The Influence of the Social Interactional Context on Test Performance: A Sociocultural View

Youyi Sun
Queen's University
Shanghai Finance University

Author's Note
I would like to thank Dr. Janna Fox for her invaluable help and support in this study.

Abstract
This study investigated the influence of the social interactional context on test performance from a sociocultural perspective. Two oral language test tasks were used and parallel task versions were developed using two test methods: the individual context and the group context. The tests were administered to 23 ESL students. Both quantitative and qualitative methods were employed to analyze the data. Results of this study, particularly the significantly different discourses generated from the individual context and the group context show that analysis of the influence of the social interactional context on test performance from a sociocultural perspective offers language test developers and researchers useful information about test development and test validation inquiry. Implications of applying sociocultural theory in second language oral performance assessment are discussed.

Résumé
Cette étude porte sur l'influence du contexte interactionnel social sur la performance aux tests selon une perspective socioculturelle. On utilise deux tâches de test de langue orale -- des versions parallèles de test ont été développées en utilisant deux méthodes d'épreuve : le contexte individuel et le contexte de groupe. Les tests ont été administrés à 23 étudiants d'anglais seconde langue. Nous avons analysé les données à l'aide de deux méthodes, quantitative et qualitative. Les résultats de cette étude, montrent que l'analyse de l'influence du contexte interactionnel social sur les performances aux tests selon une perspective socioculturelle donne aux élaborateurs et aux chercheurs spécialisés dans les tests de langue des informations utiles sur le développement et la validation de tests. En effet, des dissertations très différentes ont été générées dans un contexte individuel et dans un contexte de groupe, Les implications de l'application de la théorie socioculturelle à l' évaluation de la performance orale en langue seconde sont abordées.

The Influence of the Social Interactional Context on Test Performance: A Sociocultural View

Introduction

In the past decade there has been growing interest in incorporating a sociocultural approach in second language (L2) education (Atkinson, 2002; Lantolf & Poehner, 2008; Swain, 2000) and language testing (Chalhoub-Deville, 2003; Fox, 2002; Swain, 2001; Young, 2000). The sociocultural perspective offers language testers interesting insights and has significant theoretical and practical implications for language testing. For example, since ability, context, and performance are inextricably meshed in language use (He & Young, 1998), performance will not be seen as simply the manifestation of the individual's ability. The more dynamic aspect of social interaction will not be considered simply as a source of construct-irrelevant variance that jeopardizes validity. Rather, it will be seen as part of what we are trying to measure, the absence of which may be a threat to test validity in terms of construct underrepresentation. As such, in test validation inquiry we need to describe and interpret the influences on test performance of the dynamic social interactional context from a sociocultural perspective.

Methodologically, describing and interpreting the influences on test performance of the dynamic social interactional context from a sociocultural perspective entails qualitative analysis of test performance as a source of validity evidence. Swain (2001) noted that the turn-by-turn analysis of dialogues or conversations could reveal the way context is co-constructed by participating individuals and the way performance is situated in context. This analysis will differ from discourse analysis which focuses on linguistic and interactional features of speaking.

While sociocultural theory offers language testers interesting insights, these insights are not fully developed and much work is needed before practical benefits might be obtained (Fox, 2004). Drawing on sociocultural theory, this study aims to investigate the influence of the test method as social interactional context on test performance in a small group oral language test in the English for academic purposes (EAP) context as a source of evidence for test validation.

Sociocultural Theory in Second Language Education

Sociocultural theory (SCT) is associated with the work of Vygotsky (1978, 1986). One fundamental notion of SCT is that social interaction plays a fundamental role in the development and function of cognition including the development and use of language (Lantolf, 2000a). From an SCT perspective, language is one of the symbolic tools we use to

"mediate and regulate our relationships with others and with ourselves and thus change the nature of these relationships" (Lantolf, 2000a, p. 1). Language use is both a socially communicative act and a medium for the internal organization of experience. Therefore from the SCT point of view, language is both the result of and the tool for social interaction. It owes both its origin and its continued activation and use to social interaction.

SCT provides a theoretical framework for studies on language learning as a mediated process. According to Lantolf (2000b), current research on mediated L2 learning continues to seek to better understand how L2 learning is mediated especially in the form of peer scaffolding. He claims that "people working jointly are able to co-construct contexts in which expertise emerges as a feature of the group rather than residing in any given individual in the group" (p. 84).

Swain and her colleagues, working from the sociocultural theoretical orientation, have carried out a series of insightful studies (Kowal & Swain, 1994; Swain, 1995, 2000, 2001; Swain & Lapkin, 1998, 2001) on the collaborative dialogues among students. Their findings have shown that dialogue among learners, wherein they are able to mediate each other, can be as effective a site for learning to happen as are instructional conversations between teachers and students. Thus, Swain (2001) states "dialogue is not 'enhancing learning' or leading to learning, it is learning" (p. 288). The talk students produced provides teachers and researchers with opportunities to observe the underlying L2 learning process. As such, both language educators and second language acquisition researchers are interested in interactions in the form of conversation.

Sociocultural Theory in Second Language Oral Performance Assessment

Although Vygotsky's (1978, 1986) theory is not a theory of performance, its emphasis on social mediation and interaction makes it relevant to performance-based oral proficiency assessment. Jacoby and Ochs (1995) introduced the term co-construction, which they define as "the joint creation of a form, interpretation, stance, action, activity, identity, institution, skill, ideology, emotion, or other culturally meaningful reality" (p. 171). The co-constructed view of interaction captures the dynamic feature of social interaction. Jacoby and Ochs (1995) noted that as interaction unfolds, interactants are constantly monitoring, determining, and responding to interactional events, and that "every interactional moment is a unique space for a response to which subsequent interaction will be further responsive" (p. 178). When discussing contextual interaction, Douglas points out that context is "dynamic, constantly changing as a result of negotiation between and among the interactants as they construct it, turn by turn" (Douglas, as cited in Chalhoub-Deville, 2003, p. 374).

With regard to L2 performance assessment, McNamara (1996) points out that a weakness of current models of communicative competence is that they focus too much on the individual candidate rather than the candidate in interaction, and that the idea of performance as involving social interaction has so far featured only weakly in the work of theorists of L2 performance and researchers in language testing. McNamara (1997) points out that the exclusive focus on the ability of the candidate in the cognitive approach views the candidate in a strangely isolated light, as exclusively responsible for the performance.

The co-constructed nature of interaction in a performance-based assessment situation presents challenges to language testers in terms of construct definition, reliability, and fairness. Researchers have investigated the effect on test performance of a number of variables associated with the interlocutor in face-to-face oral language tests, such as age (Buckingham, 1997), language level (Iwashita, 1997), personality (Berry, 1997, 2007), gender (Brown & McNamara, 2004; O'Sullivan, 2000), power (Van Lier, 1989), and acquaintanceship (O'Sullivan, 2002). A major concern of these studies is to investigate systematic effects on the test score and ways to control them for the purpose of test reliability. As a result, the influences of the interactant (the examiner or other candidate) are generally considered as an additional source of measurement error.

In an early study, Shohamy (1982) found that high correlations provided necessary but not sufficient evidence for the appropriateness of substituting face-to-face oral tests with tape-mediated oral tests. In a later study, based on quantitative and qualitative analyses of data from different perspectives, Shohamy (1994) concluded that "the context of the test, either 'face-to-face' or 'tape-mediated', can affect or even dictate the type of language that is produced" (p. 118). Similarly, other researchers (Johnson & Tyler, 1998; Lazaraton, 1992) have addressed the issue of validity of language proficiency interviews, using more qualitative approaches. Their findings suggest that the one-to-one oral interview generates a special genre of language different from normal conversational speech and thus it cannot be considered a valid example of a typical, real-life conversation.

Largely because of dissatisfaction with interview as a solo means of assessing oral proficiency, researchers and test developers have shown interest in search for the small group and/or paired testing formats. There have been reports of successful use of group testing with school students in Israel (Reves, 1991; Shohamy, Reves, & Bejarano, 1986) and Zambia (Hilsdon, 1991) and with university students in Hong Kong (Morrison & Lee, 1985). Berkoff (1985) argued that using groups of students overcomes the problems of "artificial conversation" between a "distant examiner" and a "nervous examinee". However, Fulcher (1996) pointed out that reports of successful use of group oral test, claims of a reduction in "test-type" language, and claims of reduced anxiety in the literature are not well supported with empirical evidence. In a more recent study, Brooks (2009) drew

on SCT to examine and compare the interactions of L2 test takers in two oral proficiency testing formats—the individual format, where the candidates interacted with an examiner, and the paired format, where they interacted with another student. Quantitative and qualitative analyses of the resulting score and discourse data of this study suggested that the paired approach resulted in higher participant scores as well as a more complex interaction between the participants, including negotiation of meaning and consideration of the interlocutor.

Validation is an evaluation process which involves developing and evaluating evidence for proposed score interpretations and uses (Kane, 2006; Xi, 2008). Brown, Hudson, Norris, & Bonk (2002) argued that the issue of validity must be dealt with for performance assessments just as it is for any other forms of testing "– in an open, honest, clear, demonstrable, and convincing way" (p. 5), especially when the test score is used to make high-stakes decisions about students. Drawing on SCT, Swain (2001) suggests that when assessing proficiency, language testers should find a way to take account of the language that emerges during group interaction because cognitive and strategic processes are made visible in dialogue, and understanding these strategies and processes is important to an understanding of the construct being measured. Therefore, the dialogue among participants will be an important source of validation evidence.

This study, framed within SCT, aims to investigate the influence of test method as social interactional context on test performance in a small group oral language test in the EAP context as a source of evidence for test validation.The following research question was addressed:

How does test method as social interactional context influence test taker performance in the L2 small group oral performance test?

Methodology

Background

The current study was conducted in an EAP program at Carleton University, Ottawa, Canada. Two tasks from the Oral Language Test (OLT) of the Canadian Academic English Language (CAEL) Assessment were used. The CAEL Assessment is a standardized test of English in use for academic purposes. It is designed to describe the level of English language of test takers planning to study in English-medium colleges and universities. CAEL Assessment scores are used to identify test-takers who have the ability to use EAP in university classrooms. The CAEL Assessment serves as a "gatekeeper" that allows or denies access to a program in universities. There are no restrictions on the frequency of test taking for students. At Carleton University, the CAEL Assessment is also used as a

placement test in EAP support programs. Results of the test place the students into one of three main levels:

1. intensive English as a second language (ESL) course level: for students whose English level is not high enough to begin credit courses;
2. credit ESL course level: for students who are admitted to a degree program but whose language skills require some additional support;
3. English as a second language requirement has been satisfied; no ESL is required.

The CAEL Assessment OLT is a task-based tape-mediated oral language test of spoken English in use for academic purposes (Fox, 2000). It consists of five tasks which represent ways in which students talk about their academic work within colleges and universities. These tasks include:

Task 1 (2 minutes): making short presentations;
Task 2 (5 minutes): relaying information;
Task 3 (5 minutes): explaining choices;
Task 4 (5 minutes): summarizing main points;
Task 5 (8 minutes): listening and responding to group discussion.

Since the current study focused on performance in group oral discussion, Task 5 of the CAEL assessment OLT was used. In this task, test-takers are required to explain a choice for participation in a group project. The test taker is given a handout, where some topics and some prompt details relevant to these topics are listed. In the CAEL assessment OLT, the professor's instructions and the talk of other members are pre-recorded on the computer. The test taker listens to pre-recorded professor's instructions for a group oral presentation. After the instructions, the test taker has one minute to familiarize him/herself with the topics. Then, the test taker listens to pre-recorded responses of other members of a group who explain their preferences from the list of topics on the handout for participation in the presentation. After listening to the other group members, the test taker is given one minute for planning and then is asked to explain his/her own presentation choice from the topics on the handout which have not been talked about by the other group members.

The scoring of the OLT is undertaken by trained raters who listen to the test takers' recorded responses to the five tasks. Points are assigned to each task on an analytic scoring rubric on specific features of each task. The overall performance is then assessed holistically. The analytic points and the holistic points are put together to make the raw scores, which are converted to criterion-related band scores that range from 10 to 90.

Participants

Students.

Twenty-three students from the introductory EAP program at the university participated in this study. These students had taken the CAEL Assessment before they registered in this program. They had achieved an overall band score of 40 with band scores of 30 or above in all of the reading, listening, writing, and speaking band scores.

Of the 23 students, 18 were from Mainland China, 1 from Taiwan, 1 from Kuwait, 1 from Vietnam, and 2 from Japan. They had studied English as a foreign language in their native countries for 5-8 years and had studied in Canada for 1-3 years. Their ages ranged from 20 to 28 years. Twelve of the participants were male and 11 were female.

The EAP teacher.

The EAP teacher in the study had taught these students for one term. She was working on her Master's thesis at the time of the study. She was both an experienced EAP teacher and an experienced CAEL Assessment rater.

Raters.

The EAP teacher rated the students' performance on the small group discussion in the classroom. Recordings of the students performing the tasks in the computer lab were rated by the other two raters. One of these raters with a doctoral degree had taught in the EAP program and was an experienced rater. The other rater was also an experienced and trained rater.

Instruments

Tasks.

Two tasks were used in the current study. Task A was about youth and employment, Task B about violence in society. Each of these tasks was used in parallel versions by changing test methods. The recordings of Task A and Task B were similar in terms of length of time: instructions 2 minutes 17 seconds for each task; prompt conversation 2 minutes 40 seconds for Task A and 2 minutes 43 seconds for Task B; planning time 2 minutes for each task; altogether approximately 7 minutes for each task.

Two test methods were used in the study: the individual context (IC) and the group context (GC). The procedures of the IC test were identical with those of the CAEL Assessment OLT described above. In the GC test, the participants were given the handouts in the classroom and were asked to discuss the topics in small groups for ten minutes. A tape-recorder was placed on each of the desks where each group of the students sat. The

students were told that the tape-recording was for research purposes only and that scores would not be assigned to their performance on the group discussion in the classroom. The group discussion was organized by the researcher and the prompt instruction for each of the tasks was delivered to the students by the EAP teacher. Immediately after the small group discussion in the classroom, the students were led to the computer laboratory to explain their choices of the topics for the oral presentation. It took approximately 1 minute for the students to walk from the classroom to the computer lab. The test takers had no planning time in the computer lab.

Post-test questionnaire.

With the principle that test-takers' reactions to the test offer useful information about validity of the test (Brown, 1993; Elder, Iwashita, & McNamara, 2002; Fulcher, 1996), a post-test questionnaire was designed to capture the students' reactions to and opinions on the tasks and test formats used in this study (see Appendix A). The questionnaire consisted of five sections.

Part A collected the students' demographic information in terms of name, gender, and native language and identified the tasks they undertook. Questions in Part B concerned the students' reactions to and opinions on the two tasks in terms of validity (questions 1-2; questions 5-6; questions11-12), self-evaluation of task performance (questions 3-4), adequacy of timing (questions 7-8), task familiarity (questions 9-10), and task difficulty (questions 13-14).The twelve questions in Part C were designed to elicit information on the students' perception of the two test methods: the individual test and the group test, in terms of test validity (questions 15-18); test-related anxiety (questions 19-20); preference of test method (questions 21-22); test difficulty (questions 23-24); and test fairness (questions 25-26). Question 27 in Part D of the questionnaire asked the students about their preference of topic and test method. In the last part of the questionnaire, the students were encouraged to make any comments on the tests they had taken.

Split-Plot Design

The study was conducted as a part of the EAP support course final exam at the end of the term when the students had completed this course. The participants were divided into two groups on a voluntary basis as the students were usually paired in this way in their classroom. There were 10 students in Group 1 and 13 students in Group 2. Each group was required to undertake two tasks. The split-plot design (Huynh & Feldt, 1976) of this study is shown in Table 1.

Table 1
Split-Plot Design of the Study

	Individual Context	Group Context
Task A	Group 1	Group 2
Task B	Group 2	Group 1

In order to avoid any order effect, Group 1 took the Task A–IC test in the computer lab while Group 2 discussed the topics of Task A in small groups (3 groups of 3 and 1 group of 4) in the classroom. Then, Group 1 had a small group discussion (2 groups of 3 and 1 group of 4) on topics of Task B in the classroom while Group 2 made presentations individually on the computer about the topics of Task A they had chosen and did the Task B-IC test. After Group 2 finished the Task B–IC test in the lab and Group 1 finished small group discussion in the classroom, Group 1 were required to make their individual presentations in the lab. After each of the two groups had taken the two tests in the lab, the students were invited to complete a questionnaire in the classroom.

Performances of the students on the small group discussion in the classroom were assessed by the EAP teacher and performances of all the students on the tests in the lab were rated by the two raters using the CAEL Assessment holistic band score criteria. However, to allow for the rating of borderline performances, intermediate levels were included between each band score by using "+" and "-". In addition to rating the students' performance, the raters were encouraged to make comments on the candidates' performances. The recordings of the small group discussion and of the performance of the students on the computer were transcribed for analysis.

Analysis

The procedures described above produced the following data: (a) the EAP teacher's ratings of the students' performances on the small group discussion in the classroom, (b) the two raters' ratings on the students' performances in the computer lab, (c) the completed questionnaires, and (d) transcriptions of the small group discussions in the classroom and the recordings from the computer lab. The scores assigned by the two raters to all the performances of the students on the four tests ranged from 30 to 60-. These band scores were converted into a nine-point Likert scale for calculation purpose, i.e., 30 = 1, 30+ = 2, and so on. After calculation of Pearson correlation co-efficient for interrater reliability, the two sets of scores by the two raters were averaged. The averaged scores were used for all of the following calculations.

Pearson correlation co-efficient was calculated to examine the correlation of the averaged scores of the GC tasks by the two raters with the scores assigned by the EAP teacher to the students' performance on the small group discussion in the classroom. *T*-tests were conducted to compare the students' performances on the same task in the two formats: IC and GC. Because the scores were achieved by two different groups of students in the two formats, the scores were also dependent on the students' abilities. In order to compare the abilities of the two groups of participants, a *t*-test was conducted to compare the means of the scores on the two tasks assigned to the two groups. Because the two groups of students performed two different tasks under IC and GC, the means of the scores assigned to the participants' performances across these two tasks were compared using *t*-tests.

Descriptive statistics analyses were conducted to analyze the questionnaire data. The students' responses to questions in Part B of the questionnaire were analyzed to see if there was any difference between Task A and Task B in terms of the students' perceptions. To better understand the influence of the two test methods on test performance, the students' responses to questions in Section C of the questionnaire were analyzed to obtain information on the students' perceptions of IC and GC.

To explore how analysis of small group oral performance from a sociocultural perspective can inform EAP task-based performance assessment in terms of validation inquiry, transcriptions of the recordings of the students engaged in small group discussions in the classroom and of the students performing the tasks in the computer lab were qualitatively analyzed to see if they produced the same kind of language samples. In analyzing the transcriptions of the recordings of the classroom small group discussions, Swain's (2001) approach to dialogue analysis was taken. In analyzing the transcriptions of the recordings of the lab presentations, the transcriptions were read and categorized for discourse features through a process of analytical induction (Hicks, 1994). Recurrent themes in the discourse features were identified.

In analyzing the transcriptions of the recordings of the lab presentations, samples from the same test method (IC or GC) were compared. Recurrent themes emerged and these were identified as features of the discourse produced from the IC task or the GC task. Then, discourse features of the two test methods were compared to identify any differences between the IC tasks and the GC tasks in terms of discourse features they produced.

Results

It should be noted that because the sample size of this study was small, any statistically significant results from these quantitative analyses can only be interpreted as suggestive.

Comparison of Performances in the Individual Context and the Group Context

In this study two tasks were used with two test methods: the IC and the GC methods. Thus, four tasks were generated. Table 2 shows descriptive statistics for the averaged scores across the four tasks.

Table 2
Descriptive Statistics for the Averaged Scores across Tasks

Tasks	N	M	SD
Task A - IC	10	4.45	2.02
Task B - GC	10	5.75	1.85
Task B - IC	13	6.15	1.52
Task A - GC	13	6.31	1.76

Statistics in Table 2 indicate that scores of Group 2 are consistently higher than those of Group 1 across the tasks. Within each group, scores on GC tests are consistently higher than scores on IC tests. The comparison of the scores for the IC and GC formats showed that the scores on Task A in the GC format were significantly higher than those on the same task in the IC format: t (21) = 2.35, $p<.05$, but the students' performances on Task B were not significantly different in these two contexts.

Pearson correlation co-efficient calculations showed significantly high correlation of the averaged scores of the GC tasks by the two raters with the scores assigned by the EAP teacher to the students' performances on the small group discussions in the classroom (for Task A-GC, r = .83, $p < .05$; for Task B-GC, r = .89, $p < .05$).

The students' responses to questions on the questionnaire about their preferred test method showed general agreement among the students that the GC test was preferable to the IC test across the two tasks. The students' preferences for GC were related to their perceptions of the test difficulty. In responding to questions concerning test difficulty in relation to test methods, 19 out of the 23 students disagreed that the GC method made the test more difficult while 14 of the students believed that the IC method made the test more difficult. This was also consistent with the students' responses to questions concerning test anxiety. Eleven of the participants agreed that they felt nervous in taking the IC test while 5 agreed that they did so in the GC test. With regard to whether the test methods provided them with enough opportunity to show their English speaking ability, 7 students from Group 1 believed that they had enough opportunity to do so in the IC test while 8 students from Group 2 believed so. For the GC test, 9 students from Group 1 and 11 from Group 2

felt it provided them with enough opportunity to show their ability to speak English. With respect to the fairness of the test, the participants seemed to be more in favor of the individual tape-mediated test. Seventeen students believed it was a fair form of test while twelve regarded the co-constructed small group test as a fair test form.

Comparisons of Tasks and Groups of Students

The *t*-test result showed no significant difference between Task A and Task B. This result was consistent with the students' ratings of the task difficulty in the questionnaire. No significant difference was found between Task A and Task B in terms of the participants' general evaluation of task difficulty. Responses also indicated that there was great agreement among the students as to their familiarity with the two tasks. Responses showed that the majority of the students believed that both Task A and Task B were valid in assessing their ability to speak English. Most of the students felt that they had enough time to do both the tasks.

The *t*-test result showed significant difference between the two groups of students in terms of ability: $t\ (21) = 2.124$, $p < .05$. This difference was also evidenced by the participants' reactions to task difficulty. Seven out of the 10 students of Group 1 rated the difficulty of Task A 3 or above while 5 out of the 13 students of Group 2 did so. For the difficulty of Task B, 5 students of Group 1 and 7 students of Group 2 rated 3 or above.

Consistency was found between the students' perceptions of task difficulty and their performance self-evaluations. All the students that rated task difficulty as below 3 believed that they did well on the task and vice versa. The students' responses were also consistent with the scores awarded to their performances on the test by the raters. The students who felt the task was more difficult and didn't believe they did very well on the task generally obtained lower marks and vice versa. Eight students from Group 1 believed that they had enough opportunity to show their ability to speak English in doing Task A while 9 students from Group 2 believed so. For Task B, 7 students from Group 1 and 7 from Group 2 felt it provided them with enough opportunity to show their ability to speak English.

Comparison of Language Samples from the IC and the GC Tests

Analysis of the transcriptions of the computer recordings of the participants performing the two tasks under IC and GC suggested some general differences in their performances on the tasks in these two contexts in the following aspects:

Range of vocabulary.

The range of the students' vocabulary seemed to be more limited on the IC task. They used more general words and made more repetitions. For example:

And that is most common and serious problem in the family. And this is really, really bad for the children. Sometimes in the family, the parents will beat their children or their parents will beat each other. It's really, really bad in front of the children. (Task B-IC)

It is true that students spend their spare time to work. So they feel stress. It is true. (Task B-IC)

In contrast, on the GC task, the students used a relatively wider range of vocabulary and relatively more complex sentence structures. Unnecessary repetitions were significantly reduced. For example:

I work in a convenience store when I was in university. And I gained some pocket money to buy anything I like. And I also gained some good experiences such as how to solve problems in any situations and how to face people in public, and, but I also got some bad influence. I am exhausted after work. I don't have time to do my homework, prepare the test and work on my presentations and work at the same time. So at the end I gave up my job because I feared the risk to fail my study. (Task B-GC)

Organization of discourse.

In the IC task, the discourse seemed to be more locally managed and to be produced on-line. Though the students knew which topic to talk about and which aspect(s) to focus on, they didn't seem to have a global plan about how to organize their discourse. The utterance was produced more spontaneously. For example:

Well, I'd like to choose the topic of combining study and work. (pause) Why I choose this topic is because I… I…I as a student, I have a part-time job outside so I think I have something to talk about this. First, I think the stress is the big problem. (Task A-IC)

This student then talked about the 'stress' until the end. He didn't talk about a second point though there was some time left when he finished talking.

In contrast, on the GC task, the discourse seemed to be generally better planned and organized. For example:

Ok. I'm going to talk about violence in the school. There has been issues that has happened couple of months ago and it's in the US. There is a high school student use a gun to shoot for the whole classroom. [After describing the US incident] What's been a problem today is the violence, and especially in the schools. [The student went on to make some comments on the negative effects of violence on campus and on weapon control in high schools.] (Task B-GC)

Exemplifications.

On the IC test, the students used more personal experiences as examples to talk about the topic. For example:

First, I think the stress is the big problem. I have because when I work I have to think to finish my homework to prepare for the test and when I study I have to think about my work so I think this is problem and the part-time job also has the problem with the grade just because when I work, I reduce my study time for preparing something so maybe I will get low mark for test and because I am so tired of the work, I can't have enough sleep or some good sleep. [Task A-IC]

In contrast, on the GC task, students used not only personal experiences, but also experiences and accounts of other people to talk about the topic they chose. For example:

From working, we can gain some money to buy anything we like and we can learn some experience from working like how to face people, and how to solve problems in some situation…Some people work and study at the same time. They have to compete with adults for jobs because some bosses they will probably prefer to hire experienced people to work because they have more experience but usually some bosses will hire young people because their salary is much cheaper than adults. [Task A-GC]

This student used what her partner had talked about in the small group discussion to explain "the positive side" and "the negative side" of youth involved in employment.

Qualitative Analysis of Small Group Discussions in the Classroom

Qualitative analysis of small group discussions in the classroom revealed how the discourse was co-constructed by all participants. The following example is an excerpt from the transcription of the recording of a group of three students engaged in doing the "youth and employment" task in the classroom.

S1: [Reading from the handout] Attitudes towards youth and employment. Positive, experience and money. Yeah. Eh…Youth? So we have to focus youth, not old people and… (1)

S2: Adult (2)

S1: [Reading from the handout] Interfere with study. So, they has study problem. That… So the youth involved in the, involved in the employment may have problems with study. So we also can talk about this, this pers… as…aspect. [Reading from the handout] Compete with adults for jobs. So, that should be the negative…negative effect, affect. (3)

S2: Negative effect? (4)

S1: Yeah. (5)

S2: Interfere. (6)
S1: Interfere. So, interrupt. Not interrupt. (7)
S2: Interfere. (8)
S1: Effect. I say it is negative effect. (9)
S2: [Reading from the handout] Compete with adults for jobs. (10)
S1: (Reading from the handout) Combining study and work. That makes problem. That's problem comes out. [Reading from the handout] Stress and problems with grade. I think this negative negative [pause] effect. This… [Reading from the handout] Self confidence and money help…This positive. It's positive effect. [Reading from the handout] Types of work available for youth. (11)
S2: But we don't need to talk about this. (12)
S1: I think we should make a choice to… (13)
S3: Yeah, make a choice (14)
S1: To pick up one topic that we want. (15)
S2: Right. (16)
S3: I think so. So, why can't we choose this, positive? (17)
S1: Yes, I think…(18)
S3: It is really good experience and we can save money to do something else. (19)
S1: I think we talk about the whole topic. So we need, we need talk the positive and both the negative. (20)
S3: Both? (21)
S1: Yes. This topic has two points here. So, each four, each of four topics have this point. So we both, we both have to talk. (22)
S2: I think the third is easy to talk. (23)
S3: Yeah, there have so many to prepare. (24)
S1: Yeah, such as stress, problem with grade. (25)
S2: So, of the four topics, we choose one. We choose combining study and work. (26)
S1: Let's practice # 3, OK? (27)
S2 and S3: OK. (28)
S1 (to S3): Could you tell me what's the stress they may have who have work in school? (29)
S3: En…(30)
S1: The stress. (31)
S3: They, the student have to work a lot, and so they can't have enough time to sleep…(32)
S1: No, "sleep" is here. Stress. (33)
S3: What does "stress" mean? (34)
S1: Yali [The Chinese word for "stress"] (35)

In this activity, the goal of the students was to choose a topic to talk about. They achieved this goal collaboratively. S1 first took the responsibility to read aloud the second topic. His reading drew his attention to the external context: what was written on the handout. He held up the word "youth" for attention and reflection and then he externalized this process (turn 1). This externalization not only revealed S1's self-mediation process but also served as an implicit invitation to his partners for their involvement. It also framed the next turn, where S2 picked up this invitation by contributing the word "adult" to show his involvement. Similarly, the third turn was largely a self-mediation process on the part of S1 and his externalization of the process also formed the context for his partners. It revealed that S1 seemed not to be able to use the word "effect" fluently. S2 seemed to have noticed this and offered some help in the following turns. In turns 4-9, S1 and S2 "negotiated" over the words "effect" and "interfere". It was obvious that both of them benefited from this negotiation. Turn 9 suggested that this interaction facilitated the former's learning of the word "effect" and the phrase "negative effect". Turn 11 revealed the interaction between the language user (S1) and the external context (the handout) and the task (to talk about the topic). The activated phrase "negative effect" seemed to play a significant role in this interaction. Turn 12 was S2's reaction to S1's last utterance. It was also S2's contribution to the co-constructed performance of the task. S3 seemed to keep silent until turn 14. His utterance revealed his reaction to what had been going on. He made more contribution in turns 17 and 19, drawing his partners' attention to the third topic by emphasizing and detailing it. Presumably because of this, S1 turned to S3 in turn 29 asking him to talk about "the stress of combining study and work". S1 and S2 collaboratively talked about this topic in turns 29-35. What seemed to be interesting was S1's use of the Chinese word in the last turn to facilitate S3's understanding of the English word "stress". This revealed S1's intention to keep the interaction going and to keep on the topic and to elicit more information from S3 about the aspect of "stress" of this topic.

The next example is an excerpt from the transcription of the recording of a group of three students engaged in doing the "violence in society" task in the classroom.

S1: Maybe… Let's move on second topic "violence in families". (1)
S2: Families is, I think…. Just like you guys said before, the effect on children, if the… (2)
S1 and S3: Yeah? (3)
S2: I think if the father and mother fight every day in the house and the children see that. This is the problem when they grow up. (4)
S3: They consider that all the families just like that. (5)
S2: Yeah, if children… if parents… (6)
S1: Beat their children? (7)

S2: Yeah, beat their children all the day. (8)
S3: Abuse. (9)
[All the three students laughed]
[Then they had some discussion on the second aspect of the topic "violence in families" and found it was hard to talk about so decided to give up the second topic.]
S3: So, well, the first one we don't know how to talk about effect on male/female relationship. The second one we don't know how to talk about problem with the elderly people. (10)
S2: So now, let's move on to violence in schools. [pause] I think weapon control in high schools is to be easier for us because we… (11)
S1: Yes, it's easy to talk about. (12)
S2: We have the example in the US high school. (13)
S3: Actually we don't need to talk about that example. Only one minute. (14)
S2: That's enough. (15)
S1: Just some details. (16)
S3: So, effect on learning, what should we talk about? (17)
S1: Effect on children's learning? (18)
S3: Maybe if we are the classmates and I know you have a gun there, and I cannot focus on my study, feel scared or if someone… I don't know. (19)
S2: I just think weapon control in the high school. This is the problem. (20)
S3: Effect on learning can [stressed] be something like that. Or if, you know… (21)
S2: Yeah, yeah, you remember just like if other students don't like you, just for example, if they don't like you and they beat you at the time and make you unconfidence. This is the effect on learning. (22)
[9-second pause]
S2: If you don't like me and you avoid me, I'll feel unconfidence and I'll feel lonely. This is the effect on the learning. (23)
S1: That maybe a kind of violence. (24)
S2: This should not be the violence [stressed]. Violence should be the kill and… (25)
S1: Kill another people and… (26)
S2: Something like that. OK. Let's say some violence on the street. (27)
S3: No, no. How can we say weapon control? (28)
S1 and S2: Weapon control… (29)
S2: is the simple example just like what I say of the US high school. (Recounting the US incident again) (30)
S3: And it also effect on learning. (31)
S1 and S2: Yeah. (32)

S1's utterance in turn 1 drew attention to the topic of violence in families. S2's utterance in turn 2 revealed that his thinking process of this topic was mediated by what the other two students had discussed before. In turn 3, S1 and S3 displayed their connectedness with S2 by showing interest in the content of S2's utterance. In the following turns (4-9), the students collaboratively worked on the topic of the effect on children of violence in families. It seemed that S2 was the initiator of the idea and S1 and S3 contributed by expanding his idea and by offering some linguistic devices. S3's utterance in turn 10 suggested that what they had talked about had been internalized in her mind. Her summary of the first two topics suggested that she felt the first two topics were not easy to talk about and she would like to work on some other topics. S2 and S1 responded to this suggestion in turns 11-13. In turns 13-16 the students referred to the US incident, which had been recounted by S2 as an example of the effect on children of violence in movies. This showed that this example had become shared knowledge among the three students. S3's utterance in turn 19 revealed a self-mediated process in her mind. While externalizing this process, she encountered some difficulty and gave up. However, S3's utterance facilitated S2's thinking, which he externalized in turn 22. The 9-second pause between turn 22 and turn 23 indicated that S1 and S3 seemed not to understand what S2 had said. And in turn 23, S2 took the responsibility to make further explanations by paraphrasing his last utterance. S2's suggestion in turn 27 indicated that he felt they had talked enough about the topic. However, his suggestion was refused by S3, who insisted on keeping on the topic of "weapon control". Therefore, S2 took his responsibility again to make further explanations, this time, by repeating the US high school example.

Discussions and Conclusions

There are a number of limitations of this study that should be noted. First, as has been mentioned, because of the small sample size (N = 23), any results from the analysis of the current study should be only interpreted as suggestive. Second, although the study was conducted in the final examination context, stakes associated with this test should be lower compared with a live-test context. Third, the sample was taken from one class; the students were familiar with each other, and the EAP teacher, who was the rater of the classroom group work in this study, was also from the same class. Therefore, it is important to interpret the findings with these limitations in mind.

Results of this study suggest the influences of the test methods on the performances of the students. For both groups, scores on the GC test are generally higher than on the IC test. This conclusion is also supported by the students' preference for this type of test over

the IC test, as revealed by questionnaire analysis. What seems certain is that the GC test is a means of "biasing for best" (Fox, 2004; Swain, 2001) in terms of anxiety reduction.

The turn-to-turn analysis of the group work activity revealed how the students' ability to speak English interacted with the social interactional context. Context was dynamic and constantly changing as a result of interaction between and among the interactants as they constructed it, turn by turn. Each utterance was framed by what had been said and contributed to establishing a space for subsequent utterances to be produced. Utterances produced by interactants were not merely manifestations of their knowledge or ability but also manifestations of the mediated cognitive and strategic process. This process arose in the interaction. Therefore ability and context features were intricately connected and it was difficult or impossible to disentangle them. Locally examining the test taker's ability engendered by the features of specific social interactional contexts and systematically exploring the test taker's performances across context will provide the tester with a clearer picture about the test taker's ability in context. Qualitative analysis of the dynamic and mutual influences of performance and interactional context will provide useful evidence for validation inquiry of task-based L2 performance assessment.

Comparison of the performances of the students on the IC format test and the GC format test and the small group discussion in the classroom revealed that the students paid attention to different aspects of their performances. The classroom discussion seemed to be a here-and-now interactional problem-solving activity where the students paid more attention to what to talk about and the effectiveness of communication, neglecting linguistic errors to keep the flow of communication. On the IC and GC tests in the computer lab, the participants paid more attention to how to talk. On the IC test, they seemed to allocate their attention more to monitoring their language. As a result, they made more self-corrections when they were aware of errors in their utterances and made more repetitions when they noticed a gap between what they wanted to say and what they could say or when they could not work out a solution to the gap. On the GC format test in the lab, the students produced more planned language. Discourse produced in the small group discussion may form a useful basis for constructing empirically based rating scales (Turner, 2000) for small group oral language performance test.

The small group discussion helped reduce the difficulty on the part of the test takers in terms of more planning time and test enjoyment. Generally, the spontaneous support provided by the interactants positively affected the performance of the students and helped reduce test anxiety.

To the extent that the student's performance in the interaction is constrained by the social interactional context, not only the stable and global features of the task but also the dynamic influences of the interactant's performance should be taken into account in

considering task difficulty. This also implies that in developing small group discussion oral test, pairing may be a critical issue to consider.

McNamara (1996) sees the work of modeling performance in language performance assessment as opening Pandora's Box. Findings of this study suggest that taking an SCT perspective, we might not see the Pandora's Box as a "black hole". Instead, we might see it as a kaleidoscope, or a dynamic landscape of meaning and relationship. For this, further research is needed in test validation inquiry to describe and interpret the influences on test performance of the dynamic social interactional context.

References

Atkinson, D. (2002). Toward a sociocognitive approach to second language acquisition. *The Modern Language Journal, 86*(4), 525-545.

Berkoff, N. A. (1985). Testing oral proficiency: A new approach. In Y.P. Lee (Ed.) *New directions in language testing* (pp. 93-100). Oxford, UK: Pergamon Institute of English.

Berry, V. (1997). *Gender and personality as factors of interlocutor variability in oral performance tests*. Paper presented at 19th Annual Language Testing Research Colloquium. Florida, USA.

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

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

Brown, A. (1993). The role of test-taker feedback in the test development process: Test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing, 10,* 277-303.

Brown, A., & McNamara, T. (2004). "The devil is in the detail": Researching gender issues in language assessment. *Teachers of English to Speakers of Other Languages (TESOL) Quarterly, 38,* 524-538.

Brown, J.D., Hudson, T., Norris, J., & Bonk, W. J. (2002). *An investigation of second language task-based performance assessments* (Technical Report # 24). Hawaii, USA: University of Hawaii, Second Language Teaching & Curriculum Center.

Buckingham, A. (1997). *Oral language testing: Do the age, status and gender of the interlocutor make a difference?* (Unpublished Master's thesis), University of Reading, Reading, UK.

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. *Language Testing, 20*(4), 369-383.

Elder, C., Iwashita, N., & McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: What does the test-taker have to offer? *Language Testing*, *19*(4), 347-368.

Fox, J. (2000). *The Canadian Academic English Language Assessment: Test score and users' guide*. Ontario, Canada: Language Assessment & Testing Research Unit, Carleton University.

Fox, J. (2002). *Test takers' and test raters' accounts of three L2 writing tests*. Paper presented at the Language Testing Research Colloquium, Hong Kong, China.

Fox, J. (2004). Biasing for the best in language testing and learning: An interview with Merrill Swain. *Language Assessment Quarterly*, *1*(4), 235-251.

Fulcher, G. (1996). Testing tasks: Issues in task design and the group oral. *Language Testing*, *13*(1), 23-51.

He, A.W., & Young, R. (1998). Language proficiency interviews: A discourse approach. In R. Young & A.W. He (Eds.), *Talking and testing: Discourse approaches to the assessment of oral proficiency* (pp. 1-24). Amsterdam, Netherlands: John Benjamins.

Hicks, A. M. (1994). Qualitative comparative analysis and analytical induction: The case of the emergence of the social security state. *Sociological Methods and Research*, *23*(1), 86-113.

Hilsdon, J. (1991). The group oral exam: Advantages and limitations. In J.C. Alderson & B. North (Eds.), *Language testing in the 1990s: The communicative legacy* (pp. 189-197). London, UK: Macmillan.

Huynh, H., & Feldt, L. S. (1976). Estimation of the box correction for degrees of freedom from sample data in the randomized block and split-plot designs. *Journal of Educational Statistics*, *1*, 69-82.

Iwashita, N. (1997). *The validity of the paired interview format in oral performance testing*. Paper presented at 19th Annual Language Testing Research Colloquium. Florida, USA.

Jacoby, S., & Ochs, E. (1995). Co-construction: An introduction. *Research on Language and Social Interaction. 28*(3), 171-183.

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

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). Connecticut, USA: American Council on Education.

Kowal, M., & Swain, M. (1994). Using collaborative language production tasks to promote students' language awareness. *Language Awareness, 3*, 73-93.

Lantolf, J. P. (2000a). Introducing sociocultural theory. In J.P. Lantolf (Ed.) *Sociocultural theory and second language learning* (pp. 1-26). Oxford, UK: Oxford University Press.

Lantolf, J. P. (2000b). Second language learning as a mediated process. *Language Teaching, 33*, 79-96.

Lantolf, J. P., & Poehner, M. E. (Eds.). (2008). *Sociocultural theory and the teaching of second languages*. London, UK: Equinox.

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

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

McNamara, T. F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18*(4), 446-466.

Morrison, D.M., & Lee, N. (1985). Simulating an academic tutorial: A test validation study. In Y. P. Lee (Ed.) *New directions in language testing* (pp. 85-92). Oxford, UK: Pergamon Institute of English.

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System, 28*, 373-386.

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

Reves, T. (1991). From testing research to educational policy: A comprehensive test of oral proficiency. In J.C. Alderson & B. North (Eds.), *Language testing in the 1990s* (pp. 178-188). London, UK: Modern English Publications and the British Council.

Shohamy, E. (1982). Predicting speaking proficiency from cloze tests: Theoretical and practical considerations for test substitutions. *Applied Linguistics, 3*, 161-171.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11*, 99-123.

Shohamy, E., Reves, T., & Bejarano, Y. (1986). Introducing a new comprehensive test of oral proficiency. *English Language Teaching Journal*, *40*, 212-220.

Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), *Principles and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford, UK: Oxford University Press.

Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J.P. Lantolf (Ed.), *Sociocultural theory and second language learning* (pp. 97-114). Oxford, UK: Oxford University Press.

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

Swain, M., & Lapkin, S. (1998). Interaction and second language learning: Two adolescent French immersion students working together. *Modern Language Journal, 82*, 320-337.

Swain, M., & Lapkin, S. (2001). Focus on form through collaborative dialogue: Exploring task effects. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 99-118). New York, USA: Pearson Education.

Turner, C.E. (2000). Listening to the voices of rating scale developers: Identifying salient features for second language performance assessment. *Canadian Modern Language Review*, *56*(4), 555-584.

Van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *Teachers of English to Speakers of Other Languages (TESOL) Quarterly, 23*, 489-508.

Vygotsky, L. S. (1978). *Mind in society*. Massachusetts, USA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language*. Massachusetts, USA: MIT Press.

Xi, X. (2008). Methods of test validation. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd edition, Volume 7)*: Language testing and assessment*, (pp. 177-196). New York, USA: Springer Science and Business Media LLC.

Young, R. (2000). *Interactional competence: Challenges for validity*. Paper presented at the Language Testing Research Colloquium. British Columbia, Canada.

Appendix A: Questionnaire

We would like to know your reactions to the test. Please answer the questions in as much detail as possible. This will assist us in our research. Your cooperation is appreciated.

A. Please complete these details:

Name_____ Gender_____ Native language_____

Now you have taken two tests: two tasks in different formats. The two tasks you have undertaken are:

Task A: Youth and Employment
Task B: Violence in the Society

The two formats are:

Format 1: Computer-based test (as what you did on the CAEL Assessment)
Format 2: Small group discussion test (You had a group discussion in the classroom before making your presentation on the computer.)

Please specify which task you have undertaken in the two formats:

Computer-based test_____; small group discussion test _____
A. Youth and employment B. Violence in the society

B. Please note the following questions are asked for the two **TASKS** you have undertaken.

Task A: Youth and Employment
Task B: Violence in the Society

Please complete the following by placing a circle around the most appropriate answer.

1. I believe that Task A would provide an examiner with an accurate idea of my ability to speak English.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

2. I believe that Task B would provide an examiner with an accurate idea of my ability to speak English.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

3. I believe I did well on task A.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

4. I believe I did well on task B.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

5. I had enough opportunity to show my ability to speak English in doing Task A.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

6. I had enough opportunity to show my ability to speak English in doing Task B.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

7. I had enough time to do task A

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

8. I had enough time to do task B.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

9. I am familiar with the content of Task A.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

10. I am familiar with the content of Task B.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

11. I believe Task A produces the type of language required of students in studying a university course.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

12. I believe Task B produces the type of language required of students in studying a university course.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

13. Please rate Task A for difficulty.

easy				difficult
1	2	3	4	*5*

14. Please rate Task B for difficulty.

easy				difficult
1	2	3	4	*5*

C. Please note the following questions are asked for the two **FORMATS** of the test you have taken.

Format 1: Computer-based test (as what you did on the CAEL Assessment)
Format 2: Small group discussion test (You had a group discussion in the classroom before making your presentation on the computer.)

15. I understood what I was supposed to do during the Format 1 test.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

16. I understood what I was supposed to do during the Format 2 test.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

17. I believe that Format 1 would provide me with an opportunity to show my ability to speak English.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

18. I believe that Format 2 would provide me with an opportunity to show my ability to speak English.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

19. I felt nervous while I was doing the Format 1 test.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

20. I felt nervous while I was doing the Format 2 test.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

21. I like Format 1.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

22. I like Format 2.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

23. Format 1 makes the test more difficult.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

24. Format 2 makes the test more difficult.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

25. Format 1 is a fair form of test.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

26. Format 2 is a fair form of test.

Strongly agree	Agree	No opinion	Disagree	Strongly disagree

D. Your preference of the test

27. If you were going to take an oral test in an examination, which one of the four tests would you prefer to take? Put a '1' next to the task you would prefer most, a '2' next to your second choice, and a '3' next to your third choice, and a '4' the test you would least like to take.

Task A in Format 1

Task A in Format 2

Task B in Format 1

Task B in Format 2

Please add any other comments you wish to make: