Iowa State University

From the SelectedWorks of Gary Ockey

2014

The potential of the L2 group oral to elicit discourse with a mutual contingency pattern and afford equal speaking rights in an ESP context

Gary Ockey, *Iowa State University*

Available at: https://works.bepress.com/gary-ockey/10/

NOTICE: this is the author's version of a work that was accepted for publication in English for Specific Purposes. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in English for Specific Purposes, [35, (2014)] DOI:10.1016/j.esp.2013.11.003

The potential of the L2 group oral to elicit discourse with a mutual contingency pattern and afford equal speaking rights in an ESP context

1.1 Introduction

The group oral format has emerged as a possible solution to the need for a test of second language (L2) speaking ability which assesses a test taker's ability to engage in a discussion. In this format, a small group of three or more test takers is expected to discuss a topic without support or interference from a test examiner. This oral assessment format may also be amenable to English for Specific Purposes (ESP) situations, such as for an English for business communication program. These ESP contexts require assessments that measure a test taker's content knowledge of a given field of study as well as the test taker's oral language proficiency (Douglas, 2000).

The emergence of the group oral has been due in part to the increasing criticism of test tasks which include only an examiner and a test taker, such as the one-on-one oral interview, for yielding discourse which is controlled by the examiner (Johnson & Tyler, 1998; Kormos, 1999; Lazarton, 1996, 2002; Ross & Berwick, 1992).This one-on-one format makes it unreasonable to infer the degree to which test takers can engage in a discussion with peers as they commonly would in the real world or the language classroom. A group oral assessment format may be more appropriate for assessing this ability to engage with peers. Moreover, the group oral format may make it possible to measure knowledge of a content domain and a more defensible construct of oral ability, one which includes interactional competence (e.g., appropriate turn taking, opening and closing of gambits, responses to others, and negotiation and development of topics) as

well as more traditional measures of oral ability such as comprehensibility, fluency, grammar, and vocabulary (See Appendix A).

Two specific concerns about the degree to which the group oral can provide opportunities for test takers to demonstrate their abilities to discuss a topic have been raised. The first is that test takers could choose to pay little attention to what others in the group are saying, opting instead to plan what they will say next (He & Dai, 2006). This might be of particular concern in English for Specific Purposes (ESP) contexts in which learners have focused on learning content about a particular subject matter domain as well as field-specific vocabulary and are expected to demonstrate their abilities to use this knowledge on an assessment as well as their oral language ability. If test takers did choose to ignore what others say, the resulting discourse would likely not provide an indication of the extent to which test takers could participate in a discussion in which they actively connect what they say to what others say in a pertinent manner (Sacks, Schegloff & Jefferson, 1974; van Lier, 1989). As a result, it would be expected that the discourse would not have a pattern of mutual contingency, a pattern in which the talk is determined partly by each group member's preinteraction plans and partly by the other members' talk (Roloff & Ifert, 1998). If the elicited discourse were composed of disconnected monologues, it may be that the targeted content knowledge can be appropriately assessed, but a test taker's interactional competence, an important aspect of oral ability, cannot. On the other hand, discourse with a mutual contingency pattern, might suggest that both content knowledge and oral ability, including interactional competence, can be effectively assessed by a group oral.

The second concern about the group oral is that some test takers could try to

control the communication, which would likely mean that other test takers would not have equal speaking rights and duties (Berry, 2004; Bonk & Van Moere, 2004; Ockey, 2009) and as a result may not be able to demonstrate their content knowledge or oral language ability. Conversational analysis research, the focus of which is to describe the order and structure of conversations (Sacks, Schegloff & Jefferson, 1974), indicates that equality of rights and duties among speakers is a crucial feature of a natural conversation among peers (Goffman, 1981; Silverman, 1973). If the discourse elicited were dominated by a subset of the members, it would be unreasonable to conclude that the format can effectively assess a broad construct of oral ability as well as knowledge of a particular content.

Specific purpose language assessments should be developed based on an analysis of a specific purpose target language use situation, so the test task and the content are authentically representative of tasks encountered in this context. This makes it possible to make inferences about a test taker's ability to use language in the target language use situation (Douglas, 2000; Taylor & Geranpayeh, 2011). A group discussion designed to assess test takers' abilities to discuss a topic taken from the content domain of interest might be fairly well aligned with the target language use situation for students who are studying in an English for specific purposes context. For instance, a group oral assessment which requires test takers to discuss a topic that they have encountered in their business English language course might be an appropriate way to assess their English speaking proficiency as well as their mastery of the business topics covered in the course. The present study aimed to determine if the group oral format might be appropriate for ESP language learning contexts. More specifically, it aimed to investigate

the extent to which the group oral can elicit discourse with a mutual contingency pattern and afford equal speaking rights and duties when test takers are cognizant of the importance of demonstrating both their content knowledge of the assigned discussion topic and their oral language ability.

Literature review

An increasing amount of literature on the group oral has emerged. The research environments in which these studies are embedded suggest the widespread use of the group oral for various contexts. These studies have been conducted in high stakes English for general/academic purposes environments (Hilsdon, 1995; He & Dai, 2006; Turner, 2008) as well as medium stakes English for general/academic purposes environments, such as for placement in academic English programs or end of term course grading (Bonk & Ockey, 2003; Folland & Robertson, 1976; Liski & Puntanen, 1983; Van Moere, 2006). Findings from these studies coupled with relevant findings of paired oral assessments, in which two test takers are assessed together (Brooks, 2009; Davis, 2009; Ducasse & Brown, 2009; Galaczi, 2008; Iwashita, 1996; Norton, 2005) provide insight into factors that could potentially affect this assessment format.

Research indirectly related to the extent to which the paired and group oral formats can elicit discourse with a mutual contingency pattern has been conducted. Taylor (2000) found that the paired format yielded more varied language, more balanced interaction, and more communicative language functions than a one-on-one oral interview. Similarly, Brooks (2009) reports more interaction, negotiation of meaning, and complex language in the paired format as compared to a one-on-one oral interview. On the other hand, for research conducted on a high stakes group oral assessment in China,

the College English Test-Spoken English Test (CET-SET: He and Dai, 2006), few instances of disagreeing, asking for information, challenging, supporting, modifying one's opinion, persuading, developing, and negotiating meaning in group oral discourse were found, suggesting that test takers chose to produce disconnected monologs rather than co-construct meaning with each other. A follow-up study on the same exam conducted by Gan (2010), however, found that the group oral elicited these language functions in higher level students but not in lower level students, presumably because the lower level students did not have these abilities. It is also likely that the test takers had become more familiar with the expectations of a group oral discussion test than they were when the format was first introduced. The findings from these studies suggest that paired and group assessments can elicit discourse which can provide an indication of a test taker's interactional competence.

Research on the group and paired oral have indirectly addressed the concern that some test takers' speaking rights and duties could be limited by others in the group through an interlocutor effect. Berry (2004) found that test takers were awarded higher scores when they were placed in groups that were comprised of members with high levels of extroversion, while Ockey (2009) found that an assertive test taker was advantaged by being the only assertive test taker in the group and disadvantaged by being in groups which included only assertive test takers. On the other hand, Ockey did not find an effect for group membership based on assertiveness for non-assertive test takers, and Bonk and Van Moere (2004) failed to find an effect for group membership based on gender. An effect for familiarity with partner has been found in some studies (O'Sullivan, 2002; Yamada, 2006; Ying, 2009) while other studies have failed to find an effect for

differences in partner proficiency (Davis, 2009; Csepes, 2001).

Of particular importance to the present research are studies which have analyzed paired or group discourse using a conversational analysis approach. Galaczi (2008) investigated discourse of the paired format yielded from the First Certificate in English test. Based on Storch's (2002) model of dyadic interaction in an ESL classroom context, she identified three discourse patterns: parallel, asymmetric, and collaborative. In parallel talk, the partners both initiated topics, but neither developed the topic initiated by the other, suggesting discourse that did not display a mutual contingency pattern. In asymmetric talk, a dominant partner initiated topics while a passive partner minimally responded to these topics and did not initiate any topics, suggesting unequal speaking rights and duties between them. In collaborative talk, the discourse showed a mutual contingency pattern, and speakers had equal rights and duties with both partners initiating topics and responding to topics initiated by their partner. Galaczi reports that collaborative and parallel patterns and a blend of these patterns were most prominent in her data. May (2009) expanded on Galaczi's research by providing evidence that raters tend to "penalize" candidates who produce asymmetric discourse patterns by assigning them low marks for interactional competence. Nakatsuhara (2011) found that groups of three tended to create more collaborative discourse than groups of four. Groups of three, for example, were more likely to help each other express their views and encourage via "scaffolding" quieter members to participate in the discussion. An effect for proficiency level was found in both groups of three and four.

Despite concerns to the contrary, taken as a whole, the research suggests that, at least under certain conditions, the group oral can elicit discourse with a mutual

contingency pattern and afford test takers equal speaking rights and duties in English for general and academic purposes contexts. The current research was designed to determine if the group oral could also be used to assess the L2 oral abilities of students in ESP programs, in which they were expected to demonstrate both their oral English proficiency and their content knowledge of a specified domain.

1.1.2 Research questions

In an ESP context in which test takers are cognizant of the importance of demonstrating their content knowledge of the assigned discussion topic as well as their second language oral ability to what extent can the group oral elicit discourse:

1. with a mutual contingency pattern?
2. which affords speakers an equal distribution of rights and duties?

To achieve these aims, discourse elicited by a group oral designed to assess the students' abilities to engage in a discussion in English and demonstrate their knowledge of a specific subject matter domain in an international business program was used.

1.2 Material and methods

1.2.1 Context

The program of study was designed to provide English support to students who were completing international MBA degrees in an English-medium university in Japan. The English curriculum included readings and discussions on international business topics. The students were expected to demonstrate knowledge and understanding of these topics by discussing them in English. Vocabulary and sentence structures deemed to be important for this population of students were identified in the readings and students were expected to use these language items and structures during large and small group class

discussions.

1.2.2 Test takers

The test takers were students, who were studying, in the MBA program. All were university graduates, who had experience or background in international business, were approximately 30 years old, and were upper-intermediate speakers of English. When the test was administered at the end of the 8-week program, the test takers knew each other fairly well because they had studied together in classes of 12 for 15 hours per week. Grouping, gender and country of origin for each test taker are provided in Table 1. To give an indication of their English language proficiencies, this table also includes scores assigned to students on the rating scale provided in Appendix A. As can be seen, the students are judged to have intermediate to advanced levels of English proficiency. The test takers are referred to by pseudonyms throughout the paper, and all signed consent forms indicating their willingness to participate in the study.

1.2.3 The group oral

The group oral test employed in this study was used to assess test takers' abilities to discuss a topic in their major area of studies. Test takers were informed that their scores would be based on comprehensibility, fluency, grammar, vocabulary, content, and interactional competence. The rating scale had been designed and used for a number of years to assess both content knowledge of business English issues covered in the course as well as L2 oral ability. At the beginning of the semester, students had been given the rating scale (see Appendix A) to help them set goals and gage their process during the course. Throughout the course, the students were reminded that on the final oral examination, particular attention would be paid to use of the field specific vocabulary and

the ability to discuss an important topic covered during the course. They were informed that the content category would count twice as much as each of the other rating scale categories.

Table 1. Test takers

Test taker Pseudonym	Group	Gender	Country of origin	Comprehensibility	Fluency	Grammar	Vocabulary	Content	Interactional competence
Keiko	One	Female	Japan	3.75	3.50	3.25	3.25	3.50	3.50
Fumiya	One	Male	Japan	2.75	2.75	2.75	2.75	2.75	2.75
Thung	One	Male	Cambodia	3.50	3.75	3.75	3.75	3.50	3.75
Phil	Two	Male	Kenya	3.50	4.25	4.50	4.50	3.50	3.75
Yasuhiro	Two	Male	Japan	3.25	3.50	3.25	3.50	3.50	3.25
Tatsuki	Two	Male	Japan	3.50	3.50	3.50	3.50	3.25	3.25
Than	Three	Male	Indonesia	4.00	4.00	3.50	4.00	4.00	4.00
Shota	Three	Male	Japan	2.00	3.00	3.00	3.00	3.50	3.00
Nao	Three	Male	Japan	3.00	3.50	3.00	3.00	4.00	3.00
Kara	Four	Female	Indonesia	2.75	3.00	2.75	3.00	2.75	3.25
Ban	Four	Male	Laos	3.25	3.25	3.25	3.50	3.25	3.25
Makoto	Four	Male	Japan	3.25	2.50	2.75	2.75	3.00	2.75

Scores based on average of two ratings except for Group Three which was rated by only one rater

Prior to the assessment, classroom teachers assigned students to testing triads with the goal of creating heterogeneous groupings based on country of origin and gender. These principles for grouping mirrored classroom teaching practice, and it was felt that following them would enhance the authenticity of the assessment. These teaching practices had been adopted because teachers believed that the use of heterogeneous grouping by country of origin and gender provided students with opportunities to hear diverse viewpoints and best prepared students to work in an international business

context in which it would be expected that they would encounter various English dialects. On the day of the test, predetermined triads were invited into the testing room and seated in a semi-circle. Test takers were not told who their group members would be before this time, nor the topic that they would be assigned to discuss. They had, however, been told they would be assigned a topic based on the readings they had discussed during the course. All groups of test takers were assessed simultaneously in different rooms to ensure that none of them had an advantage over others by hearing about the topic from their classmates. To begin the assessment, a test administrator read the test prompt out loud and then gave it to the test takers, so they could also read it if they desired. The prompt was: 'Is it true that international investment in developing countries leads to unfair treatment of labor and environmental destruction?' After they had read the prompt, test takers were given one minute to think about how they would respond. After one minute, the test administrator said, 'Please begin,' and subsequently did not speak or interact with the group in any way during the 12 minutes that was allotted for the discussion. After 12 minutes, the group was asked to stop speaking and informed that the test was finished.

1.2.4 Data analysis

Because of its potential to provide a strict test of the degree to which discourse has a pattern of mutual contingency and indicates that test takers have equal speaking rights and duties, the analysis was based on the notion of topic, or "what the discourse is about" (Kormos, 1999; Reinhart, 1982). The analysis employed a modified framework of Keenan and Schieffelin (1976) appropriate for analyzing non-native discourse produced by three speakers. Keenan and Schieffelin's framework requires that for a topic to be

established a speaker must: 1) obtain attention of hearer, 2) speak sufficiently clearly for hearer to decipher what is said, 3) indicate the information that is important for understanding, and 4) indicate semantic relations between referents in the discourse topic. The hearer must: 1) attend to what the speaker says, 2) decipher what the speaker says, 3) identify the information that is important for understanding, and 4) identify the semantic relations between referents in the discourse topic. Topic nomination is based on evidence from the discourse that the speaker has completed the four steps required of the speaker while topic establishment is based on evidence that hearers have completed the four steps required of the hearer. Since three test takers were in each group, topics were considered to be established when all three test takers appeared to accept a discourse topic. Because, in an L2 testing situation, it may be beneficial to fake understanding by providing positive feedback with gestures and agreement fillers when another test taker is speaking, gestures and agreement fillers, i.e., simple agreement or disagreement words such as 'uh huh' or 'I agree with you,' were not considered evidence of topic establishment in the analysis. This decision was based on the need for a strict test of whether or not test takers connected their speech to that of their peers. It should be noted that these procedures would tend to underestimate the number of topics established.

The data analysis procedures were as follows. The analyst conducted preliminary analysis of one of the groups. A focus group of about 12 individuals, who had advanced degrees in applied linguists and training in discourse analysis, discussed the analytic procedures and provided feedback to the analyst. The most notable feedback related to the role of gestures and agreement fillers in determining whether or not a listener had accepted a topic. For reasons discussed above, it was determined that they should be

excluded from the analysis. After feedback from the focus group, the procedures were revised and the analyst completed the analyses. Both transcripts and videotapes were repeatedly analyzed before arriving at a final result. The videotaped discourse was transcribed based on the symbols shown below.

[ ]	overlapping talk
-	self interruption
?	rising intonation
(( ))	non verbal action
(.)	a notable pause of less than half a second
(.5)	a pause of a half second or slightly more

It was assumed that the prompt assigned to the examinees constituted a theme or the more general topic that encompasses related topics. Thus, theme is used to indicate the general topic that was assigned to test takers to discuss and not considered an instance of topic nomination. Theme is, however, indicated in the tables to show how the assigned prompt steered the discussion.

1.3 Results

The results section begins with evidence of test takers co-constructing meaning by connecting their talk to that of others. The example is also used to demonstrate the way the data were coded. The section then provides summary data of the discourse from each group based on the coding scheme.

1.3.1 Evidence of the pattern of mutual contingency and data coding examples

The discourse of each group followed a similar pattern, one that could be identified as one of mutual contingency. Following is an example of a topic, ‘reasons for increased

labor standards,' which was successfully established by Group Two.

59 Tatsuki: As you mentioned in the WTO case ((looking at Yasuhiro))
60 like the consumer - like Rebook or Nike Corporation -
61 American's consumer's pressure on the factory
62 based on Shanghai or Thailand -
63 oh no Hong Kong and Taiwan (.5)
64 the labor standard will improve (.) by their - by the consumer

In this sequence, Tatsuki nominates the topic of 'reasons for increased labor standards.' Step one of the modified Keenan and Schieffelin framework appears to be satisfied since Tatsuki looks at both hearers when he speaks, and step two appears to be met since the analyst was able to decipher what the speaker said. Tatsuki completes steps three and four of topic nomination by introducing his opinion that labor standards will improve as a result of consumer pressure. In the following sequence, which begins 15 lines later, there is evidence that Phil accepts the topic

79 Phil: I think you raise a good point ((looking at Tatsuki))
80 that those foreigners may introduce a new labor standards
81 that will appear favorable
82 and might be appropriated by domestic organizations
83 however many of those labor standards in my view
84 they're not more often -
85 they are created without sensible economic constraints of country

In this sequence, Phil completes the four steps of a hearer for the topic of 'reasons for increased labor standards.' Step one appears to be completed since Phil is looking at

Tatsuki both when Tatsuki speaks and when he speaks, and step two appears to be satisfied since Phil looks as if he is able to hear what Tatsuki said and is facing him. Phil completes steps three and four of topic establishment when he indicates that he thinks labor standards introduced by foreigners may appear favorable but are not created with sensible economic constraints. That is, Phil identifies what is important and identifies the semantic relations between referents in the discourse topic. The nominated topic of 'reasons for increased labor standards' is established when Yasuhiro accepts it in the sequence below, 23 lines later.

108	Yasuhiro:	so your point is ((looking at Phil)) that the local –
109		no no that the international investments do not -
110		is not the real reason for increasing the labor standards
111	Phil:	No
112	Yasuhiro:	in fact?
113	Phil:	No
114	Yasuhiro:	No?
115	Phil:	No.
116	Yasuhiro:	Yes?
117	Phil:	That's what I mean
118	Yasuhiro:	oh (.5) okay I agree with this idea
119		and I also would like to add that (.5)
120		uhh international investment - I mean the factories -
121		I mean the companies can be in the developing countries
122		but the number should be very small

123		and also the people
124		who work for these companies have very low proportion
125		in this country
126	Phil:	[right]
127	Yasuhiro:	[so] I don’t think umm ((looking at Tatsuki)) (.5)
128		I really agree with your idea partly
129		that the national –
130		international investments increase the labor standard
131		in somewhat - to some extent
132		but still I believe in his ((pointing to Phil)) idea
133		that the whole economy growing –
134		growth will raise their labor standard
135		in that country

Yasuhiro completes steps 1 and 2 in the framework by looking at both Phil and Tatsuki and appearing to hear them. Yasuhiro completes steps three and four of topic establishment when he confirms Phil's view on the topic in line 110, and partially agrees with Tatsuki's view on the topic in lines 128 to 131.

The example illustrates that the test takers were actively connecting their talk to that of the others' talk in the group. Of particular note is the exchange between Phil and Yasuhiro in lines 111 to 117 in which Yasuhiro was sure to clarify what Phil meant before he agreed with his position. This example also provides some indication of the extent to which test takers can demonstrate their knowledge of the topic: international investment in developing countries leads to unfair treatment of labor and environmental destruction, while participating in a discussion. Each indicates a view of the topic and provides a defense of this position based on knowledge of the topic.

1.3.2. Discourse topics and patterns

Tables 2 through 5 present the topics which the analyst was able to identify in the discourse of the four groups of three test takers. The theme, that is, the general topic introduced by the assessment prompt, is in bold. In each of the tables, the left column displays a letter assigned to each topic, and the second column indicates the topic nominated. The the third and fifth columns display numbers which indicate the order in which a nominated or established topic was present in the discourse. For instance, for group One (Table 2), the '1' shows that the first move was for Fumiya to nominate the theme, 'Foreign investor's treatment of local workers'. The '2' shows that the next move was for Fumiya to nominate the topic 'Wages provided by international companies'. The '3' signifies that after this, Keiko accepted the topic, 'Foreign investor's treatment of local workers', and the '4' indicates that the next move was for

Keiko to accept the topic, ‘Wages provided by international companies’. The fourth column identifies the test taker that nominated a topic, and the last column indicates the person(s) that accepted a topic. Instances in the fourth column where a person’s name appears twice indicate that the person tried to nominate the topic more than once. Blanks in the fifth column indicate that the nominated topic was not accepted by the other test takers while blanks in the sixth column make it possible to determine which test takers did not accept the nominated topic.

The way in which topics were established indicates that this discourse displayed features of mutual contingency. That is, the discourse pattern that developed was partly determined by preinteraction plans and partly by the conversational behavior of the other participants. Speakers tended to nominate a number of topics in one turn, so it was often not clear whether a topic was established until many lines later, and usually after a number of topic nominations or renominations. This pattern is somewhat expected in academic discourse because participants commonly nominate more than one topic to support their views. Other participants have a tendency to establish these topics by addressing them one at a time. This pattern is indicated in the ordering of the respondents’ establishment of the topic in column 5.

(Insert Table 2 about here)

For instance, for Group One (Table 2), the second topic (B) was nominated (labeled 2 to indicate the second turn) by Fumiya and accepted by Keiko after one other topic had been nominated (indicated by the 4 in column 5 next to Keiko). However, Thung did not accept the topic until nine other topics had been nominated, renominated, or accepted as topics. This is indicated by the number 14 in column five next to Thung. It is also apparent that at the beginning, a number of topics were often nominated or established in one turn, while toward the end of the discourse, a number of sequential turns tended to focus on one topic. In other words, preinteraction plans

tended to have more effect on the discourse during the early part of the 12 minute discussion while other group members' conversational behavior tended to have more effect on the discourse as the discussion progressed. Both of these patterns held for all four groups.

1.3.3 Topics nominated and established

The first research question was: In a group oral to what extent can discourse show a mutual contingency pattern? The degree to which test takers were able to nominate topics and the extent to which these topics were established were used to answer this question.

The results for Group One are presented in Table 2. The data show that Group One test takers nominated the two themes in the prompt. Nine topics were nominated within these two general themes, two within 'Foreign investor's treatment of local workers' (A) and seven in relation to the theme of 'Foreign investments' effect on environment' (D). Of the nine nominated topics, four were established (B, E, F, and H) since a speaker introduced each topic and both listeners accepted it, and one (G) was incompletely established since only listener, Thung accepted the topic that a speaker, Fumiya nominated—Keiko did not accept this topic.

Group Two's results are shown in Table 3.

(Insert Table 3 about here)

Group Two test takers nominated the two themes in the prompt and six topics. The theme of 'Foreign investor's treatment of local workers (A),' was established as a topic, while the theme of 'Foreign investments' effect on environment (D),' was not. Six topics were nominated, all within the established theme (A). Three of the six topics (E, F, and G) were established, and one (H) was incompletely established.

The results for Group Three are shown in Table 4.

(Insert Table 4 about here)

Group Three nominated the two themes in the prompt and eight topics within these themes. The theme of ‘Foreign investor’s treatment of local workers (A),’ was established as a topic, while the theme of ‘Foreign investments’ effect on environment (D),’ was incompletely established. Three of the nominated topics (B, C, and E) were established, and two (G and I) were incompletely established.

Group Four’s results are presented in Table 5.

(Insert Table 5 about here)

The group nominated the two themes in the prompt and eight topics within the themes. Both themes (A and C) and five of the eight topics (B, D, E, G, and I) were established, while two other nominated topics (F and H) were incompletely established.

1.3.4 Individual test takers’ rights in topic nomination and establishment

The second research question was: In a group oral, to what extent can there be equal distribution of speaking rights and duties in an ESP context? The extent to which each test taker in a group was able to nominate topics and the degree to which each test taker’s nominated topics were established was used to answer this question.

For Group One, all members nominated at least two topics, but there was a fair amount of variation in the number of topics nominated by each test taker that became established. Fumiya nominated theme A, and Keiko nominated theme D, both of which were established as topics. Fumiya nominated topics B, F, G, H, and K of which B, F, and H were established. Keiko nominated C and E of which E was established. Thun nominated I and J, neither of which was established as a topic. Both of Thun’s topics were nominated toward the end of the 12-minute testing time, so it is not clear whether the other test takers chose not to accept either of Thun’s topics or whether there was not enough time for them to accept these topics because the test

administrators stopped the test at the end of the allotted 12 minutes for the assessment. That is, if the test had been longer, it may have been that Thung's nominated topics would have been established.

In Group Two, each test taker nominated two topics, and each had one nominated topic established. Yasuhiro nominated the theme of Foreign investor's treatment of local workers, which was established as a topic, and Phil nominated the theme of Foreign investment's effect on environment, which was accepted by Tatsuki, but not Yasuhiro. Yasuhiro nominated two topics (B and G), one of which was established. Phil also nominated two topics (C and F), one of which was established and one of which was incompletely established. Like Yasuhiro and Phil, Tatsuki nominated two topics (E and H), one of which was established and one of which was incompletely established.

In Group Three, a very unequal distribution of number of topics nominated by each test taker and number of established topics for each test taker was present. Shota nominated the theme 'Foreign investor's treatment of local workers' (A) which became established, and Than nominated the theme 'Foreign investment effects on the environment' (F) which Shota accepted but Nao did not. Shota nominated two topics (B and E), both of which were established. Than nominated five topics (C, D, G, H, and I), one of which was established and two of which were incompletely established. Nao nominated the last topic, and neither Shota nor Than accepted it as a topic.

In Group Four, all test takers nominated at least two topics, and each had at least one topic established. Kara nominated both the theme of 'Foreign investment effects on environment' and 'Foreign investor's treatment of local workers,' and both were established as topics. Kara also nominated three topics (B, D, and J). Topics B and D were established while J

was not; it was nominated in the last turn in the discourse. Makoto nominated three topics (E, G, and H). Topics E and G were established while H was incompletely established. Ban nominated two topics (F and I). Topic I was established, and topic F was incompletely established.

1.4 Discussion

1.4.1 Extent to which group oral can elicit discourse with a mutual contingency pattern

In discourse with a mutual contingency pattern, test takers need to connect their talk to that of others in a 'here and now' way. In this study, the degree to which the groups' discourse had a mutual contingency pattern was measured by the extent to which groups were able to nominate and establish topics. The study found that all groups were able to nominate topics related to the two themes in the assigned discussion prompt, and roughly half of each group's nominated topics became established. These findings suggest that much of the discourse co-constructed in the groups had a mutual contingency pattern, that is, in addition to being affected by an individual group member's preinteractional plans, communication was affected by all members of the groups conversational behavior.

One possible reason that more of the nominated topics did not become established is that there were too many competing topics nominated at the beginning of the discourse for hearers to remember to respond to all of them. This might have resulted from test takers having time (one minute) to think about how they would respond before they began the discussion. This finding is in line with that of Van Moere (2007) who analyzed groups of four Japanese speakers in an English for academic purposes context. While at the beginning of the discussion, test takers tended to initiate a number of topics in one turn, toward the end of the discourse, a number of sequential turns tended to focus on one topic. A possible reason for this result is that in an ESP context, test takers want to be sure that they demonstrate at least some of the content knowledge

that they know about an assigned topic early on in an assessment. As a result, they nominate various topics without providing an opportunity for others to establish these topics. As time progresses, they either become satisfied that they have demonstrated enough content knowledge to do well on the assessment, or their involvement in the discussion becomes more important than their concern about demonstrating their content knowledge. This increased turn taking as the discussion progresses provides an opportunity for the raters to assess a test taker's interactional competence, an important component of oral ability. Given the need to provide test takers with opportunity to demonstrate both content knowledge and interactional competence, it is therefore crucial that sufficient time is allotted when the group oral is used in an ESP context.

In discourse with a mutual contingency pattern, it would not be expected that all nominated topics would be established. For instance, some topics which were nominated right before the 12-minute time limit elapsed could not possibly be established based on the criteria for topic establishment. Furthermore, the criteria used for establishing a topic may have been unreasonably stringent. In some cases, one or two of the test takers would talk about a nominated topic for a number of lines with the other one or two test takers seemingly understanding (based on body language and verbal cues such as 'uh huh'); however, it was not judged as an established topic since the other test taker(s) did not accept the topic by completing the four steps in the Keenan and Scheifflin (1976) framework. As mentioned in the methodology section, this strict criteria was deemed necessary given the possibility that test takers might pretend to understand what others were saying as a strategy for increasing their test scores. It is also possible that hearers had nothing to say about a topic or felt that what the speaker said was complete, and chose not to respond to the nominated topic, even though they were engaging with the topic in a 'here and now' way. Given these strict criteria for topic establishment, the fact that

roughly half of the nominated topics became established is rather convincing evidence that the group oral does have potential for eliciting discourse with a mutual contingency pattern. The possibility that test takers were often planning what they would say next rather than listening to what others were saying does not seem likely given the time constraints of the task, the stringent criteria for topic establishment, and that many of the nominated topics were established.

Determining the degree to which the group oral discussion has the potential to elicit a discussion in an L2 ESP context is challenging because it may be that the test takers do not have the communicative skills that the test aims to elicit (Gan, 2010). That is, if test takers do not demonstrate the target language ability, it may not be possible to determine if they had the ability or not, making it difficult to judge the effectiveness of the test task at eliciting the desired abilities. The context of this study was deemed appropriate for investigating this question because the test takers all had upper-intermediate levels of English and had spent time developing their abilities to sustain a discussion on a topic similar to the one that the prompt was designed to elicit. These different contexts may help to explain why He and Dai (2006) found little evidence of communication skills elicited by the group oral, since the test takers in He and Dai's study may not have had the communication skills necessary to sustain a discussion with their peers.

It is important to note that this study did not directly investigate the extent to which the group oral provided a valid estimate of test takers' content knowledge of the international business topic that they were assigned to discuss. While the research suggested a pattern of mutual contingency in which test takers did have some opportunity to demonstrate their content knowledge, it is not clear how much of their content knowledge the test takers actually were able to demonstrate. Future research should aim to determine how much of a test takers content

knowledge can be assessed by a group oral discussion and how much time is needed for this purpose in a given context.

1.4.2 Extent of equal distribution of speaking rights and duties

To determine the extent to which test takers had an equal distribution of speaking rights and duties when taking the group oral in an ESP context, number of topics nominated and established for each member of a group was investigated. The findings indicate that every member of each group nominated at least two topics, and the number of topics nominated by each member of each group was fairly evenly distributed. This would indicate that group members had equal rights as speakers to introduce and change topics suggesting that the group oral may be better suited for assessing the oral ability of topic nomination than assessments such as the one-on-one oral interview which has been criticized for failing to assess this ability (Kormos, 1999).

In two of the groups (Group Two and Group Three), members had a roughly equal distribution of topics nominated and established, while in the other two groups (Group One and Group Four), there was a fairly even distribution of number of topics nominated but an uneven distribution in number of topics established. A partial explanation for this finding is that some topics were nominated toward the end of the 12-minute discussion, and opportunity for other members to respond to them was not available. The results of the study suggest the importance of designing the group oral so all members of the group have sufficient time to nominate topics and have their ideas acknowledged by the other members of the group. It is also important to point out that while it is desirable that speakers' rights and duties are the same, only potential for such equality among speakers is necessary for authentic communication (van Lier, 1989).

1.5 Conclusions

The findings of this research indicate that the group oral can elicit discourse with a mutual contingency pattern, and can generally afford test takers equal speaking rights and duties in ESP contexts. The research suggests, however, that various factors can impact the extent to which this is the case. One such factor is the amount of time assigned for the discussion. Enough time must be allotted for test takers to demonstrate both their knowledge of the content and their interactional competence. This finding coupled with the results of other group oral studies underscores the importance of carefully determining the appropriate test task characteristics for a given ESP context and appropriately implementing them when using the group oral to assess both field-specific content knowledge and a rich construct of oral language ability, which includes interactional competence.

References

Berry, V. (2004). A study of the interaction between individual personality differences and oral performance test facets. Unpublished Doctoral Dissertation. King's College, University of London.

Bonk, W., & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing, 20(1)*, 89-110.

Bonk, W., & Van Moere, A. (2004, March). L2 group oral testing: The influence of shyness/outgoingness, match of interlocutors' proficiency level, and gender on individual scores. Paper presented at the annual meeting of the Language Testing Research Colloquium (LTRC), Temecula, CA.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing 26(3)*, 341-366.

Csepes, I. (2001). Is testing speaking in pairs disadvantageous for students? A quantitative study of partner effects on oral test scores. *novELTy, 9(1)*, 1-17.

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

Douglas, D. (2000). *Assessing languages for specific purposes*. Cambridge: Cambridge University Press.

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

Folland, D., & Robertson, D. (1976) Towards objectivity in group oral testing. *English Language Teaching Journal, 30 (2)*, 156-167.

Galaczi, E. D. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in

English examination. *Language Assessment Quarterly, 5*, 89-119.

Gan, Z. (2010). Interaction in group oral assessment: A case study of higher- and lower-scoring students. Language Testing, 27(4), 585-602.

Goffman, E. (1981). *Forms of talk.* Oxford: Basil Blackwell.

He, L., & Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing 23 (3)*, 370-402.

Hilsdon, J. (1995). The group oral exam: Advantages and limitations. In J. Alderson & B. North (Eds.), *Language testing in the 1990s: The communicative legacy* (pp.189-197). Hertforshire: Prentice Hall International.

Iwashita, N. (1996). The validity of the paired interview format in oral performance assessments. *Melbourne Papers in Language Testing 5(2)*, 51-66.

Johnson, M., & Tyler, A. (1998). Re-analyzing the OPI: How much does it look like natural conversation? In R. Young & W. He (Eds.). *Talking and Testing: Discourse approaches to the assessment of oral proficiency* (pp. 27-51). Philadelphia: John Benjamins Publishing Company.

Keenen, E., & Schieffelin, B. (1976). Topic as a discourse notion: A study of topic in the conversations of children and adults. In C. Li (Eds.), *Subject and Topic* (pp. 66-113). New York: Academic Press.

Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: A conversation analysis of roleplay and non-scripted interviews in language exams. *Language Testing (2)*, 16, 163-188.

Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing, 13 (2)*, 151-172.

Lazarton, A. (2002). *Studies in language testing 14: A qualitative approach to the validation of oral language tests.* (M. Milanovic, Ed.). Cambridge University Press.

Liski, E., & Puntanen, S. (1983). A study of the statistical foundations of group conversation tests in spoken English. *Language Learning, 33 (2),* 225-246.

Norton, J. (2005). The paired format in the Cambridge speaking tests. *ELT Journal 59(4),* 287-297.

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

Nakatsuhara, F. (2011). Effects of test taker characteristics and the number of participants in group oral tests. *Language Testing, 28(4),* 483-508.

Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test score. *Language Testing, 26(2),* 161-186.

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

Reinhart, T. (1982). *Pragmatics and linguistics: An analysis of sentence topics.* Bloomington: Indiana University Linguistics Club.

Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition, 14 (2),* 159-176.

Roloff, M., & Ifert, D. (1998). Mutual influence in persuasive contexts. In M. Palmer & G. Barnett (Eds.), *Volume XIV: Progress in communication sciences*. Stamford CN: Albex Publishing Corporation.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematic for the organization of turn-taking for conversation. *Language, 50 (4),* 697-735.

Silverman, D. (1973). Interview talk: Bringing off a research instrument. *Sociology, 7 (1)*, 31-48.

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

Taylor, L. (2000). Investigating the paired speaking test format. University of *Cambridge ESOL Examination Research Notes, 2*, 14-15.

Taylor, L., & Geranpayeh, A. (2011). Assessing listening for academic purposes: Defining and operationalizing the test construct. *Journal of English for Academic Purposes, 10*, 89-101.

Turner, C. (2008). The specificity of the "research approach" in classroom studies: Probing the predictability of washback through teacher conceptual and instrumental evidence in Quebec high schools. Language Testing Research Colloquium, *Hangzhou, China.*

van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly, 23 (3)*, 489-508.

Van Moere, A. (2006). Validity evidence in a group oral test. *Language Testing, 23(4)*, 411-440.

Van Moere, A. (2007). Group oral tests: How does task affect candidate performance and test scores? Unpublished Ph.D. thesis, Lancaster University, Lancaster, UK.

Yamada, H. (2006). An analysis of oral testing in pairs. Research Journal of *Jin-ai University, 5*, 69-84.

Ying, B. (2009). The impact of familiarity on group oral proficiency testing. *CELEA Journal, 38(2)*, 114-125.

Appendix A

Rating scales for speaking tests

	Comprehensibility	Fluency	Grammar	Vocabulary	Content	Interactional competence
5	Rarely mispronounces, able to speak with nearly perfect pronunciation, intonation, and rhythm, little or no foreign accent	Fluent speech, speaks confidently and effortlessly, speech is smooth and natural	Uses high level discourse with near perfect accuracy, shows an ability to use the full range of grammatical structures effortlessly and accurately which are needed to achieve the task	Confidently uses wide range of technical and general vocabulary precisely and effectively	Is able to express opinion and support it with strong logical explanations and facts, is confident discussing subject matter in a sophisticated manner	Shows confidence and naturalness, shows ability to negotiate meaning, shows how ideas or opinions are related, may initiate conversation, shows near perfect pragmatic usage, completes task effectively
4.5 4.0	Pronunciation is clear, occasionally mispronounces or has non-perfect intonation or rhythm, articulation is clear, has mastered all sounds, accent may sound foreign, but does not interfere with understanding	Speaks with confidence, but has a few unnatural pauses, occasionally gropes for words unnaturally	Shows ability to use nearly the full range of grammatical structures, but may make some errors when using some complex sentence types, errors do not interfere with meaning	Shows range of technical vocabulary which is sufficient for task, but fine shades of meaning are occasionally inappropriate	Is able to express opinion and support it with logical explanations and facts, support may be considered moderately strong, is fairly confident discussing subject matter in a sophisticated manner	Generally confident, responds appropriately to an opinion, shows ability to negotiate meaning, shows how ideas are related, may not always be pragmatically acceptable, completes task effectively
3.5 3.0	Pronunciation is not perfect but can be understood without concentrated listening, articulation is generally clear, may not have mastered all sounds	Speech is a little hesitant, has some unnatural rephrasing and groping for words	May not have mastered full range of structures, but uses complex and simple sentences, may make a few global errors, has no trouble completing task	Has sufficient vocabulary to complete task, but may not use it appropriately, may use technical vocabulary, but not always effectively	Is able to express opinion and support it somewhat effectively, support is generally logical and clear but may not be convincing, may not be confident discussing subject matter	Somewhat confident, responds appropriately when asked for opinion, may sometimes be pragmatically unacceptable, completes task somewhat effectively
2.5 2.0	Sometimes mispronounces, may require concentrated listening, but is completely understandable, may not articulate clearly, may not have mastered some sounds	Speech is often hesitant, frequent unnatural rephrasing and groping for words,	May use simple sentences to express meaning, complex sentences are used but often inaccurate, can express desired meaning, errors may occasionally interfere with meaning, is able to complete task	Vocabulary is adequate for achieving task, but often used inappropriately. Does not accurately use technical terms used in the field	Is able to express opinion, but supporting claims may not be clear or completely logical, may not have solid reasons or evidence for opinions, arguments may not be convincing	Not confident, shows agreement or disagreement to opinions at the surface level but not at the discourse level, may often be pragmatically unacceptable, completes task but not effectively
1.5	Frequently mispronounces, accent impedes comprehensibility, requires concentrated listening but is generally comprehensible	Strained speech, often groping for words, some long unnatural pauses (except for routine phrases)	Relies mostly on simple sentences which are often inaccurate, cannot control complex sentences, mistakes often impede meaning,	Lacks the necessary vocabulary to discuss the topic with any sophistication	Has a difficult time expressing opinions clearly, supporting claims are extremely weak, has no solid reasons or evidence to support claims, is	May use simple phrases to show agreement or disagreement, but does not relate ideas at discourse level, may often be pragmatically unacceptable,

1.0			has difficulty completing task		not convincing at all, does not appear to know much about topic	task may not be completed
0.5 0.0	Frequently mispronounces, heavy accent, even with concentrated listening often incomprehensible	Fragmented speech that is so halting that conversation is virtually impossible	Cannot control even simple sentences, grammar is not sufficient to complete task	Vocabulary is inadequate to achieve the task	Shows frustration in trying to communicate opinions, has no clear claim or does not support claim, may know little or nothing about the topic	May require prompting, produces monologues which are unrelated, is not aware of appropriate pragmatic usage, does not complete task

If a student consistently fulfills the criteria being tested, s/he receives the score at the top of the box; whereas if s/he sometimes achieves the expected level, but sometimes slips to a lower criteria, s/he is given the lower score in the box.