

There were two main reasons for selecting the particular participant population for this research. Firstly, as further explained in 3.2, it was considered that possible confounding variables related to the participants could be controlled to the minimum. Namely, in addition to their first language and age, Japanese upper-secondary school students are relatively homogeneous in terms of their educational backgrounds, since all primary and secondary schools follow the educational guidelines published by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) strictly.

Secondly, it was hoped that the research process and results might promote greater interest in the testing of English speaking skills in Japanese upper-secondary schools. As described in Sasaki's (2008) summary of the 150-year history of English language education and assessment in Japan, great emphasis is now placed on teaching speaking skills as *practical communication*

---

3   This is different from the current English educational practice in Japan. Obligatory English education has started for the 5th and 6th graders at primary schools since 2011 (MEXT, 2008).

ability, and the current course of study encourages the use of communicative activities in pairs and groups in the classroom more than ever (MEXT, 2003). Several innovations have recently been introduced to make English language education more communication-oriented. For example, 169 upper-secondary schools were nominated and were supported by MEXT to develop their unique English teaching programmes. In 2006, a listening component was introduced into a nationwide entrance examination for university (University Entrance Centre Examination) which is taken by about half a million students every year. However, despite these innovations, how to assess students' speaking ability has not been transparent at all. MEXT has provided no guidelines or rating scale for speaking assessment in upper-secondary schools, and there is no plan to introduce a speaking component in the University Entrance Centre Examination (personal communication with the chief researcher at the NCUEE, Ishizuka, 2004). Therefore, to provide a possible solution to this discrepancy between the changes in teaching practice and testing practice, it was hoped that conducting group oral test research in Japanese upper-secondary schools could demonstrate one possible method of assessing students' speaking ability, and might potentially be helpful in promoting the testing of speaking ability in their educational context.

1.3 Structure of the Book

This book is structured as follows: Chapter 2 contextualises the study by reviewing different aspects of validity evidence and issues concerning the use of paired and group oral tests. Based on Weir's (2005) socio-cognitive framework for validating speaking tests, it also reviews studies investigating the factors affecting test-takers' performance and the scores awarded in paired and group oral tests. The review focuses on factors of direct relevance to this study, including test-takers' characteristics (in particular, test-takers' extraversion and proficiency levels) and task types and conditions. At the end of this chapter, the research questions are established.

Chapter 3 describes the methodology for data collection and analysis, using both quantitative and qualitative approaches. The research design took the findings of preliminary studies into account, and special care was taken to make sure that intervening factors were controlled as much as possible. The reasons for the methodological selections are also explained, while acknowledging the issues and concerns related to these choices.

The results of the main study are presented in Chapters 4 and 5. Chapter 4 reports the results of the quantitative analysis. Multiple regression methods were utilised with three dependent variables to measure conversational style: *goal orientation, interactional contingency* and *quantitative dominance*. As well as examining to what extent the dataset satisfies the assumptions of multiple regression analysis, the chapter also describes the data collected, including test-

takers' extraversion and oral proficiency levels and a student feedback questionnaire. Chapter 5 interprets and elaborates the statistical results by examining actual interactions with the use of Conversation Analysis (CA). The findings are reported while also discussing the relationship between two test-taker characteristics and group interactions in general, allied to different task types and group sizes.

Chapter 6 summarises and synthesises the findings of this research. The implications of the findings and contributions of this study are considered, and the chapter concludes by describing the limitations of the present study and offering suggestions for future research.