Group Discussion Tests: investigating the construct of interaction

D. A. Leaper

Hankuk University of Foreign Studies

davidleaper@gmail.com

A Group Discussion Test (GDT) is an assessment of conversational ability in which three or four test-takers discuss a prompt for a specified period of time. Unlike other speaking assessment methods, apart from starting and stopping the conversation, the rater does not take part, meaning that the interaction is between relatively equal status participants and should result in a more conversation-like sample of language. This presentation examines this assumption by investigating the construct of interaction in 82 videoed GDTs that were part of a proficiency test that took place at a university in Japan. Using a speech functional analyses framework, the tests' discourse structure was analyzed and the extent to which the turn-taking, topic nomination and involvement resembled conversation was investigated. The investigation revealed that for the most part, features of conversation were represented in the GDTs. However, there were also cases where the resulting interaction was not very conversation-like at all. These cases are discussed and recommendations made that should decrease the likelihood of dysfunctional GDTs.

Key words: oral assessment, group discussion test

I. Introduction

The dominant method of assessing speaking ability has been through the use of the structured interview. The assumption was that interview tests were measuring the ability to converse (Lazaraton, 1992). However, the limits of the interview format to elicit fundamental aspects of communicative behavior soon came to be questioned, most notably by van Lier (1989) pointing out that conversations are unplanned, participants have equality of status and there is an unpredictable sequence and outcome, while interviews are "almost inescapably asymmetrical" (P501) being planned and dominated by the interviewer's asking of questions and evaluation of the answers.

The concerns of van Lier (1989) were bourne out by a conversation analysis of an OPI test used in workshops for training OPI examiners. Johnson and Tyler (1998) based their analysis on procedures of Schegloff and Sacks (1973), Sacks et al (1974) and Schegloff et al (1977) to examine features of conversation. The researchers expected that participation in a naturally occurring conversation would be approximately balanced between the participants, and that there is a need to maintain the face of those who take part. However, in the analysis they found that turn-taking, length of turns and their distribution were determined entirely by the two interviewers.

Johnson and Tyler (1998) also looked into the role of topics in the conversation as spontaneously created and collaboratively constructed by conversationally involved participants. They found that the topics are decided in advance and contrived for the purpose of the interview. This is shown whenever the interviewer shifts to a new task and is particularly obvious when the interviewer introduces and ends the role.

This work was supported by Hankuk University of Foreign Studies Research Fund of 2010

A later study that used data from 35 OPI interviews confirmed interviewer dominance of topic (Johnson, 2001). In a typical example, the interviewer changed the topic from a description of a room to a question designed to elicit a response about current events in South-East Asia, topics that bear little relationship to each other (Johnson, 2001).

It is clear then, that in terms of turn taking, turn sequence and topic nomination, OPIs depart from what is expected in conversational interactions. This makes it unlikely that what Gumperz (1982) calls 'conversational involvement' can be achieved by the participants of an interview. Participants of a discussion show conversational involvement by engaging actively with each other and showing interest in the topic. This involvement may be expressed by the participants evaluating or reacting to what the other participants are saying and avoiding open disagreement and face threatening acts – or at least using indirectness and hedges to present them. The crucial difference is that the role of the interviewer is pre-assigned and this introduces a power difference that is not conducive to producing conversational language (Young 1995).

If the construct of interaction in conversation is what we are trying to measure in a speaking test, then a solution would be to remove the interviewer from the interaction. This is what a Group Discussion Test (GDT) does.

II. Group Discussion Tests: a literature review

GDTs have appeared in the literature from time to time as a solution to large scale oral testing (Folland & Robertson, 1976, Hilsdon, 1995). In these tests, a group of students are given a subject to discuss, and the rater only assesses the language they produce without intervening. An advantage of using the GDT is that raters can focus on content alone, and students may feel less stressed than in interview style tests (Fulcher, 1996, He & Dai, 2006). They are also an efficient method of judging the oral ability of large numbers of students in a short time (Bonk & Ockey, 2003). Further research has found that statistically, the number of turns had no impact on student scores as a main effect and neither did the interaction between words spoken and incoming proficiency level of the student. This shows that high level students do not improve their scores by talking more and that low level students are not penalized for speaking less (Kobayashi, Van Moere & Johnson, 2005). At the same institute the effects of gender and shyness were researched (Bonk & Van Moere, 2004, Van Moere, 2006). In the 2004 study, gender was found to have no significant impact on scores, and shyness a minimal one. Shyness was determined by a survey given immediately after the test to 1,150 students in over 300 tests and compared the results to teacher predictions of student ability. At the most extreme ends of the scale, the difference between the shyest and most outgoing students could be quantified as a difference of 2.5 points out 20. Van Moere (2006) further investigated shyness in a questionnaire on test-taker perceptions and found that shy students were no more likely to feel they did not perform to their ability than non-shy students, and a substantial majority of the students did not feel that their contribution to the discussion was inhibited by the other members in their GDT.

In the literature on GDTs, studies that investigated the language elicited by the task are rare. He and Dai (2006) examined the group discussion task that is part of the College English Test – Spoken English Test (CET-SET). The study analyzed 48 transcripts of 197 candidates and found that the number of candidates who engaged in negotiating opinions numbered only six out of the 144 candidates. Since the purpose of the task was to elicit this kind of language, and candidates are apparently aware of the requirements, a mismatch was found between the design of the test and group discussion as a means of carrying it out. As He and Dai state, "conversational features do not appear in speaking tests just because we introduce speaking partners with equal social power" (2006:393). It seemed that candidates were concentrating on

producing their response rather than listening to another's opinion and reacting to it. Another reason they put forward was lack of interest in the prompt. Their survey showed that 60.2% of the candidates found the topics of the discussion uninteresting or dull. This contrasts with the Van Moere (2006) survey in which 80% of participants thought the prompts used were "effective for making people talk". The contrasting results probably reflect the different way the prompts were constructed. The CET-SET prompts are a single question ("Is it desirable to live in a big city?") whereas the Van Moere (2006) prompts have multiple questions and are presented in the form of a paragraph, along with a translation in the students' native language (Van Moere's prompts used the same format as this study, see appendix 1 for examples).

From this literature survey it can be seen that when it comes to eliciting conversation, the exchanges in interviewer-led tests are affected in terms of the turn-taking, topic and type of response. Having group discussions without a pre-assigned interlocutor is an alternative that seems promising. It offers practical advantages in the way it can assign large numbers of students scores for their oral ability, neither gender nor placement with different proficiency levels have been shown to affect scores, and shyness seems to have a limited impact. However, He and Dai's study raises concerns about the language that this format of test can elicit.

III. Analyzing Group Discussion Tests as conversation

To investigate the interaction of the GDT the Speech Function Analysis as explained in Eggins and Slade (1997) was chosen. The SFA is based on Halliday's (1984, 1997) functional-semantic account of dialogue and its objective is to account for the way "the participants are constantly negotiating relationships of solidarity and intimacy" (Eggins and Slade, 1997:179). It was deemed appropriate because it attempts to provide a full account of conversation and lends itself to a quantitative handling of the data more readily than other forms of analysis.

The SFA analyzes the conversation into 'moves' that are classified by function. A move is a unit of discourse that can be regarded as complete if it is possible that another speaker could begin talking, and not be seen as interrupting. They can be recognized by both grammatical and prosodic factors. In terms of grammar the criteria for recognizing a move is one of independence. Those that are dependent on another clause for mood are generally not recognized as a separate move. Thus, dependent clauses in which the subordinate clause comes first suggest that the speaker has not finished the move, and so is counted as part of the move represented by the main clause. Where the main clause comes before the subordinate clause, they too can usually be counted as a single move unless prosodic features dictate otherwise. Embedded clauses and quoting or reporting clauses too take on the function of the move that they are part of.

Identifying a separate move is also dependent on the prosodic features of the utterance. Here the signal of the point where a transfer may be made can be found in the place where the speaker chooses to pause. While the grammatical clause can signal a potential end to the turn, speakers can maintain the move by rushing on to the next clause without pausing. In these cases, two independent clauses are considered part of the same move. And as noted above, where the dependent clause comes first, it would be a separate move if the speaker chooses to pause between it and the main clause.

After the transcripts are divided into moves, they are analyzed for function. In this study I am interested in the construct of interaction as outlined in the literature review above, so I will explain two functions that are relevant to this study.

The first feature is the 'Open' move. These set the topic and give the participants something to talk about. They are not necessarily the first words of the conversation, since participants may start by greetings and introductions. Open moves may be

carried out by either a statement or a question. As the conversation develops they may be used to change the topic, or if the participants run out of things to say, restart the conversation.

This is a conversation test, so the initial topic is set by the prompt. However, the test-takers are assessed on their talking ability, not on how they deal with the topic, so they do not lose marks if they change the topic. Nonetheless, that the topic is set by the prompt runs counter to the notion of the topic of conversation being spontaneously created. In the analysis then, the Open moves will be analyzed by examining those that come directly from the prompt, those that are influenced by the prompt, and those that are unrelated. If the interaction of the GDT resembles conversation it would be expected that the participants would adapt the topic of the prompt to the context of the discussion and new topics will collaboratively emerge.

The other feature to be examined is 'conversational involvement', that is, the extent to which the participants are committing to the conversation. A feature identified in the SFA as the Register move seems to be the obvious choice for this. These moves add no new information, but are how the listener indicates support for the speaker. The most minimal example of this move is the 'mm' which shows that the listener is paying attention, but they also include such utterances as 'That's interesting'. Although there are visual manifestations of this move, like nodding for example, this study will use the verbal Register moves as an indication of the listeners' level of involvement in the conversation.

Finally, turn-taking will be examined as a signal of how equal the participants are in the conversation. In this analysis, every move is counted as a turn, and this includes minimal Register moves. As a conversation should take place between agents with equal status, it should be expected that the number of turns in the GDT be reasonably equal. To determine how equal they are, the standard deviation of the number of turns by the participants will be calculated as an indication of spread.

This paper will proceed by identifying the GDTs in the corpus that are rich and poor in these features, and then examine its circumstance so that recommendations can be made for using GDTs as a form of assessment.

IV. Data for the Study

The data was collected at a private languages university in Japan. The students at are a uniform group, all having studied English for six years at senior and junior high schools in Japan, all choosing to major in English, with an average TOEIC score of 572. The GDTs are the oral component of a four skills proficiency taken as incoming first year students and at the end of their first and second years. Every year more than 1500 students take the test over two days in January and another 700 incoming freshmen in March. For incoming students it is used as a placement test, and for first and second year students it makes up 20% of their grades for some subjects.

The format of the GDTs is to have either three or four participants assigned randomly to take the test together, but not with students who were their classmates or raters who were their teachers. The data for this study is taken from videos taken as a matter of procedure during two administrations of the test. The first one was in 2004 when 113 incoming first year students were recorded in 30 GDTs, and the second was in January 2005 when 188 students in 52 GDTs were recorded after their first year of study.

For each GDT one of four different prompts is chosen at random by the raters. The test-takers have a minute to read and consider what they will say before the discussion starts. A Japanese translation appears beneath each prompt to ensure understanding. The prompts used in the 2005 administration can be found in the appendix

During the test, the raters indicate when to begin and finish and then rate the test independently using a five band rating that includes fluency, pronunciation, grammar,

vocabulary and communicative skills. They do not intervene in the discussion unless a candidate has not spoken enough to give a score. This happens seven times in 2004 GDTs, and only two times in the 2005 administration.

The transcripts were coded using the Speech Function Analysis (SFA) of Eggins and Slade (1997), and a database was created using UAM Corpus Tool, which can be used to count and identify the features of interaction that this study is interested in.

V. Results

Table 1 below summarizes the data of the corpus. The numbers seem to indicate improvements made by the test takers after one year of study. However, this requires further analysis as other factors such as level of participants and time need to be taken into account.

Table 1. The corpus formed from two administrations of the test

	March 2004 Admin 30 GDTs		Jan 2005 Admin 52 GDTs	
	No.	Avg per Grp	No.	Avg per Grp
Stds	114	3.8	188	3.6
Words	12905	430.2	30732	591
Moves	2033	67.8	5536	106.5

The first feature to be examined is the Open move. These were identified and coded as either directly from the prompt (either as a question or as a statement in response to the prompt), influenced by the prompt, or unrelated to the prompt. Formally defining the topic of a conversation has long been recognized as a difficult task (Brown & Yule, 1983). In this study, the Open moves that came from the prompt were recognized by being able to match the words used by the test-taker with the prompt. Open moves influenced by the prompt had to be on the same general subject. For example, from the discussion on cell phones a common question was “How much is your cell phone bill?” even though there is no mention of expense in the prompt. To be considered to be unrelated to the prompt, the new topic has to move the conversation away from the prompt’s subject. For example, on the discussion on indoor and outdoor sports, the Open move “do you belong to some circle or club activity?” moved the discussion onto what club the participants belonged to, and hence was considered an unrelated Open move. Although it was sometimes difficult to decide the difference between ‘influenced by the prompt’ and ‘unrelated subject’, for the purposes of this study the distinction is not so important since in both cases the participant in the test is engaging in an interactive way with the topic, as we would expect in a conversation.

In Table 2 it can be seen that the participants relied on Open moves from the prompt, but overall, there are more Open moves influenced by the prompt and there were a substantial number of unrelated Open moves. This shows that the test-takers were adapting the subject of the conversation and is an indication of the collaboration that we might expect to find in conversational interaction.

Table 2. Open moves

	2004	2005	Total
Question from Prompt	37	72	109
Statement on Prompt	8	23	31
Influenced by Prompt	35	92	127
Unrelated to Prompt	20	31	51
Total Open moves	100	219	319

Looking more closely at how Open moves are distributed among the GDTs it is not surprising that the vast majority first Open moves in the GDTs were either read out directly from the prompt, or were statements in response to it. Of the 52 initial Open moves in the 2005 administration, 28 of them were started by one of the participants using a question from the prompt, while 18 started by responding to the prompt with a statement. Of the remainder, six used a question that was influenced by the prompt and the last three were questions about the test itself. Given that the test-takers have little choice but to start the conversation in this way, it suggests that test-takers then collaboratively engage in the topic, which evolves in a conversation like way.

The second feature to examine is the Register move. Table 3 below gives the statistics of its use in the test. The Register moves stand out as being dramatically higher for the students at the end of their first year of study than as incoming first year students.

Table 3 Register moves

	Register	
	2004	2005
Tot.	278	1162
Av.	9.3	22.3
Max	28.0	86.0
Min	2.0	8.0
Stdv	6.5	19.2

The final feature to examine is turn-taking. Table 4 below shows the general statistics for the corpus. It can be seen again that there is an appreciable increase in the number of turns from 2004 to 2005. It is likely that most of these can be accounted for by the increase in Register moves

Table 4. Turn-taking

	Turns	
	2004	2005
Av.	58.3	94.8
Max	134	300
Min	28	30
Stdv	30.2	52.9

When examining the Open moves, it seems that it is a matter of quality rather than quantity. Among the most balanced GDTs in 2005 in terms of turns taken, the number of Open moves varied from as few as two to just five, compared to the maximum number of 13 found in one GDT. Having more Open moves in a GDT is more likely a sign that one speaker is dominating. In 2004 the GDT with the highest number of Open moves was also one of the least balanced in turns. Amongst the three participants, one person accounted for 47.8% (the other two 35% and 15.93%). This GDT also had the highest number of Register moves in the data (28), showing that despite the imbalance the participants were involved. A closer look at the transcript reveals that although one speaker was dominant, another of the test-takers played a part by asking questions, but the other member of the group did not do once. One factor in this may have been that the two participants who carried the conversation were male, and the one who did not was female and may have felt inhibited.

The role of Register moves can be seen when comparing this to another GDT in 2005 which had a dominance speaker (47% of the turns). This GDT is conducted like an interview, with all interaction passing through the dominant test-taker. Not once

did the other two participants interact with each other, and all questions apart from a single transfer of turn move was asked of the dominant test-taker. In this test it is apparent that there is a difference in ability between the participants that prevents the interaction from seeming conversation-like. This GDT had one of the fewest numbers of Register moves in the corpus (12), and it shows in the cold atmosphere of the test.

Another characteristic of some GDTs in 2005 that had low Register moves was that the participants took fewer longer turns with fewer Open moves. These GDTs were carried out by each speaker taking a long turn and when finished the next person starts talking for their turn. One of these GDTs had a single Open move at the start and then proceeded with each participant developing their story without paying much heed to what the other was saying, perhaps similar to the students in the He and Dai (2006) paper. The lack of involvement showed in the fewest number of Register moves in the data for that administration (8). Among the GDTs with the most Register moves, on the other hand, a strong relationship was found with Open moves that were unrelated to the topic or influenced by it.

Another element having an impact on these figures is the prompt. In the five most balanced GDTs four of the prompts were the lighter topics of 'cell phones' or 'outdoor vs indoor activities', while four of the five least balanced GDTs were on the topic of traditional families. This suggests that some test-takers are being excluded by the difficulty of discussing a more serious topic. However, there was no relationship between GDTS with the highest number of Register moves and the prompt itself, suggesting that the Register moves in these tests are dominated by a few of the participants.

V. Conclusion

From this study some features of the construct of conversation-like interaction have been investigated. The corpus shows that many of the participants are adapting the prompt to their needs, as shown by the high percentage of Open moves influenced by the prompt, and a considerable number of new topics are being produced in these GDTs. Indeed, the relationship between topics unrelated to the prompt and plentiful Register moves is not accident, it shows that students get involved in the conversation of such GDTs. Finally, when administering a GDT assessment, the role of the prompt is significant, for being able to talk about the prompt suggests that test-takers are more likely to have a balanced discussion in terms of the number of turns.

On the other hand, this study also shows that unconversation-like interactions can result from GDTs. This seems to happen more when the prompt is more difficult, the participants are not equal in terms of ability, and when the participants do not assert themselves. It is plain then, that administrators need to take such factors into account when hoping to take advantage of the many advantages that the GDT offers.

Appendix 1: 2005 Speaking Prompts

Prompt 1

In the traditional Japanese family, men earned the money and women did the housework. Is your family traditional or not? Why do you think so? What are the advantages and disadvantages of the traditional family? Do you think the situation in Japan is changing? Why?

Prompt 2

Around the world, people are marrying later and later. What are the advantages of being single? Why? What are the advantages of being married? Why? Do you want to get married or would you prefer to be single? In future, do you think people will still want to get married?

Prompt 3

More and more young people are spending their free time inside the house watching TV, using the Internet and playing computer games. Do you like to spend your free time inside or do you prefer to do outdoor activities? Should we all try to spend some time doing outdoors activities? Why or why not? Is it healthy to spend all our time indoors?

Prompt 4

These days, lots of people have mobile phones and they are becoming very important in our lives. Do you have a mobile phone? Why? Do you often use it? How do you feel about mobile phones? What are some good points and bad points about them? Why? Could you live without your mobile phone?

References

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

Bonk, W. J. and Van Moere, A. (2004). *L2 group oral testing: the influence of shyness/outgoingness, match of interlocutors' proficiency level, and gender on individual scores.* Paper presented at the Language Testing Research Colloquium.

Brown, G. and Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.

Eggins, S. & D. Slade. 1997. *Analysing casual conversation*. London: Cassell

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

Fulcher, G. (1996). Testing tasks: issues in task design and the group oral. *Language Testing, 13*, 23–51.

Gumperz, J. J. *(*1982*)*. *Discourse strategies*. Cambridge: Cambridge University Press.

He, L. and Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing 23*, 370—401.

Hilsdon, J. **(**1991). The group oral exam: advantages and limitations. In J. C. Alderson and B. North (Eds), *Language testing in the 1990s* (pp. 189–197). London: Modern English Publications and the British Council.

Johnson, M. (2001). *The art of non-conversation: A reexamination of the validity of the oral proficiency interview*. New Haven: Yale University Press.

Johnson, M. and Tyler, A. (1998). Re-analyzing the OPI: how much does it look like natural conversation? In Richard Young and Agnes He (Eds), *Talking and testing: discourse approaches to the assessment of oral proficiency* (pp. 27–52). Amsterdam: Benjamins.

Kobayashi, M., Van Moere, A. and Johnson, K. (2005). *Is silence 'golden'?: An investigation of group oral tests.* Paper presented at 14th World Congress of Applied Linguistics, Madison, Wisconsin.

Lazaraton. A. (*1992*). The structural organization of a language interview: A conversation analytic perspective. SYSTEM, 20, 373-386.

Sacks, H., Schegloff, E. A. and Jefferson G. (1974). A simplest systematics for the organization of turn-taking in conversation. *Language, 50*, 696-735.

Schegloff, E. A., Jefferson, G. and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language 53,* 361-82.

Schegloff, E.A., and Sacks, H. (1973). Opening up closings. *Semiotica, VIII,* (4), 289-327.

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

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