

A corpus-based investigation into the validity of the CET–SET group discussion

Lianzhen He and Ying Dai *Zhejiang University, PR China*

This article reports the results of an investigation, based on a 170 000-word corpus of test performance, of the validity of College English Test–Spoken English Test (CET–SET) group discussion by examining the degree of interaction among candidates in the group discussion task with respect to a set of interactional language functions (ILFs) to be assessed. The results show a low degree of interaction among candidates in the CET–SET group discussion. Consequently, the inadequate elicitation of ILFs from candidates may well pose a problem for measuring their speaking ability in this regard.

I Introduction

In recent years, a substantial body of work has been produced dealing with interaction in language testing, including conversation analysis as well as other kinds of discourse analysis. The multidimensional models of communicative competence (e.g. Bachman, 1990; Bachman and Palmer, 1996) have identified oral discourse competence as a distinct component of L2 (second-language) speakers' communicative language ability. Conversation is one of the basic means of oral interaction, and being able to participate actively and appropriately in a conversation is a skill that many language learners would like to and need to acquire (Kormos, 1999). The assessment of test-takers' conversational competence has thus gained in importance in many language tests (e.g. Cambridge First Certificate, Cambridge Certificate of Proficiency, College English Test–Spoken English Test and Public English Test Systems in China).

Van Lier (1989) argues that conversation analysis and micro-ethnography must be used to identify and describe performance features that determine the quality of conversational interactions on

Address for correspondence: Lianzhen He, Linguistics Department, School of Foreign Languages, Zhejiang University, Hangzhou, 310058, People's Republic of China; email: hlz202@yahoo.com.cn

tests (Shohamy, 1991), so conversation analysis and discourse analysis have been used to investigate the nature of the language sample that is created in language assessments. Studies on language use in tests focus on the nature of the language sample on which the scores are based. The issues that have been addressed so far in these studies relate to the nature of test discourse, differences in language use between different tests and tasks, and the relationship between the nature of the language sample and the grade awarded.

The key issue in studies of the nature of test discourse, as far as tests of speaking are concerned, is the comparability between the language use in testing situations and the language use in non-testing situations. Lazaraton (1992; 1996) and Ross (1992) analysed the turn-by-turn structure of the interview test interaction, with a focus on the role of the interlocutor in particular. They found some similarities and differences, and the differences were most clearly related to the power distance between the interviewer and the test-taker and the different roles played by each.

Research on the variation in language use between different tasks has been done by Shohamy *et al.* (1993, in Swain, 2001), focusing on the analysis of linguistic features of candidate performances such as lexical density, fluency, turn-taking mechanism, and backchanneling. Their research showed that different tasks resulted in performances with different linguistic features and this raises the issue of the validity of the inferences made about the candidate's communicative ability based on one single task.

Ross and Berwick (1992, in Malvern and Richards, 2002), in investigating the relationship between test discourse and the grade awarded, found that the frequency and extent of interviewer accommodation was related to the proficiency level of the interviewees, and they suggested that the degree of accommodation could be used as an additional dimension in the assessment of candidates (Malvern and Richards, 2002). They also found that the 'accommodation behavior' was closely related to the grade given (Ross and Berwick, 1992). Young's (1995) focus was on candidate language, and results indicated that linguistic features of candidate performance such as rate of speech or amount of elaboration clearly distinguished between two proficiency levels, while discourse features such as the frequency of initiation of new topics or reactivity to topics did not. Using their model of dyadic native–nonnative speaker (NN–NNS) discourse, in which discourse is described in terms of three features – interactional contingency, the goal

orientation of participants and dominance – Young and Milanovic (1992) studied discourse of dyadic oral interviews of the Cambridge First Certificate in English (FCE) exams. They found that variation in the structure of the discourse was related to the examiner, the theme of the interview, the task in which the participants were engaged, and the gender of the examiner and candidate.

Also using the methods of conversation analysis, Egbert (1998, in McNamara *et al.*, 2002) compared OPI (Oral Proficiency Interview) interactions involving American learners of German and an extensive corpus of naturally occurring conversation in German. She found that ‘the organization of conversational repair was explicitly explained by the interviewer, and thus the students initiated repair by means of the forms taught to them, which are not found in interaction between native speakers’ (Young, 2002: 251–52). As a result, she concluded that ‘the OPI differed from conversation in its handling of means of repair.’ (McNamara *et al.*, 2002: 223).

Recently Johnson (2001, in Young, 2002: 250) did an analysis of 35 LPis (Language Proficiency Interviews) conducted over the phone. One of her major findings was that ‘the distribution and allocation of turns in LPis differs markedly from the way that turns are distributed and allocated in ordinary conversation.’ In ordinary conversation, the speakers have equal right to take the turn or to select the next speaker, while in her data the selection of the next speaker is largely controlled by the interviewer. Based on her findings, she concluded that the LPI does not test speaking ability in the real-life context of a conversation as claimed in publicity for the test.

Compared with oral interviews, little validation work has been done on the relatively innovative formats, the paired and group oral, where candidates interact with each other rather than with a native speaker interlocutor and are observed by one or more assessors, and empirical findings are few. The group oral format is not entirely new to the L2 testing research literature (Bonk and Ockey, 2003). Folland and Robertson (1976) were among the first writers to recommend using the group discussion in oral testing. Since then there has been reports of group testing being used successfully in Israel and Zambia with school students and in Italy with university students (Fulcher, 1996). Research by Whiteson, Fulcher and others has come up with results favoring the use of group oral format. Whiteson and Fulcher have reported that ‘test-takers considered the group oral as a valid form of L2 testing’ (Bonk and Ockey, 2003: 91), and Folland, Robertson and Fulcher have found that ‘examinees felt more comfortable and confident speaking to

one another than to an examiner' (Bonk and Ockey, 2003: 91). Fulcher also found that:

the group discussion format was the easiest task when compared to a picture-based discussion with an interviewer or a text-based discussion with an interviewer, suggesting that it might be appropriate for use with learners of lower levels of proficiency. (Bonk and Ockey, 2003:91)

According to Swain (2001: 277):

there are a number of reasons why some tests now include pairs or small groups of individuals interacting together to debate an issue or to solve a problem. Dissatisfaction with the oral interview as the sole means of assessing oral proficiency and a search for other tasks that elicit different aspects of oral proficiency are concomitant reasons.

With the movement of communicative language teaching, pair and group work have been advocated and adopted in the classroom. An attempt to influence teaching practices or, alternatively, to mirror teaching practices have also played an important role (Swain, 2001). 'Economic reasons, too, have played their part: where there are many students to be tested, it can be less expensive to test them in pairs or groups' (Berry, 2000 in Swain, 2001: 277). Berkoff (1985) argues for the use of group oral format stating that it overcomes the problems of 'artificial conversation' between a 'distant examiner' and a 'nervous examinee'. Bonk and Ockey (2003: 90) contend that 'providing the students with opportunities to initiate and control conversation during the test may mean an enhancement of the validity of the score-based inferences.'

Among the few studies of the paired format, 'Iwashita (1998) carried out a small-scale study of a paired oral assessment in Japanese, with candidates matched with higher proficiency and lower proficiency candidates respectively. She found that being matched with a higher proficiency candidate generally led to the production of more talk, but that this did not necessarily have an impact on scores' (McNamara *et al.*, 2002: 226). An extension of the paired oral format is the group oral, and one of the few attempts in this area is made by Berry (2000, in Swain, 2001), who examined the relationship between extraversion and performance on a group oral test. She has found:

a complex relationship between the characteristics of the test-taker and those of the rest of the group... The scores of introverts are suppressed when they take part in a discussion... in a group with a low mean level of extraversion, and are elevated when in a group with a high mean level of extraversion. The reverse holds for extroverts (Swain, 2001: 278).

The problem of grouping found by Berry is echoed by Foot (1999), who expresses his concern about how difference in the ability of paired candidates can affect performance. In a similar vein, Swain (2001: 277–78) notices that little validation work has been carried out on small group oral testing. Based on Berry's findings, she wonders:

How are scores based on interaction among participants to be interpreted as an indication of individual performance ability? Can they be interpreted as individual performance ability at all? [She thus emphasizes that] this potential for unfair biases if students of differing compatibilities (or of differing abilities) are grouped together needs further investigation.

In a sense, Foot and Swain have identified some areas for further validation work regarding the paired and group oral such as the sampling problem and to what extent the candidate's performance in the paired and grouped situation can be interpreted as individual ability. More recent work was done by O'Sullivan (2002), who explores the effect of test-takers' familiarity with their partner on pair-task performance and calls for 'the need of urgent and extensive study' of any test format that employs tasks requiring interaction between individuals.

The present study is an attempt to examine the degree of interaction in the group discussion section of College English Test–Spoken English Test (CET–SET), a national oral proficiency test for non-English majors in China, and to validate the match between intended and actual test-taker language with respect to a checklist of interactional language functions (ILFs) representing the construct of spoken language ability.

Messick's (1980; 1989) unified framework of validity has been gaining wide acceptance from researchers and has become the cornerstone for most validation work. According to Messick (1980: 1015), 'construct validity is indeed the unifying concept that integrates criterion and content considerations into a common framework for testing rational hypotheses about theoretically relevant relationships'. Bachman (1990: 254–56) accepts the position of Messick that validity is a unitary concept pertaining to test interpretation and use and emphasizes the centrality of construct validity to test interpretation and use. He views constructs as 'definitions of abilities that permit us to state specific hypotheses about how these abilities are or are not related to other abilities, and about the relationship between these abilities and observed behavior.' Furthermore, Bachman points out that 'construct validity concerns the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs.'

In other words, 'the fundamental issue in construct validity is the extent to which we can make inferences about hypothesized abilities on the basis of test performance.' In construct validation, therefore, we seek to provide evidence that supports specific inferences about relationships between constructs and test scores.

II CET-SET and the present study

First administered in June 1987, College English Test (CET) is a test of English proficiency of non-English majors at Chinese universities,¹ testing candidates' listening, reading and writing ability. Every year, several million Chinese college students take the test and CET, especially CET4, has become a high-stakes test since a certificate of CET4 is a prerequisite for a Bachelor's degree in many Chinese universities. To answer the call for more emphasis to be put on the oral English proficiency of Chinese college students from the Ministry of Education, as is specified in the revised *College English Teaching Syllabus*², and to bring about positive backwash effect on classroom teaching, National College English Testing Committee began to administer College English Test-Spoken English Test (CET-SET) on a nationwide scale in 1999. The test is administered twice a year, in May and in November, and only students with a score of 80 in CET4 or 75 in CET6 can take the test.³ The holders of CET-SET certificates stand a better chance than those who do not have the certificates of being employed by government organizations, foreign companies or joint ventures, positions for which there is intense competition.

CET-SET adopts a face-to-face small group format involving an interlocutor, an additional examiner and three or, in rare cases, four candidates. The format of CET-SET is summarized in Table 1. Part 1 serves as a 'warm-up', with a brief self-introduction by each candidate followed by a question to each candidate related to the

¹In China, there are two national English tests for college students: Test for English Majors (TEM4 to be taken in the 4th semester and TEM8 in the 8th semester) and College English Test for Non-English Majors (CET4 as the basic requirement and CET6 as a test of higher proficiency). A certificate of TEM4 or CET4 is, in many Chinese colleges and universities, a prerequisite for a Bachelor's degree.

²*College English Teaching Syllabus* was first compiled and published in 1985 and revised in 1999.

³The reason for the restriction is the possible mismatch between the number of trained and qualified examiners and the potentially large testing population and the ensuing logistical problems such as space, equipment, etc.

Table 1 Format of CET-SET

Part	Time	Participants	Task format
1	5 min	Examiner-candidate	Verbal questions to each candidate
2	5.5 min	Examiner-candidate	Visual stimulus
	4.5 min	Candidate-candidate	Group discussion
3	5 min	Examiner-candidate	Verbal questions to each candidate

topic of the oral test (e.g. City Traffic). In Part 2, each candidate is first given a visual prompt (pictures, diagrams, etc. on a card) and is asked to describe the pictures or diagrams briefly and comment on them for 1.5 minutes after a 1-minute preparation. The pictures, diagrams, etc. given to the candidates in a group belong to the same topic area. After the individual presentations, the candidates are asked to have a 4.5-minute group discussion on the given topic related to the visual prompt, and the candidates' performance is judged according to their contribution to the discussion. In Part 3, the interlocutor asks each candidate one last question about the topic to further test the candidates' oral proficiency (for a sample of the test, see Appendix 1). Examinees are scored on accuracy and range, size and discourse management, flexibility and appropriacy (for CET-SET Rating Scale, see Appendix 2).

The focus of the present study is the 4.5-minute group discussion in Part 2 where candidates are asked to have a discussion on a given topic among themselves. This task is designed to measure the candidates' communicative language ability (CLA), and it is in this part that the interactional language functions (ILFs) included in the *CET-SET Syllabus* are to be assessed.⁴ The candidates are expected⁵ to engage in communicative interaction while arguing with each other, asking each other to clarify a point and trying to reach an agreement.

Kramsch (1986) defines communicative interaction as entailing negotiating intended meanings, i.e. adjusting one's speech to the effect one intends to have on the listener, anticipating the listener's response and possible misunderstandings, clarifying one's own and the other's intentions, and arriving at the closest possible match between

⁴CET-SET Syllabus lists a number of language functions to be assessed and, for the purpose of the present study, only those functions related to the group discussion part are examined.

⁵Before taking CET-SET, testing centers organize training sessions to familiarize the candidates with the test format. And during the test, clear instructions are given to the candidates about what they are supposed to do during the group discussion (see Appendix 1).

intended, perceived and anticipated meanings. And in defining and explicating interactional competence, Kramsch (1986: 367) writes:

successful interaction presupposes not only a shared knowledge of the world, the reference to a common external context of communication, but also the construction of a shared internal context or 'sphere of inter-subjectivity' that is built through the collaborative effort of the interactional partners.

CET-SET designers hold the view that it is a direct assessment of the candidate's ability in interactional competence in that speaking is a productive skill and its outcome can be directly observed. Based on this, they have argued that CET-SET is sure to be valid as long as it is properly designed (Yang, 1999). It is true that test performance can be directly observed, but it does not ensure that the interpretation made on the basis of it is valid because it is essential to take the quality of the elicited performance into consideration. In other words, to justify the validity of the interpretation based on test performance, evidence needs to be provided that the elicited test performance reflects the areas of speaking ability to be measured and very little else. The group discussion task in CET-SET is designed to involve the candidates in communicative interaction. However, attention must be drawn to the possible discrepancy between the test designers' intentions and the candidates' actual performance. Spence-Brown's (2001: 463) study of authenticity in an embedded assessment task reveals that the degree of authentic interaction, and the ways in which individuals' abilities are engaged, vary widely. She therefore suggests that 'authenticity must be viewed in terms of the implementation of an activity, not its design.' Van Lier (1989) also contends that there is a need to look at oral tests from within and to analyse the speech events used in order to address issues of validity. The possible discrepancy indicates the need for validation studies, that is, to what extent the elicited test performance reflects the areas of speaking ability the test designers intend to assess, i.e. the candidates' ability to engage in communicative interaction. Since its first administration in 1999, little validation study has been done, and the present study is thus an attempt in this regard.

III Research method

1 Data

The data for the present study were taken from the November 2001 administration of CET-SET. The test was administered on 4 days

(two weekends), and for security reasons, a total of 8 topics were used, one for each half day (for the eight group discussion topics, see Appendix 3). All the tests were audio recorded, and altogether 102 tapes were obtained. Tapes of poor quality were discarded and finally 67 tapes were transcribed. The transcription was first done by 34 senior students and then carefully reviewed by the 3 research team members. The result was a 170 000-word corpus consisting of eight sub-corpora, each on one topic area.

A questionnaire involving 196 candidates was conducted after the test to obtain some qualitative data for further validation evidence. The questionnaire (see Appendix 4) was designed to elicit information regarding the candidates' perception of the CET–SET group discussion and how they dealt with it. Alderson (1985) argues for the use of introspective data and, more practically in the field of oral testing, retrospective data from the candidates. He contends that such data would provide valuable information about the ways in which the candidates deal with test items. 'Messick (1989) explicitly recommends including test-taker perceptions as a crucial source of evidence for construct validity' (Elder *et al.*, 2002: 363–64). Elder *et al.* (2002) caution, however:

that we should not rely too heavily on test-taker feedback, either as a basis for test design or in mounting test validation arguments. Testtaker reactions and attitudes may be conditioned by a range of different attributes (e.g. gender, social class, professional experience, proficiency) as our earlier review would indicate, as well as by features of the task itself.

2 Code scheme

For the purpose of the present study, the data of 48 CET–SET group discussions, six randomly selected from each sub-corpus, were analysed by the research team members in terms of the occurrence of a set of eight ILFs included in *CET–SET Syllabus* (see Note 4 above). The eight ILFs examined were:

- a) (Dis)agreeing: Express (dis)agreement with what another speaker has said;
- b) Asking for opinions or information: Ask for opinions or information;
- c) Challenging: Challenge opinions or assertions made by another speaker by giving countering reasons or evidence;
- d) Supporting: Support opinions or assertions made by another speaker by providing more reasons or evidence;
- e) Modifying: Modify arguments or opinions in response to another speaker;

- f) Persuading: Attempt to persuade another speaker to accept one's view;
- g) Developing: Express ideas building on what another speaker has said;
- h) Negotiating meaning:
 - Asking for clarification: Ask for explanations of words, expressions or opinions that may not have been understood;
 - Giving clarification: Give clarification as required by another speaker or correct another speaker's misunderstanding of one's own message;
 - Asking for confirmation: Restate what another speaker has said, or part of it, to confirm his/her own understanding of the message;
 - Checking for comprehension: Check the listeners' understanding of the message to find out whether as a speaker he/she is understood by others.

The coding was done by the research team members. Prior to the coding procedure, the team agreed in general terms upon the steps of coding and the coding scheme and consensus was reached in identifying and classifying the ILFs. In cases of disagreement, the team members would examine the transcripts closely and from time to time listen to the tapes for phonological cues. At certain points help from native speakers who were teaching at Zhejiang University was enlisted.

In the following section, each type of ILF is illustrated by an example from the data. The analysis of group discussion data is a complex undertaking. To examine the degree of interaction among candidates in the group discussion task, it is useful to establish a checklist of ILFs. Yet it must be acknowledged that this method has its limitations. For example, there is no clear-cut boundary for each language function and some utterances involve multiple language functions, and therefore this method has certain limitations. However, it can provide some insight into what really goes on in the group discussion task, which is difficult to access by other means.

a (Dis)agreeing:

Excerpt 1:

B: Well, you (*referring to Candidate A*) said that people ... when people are lonely they can keep pets to change their feelings. Mm ... but I think that why they just not go outside, make friends with others? Why they have to ... er ... disturb the little animals' life to change their ... to let themselves better? Why they just can ... can make friends with others?

- A: Because a pet can be your ... can company you all the time. But friends cannot. Maybe a very good pet can make you feel that it's your relative. It's different, I think, not friend.
- C: I will agree with her (referring to Candidate A). I think ... mm ... friend can't accompany you, but pets can accompany you all the times. Mm.

The topic for discussion is 'Should people be encouraged to keep pets?' Excerpt 1 is one example of agreeing. Candidate B thinks that people can relieve their loneliness by going outside and making friends with other people. Candidate A contradicts Candidate B by saying that friends cannot accompany people all the time, while a pet can. And Candidate C expresses agreement with Candidate A's opinion that a pet is different from a friend by repeating the reason given by Candidate A.

b Asking for opinions or information:

Excerpt 2:

- A: Ok, I'll say first. In my opinion, en ... I think the habit that en ... 'early to bed and early to rise' is a good health especially for the college students. En ... because college students en ... is also en ... is also young people, they should enough rest to make the refreshing, en ... to refreshing the next day. So that they have enough energy to ... to ... mm ... to afford their ... the next day's study and life. And if they and if they sleep too late, they may feel tired and en ... can't have enough energy to ... mm ... to make them ... to make them have a good ... er ... to make them have a good health, life in the next day, sorry. What's your opinion?

In Excerpt 2, the topic for discussion is 'Which aspect is most important for university students to stay healthy?' After giving her opinion that early to bed and early to rise is most important, Candidate A asks for another candidate's opinion by asking, 'What's your opinion?'

c Challenging:

Excerpt 3:

- A: En ... I think it's essential. Because if you are high educated, you'll gain more knowledge. En ... I think ... en ... you may gain more skills. Mm ... why people ... people learn more, they can do more things. En ... and the company or some government, they ... if they want to en ... provide some career, they always ask for higher ... en ... education people.
- C: I don't agree with you. How do you explain that Bill Gates and other IT ... en ... CEOs? En ... the ... the many of them don't graduate from universities.

Here the candidates are having a discussion on the topic 'Is graduate education essential for a successful career?' Candidate A thinks it is essential and justifies her statement with evidence from the job market. Candidate C challenges Candidate A's opinion by giving

counter evidence of Bill Gates and some others who do not have graduate education but have a very successful career. One point to note here is that in this excerpt Candidate C first expresses disagreement and then gives counter evidence. Since the disagreement here serves as a prelude to further argument, Candidate C's contribution, rather than being double counted, was counted in terms of Challenging only.

d Supporting:

Excerpt 4:

- A: Ok, and I think maybe the most popular en... festival in China is Lover's Day. Because among young people the... er... most of them maybe are lovers...
- C: I disagree. I think the most popular festival between our college students are young people, is the Christmas Day. Er... er... western... western festival is gaining much popularity between our young people. Er... you see every Christmas Day coming, we'll organize... organize some activities during the... during the day. Er... they also go to the club to have some singing and dancing. They also can go with their boyfriend and girlfriend to have a... er... to have a celebration. So... er... I think the... er... Christmas Day is more popular than other day... other festival.
- B: I agree with you (referring to Candidate C). And... and er... and there are another factor. Because the Christmas Day are... are... are near of the New Year's Day, you can celebrate the two... the two festival together. And... and the... um... the atmosphere is very... is very... very delight. I think the... the Christmas Day is more popular than Valentine's Day.

In Excerpt 4, the topic is 'Which festivals are most popular with young people in China?' Candidate A says that it is Lover's Day (Valentine's Day). After expressing disagreement, Candidate C states that it is Christmas Day and gives justification. Candidate B supports Candidate C's opinion by giving further evidence, that is, Christmas Day is close to New Year's Day. Here Candidate B's expression of agreement is also a prelude, so his contribution was counted in terms of supporting only.

e Modifying:

Excerpt 5:

- A: My ideal family model is a family with... er... with two children. The best is they are twins – one boy, one, one girl...
- C: I think in conditions of China, I think it's better for... mm... family to have one child. Er... because I think though the family has only one child, he or she could have friends. En... they also can help each other. And... mm... first of all, it's very important for the condition of China, I think.
- A: Yes, I think from the... en... view from the point of the country, we had better have only one children; but... en... view from the development

of the children, we had better have more than one children ... er ... for example, two children. And the compromise is ... may be difficult.

The topic for discussion here is ‘The ideal size of a family’. Candidate A expresses her view on the ideal size of a family, i.e. a family with two children. Candidate C voices her opinion that, considering China’s current situation, it is better to have one child. In response to what Candidate C has said, Candidate A modifies her opinion by taking China’s current situation into consideration and says it may be difficult to reach a compromise.

F Persuading:

Excerpt 6:

B: ... I think ... mm ... just through ... mm ... mm ... physical exercise ... en ... can make you much stronger and ... mm ... much energetical and make you more healthy ...

C: I really agree with her, I also think en physical exercise is very important, the most important ...

A: Ok, I would like add something that ... en ... I also agree with you that ... en ... make as much as exercise can make our health. But I think, first of all, we maybe should get up early, so they have enough time to take exercises. Because we know in the rest of the day maybe we have not enough time to take exercise, but in the morning, if we get up early we have ... en ... we can breathe the fresh air and we have ... and so that we can ... en ... take exercises. And we also can get play ... get play for a new day’s study and work ... en ... and so on. And ... and I think, mm ... , get up early and sleep early is a good health, is a good habit for our health.

In this excerpt, the candidates are asked to have a discussion on the topic ‘Which aspect is most important for university students to stay healthy?’ Both Candidate B and Candidate C hold the view that doing exercises is most important for college students to stay healthy. Candidate A first admits that doing exercises is important to health; she then tries to persuade the other two to accept her view that early to bed and early to rise is most important because when doing so people have enough time to do exercises. One thing to note here is that although Candidate A starts her turn with ‘Ok, I would like add something that ...’, seemingly suggestive of developing a theme, she is not expressing her ideas building on what other speakers have said. So instead of classifying it as developing, the excerpt was counted as persuading.

g Developing:

Excerpt 7:

A: I think we should behave, behave ourselves politely in action. And we should, we should guide the young men in practice such as we should ...

er ... we should get the young men ... get the young men together and play with them and we behave politely. I think it will affect them.

C: I think if I find their wrong doing of young men, we should point it out and tell them to correct.

A: I think the ... I think the approach should be ... should be polite and they can accept it and not to rebel.

In Excerpt 7, the topic is ‘What can be done to educate people to conduct themselves properly?’ Candidate A starts by saying that we should try to influence or set examples for the young men, by behaving politely ourselves. Candidate C furthers the discussion by saying that we should point out people’s bad behavior and tell them to behave themselves. Building on what Candidate C has said, Candidate A suggests that it should be done in a polite way to make people willing to accept that view. Here both Candidate A and Candidate C have successfully used the introductory expression, ‘I think’, as a conversational opener. And when Candidate A talks about what ‘the approach should be’, he is using an abstract noun, a generalization, which covers ‘should point it out and tell them to correct’ and ‘should be, should be polite and they can accept it and not to rebel’. Thus, Candidate A builds on Candidate C’s statement by adding another dimension, a generalization, which covers both his response and that of Candidate C. Then he, like Candidate C, proceeds to make a specific recommendation.

h Negotiating meaning:

- Asking for clarification

Excerpt 8:

A: Though a good exercise is very important ... en ... we mustn’t eat too much food like ... en ... rich in fat. We must eat something green, such as apple, an apple a day keeps doctor away. En ... and we must have vegetable and meat every day, and vegetable is ... en ... are rich in vitamin, and meat rich in protein ... en ... so I think exercise and good diet is very important, if we ... en ... have strong body, we can study well.

B: I want to know, do you think a healthy diet is more important or doing exercises more important? I want to know your opinion.

In this excerpt, Candidate A first says that doing exercises is important for college students’ health and then says a good diet is very important, with examples of the kind of food that is good for health. In his turn, Candidate B is seen to bring in the two points dealt with earlier by Candidate A, who has, however, not clarified which he considers the more important. Candidate B draws out both points from the previous utterances and presents them in contrast, in order to seek this clarification.

- Giving clarification

Excerpt 9:

B: Personally I think the mobile phone is the most convenient, because you can use it everywhere. And now mobile phone can link to the Internet. You can send e-mail by mobile phone. I think if I have enough money, I'll buy one.

C: You mean you will buy a computer. But I think nowadays there is still some ...

B: No, I mean I'll buy a mobile phone.

In this excerpt, the candidates are having a discussion on the topic 'Which is the best means of communication?' Candidate C, by using the expression 'You mean', is trying to follow what Candidate B says in the previous turn and give his comment, but he obviously fails to understand what Candidate B means. B therefore interrupts to correct Candidate C's misunderstanding and makes it clear that he will buy a mobile phone, not a computer.

- Asking for confirmation

Excerpt 10:

B: Oh, I like to live in big cities too, especially downtown because there're a lot of department I can go there to shopping. En ... you see the young ... the young people like to dance and go to the wine bar and disco ... disco ... disco, er ... so ... so it's a heaven for the young people.

A: So you mean the big cities' atmosphere cater to young's taste?

The topic for discussion here is 'Is it desirable to live in a big city?' In this excerpt, Candidate B voices her preference and gives justifications for her preference, and ends her turn with a concluding remark '... so it's heaven for the young people'. Candidate A, in order to confirm her understanding of what Candidate B says, asks the question 'So you mean the big cities' atmosphere cater to young's taste?'

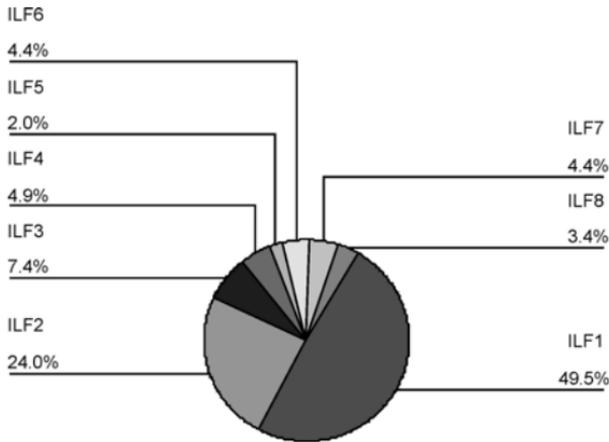
- Checking for comprehension

This subtype of negotiation of meaning, such as 'Do you see what I mean', is absent from the data.

IV Findings

1 Distribution of ILFs in the CET–SET group discussion

The distribution of the eight ILFs elicited by the CET–SET group discussion is presented in Figure 1. We can see from the pie chart



Notes: ILF1 = (dis)agreeing; ILF2 = asking for opinions or information;

ILF3 = challenging; ILF4 = supporting; ILF5 = modifying;

ILF6 = persuading; ILF7 = developing; ILF8 = negotiating meaning;

Figure 1 Distribution of 8 interactional language functions

that, among the eight ILFs, the most frequently elicited one is (Dis)agreeing, which accounts for 49.5% of the total number of occurrences, followed by Asking for opinions or information, 24.0%. Each of the remaining six – Challenging, Supporting, Modifying, Persuading, Developing and Negotiating meaning – accounts for a low percentage respectively. Figure 1 shows that in the group discussion, the forms of interaction among candidates center on expressing (dis)agreement to what other candidates have said and asking for opinions or information to a lesser degree. The forms of interaction such as presenting counter-arguments, providing further evidence to support another candidate, modifying one's own opinions in response to other candidates, trying to persuade others to accept one's own point of view and negotiating mutual understanding when exchanging opinions occur far less frequently in the discussion.

2 Elicitation of ILFs across group discussions and candidates

More insights are gained when we look into the elicitation of ILFs across CET-SET group discussions and candidates (Table 2 and

Table 3). From Table 2, we can see that (Dis)agreeing occurs in 45 out of the 48 group discussions (93.8%) and in six group discussions it occurs four times. The function of Asking for opinions or information is elicited by 27 out of the 48 group discussions (56.2%). With regard to the remaining six, there are very few instances: Challenging occurs only in 13 out of the 48 group discussions (27.1%), Supporting in nine (18.8%), Modifying in four (8.3%), Persuading in nine (18.8%), Developing in eight (16.7%) and Negotiating meaning in five (10.4%).

Table 3 shows that only a small number of the candidates produced the ILFs such as Challenging, Supporting, Modifying, Persuading, Developing and Negotiating meaning in the discussion. The functions of Challenging, Supporting, Modifying, Persuading and Developing are signals of a candidate's reaction to other candidates' contributions and the degree of his or her involvement in the discussion. That is, engaging in communicative interaction is quite different from giving an individual presentation, in that the former requires the candidate to pay attention to what other members have said and to react to the contributions of others. This reactivity, or contingency, is a property of adjacent turns in dialogue in which the topic of the preceding turn is co-referential with the topic of the following turn (Young and Milanovic, 1992). To be reactive to other

Table 2 Elicitation of ILFs across the 48 CET-SET group discussions

Frequency	ILF1	ILF2	ILF3	ILF4	ILF5	ILF6	ILF7	ILF8
0	3	21	35	39	44	39	40	43
1	11	13	11	8	4	9	7	3
2	18	8	2	1	0	0	1	2
3	10	4	0	0	0	0	0	0
4	6	2	0	0	0	0	0	0

Table 3 Elicitation of ILFs across the 144 candidates

Frequency	ILF1	ILF2	ILF3	ILF4	ILF5	ILF6	ILF7	ILF8
0	64	106	129	134	140	135	135	138
1	60	28	15	10	4	9	9	5
2	19	9	0	0	0	0	0	1
3	1	1	0	0	0	0	0	0

candidates' utterances in the discussion, the candidate should be able to challenge or support others' views, modify one's own views, or try to persuade others to accept one's own views and so on. Disappointingly, among the 144 candidates, only 47 of them – less than one-third of the total – produced some of these five functions.

Moreover, the lack of Negotiating meaning is evident in the CET–SET group discussions. The negotiation of meaning refers to the collaborative work that speakers undertake, motivated by the need and desire to understand each other, to overcome trouble in communication. It is necessary to negotiate mutual understanding when exchanging opinions. However, only six out of the 144 candidates negotiated meaning in the discussion. Most candidates fail to negotiate mutual understanding even though they are encouraged to do so by the interlocutor before the discussion.

3 Results of questionnaire analysis

Answers to Question 1 of Part 2 on the questionnaire – i.e. how they felt when they were having the group discussion – show that 5.6% of the candidates felt very nervous during the group discussion, followed by 19.9% (quite nervous), 56.6% (not nervous) and 17.9% (relaxed). This seems to provide evidence for the claim that the group oral may reduce test anxiety (Fulcher, 1996). In responding to Question 2, the candidates' perception of the difficulty of the group discussion, nobody chose A (very difficult). 8.7% of the candidates considered the group discussion quite difficult, and the rest chose C, not difficult (60.2%), or D, easy (31.1%). When asked about the discussion topics (Question 3), on average 46.4% of the candidates found the topics for discussion not interesting and 13.8% thought the topics were dull.⁶ And over half of the candidates, 60.7%, regarded the examiners as their target audience. When asked about attention to other candidates' talking (Question 6), only 32.1% of the candidates chose A and the rest all chose B. In other words, most of the candidates concentrated on organizing their own ideas during the group discussion. In responding to the question regarding meaning negotiation (Question 7), only 7.1% of the candidates chose A, 'asked for clarification or explanation'. And when asked about whether they tried to reach an agreement with the other candidates

⁶The topics for different groups are different, so the average is given here.

during the discussion, over two-thirds (74%) of the candidates acknowledged that they failed to do so since they only tried to express their opinions clearly when their turn came.

V Discussion

The results of an analysis of the 48 CET–SET group discussions show low frequency of the occurrence of the ILFs such as Challenging, Supporting, Modifying, Persuading, Developing and Negotiating meaning. In the group discussion, candidates are required to argue with each other or ask each other questions to clarify a point, and try to reach an agreement on a given topic, but there is little evidence from the data in this regard. The research team members consider the following to be the possible reasons for the lack of communicative interaction in the group discussion.

First, the candidates' interpretation of and approach to the group discussion task contributes to this relatively low degree of interaction. A useful concept in characterizing the students' interpretation of and approach to the task is that of framing. Frame, or script, in cognitive psychology, refers to 'units of meaning consisting of sequences of events and actions that are related to particular situations.' (Richards *et al.*, 1992). It implies both structures of expectation and the constantly shifting and active construction of an event by participants. The employment of a certain frame has a direct bearing on how a student approaches a certain task. In employing an assessment frame, for example, a student approaches the task as an assessment event and behaves as is expected and required of an assessment event. Framing thus involves interpretation of events, an orientation towards them and a basis for behavior. Candidates' responses to the questions in the questionnaire show that for most candidates the assessment frame seems to predominate. They are strongly aware of the testing situation and interpret the CET–SET group discussion as an assessment event rather than a meaningful discussion with group members. The orientation to the group discussion task as an assessment event serves as a basis for candidates' behavior in the test. Accordingly, candidates try to display their best possible performance with an emphasis on language abilities by resorting to some test-taking strategies. For example, they may take advantage of the time while other candidates are speaking to think about and organize their own ideas in their next turn. They concentrate on expressing their own line of thought well rather than

responding actively and relevantly to what other candidates have said in developing the discussion. Moreover, the candidates consider the examiners, rather than the other candidates in the group, to be their target audience, which is evidenced by the high frequency of such expressions as 'I agree with what he/she says' and 'I want to add to what they have said' in the data. Their heavy emphasis on individual performance interferes with the development and success of the discussion. This, to some extent, helps to explain the low frequency of Challenging, Supporting, Modifying, Persuading, Developing and Negotiating meaning elicited by the CET-SET group discussion.

Second, another factor which helps to explain the low degree of interaction among candidates in the CET-SET group discussion task lies in the candidates' lack of confidence in their ability to handle the group discussion task as communicative interaction. The group discussion is a spontaneous speech event in which candidates must organize and express their opinions on the spur of the moment. Reacting to what other candidates have said under the pressure of time can be a problem for the candidates who put emphasis on accuracy and fluency. When the candidates feel that they are not capable of reacting to what others have said in their most accurate and fluent English, they choose to avoid that and instead express their own line of thought to display their best possible performance for assessment purposes. Excerpt 11 sheds some light on this.

Excerpt 11:

- C: En ... I agree, I agree up to a point, but I think it is not essential. Mm ... as we all know, students in university ... en ... not all the students in university are working hard, study hard. They ... somebody often ... often ... mm ... absent from ... often absent from class. And I don't think the students in university is better than ... en ... is better than the people didn't enter into the university.
- A: ...
- C: ...
- A: ...
- B: If you are boss, en...and the ability of the two person may be equal, but one is higher educated, and the second is middle school student. What do you prefer?
- C: Mm ... I ... if I ... en ... I prefer ... mm ... how can I ... en ... students ... I like study, but I don't think the entire student are like study. En ... just in my classmate, some of my classmates ... en ... they ... I don't think they study hard. I think it is waste of time for them. They ... all they can do is ... en ... watching TV, go dancing. I don't think they concentrate on study. They I think they can get a good chance, if they ... mm ... go to society.

In this excerpt, the topic for discussion is 'Is graduate education essential for a successful career?' Particular mention should be made

about Candidate C's contribution. At first he makes an attempt to answer Candidate B's question, saying 'Mm ... I ... if I ... en ... I prefer ... mm ... how can I ...'. But feeling an inability to cope with that on the spot, as is evidenced by the false starts and discourse fillers in his first utterance, he gives up after a short pause and turns to continue his previous line of thought that not all the students in the university work hard.

On the other hand, there are also cases where candidates, rather than reacting to what others have said, continue their own line of thought, as is illustrated in Excerpt 12.

Excerpt 12:

A: I think for a university student, a good health ... body is very useful to us. Because as we all know, the competition ... a competitive society is come and the 21st century is come. And we, as we know, weal ... a health is a wealth. And so we have ... if we have a good body we can face the competition ... er ... in front of us. And we can, we can study hard, study more better and work better.

C: Yes, I think so.

B: Er ... I agree with you. And to stay healthy ... er ... we know that good living habit is ... a healthy diet and adequate physical exercises is both ... is all very important. But I still think good living habits is the most important. Because as a student perhaps we don't have much time for exercise and we ... we didn't have so much knowledge to be ... er ... to on a good diet. So a good living habits is so easy to do and we only need to sleep well, and eat well and not do, not ... er ... something excessive and I think ... er ... we can get a good state of health.

C: Er ... I agree with you. And I have my opinion. I think, yes, we first student ... er ... we have many exercises but I think ... er ... many of us spent too much time to study. And we see get up very earlier and go to sleep very late. And I think this don't help to our health ... healthy. I think ... er ... the normal time to go to bed is very important to our health.

A: I want to add something. That we not only need a good body and we also need a good mind ... er ... especially a health mind. So because as the WTO, China enter ... enter ... enter into the WTO, we are confront more competition and we have a good body is the first step we need. And second step we need the health mind. Only we need wide knowledge and we can face the competition and compete with the foreign country and foreigners. And so ... er ... I think good minded is useful for us.

In this excerpt, the three candidates are having a discussion on the topic 'Which aspect is most important for university students to stay healthy?' In his first turn, Candidate A states the importance of a healthy body in a competitive society. Then the other two candidates express their views on the topic. Both of them agree with what Candidate A says using expressions like 'Yes, I think so' and 'I agree with you ...', and provide additional information about the important factors contributing to a healthy body. In his second turn, rather

than following up what Candidate B and Candidate C have said, Candidate A develops his own ideas presented in the first turn by adding the importance of a healthy mind. We can see that Candidate A's contribution in these two turns is rather like a monologue. Obviously, Candidate A intends to express his own ideas rather than react to what others have said to engage in the discussion. It is likely that he does not have much confidence in handling the group discussion interactively in his best possible English, especially with regard to accuracy and fluency. Therefore he is probably using the time when other candidates are speaking to plan and organize what he is going to say in his next turn.

Third, lack of interest in the group discussion task itself is also likely to have an effect on the candidates' behavior in the test. The result of questionnaire analysis indicates that on an average 46.4% of the candidates think that the topics for discussion are not interesting and 13.8% think the topics are dull. Although great efforts are made to soften the testing situation on the part of the test designers, candidates are generally aware of the testing situation and approach the test as an assessment event. A test task thus faces the problem of a lack of communicative purpose, especially when the task fails to engage the candidates' interest. The eight CET-SET group discussion topics do not seem very engaging and, as a result, the candidates' attempt to work with the task and their motivation to develop the topic to any significant degree with other members are likely to be minimal.

Finally, the tendency of the candidates to produce long turns may also play a part. Before the group discussion begins, candidates are informed that their performance will be judged according to their contribution to the discussion. It seems that many candidates interpret contribution in terms of quantity rather than quality. As a result, they try to talk as much as possible rather than make relevant contributions to the discussion. Furthermore, in the discussion, the candidate who has the turn is usually allowed to express his or her ideas fully before the next candidate takes the floor unless he or she keeps talking for too long. Accordingly, turns in the group discussion are often long and resemble short monologues, as shown in Excerpt 13.

Excerpt 13:

B: I think in China, the ... the western festival are more popular er ... er ... because I think er ... in nowadays the ... the international interaction are more and more popular. And young and now young people are ... er ... our young people are dreaming of going to ... going to ... going abroad. And they ... er ... they may they must be familiar with the with the western festival. I think it's a ... it's a very important festival.

- A: Ok, and I think maybe the most popular en festival in China is Lover's Day. Because among young people the ... er ... most of them maybe are lovers. So in that day, they can er ... express their heart to ... to ... to their lover and they often ... they often give flowers such as roses to their lover. And I think it will become more and more popular.
- C: I disagree. I think the most popular festival between our college students are young people, is the Christmas Day. Er ... er ... western western festival is gaining much popularity between our young people. Er ... you see every Christmas Day coming, we'll organize ... organize some activities during the ... during the day. Er ... they also go to the club to have some singing and dancing. They also can go with their boyfriend and girlfriend to have a ... er to have a celebration. So er ... I think the ... er ... Christmas Day is more popular than other day, other festival.

In this excerpt, the three candidates are having a discussion on the topic 'Which festivals are most popular with young people in China?' They take turns to fully express their opinions in long turns. Candidate B gives a general view that '... the western festival are more popular ...' since 'our young people are dreaming of going to ... going to ... going abroad'. Candidate A thinks that it's Lover's Day that is the most popular. Candidate C, although starting his turn with 'I disagree', does not provide any counter-argument. He simply voices his own opinion that Christmas is the most popular festival for young people and mentions some of the popular Christmas activities. Obviously, within the 4.5 minutes allowed for the discussion, the longer the turns are, the fewer the turns for each candidate. This, to some extent, results in the low frequency of the occurrence of interactional language functions.

VI Conclusions

The present study is a corpus-based investigation of the validity of the CET–SET group discussion. The comparison between the candidates' actual performance and the test syllabus was examined with respect to the degree of interaction among candidates in the group discussion task. The degree of interaction was analysed by means of a checklist of eight interactional language functions included in *CET–SET Syllabus*. Quantitative analysis shows that the frequency of the occurrence of ILFs, especially the functions of challenging, supporting, modifying, persuading, developing and negotiating meaning, are very low. There is a discrepancy between the elicited performance and the test designers' intentions regarding the elicitation of ILFs. With further evidence from the qualitative analysis, we find that a

variety of factors may help to explain this low degree of interaction among candidates in the CET–SET group discussion. These may involve the candidates' interpretation of the discussion task as an assessment event rather than communicative interaction with other members, lack of confidence in their ability to handle the discussion task as communicative interaction, lack of interest on the part of the candidates in the discussion topics and a tendency to produce long turns in the discussion. Thus, the inadequate elicitation of ILFs from the candidates may well pose a problem for measuring their speaking ability in terms of the ability to engage in communicative interaction.

Evidence from the present study seems to suggest that conversational features do not appear in speaking tests just because we introduce speaking partners with equal social power. The research team members would like to know whether it happens in other tests that now use the pair or group format, such as those provided by UCLES or the National Educational Examination Authority in China (NEEA). Another issue that the present study raises and which needs further investigation is the grouping of the candidates. The research team members are interested in knowing whether and to what extent grouping influences candidates' performance. The issue of grouping cannot be investigated in the present study because the grouping was done by a computer program and was beyond the investigators. One other possible avenue of further research is the investigation of topics. The result of the present study shows a low degree of interaction among candidates in the CET–SET group discussion in general. The research team members will take a closer look at the data to see if there are any task specific trends, hoping that the findings will give the test designers some idea of the topics that are engaging and able to elicit more ILFs included in the test syllabus.

Acknowledgements

We would like to acknowledge with thanks the invaluable input we received from the anonymous Language Testing reviewers, whose questions, comments and suggestions helped to make this a better article. Our thanks also go to the editors of *Language Testing* for their constant support and encouragement, and to Lyle Bachman and the LASSLAB group at UCLA for their questions and comments on the present study. We remain solely responsible for the remaining errors in this article. This article was supported by an MOE project of

the Center for Linguistics and Applied Linguistics at Guangdong University of Foreign Studies.

VII References

- Alderson, J.C.** 1985: Innovations in language testing? In Portal, M., editor, *Innovations in language testing*. Slough: NFER-Nelson, 93–105.
- Bachman, L.F.** 1990: *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. and Palmer, A.S.** 1996: *Language testing in practice*. Oxford: Oxford University Press.
- Berkoff, N.A.** 1985: Testing oral proficiency: a new approach. In Lee, Y.P., editor, *New directions in language testing*. Oxford: Pergamon Institute of English, 93–100.
- Berry, V.** 2000: An investigation into how individual differences in personality affect the complexity of language test tasks. Unpublished PhD thesis, King's College, University of London.
- Bonk, W.J. and Ockey, G.J.** 2003: A many-facet Rasch analysis of the second language group oral discussion task. *Language Testing* 20, 89–110.
- Egbert, M.M.** 1998: Miscommunication in language proficiency interviews of first-year German students: a comparison with natural conversation. In Young, R. and He, A.W., editors, *Talking and testing: discourse approaches to the assessment of oral proficiency*. Amsterdam and Philadelphia: John Benjamins, 147–69.
- Elder, C., Iwashita, N. and McNamara, T.** 2002: Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer? *Language Testing* 19, 347–68.
- Folland, D. and Robertson, D.** 1976: Towards objectivity in group oral testing. *The Modern Language Journal* 30, 156–67.
- Foot, M.C.** 1999: Relaxing in pairs. *English Language Teaching Journal* 53, 36–41.
- Fulcher, G.** 1996: Testing tasks: issues in task design and the group oral. *Language Testing* 13, 23–51.
- Kormos, J.** 1999: Simulating conversations in oral proficiency assessment: a conversation analysis of role-plays and non-scripted interviews in language exams. *Language Testing* 16, 163–88.
- Kramsch, C.** 1986: From language proficiency to interactional competence. *The Modern Language Journal* 70, 366–72.
- Lazaraton, A.** 1992: The structural organization of a language interview: a conversation analytic perspective. *System* 20, 373–86.
- 1996: Interlocutor support in oral proficiency interviews: the case of CASE. *Language Testing* 13, 151–72.
- Malvern, D. and Richards, B.** 2002: Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing* 19, 85–104.

- McNamara, T.F., Hill, K. and May, L.** 2002: Discourse and assessment. *Annual Review of Applied Linguistics* 22, 221–43.
- Messick, S.A.** 1980: Test validity and the ethics of assessment. *American Psychologist* 35, 1012–27.
- 1989: Validity. In Linn, R.L., editor, *Educational measurement*. 3rd edition. New York: American Council on Education/Macmillan.
- Ministry of Education** 1999: *College English Teaching Syllabus*. Shanghai: Shanghai Foreign Language Education Press.
- National College English Testing Committee** 1999: *CET–SET Syllabus*. Shanghai: Shanghai Foreign Language Education Press.
- O’Sullivan, B.** 2002: Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing* 19, 277–95.
- Richards, J.C., Platt, J. and Platt, H.** 1992: *Longman dictionary of language teaching and applied linguistics*. London: Longman.
- Ross, S.** 1992: Accommodative questions in oral proficiency interviews. *Language Testing* 9, 173–86.
- Ross, S. and Berwick, R.** 1992: The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition* 14, 159–76.
- Shohamy, E.** 1991: Discourse analysis in language testing. *Annual Review of Applied Linguistics* 11, 115–31.
- Spence-Brown, R.** 2001: The eye of the beholder: authenticity in an embedded assessment task. *Language Testing* 18, 463–81.
- Swain, M.** 2001: Examining dialogue: another approach to content specifications and to validating inferences drawn from test scores. *Language Testing* 18, 275–302.
- van Lier, L.** 1989: Reeling, writhing, drawling, stretching, and fainting in coils: oral proficiency interviews as conversation. *TESOL Quarterly* 23, 489–508.
- Yang, H.Z.** 1999: CET–SET design principles. *Wai Yu Jie* 75, 48–57.
- Young, R.** 1995: Conversational styles in language proficiency interviews. *Language Learning* 45, 3–42.
- 2002: Discourse approaches to oral language assessment. *Annual Review of Applied Linguistics* 22, 243–62.
- Young, R. and Milanovic, M.** 1992: Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition* 14, 403–24.

Appendix 1 CET–SET Sample test (examiners’ material)

Topic A–1

Topic Area: City Life

Topic: City Traffic

Part 1 (5 minutes)

Examiner:

Good morning (Good afternoon), everybody. Could you please tell me your name and the number of your admission ticket? Your name, please. And your number? ... Your name? ... And your number? ... Thank you.

Now would you introduce yourselves briefly to each other, please? (1.5 minutes) [Interrupt him/her if he/she talks for more than 30 seconds: OK, [C?], please.]

OK, now that we know each other we can do some group work. The topic area for our group work is 'City Life'. I should remind you that Chinese should not be spoken during the test. First of all, I'd like you to say something about your life in the city.

[C1 or C2 or C3] [Start preferably with the one who performs the best in the introduction part]

- 1) How do you like living in Beijing (Shanghai, Nanjing ...)?
- 2) What do you think is the most serious challenge of living in a city like Beijing (Shanghai, Nanjing ...)?
- 3) How do you like shopping in a supermarket?
- 4) Where would you like to live, downtown or in the suburbs, and why?
- 5) What measures do you think we should take to reduce air pollution in Beijing (Shanghai, Nanjing ...)?
- 6) Can you say something about the entertainment available in your city?
- 7) Where would you like to find a job after graduation, in a big city like Beijing or Shanghai or in a small town and why?
- 8) What's your impression of the people in Beijing (Shanghai, Nanjing ...)?/How do you like the people in Beijing (Shanghai, Nanjing ...)?

Part 2 (10 minutes)

Examiner:

Now let's move on to something more specific. I'd like you to talk about city traffic. You'll have a picture (some pictures) showing two different types of transport. I'd like each of you to give a brief description of each type and then compare the two types. You'll have one minute to prepare and each of you will have one and a half minutes to talk about the picture(s). Don't worry if I interrupt you at the end of the time limit.

[1 minute later. Start preferably with the one who performs best in the first part.]

Now, [C1], would you please start first? [C2] and [C3], would you please put your pictures aside and listen attentively to what [C1] has to say?

[1.5 minutes later] Thanks. [C2], now it's your turn.

[1.5 minutes later] Thank you very much. OK, [C3], now it's your turn please.

Right. Now we all have some idea of the various kinds of city transport. I'd like you to discuss this topic further and see if you can agree on which is the best type of transport for a big city like Beijing (Shanghai, Nanjing ...). During the discussion you may argue with each other or ask each other questions to clarify a point. You will have about four and a half minutes for the discussion. Your performance will be judged according to your contribution to the discussion. Don't worry about the time. I'll remind you when time runs out.

[If one candidate keeps talking for too long]

Sorry, I'll have to stop you. Let's listen to what [C?] has to say.

[If one candidate keeps silent for a long time] OR:

[If the group keep silent for some time, choose one of the three candidates to start the discussion.]

Now, [C?], could you please say something about your view of ... ?

[4.5 minutes later]

That's the end of the discussion. Thank you.

Part 3 (5 minutes)

Examiner:

Now I'd like to ask each of you just one last question on the topic of 'City Traffic'.

[Select any of the following questions as appropriate to ask each of the three candidates. More time might be given to whoever fails to give a satisfactory performance in the presentation and discussion part.]

[C1 or C2 or C3]

- 1) During the discussion, why did you say that ... ?
- 2) What kind of transport do you usually use in your city?
- 3) Do you have any suggestions as to how traffic conditions can be improved in big cities?
- 4) Do you think private cars should be encouraged?
- 5) Why do you think some Western countries encourage people to ride bicycles?

That's the end of the test. Thank you, everybody.

Appendix 2 CET–SET Rating Scale

	Accuracy and range	Size and discourse management	Flexibility and appropriacy
5	<ul style="list-style-type: none"> • Basically correct use of grammatical and lexical items. • Adequate vocabulary and a fair range of grammatical structures for the given task. • Fairly good pronunciation though some residual accent is acceptable. 	<ul style="list-style-type: none"> • The ability to produce extended and fairly coherent discourse concerning the given task, though with occasional pauses due to loss of words. 	<ul style="list-style-type: none"> • Natural and active participation in the discussion. • Use of language generally appropriate to context, function and intention.
4	<ul style="list-style-type: none"> • Errors in the use of grammatical/lexical items that do not 	<ul style="list-style-type: none"> • Manifestation of ability to produce coherent and more 	<ul style="list-style-type: none"> • Frequent contribution to the discussion but sometimes not to

	seriously interfere with communication are permissible.	complex utterances, though most contributions are short.	the point or without directly interacting with other participants.
	<ul style="list-style-type: none"> • A basically satisfactory range of vocabulary to deal with the given task. • Acceptable pronunciation. 	<ul style="list-style-type: none"> • Frequent pauses while organizing thoughts and searching for words, which sometimes interfere with communication. 	<ul style="list-style-type: none"> • Use of language basically appropriate to context, function and intention.
3	<ul style="list-style-type: none"> • The use of grammatical/lexical items may be incorrect and sometimes impede communication. • A minimum range of vocabulary and grammatical structures to cope with the given task. • Pronunciation may be faulty and sometimes impede communication. 	<ul style="list-style-type: none"> • Mainly short utterances. • Long and frequent pauses while organizing thoughts and searching for words, which often interfere with communication, though basically fulfilling the given task. 	<ul style="list-style-type: none"> • Less active participation in the discussion, and occasional inability to adapt to new topics or changes of direction.
2	<ul style="list-style-type: none"> • Unintelligibility caused by grammatical/lexical errors. • Insufficient grammatical/lexical items to cope with the given task. • Poor pronunciation that causes breakdowns in communication. 	<ul style="list-style-type: none"> • Short utterances and disconnected speech, which is difficult to follow, making communication almost impossible. 	<ul style="list-style-type: none"> • Inability to take part in group discussion.

Appendix 3 The eight group discussion topics of the November 2001 administration of CET-SET

A. Which aspect is most important for university students to stay healthy?	Morning, Day 1
B. Is graduate education essential for a successful career?	Afternoon, Day 1
C. The ideal size of a family	Morning, Day 2
D. Which festivals are most popular with young people in China?	Afternoon, Day 2
E. What can be done to educate people to conduct themselves properly?	Morning, Day 3

F. Is it desirable to live in a big city?	Afternoon, Day 3
G. Should people be encouraged to keep pets?	Morning, Day 4
H. Which is the best means of communication?	Afternoon, Day 4

Appendix 4 Questionnaire

- I. Please tick the topic of your group discussion:
- Which aspect is most important for university students to stay healthy?
 - Is graduate education essential for a successful career?
 - The ideal size of a family
 - Which festivals are most popular with young people in China?
 - What can be done to educate people to conduct themselves properly?
 - Is it desirable to live in a big city?
 - Should people be encouraged to keep pets?
 - Which is the best means of communication?
- II. Please complete the following by placing a circle around the most appropriate answer.
- During the group discussion, you felt
 - very nervous
 - quite nervous
 - not nervous
 - relaxed
 - You found the group discussion
 - very difficult
 - quite difficult
 - not difficult
 - easy

Specify if you chose A or B: (multiple answers OK)

 - Failing to catch the topic for discussion
 - Not familiar with the topic for discussion
 - Not knowing the words related to the topic for discussion
 - Failing to understand what the other candidates were saying
 - Others, please specify: _____
 - You found the group discussion topic
 - very interesting
 - quite interesting
 - not interesting
 - dull
 - You found the time for the group discussion
 - too long
 - appropriate
 - too short
 - During the group discussion, your target audience were
 - the examiners
 - the other candidates
 - neither the examiners nor other candidates

6. During the group discussion, when other candidates were talking, you
 - A. listened attentively so as to react accordingly
 - B. thought about what you were going to say when your turn came
 - C. Others, please specify: _____
7. During the group discussion, when you failed to understand what other candidates had said, you
 - A. asked for clarification or explanation
 - B. ignored it
 - C. pretended to understand it and went on to express your own views
 - D. Others, please specify: _____
8. During the discussion, you
 - A. tried to reach an agreement with the other candidates on the given topic
 - B. tried to express your opinions clearly
 - C. Others, please specify: _____