

A Validation Study of the ECCE NNS and NS Examiners' Conversational Styles from a Discourse Analytic Perspective

Yang Lu
University of Reading

This study explores the conversational styles of the Examination for the Certificate of Competency in English (ECCE) native and nonnative speaker (NS and NNS) examiners when responding to candidates' replies and eliciting questions and justifications and their effect on the assessment of the test takers' oral proficiency. A discourse analytic approach following the systematic and functional tradition was implemented to analyze twenty audiotaped ECCE speaking test events. The findings show that certain conversational styles such as informing, commenting, back-channeling, interrupting, and so forth, act as non-eliciting discourse elements in facilitating sufficient oral samples from candidates. Differences between the amount and types of discourse features produced by the NS and NNS examiners seem to suggest a variability caused by examiners' linguistic and cultural backgrounds.

Face-to-face and multitask oral proficiency tests have been widely implemented in EFL oral assessment development. The interviewers, as they have been called in the traditional and still prevalent form of oral proficiency interviews (OPI), are now playing not only the role of examiner who conducts the test and asks questions but also the role of interlocutor who interacts with the examinees to assist them in completing the tasks. Consequently, the spoken discourse resulting from these exchanges has also been changed from an examiner-dominated nature to a co-constructive nature of discourse jointly shaped and developed by both examiners and test takers.

One of the possible sources of construct-irrelevant variance in speaking tests is examiners' misrepresentation of the test developer's intended constructs due to personal and sometimes unconscious discursal styles during the spoken interaction with test takers. Some speaking-test developers have used structured face-to-face oral tests and interlocutor frameworks to minimize the negative effect on the validity of their speaking tests. Nevertheless, it has been indicated that even preformulated and scripted speaking tests have to accommodate the examiner's deviation from the interlocutor framework through conscious or subconscious individual styles that change the way in which test takers are examined (Lazaraton, 1996a; O'Loughlin, 1997; O'Sullivan & Lu, 2004). This issue becomes particularly crucial when a speaking test recruits both NS and NNS examiners who may bring their own linguistic and cultural backgrounds to their double role of assessor and interlocutor.

The ECCE Speaking Test is a structured but not scripted face-to-face oral examination targeted at the Independent User level on the Common European Framework Scale. The three tasks in the test are designed to represent different interaction patterns and discourse styles in order to assess candidates' competence to convey and elicit information and to support their decisions. The examiners have to take not only a dominant role as interviewer in Task 1 but also a passive role as information provider in Task 2 to enable the test taker to make a

decision. Though Task 3 returns the role of initiator to the examiner, the role is not as overbearing as it is in Task 1, as the goal in Task 3 is to encourage candidates to elaborate on the reasons for their decisions.

Preferably, if the examiners faithfully carry out the three tasks and ensure that their own conversational styles and discourse behaviors do not alter the conditions under which the candidates perform and are examined, the test can successfully assess what it is designed to assess, thereby functioning as a valid and reliable test. However, is that the case? Apart from the factors brought by the test takers themselves that may affect performance, have the ECCE examiners, native or nonnative, strictly followed the ECCE Oral Examiner's Manual instructions and conducted the test unvaryingly so that the candidates can be assessed equally? Furthermore, have they guaranteed that the test takers have supplied them with sufficient samples of spoken language to allow them to make fair judgments? If they have not succeeded in doing so, is it their divergent discourse features or conversational styles that have prevented the test takers from performing to the best of their possible ability? Have the NS and NNS examiners varied in this aspect? These are the concerns and inquiries investigated by this study.

Background

The reliability and construct validity of oral assessment have been thorny issues for the language testing community. On the one hand, as Luoma (2004) summarizes, quantitative approaches such as correlation coefficients and standard error measurements (SEM) have been widely applied by testing boards to improve the estimates of test scores, so that stakeholders' confidence in the test can be maintained. On the other hand, as Lazaraton (2002) observes, only since the last decade have process-based or discourse-based studies on oral language assessment been attempted to examine the nature of the speech event and its quality in relation to the validity and reliability of oral assessment.

This qualitative and empirical approach was first called for by Van Lier (1989) to investigate the "turn-by-turn sequential interaction" (p. 497) so that the practice of designing the procedures and rating scales of OPIs could be evaluated. This need, as Fulcher (1987) had remarked on earlier, could be found in discourse analysis, a then-new approach to construct validation by which the construct can be empirically tested. Notably, systematically transcribed speaking tests have since been used to scrutinize spoken discourse using, according to He and Young (1998), mainly three approaches: Conversation Analysis (CA), Ethnography of Speaking and Speech Acts, and Gricean Pragmatics.

Extensive discourse studies have been conducted to investigate the interviewer's and the interviewee's behavior in OPIs regarding test validity, task effect, and the effect of the interlocutor on the candidates' rating. The first question researchers asked was if interviewers conversed in OPIs in ways similar to natural conversation. Young and Milanovic's (1992) study was one of the first to investigate this question. The study analyzed features of dominance, contingency, and goal orientation, as well as contextual factors in the data and suggested that the discourse was highly asymmetrical, which constrained both the interviewers and test takers in terms of what they could contribute in the oral interaction. These styles remained stable over time even in structured and scripted oral proficiency tests (Brown & Lumley, 1997; Lazaraton, 1992, 1996a; Reed & Halleck, 1997).

On the subject of the impact of interviewers' discourse styles on examinees' ratings, Ross and Berwick's research (1992) investigated whether the interviewer's control and accommodations in OPIs affected ratings and the degree of such effects. Their findings are that test taker's ratings could be predicted from the amount and types of accommodation that interviewers have to make. Subsequent studies have expanded the scope of investigation from looking at the interviewer's discourse to looking also at that of the interviewee's. Brown and Hill (1998) analyzed the co-constructed discourse in the IELTS Speaking Test based on the results of FACETS analysis in terms of the "easy" or "difficult" interviewer. They revealed that the easiest interlocutor shifted topics more frequently, asked simpler questions, and engaged in more question-answer exchanges, while the most difficult interlocutor challenged candidates more and acted more like a natural conversation partner. In a subsequent study, Brown (2003) applied CA to examine the impact of two different interviewers' discursal styles—"teacherly" and "casual" (p. 17)—on the same candidate's performance. Raters were employed to comment on the candidate's oral production resulting from the two interviews. The results show that the test taker is judged an effective communicator when taking the test with the teacherly interlocutor who, among other things, developed and extended topics skillfully. With the casual interviewer, who used more nontest conversational eliciting strategies, the candidate was judged as unforthcoming or uncooperative in communication.

As a result, to minimize the stable but unpredictable individual interviewer styles, test developers became interested in the application of an interlocutor frame to guide and constrain oral examiners from changing the ways that test takers are assessed. A series of studies by Cambridge ESOL (Lazaraton, 1996a, 1996b; Lazaraton & Saville, 1994) about the effect of interlocutor frames (or test scripts) have shown that deviation from interlocutor frames is frequent. The results of FACETS analysis in these studies indicate that this problem affects reliability of the ratings. Since oral examiners who also have to act as interlocutors cannot be considered a neutral factor, a choice between the face validity and reliability of the OPI procedure has to be made.

Along the same lines, O'Sullivan and Lu (2004) analyzed 30 seconds of pre- and post-deviation oral production by examinees in 62 audiotaped IELTS Speaking Test events. They identified the four most frequent deviations from the interlocutor frame: paraphrasing questions, interrupting with questions, asking improvised questions, and commenting after test takers' replies. They found that there was a task factor in terms of frequency of deviation. But, because deviations were not frequent in the data, there were no systematic changes between the pre- and post-deviation spoken samples regarding accuracy, complexity, and fluency except *expanding* in the discourse, one of the three specific features of prolonging. Therefore, O'Sullivan and Lu suggested that the interlocutor frame can be flexible with deviations, such as paraphrasing questions, if the nature of the question is abstract or cognitively challenging.

With regard to the differences between the NNS and NS examiner discursal performance and their impact on rating and test takers' discursal performance, there have been few studies. The difference in rater harshness between the two groups seems to have been the most interesting area for previous research (Brown, 1995; Fayer & Karshiski, 1987; Sheorey, 1986; Van Meale, 1994 as quoted in Reed & Cohen, 2001). To date, the present study can only draw valuable insights from the study by Berwick & Ross (1996), which analyzed the discourse of six Japanese as second language (JSL) interviews and six English as second language (ESL) interviews conducted by two trained male examiners. The Japanese

JSL examiner and the American ESL examiner varied systematically in terms of approaches in the spoken discourse with the test takers. Statistical analyses based on the 12 interviews revealed that the JSL examiners offered significantly more accommodation such as display questions, overarticulation, lexical simplification, and more control in terms of topic shift. In contrast, the ESL examiner responded more to the content and gave the test takers more chances to elaborate on the topics. Therefore, Berwick and Ross suggested that there is “a degree of cultural/pragmatic relativity” (p. 48) in the OPI procedure and called for further research with larger amounts of data to verify this phenomenon.

Though discourse-based study is a recently established direction for research in oral assessment, discursal performance has been a subject for a substantial amount of investigation. Carroll claimed in 1980 that the expert speaker can initiate, expand, and develop a theme, while modest speakers lack flexibility and initiative, and marginal speakers rarely take initiative and maintain dialogue in a rather passive manner. How and to what extent examinees of different levels of oral proficiency perform distinctively has also been especially important to testing organizations that are concerned with the inclusiveness and efficiency of their rating scales (see Hughes, 1989; Weir & Bygate, 1990). Later studies have suggested that when performing the same language elicitation task, higher level test takers may be more likely to produce more complicated discourse features such as initiating, elaborating, supporting, challenging, speculating, and developing topics than low- or lower-level candidates (Hasselgren, 1997; Lazaraton, 2002; Lazaraton & Wagner, 1996; Shohamy, 1994; Young, 1995), which parallels the findings of spoken discourse analysis of learner’s speech by the systematic and functional approach (Hoey, 1991; McCarthy & Carter, 1994).

To conclude, studies about how examiners as both assessor and interlocutor initiate and manage the discourse in face-to-face oral proficiency tests have resulted in implementation of structured or scripted face-to-face oral tests with, sometimes, an interlocutor frame. It seems that these frames have constrained but not totally succeeded in restricting examiners from using their individual discourse styles, cultural-specific or not. Further research has suggested that interlocutor frames can be made flexible to allow space for examiners to adjust to examinees with different ability levels and cognitive maturity. Elaborating, taking initiative, and so forth, in spoken discourse have been recognized as salient features or indicators for high-level oral communicative language ability, and some tests have been trying to differentiate these features in order to assess accurately and fairly. However, how the interlocutor examiners respond after test takers’ replies, and their questions to elicit demonstration of discourse competencies, have not been fully investigated. Furthermore, to validate Berwick and Ross’s proposal of a “cultural/pragmatic relativity” in oral assessment (1996), further research with more data from live EFL speaking tests with several types of tasks, conducted by not just one NS or NNS examiner, is needed to investigate this validity issue.

Aims of the Study

In a structured but unscripted direct oral exam—the ECCE Speaking Test—discourse analysis was carried out to examine the oral interaction between the examiners and examinees

1. to see if there are overall differences between the amount and types of the eliciting and non-eliciting moves in discourse produced by the NNS and NS examiners;

2. to identify the non-eliciting specific discourse features in the examiners' follow-up moves that do not encourage the examinees to elaborate or prolong their replies, decisions, or choices;
3. to identify the non-eliciting specific discourse features that do not encourage initiation from the examinees to seek information; and
4. to find out if there are differences between the NNS and NS examiners in the amount and types of non-eliciting discourse features that do not encourage the examinees' elaboration and initiative.

Methodology

ECCE Speaking Test

The speaking section is an integral part of the Examination for the Certificate of Competency in English (ECCE), produced by the English Language Institute, University of Michigan (ELI-UM). Its purpose is to assess the candidates' basic operational competence in giving and asking for information, and justifying decisions and choices, and so forth. According to the *ECCE Oral Examiner's Manual* (English Language Institute, 2004), the ability to elaborate and to take initiative are salient features of discursal performance.

In Task 1, the competence of conveying nonsensitive personal information is assessed, as examiners are required to use a variety of questions (open and closed) to elicit speech from candidates. Task 2 is for eliciting initiations in order to assess the ability to ask for information to make a decision or give a suggestion based on a prompt that presents a situation and the candidate's task. Pictures or photographs are provided to illustrate the task. Task 3 continues the topic in Task 2, and examiners are instructed to encourage examinees to elaborate the reasons for their decision or suggestion. To obtain more oral samples so that ratings can be as accurate as possible, examiners are also provided elaboration questions to prolong the spoken interaction.

Though the examiners are not provided with scripts, the ELI-UM gives fairly detailed guidelines for conducting the three tasks and specific Dos and Don'ts for how to behave and speak in the oral interaction (English Language Institute, 2004). These guidelines and instructions will be presented in the Discourse Analysis (DA) section because of their importance to the DA approach of this study. The examiner's manual also offers "Descriptors of Salient Features" and a section that explains the indicators for the salient features. A checklist for decision-making on the five criteria—fluidity of delivery, elaboration and initiative, vocabulary, grammar, and intelligibility—is also given to guide the examiners. The manual explains that fluidity of delivery, elaboration and initiative, and vocabulary have proved to be the best indicators to distinguish levels on the test. Overall ratings are Competent Speaker, Moderately Competent Speaker, Marginal Speaker, and Limited Speaker. A candidate rated Limited Speaker fails the speaking test.

Data

Twenty ECCE live audiotaped speaking tests administered in May and June 2004 were provided by the ELI-UM. Nine were administered by NNS examiners and 11 by NS examiners. Because one examiner from each group failed to tape Task 1, the data consist of 18 recorded Task 1s for analysis. The analysis of Tasks 2 and 3 by one of the NS examiners is

not included because the examiner's repeated effort to enable the candidate to understand the tasks failed. As a result, the data have 19 recorded Task 2s and 3s for analysis. The NNS examiners are numbered NNS1 to NNS9, while the NS examiners numbers are from NS01 to NS11. Information about the examiners was provided by the ELI-UM regarding nationality, age, native language, other language, length of time as ECCE oral examiner, and the training that they had been provided. The examinees' ratings were also provided, which shows that 13 out of 21 were given the rating of Competent Speaker, one was judged as Marginal Speaker, with the rest of them as Moderately Competent Speakers.

Because Tasks 2 and 3 of the test are based on one of the eight prompts provided by the ELI-UM, I requested tests that used the same prompt so that content would not be a source of irrelevant variation in the examiners' and test takers' spoken discourse. As a result, tests that used the same prompt, which required the candidates to make a decision, support that decision and elaborate their reasons, were provided for the study.

Transcripts

Transcribing the live tests is orthographical as long as it can reflect the discourse sequence and consequence of the spoken interaction between the examiner and candidate. Therefore, length of pausing, stressed syllables, loudness of speech, and overlapping were not transcribed. The following are the speech features depicted in the transcripts and the conventions employed when necessary:

1. Filled pauses are transcribed as "er" or "um."
2. A question mark is put at the end of a statement with a rising tone that functions as a question in the discourse.
3. A comma after a word or phrase shows an unfilled pause with either rising or falling tone.
4. A full stop is put at the end of a completed sentence with a pause.
5. A circumflex ^ after "okay," "yeah," or "yes" shows a rising tone, while a backslash \ denotes falling tone.
6. One x expresses one syllable of an untranscribable word.
7. Xs are substituted for language other than English, depending on the number of syllables heard.
8. Nonverbal discourse features such as laughing are put in parentheses.

Discourse Analysis

A task-specific model of the systematic and functional approach developed for investigating the discourse of the Oral Proficiency Test (OPT), which usually consists of several tasks and assigns the examiner as interlocutor (Lu, 2003) was employed for the research. The underlying principles for developing this DA model are different from those of CA, as stated by Lazaraton (2002), in the following two aspects: (1) CA insists on unmotivated looking rather than pre-stated research questions, while the task-oriented approach is prescriptive and has specified framework for tasks that elicit different discourse patterns; and (2) CA insists on employing the "turn" as the unit of analysis, while the task-oriented approach takes "exchange" as the unit of analysis to reflect the chaining together of functions.

In contrast, the DA approach is prescriptive and selective by nature because it targets the initiating and sustaining discourse features used by the test takers as indicators for

communicative language ability. The unit for analysis in this approach is the Topic Exchange and its subsequent levels, Move and Act, as proposed by Sinclair and Coulthard (1975) and modified by Burton (1981), Coulthard and Montgomery (1981), Francis and Huston (1992), and Hoey (1991). This approach also borrows categories of moves from Eggins and Slade (1997) and refers to studies by Hoey (1991) and McCarthy and Carter (1994) for differences between native-speaker discourse and learner discourse. Specifically, the DA for Task 1 and part of Task 3 is to examine the extended structure of a topic exchange [I (R/I) R (Fⁿ)], which shows that a topic exchange can be longer than the basic Initiation-Response-Feedback structure, and consists of a Response treated as an Initiation, then a Response, then a Follow-up, and maybe more Follow-up moves. But, because the approach is task-specific, chaining of adjacency pairs with the second move treated as Initiation is also examined for Tasks 2 and 3.

As is shown in Figure 1, the task-specific DA approach integrates three fundamental factors in oral assessment: task-specific discourse features, expected discourse features, and targeted features indicating high or higher-level proficiency. Models based on the basic principle for analyzing interactive and monologic discourse are established and applied to live video- or audiotaped speaking test events. Because ECCE speaking tests do not have a monologue task, the interactive model by Lu (2003) is presented in Figure 2 (next page).

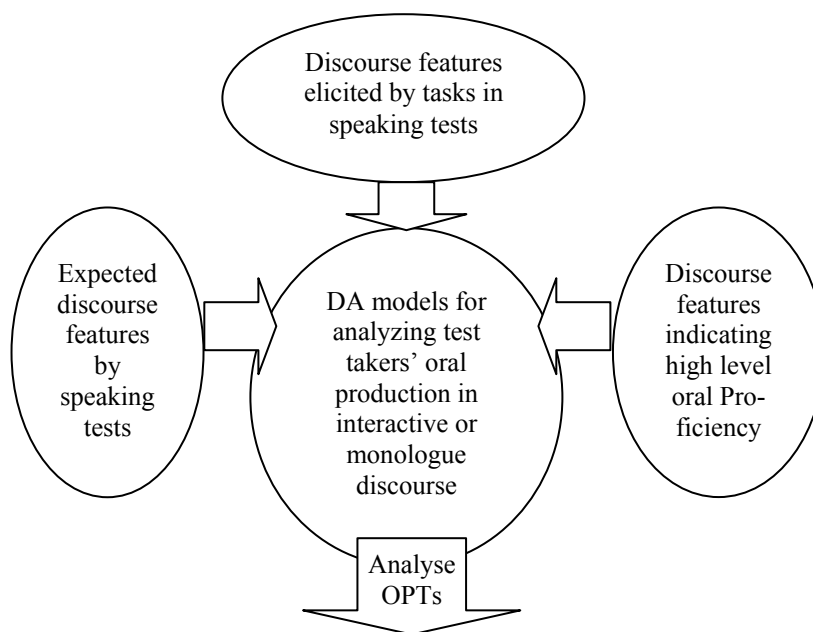


Figure 1. Elements Integrated in DA Models for Analyzing OPT Discourse.

As Lu (2005) explains, this is an overall model for analyzing an exchange in the interactive discourse in OPTs. Specific models are formulated based on the overall model, depending on task type, discourse construct, and expected oral output by the test takers. Therefore, a specific model for analyzing a particular task is sometimes based on only part of the overall model. For example, a specific model for analyzing test taker response in an interview discourse, as in Task 1 of the ECCE Speaking Test, will only adopt Prolonging

under Sustaining to see if examinees have responded to show their oral proficiency level is high.

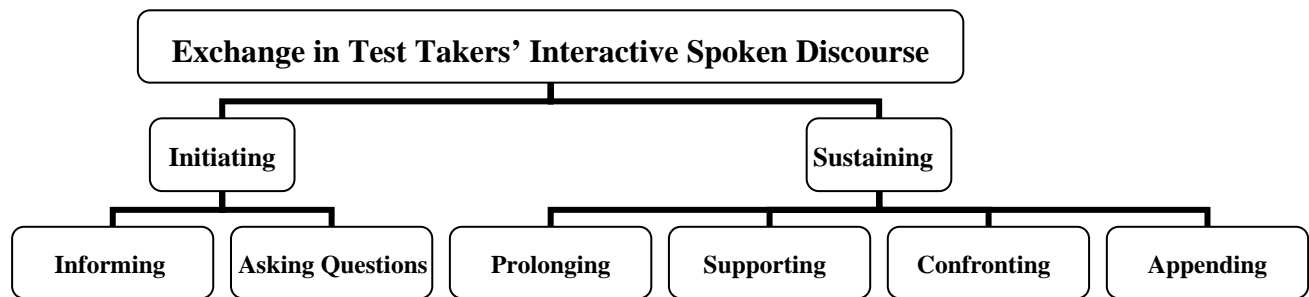


Figure 2. Overall Model for Analyzing Interactive Discourse in OPT.

Applied to the present study, two essential modifications were made for developing specific models to suit the research objectives: (1) Although test takers' discourse is not the focus of DA, it is the indicator for the effect of the examiner's discourse and regarded as the starting point for analyzing the examiner's previous and subsequent turn. Therefore, if an examiner is expected to elicit prolonged speech in a follow-up move after a candidate's answer to her/his question in Task 1, the candidate's turn after the follow-up move will be looked at, rather than the examiner's, before a decision is made as to the examiner's discourse function in terms of being eliciting or non-eliciting. (2) The general guidelines and the Dos and Don'ts that represent the testing organization's requirements and expectations of the examiners in order to implement valid tests are integrated in the models for analyzing the three tasks. Therefore, discourse behavior such as interrupting, correcting mistakes, and so forth are regarded as non-eliciting discourse features, since they divert from the expectations of the ELI-UM.

Based on the two elementary modifications, the task-specific DA models for analyzing the spoken discourse in the ECCE were developed with the general guidelines and the Dos and Don'ts from the examiner's manual related to the individual task incorporated. Examples from the data and explanations are given when necessary.

Task 1

For this task, ECCE Speaking Test examiners are instructed to elicit talk and longer responses from the candidates. Specific Dos and Don'ts related to this task are:

- a. elicit longer responses by asking questions that establish context followed by requests for more specific information;
- b. follow up on the examinees' replies;
- c. foster coherence and continuity by using content provided by the candidates; and
- d. respond naturally to what the examinees say by using utterances such as "Uh-huh," "Yes," and "Oh, I see."

As a result, DA for this task is to first find the divisions of the topic exchanges, then to locate the follow-up moves made by examiners after candidates' replies that elicit or do not

elicit elaboration of the topic by candidates. The locating process will carry on until a new topic is raised by the examiners. The specific DA model for Task 1 is shown below in Table 1.

Obviously, the question-answer adjacency pair in spoken discourse is excluded for this study, as there is no third turn to initiate more talk or longer responses from the candidates. Furthermore, the topic exchange may terminate at Turn 4 and the analysis will conclude accordingly. Therefore, Turns 5 and 6 are not the compulsory parts of an exchange for analysis. However, there may be more than six turns in the exchange dealing with the same topic and consequently included for the analysis. Example 1 illustrates how DA is conducted for analyzing oral interaction in Task 1.

Table 1. DA Model for Analyzing Examiner's Eliciting or Non-Eliciting Move in Task 1

Turn	Speaker	Discourse Feature by Examiner and Test Taker	Discourse Analysis
1	Examiner	Opening : eliciting	Initiating a topic exchange.
2	Candidate	Responding : replying	Identified.
3	Examiner	Follow-up move	Eliciting or non-eliciting candidate's prolonging move? Specific discourse features are analyzed.
4	Candidate	Responding	Prolonging or not prolonging? (providing evidence for Turn 3)
5	Examiner	Same as in Turn 3	Same as for Turn 3.
6	Candidate	Same as in Turn 4	Same as for Turn 4, but providing evidence for Turn 5.

Example 1:

Examiner: Uh-huh, very interesting, so you, you put like information about the cartoon on the website?

Candidate: I, yes, I I I have got er, the series the whole series on our computer, download it, and put it in the cds. [reply]

Examiner: Uh-huh [eliciting follow-up move: engaging]

Candidate: like er, 80 cds. [prolonging: elaborating]

Examiner: wow. It's a lot of cds.(Laugh)[non-eliciting follow-up move: commenting]

Candidate: yeah [no prolonging]

From NS03 & TT03

Task 2

For Task 2, the test developer requires the examiner to change roles, from the interviewer to a comparatively passive information supplier about the pictures or situation in the prompt. They are reminded that the examinees should take an active role in this task. The model for analyzing Task 2 is presented in Table 2.

Table 2. DA Model for Examiner’s Eliciting or Non-Eliciting Responding Move in Task 2

Turn	Speaker	Discourse Feature by Examiner and Test Taker	Discourse Analysis
1	Candidate	Opening : eliciting	Identifying a topic exchange.
2	Examiner	Responding : answering	Treated as initiation? Eliciting or non-eliciting the next question? Specific discourse features are analyzed.
3	Candidate	Opening : eliciting	Identified.
4	Examiner	Responding : answering	Same as in Turn 2.
5	Candidate	Opening : eliciting	Identified.
6	Examiner	Responding : answering	Same as for Turn 4.

Although the table demonstrates the chaining of only three pairs of question-answer adjacency pairs, the analysis should end when all the question cues provided to the candidates have been asked, with, if necessary, extra questions. Example 2 illustrates how DA is conducted for analyzing oral interaction in Task 2.

Example 2:

Candidate: where they live? [eliciting]

Examiner: well, the leopards are living in Africa. And the pandas, they’re in china [eliciting responding move: unelaborated answer]

Candidate: how many are left? [eliciting]

Examiner: there’s about 20000 leopards left today, and the pandas, there’s only like 1000 pandas left. So there’re not very many [eliciting responding move: unelaborated answer]

Candidate: how many can we save this year? [eliciting]

From NS03 & TT03

Task 3

The *ECCE Oral Examiner’s Manual* states that the expected language functions from the candidates for Task 3 are to express a choice, preference, or opinion and support it/them. As a result, the examiners are instructed to encourage the candidate to elaborate the reasons for the decision, choice, and so forth, and also to encourage the candidate to discuss why something was not chosen (termed as non-choice in this study).

The discourse pattern for which examiners complete this task is shown in Table 3. The eliciting or non-eliciting moves in the framework for elaboration of the candidate’s choice or non-choice are put together in one turn due to their similarity in nature and function. In reality, they should be independent in separate exchanges as shown in Example 3. It is evident that the follow-up move by the examiner in this task is ideally treated as an initiation to trigger the next elaboration on choice or non-choice.

Table 3. DA Model for Analyzing Examiner’s Eliciting or Non-Eliciting Moves for Elaboration of Choice or Non-Choice in Task 3

Turn	Speaker	Discourse Feature by Examiner and Test Taker	Discourse Analysis
1	Examiner	Opening	Identifying a topic exchange.
2	Candidate	Responding	Elaborating or not elaborating choice or non-choice? Providing evidence for Turn 1.
3	Examiner	Follow-up move	Treated as Initiation? Eliciting or non-eliciting elaboration of choice or non-choice? Specific discourse features are analyzed.

Example 3:

Candidate: but I this case I chose pandas because its more difficult to save.
 Examiner: yeah [eliciting move for elaboration of choice]
 Candidate: and you have to do er something um quickly [elaboration of choice]
 Examiner: uh-huh, uh-huh [eliciting move for elaboration of choice]
 Candidate: because it’s more, its, its um low, er high, low, low and its more low to er to improve their lives, so er [elaboration of choice]
 Examiner: okay. That’s okay so let’s help pandas. Okay? [non-eliciting elaboration of non-choice]
 Candidate: uh-huh [no elaboration]
 Examiner: thank you very much [non-eliciting elaboration of non-choice]

From NNS8 & TT8

Because there are also elaboration questions in the prompts provided by the ELI-UM to elicit more spoken samples from candidates in order to assist rating, part of Task 3 will have the same discourse pattern found in Task 1, where a question-answer sequence exists. This is to say that the eliciting and non-eliciting discourse features by the examiners in Task 1 should also be present here for analysis. As a result, the specific model for examining this part of the task is the same as for Task 1.

Tagging the Transcripts

Auto Text in Word was employed for the DA of the data, and the tagging process was repeated twice to ensure consistency of the analysis. In the initial tagging, the eliciting or non-eliciting discourse features were identified, and specific moves under the two broad categories were roughly tagged as they appeared in the transcripts. For the second tagging, focus was on the specific moves of the non-eliciting discourse feature. The classification was completed and exact analysis was given. For the final tagging, the previous analysis was checked and some specific moves with very low occurrences were abandoned.

Counting the Occurrences

An Excel file was used for the frequency counts. First, the analyzed discourse features were tallied for each type in each task for each examiner involved in the study. Then, totals and averages of each type of the eliciting and non-eliciting features were calculated for the

NNS Group and for the NS Group. Finally, comparisons of the totals and averages were made for the two groups.

Results

Non-Eliciting Discourse Features Employed by Both NS and NNS Examiners

DA resulted in the following list in answer to two of the objectives of the research: (1) to identify the specific discourse features in the examiners' follow-up moves that do not elicit the examinees' elaboration or prolonging of replies and decisions, choices, and so forth; and (2) to identify the specific discourse features that do not elicit initiation from the examinees to seek information.

A brief explanation for each of the listed features is given in reference to the specific requirements and conditions of the tasks. Examples from the transcripts are also provided when necessary to illustrate. The underlined utterances or turns are the examples that illustrate the discourse features in discussion.

Agreeing: follow-up moves that show the examiners' sharing the same opinion, idea, feelings, and so forth that had been given by the candidate. They differ from utterances such as "Yes," "True," "That's right," etc., in that they express explicitly the agreement without functioning as engaging the candidate in the spoken conversation (see Example 4 below).

Example 4:

Examiner: are, are, are there any pandas in Brazil?

Candidate: I don't think so.

Examiner: I don't think there are, either

Candidate: uh.

From NS03 & TT03

Answering questions: does not refer to what the examiners are required to do in Task 2—answering the candidates' questions in order to impart the information requested by the examinee. They are only included for analyzing Task 3, where the examiners are usually in the role of asking questions and eliciting expanded speech from the test takers.

Asking questions: employed by examiners while administering Task 2 when they are supposed to answer questions raised by the candidates. They can be seen as diverting from the assigned role of the examiners.

Back-channeling: follow-up moves which are repetitions of whole or part of candidates' turns with a falling intonation, as shown by Example 5. The discursal behavior is an echoing of what the candidate has just said in reply to the examiners' initiation for a new topic.

Example 5:

Candidate: I'm a marketing researcher

Examiner: marketing researcher

Candidate: yes

From NS10 & TT010

Challenging: as Burton (1981) defines it, “challenging moves function to hold up the progress of that topic or topic-introduction in some way” (p. 71). Its occurrence is present only in Task 3 when the examiners are responding to candidates’ justification for their choice. It can be realized by a statement or question (see Example 6).

Example 6:

Candidate: because er, they seem to be very cute, but I think that in this moment the leopards needs more help than pandas, because 20000 are left

Examiner: yeah.

Candidate: er

Examiner: 20000 left, pandas only 1000

Candidate: yeah.

From NNS1 & TT1

Changing topic: a task-specific discourse feature which is an opening move that initiates another topic exchange in spoken discourse. It is realized by a statement or a question. Because in Task 3 the examiners are instructed to encourage the test takers to tell why something was not chosen, this discourse feature is singled out and analyzed specifically for this study. It can also be regarded as misrepresenting the construct of the task in that examiners should not start asking the elaboration questions before they have tackled the reason or justification for not choosing something. Example 7 illustrates how the examiners employed the discourse feature and consequently terminated a necessary phase in administering Task 3.

Example 7:

Candidate: so I think we’re going to help the pandas.

Examiner: wow, that’s some good decision, um although both are in need but your organization can only help one kind, right? Um, tell me about this, have you ever seen leopards or pandas in real life?

From NNS2 & TT2

Clarification requests: follow-up moves that indicate non-understanding or lack of comprehension after candidates’ replies or answers. They are usually realized by either questions or repetitions with a rising tone of the part not understood or whole of the previous turn made by a candidate (see Example 8).

Example 8:

Candidate: I think er, people should er, think about it and er, er make er lots of er effort to help them, or so on to an organization.

Examiner: you mean people, er common people ^

Candidate: yes, of course.

From NNS3 & TT3

Commenting: follow-up moves, statements, or tag questions made by examiners to elaborate, expand, justify, evaluate, and so on, in responding to candidates’ replies to their initiations. Since the nature of the spoken interaction in a speaking test is different from that

in a natural conversation, the candidates tend to be more sensitive to what the examiners comment on regarding what they say. The discourse analysis in this study includes words and phrases such as “good,” “interesting,” “nice,” and so on. Examples 9 and 10 exemplify cases.

Example 9:

Candidate: its, its, my my father can, can pay, it's a very, very good school, so I try to, to use all I can use there you know, because I don't feel that in school, to see my father paying what he's paying, I just go there to, go there go
Examiner: good, er, I'm sure your father is very happy to hear that.
Candidate: (laugh)

From NS04 & TT04

Example 10:

Examiner: are you working?
Candidate: er, no, I am a mother, I'm married and a mother of two children.
Examiner: that's nice.
Candidate: yeah, I know.

From NNS5 & TT5

Concluding: follow-up moves of statements or questions marked with “so” or “then” at the beginning that function to summarize what has been talked about between the examiner and test taker on a topic. Example 11 shows this discourse feature, which usually demands a response from the candidate.

Example 11:

Examiner: but how exciting you got to visit London, where else did you say?
Candidate: France, just one week, Italy, one week, and last year I was a au pair for a year in New York, it was wonderful.
Examiner: so you had good experiences.
Candidate: yea, I think.

From NS06 & TT06

Confirmation requests: different from clarification requests in that, as a follow-up move, they demand affirmation of what the examiner has understood but is not certain of. They are usually realized by either “yes,” “no,” or a word or phrase with a high key rising intonation (see Example 12).

Example 12:

Examiner: okay, have you ever seen pandas before?
Candidate: no, no
Examiner: yes?
Candidate: never, no.

From NNS4 & TT4

Correcting mistakes: follow-up moves in which the examiners correct grammatical or lexical mistakes instead of carrying on with the normal flow of the spoken interaction, which is one of the Don'ts given by the ELI-UM.

Engaging: after candidates' replies to the initial elicitation and in the position of follow-up move, they are realized by "Uh-huh," "yeah," or "okay" with a mid-key rising tone to acknowledge or show attention to what has been said by the candidates without interrupting or stopping their utterances.

Exclamations: another discourse feature to indicate acceptance and interest in the test taker's talk, but realized by utterances such as "wow," "ah," "oh," "really?," or laughter, which show surprise, amazement, disbelief, amusement, etc.

Informing: follow-up move of a statement made solely to provide information new to candidates (see Example 13).

Example 13:

Examiner: good, um, there just some general questions about animals, um, do you enjoy going to the st. pauval zoo? Have you been there?

Candidate: yes, I have already been there er, twice with my daughters of course.

Examiner: I just went last week.

Candidate yes.

From NS01 & TT01

Interrupting: this discourse behavior terminates candidates' responses and replies to the examiners' initiation. It is one of the diverting discourse behaviors that the ELI-UM advises the examiners not to do.

Marker: follow-up move that is realized by "okay," "right," "alright," etc., at the beginning of a turn with a falling intonation. The consequence of employing such a discourse feature intentionally or unintentionally is usually the termination of a topic exchange, as illustrated in Example 14.

Example 14:

Examiner: yes, OK, and how do you believe that English er, has changed your life and work?

Candidate: yeah, it's very important, I think the English is very important, you know, because er, you know businesses needs English, it's the er business language

Examiner: okay\

Candidate: and er, I don't know

From NNS5 & TT5

Supplying elaborated answers: another task-specific discourse feature that acts as a responding move in Task 2 after candidates' initiation for information needed to make a decision. An answer is regarded as elaborated if more than the required information is provided, which results in a non-eliciting turn to the examinee's next initiation when it is still needed (see Example 15).

Example 15:

Candidate: um, which is more easy to protect, protect?

Examiner: well, okay, the, the leopards there are 20000 in Africa, and the hunters are killing them for their furs. The pandas are living in forests their habitat is being invaded because people are cutting down the trees, right, they eat bamboo.

Candidate: yes.

From NS07 & TT07

Supplying unelaborated answers: in contrast to supplying elaborated answers, examiners are considered to be complying with the guidelines and representing the construct designed in Task 2 when they provide needed information only (see Example 16).

Example 16:

Examiner: um, hun, we would, er which is cheaper to help? To protect, to help to protect.

Candidate: okay, you have enough money to protect 200 leopards or 10 pandas.

From NS07 & TT07

Supplying vocabulary: similar by nature to the diverting discourse behavior of correcting mistakes, this unit of discourse analysis interrupts candidates' turns and causes disruption in the spoken discourse. As a follow-up move it is usually a word or phrase unknown to the test taker, as shown by Example 17.

Example 17:

Examiner: er, apart from these leopards and pandas, have you ever heard about other animals that are recently in danger? Think about other

Candidate: yes,

Examiner: species.

Candidate: yes, the wise people want to kill, wise, wise

Examiner: the whales, yeah.

Candidate: the whales, yes.

From NS07 & TT07

Comparing Overall Results between the NS and NNS Examiner Groups

This section reports the results regarding: (1) whether the eliciting and non-eliciting discursal features by the ECCE Speaking Test examiners are the same or different for the NNS and NS examiners, and (2) whether the amount and types of discourse features that do not elicit the examinees' elaboration and initiative by the NNS and NS examiners are the same or different. First, a comparison of the amount of eliciting and non-eliciting discourse features by both the NNS and NS examiners is presented in Table 4 to show the examiners' overall discursal performance in conducting the speaking test.

Table 4 shows that in general the ECCE Speaking Test examiners elicited significantly more in the discourse for Tasks 1 and 2 as compared to Task 3, when they were supposed to encourage the examinees to elaborate the reason for their choice and non-choice. As a whole,

the examiners produced considerably more eliciting and non-eliciting features in their follow-up moves in respect to promoting elaboration replies in Tasks 1 and 3.

The table also reveals that the NS examiners produced more eliciting moves, particularly in the cases of initiating elaboration on replies in Tasks 1 and 3 and getting examinees to ask questions in Task 2, while the NNS examiners produced more non-eliciting moves. However, the NNS examiners' discursal performance in Task 3, when trying to encourage elaboration on reasons for choice or non-choice, showed no significant difference from those of the NS group.

Table 4. Eliciting and Non-Eliciting Discourse Features in Tasks 1, 2, & 3 by NNS and NS Examiners

Task	Discourse Feature	No. of Occurrences		
		Total	NNS	NS
1	Eliciting Follow-up Move	166	79	87
	Non-Eliciting Follow-up Move	88	60	28
2	Eliciting Responding Move	88	35	53
	Non-Eliciting Responding Move	38	22	16
3	Eliciting Moves for Elaboration of Choice	26	12	14
	Non-Eliciting Moves for Elaboration of Choice	13	6	6
	Eliciting Moves for Elaboration of Non-Choice	7	3	4
	Non-Eliciting Moves for Elaboration of Non-Choice	23	11	12
	Eliciting Follow-up Move	106	44	62
	Non-Eliciting Follow-up Move	50	32	18

Table 5 (on the next page) provides the top three types of eliciting moves in each task and the number of occurrences produced by the NS and NNS examiners. As shown in the table, there is not much difference between the NNS and NS examiners for the amount and types of specific discourse features by which they elicited elaboration and initiative from the candidates. Differences such as the NS examiners' tendency to use commenting in their follow-up moves after test takers' replies and confirming to elicit questions in Task 2 seem to suggest some characteristics in the NS examiners' discourse. The results also give evidence that engaging and exclamation as follow-up moves are most effective in eliciting elaboration.

Table 5. Eliciting Follow-up Moves by NNS and NS Examiners in Task 1

Task	Type of Discourse Features	No. of Occurrences	
		NNS	NS
1	Eliciting Follow-up Moves:		
	• Engaging	50	55
	• Exclamation	13	15
	• Commenting	3	9
	• Back-channeling	3	2
	• Acknowledging	3	0
	• Informing	3	0
2	Eliciting Responding Moves:		
	• Unelaborated answer	22	26
	• Confirming	6	13
	• Elaborated answer	4	1
	• Non-Informing answer	0	3
3	Eliciting Moves for Elaboration of Choice:		
	• Engaging	8	4
	• Asking question	4	5
	• Confirmation request	0	2
	Eliciting Moves for Elaboration of Non-Choice:		
	• Challenging	1	2
	• Asking question	1	1
	• Marker	0	1
	• Prompt	1	0
	Eliciting Follow-up Moves:		
	• Engaging	30	37
• Exclamation	5	6	
• Agreeing	2	0	
• Commenting	1	4	

Comparing the Non-Eliciting Discourse Features by NNS and NS Examiners

Differences between the amount and types of the non-eliciting discourse features produced in each task by NNS and NS examiners are presented in this section. These differences are shown by the numbers and percentages of all occurrences of each type.

Table 6 shows that the NNS examiners' non-eliciting follow-up moves are double those produced by the NS examiners. The two groups discouraged the test takers to prolong their replies and elaborate on the topic being dealt with in different ways as well. The NNS examiners are more likely to do so by back-channeling, requesting confirmation, and informing, and marking boundaries in the discourse, while the NS examiners tend to do so by commenting and concluding the candidates' replies.

Table 6. Non-Eliciting Follow-up Moves by NNS and NS Examiners in Task 1

Types of Non-eliciting Follow-up Moves: Task 1	No. of Occurrences		% of All Occurrences	
	NNS	NS	NNS	NS
Agreeing	1	0	1.67	--
Back-channeling	8	1	13.33	3.57
Challenging	3	1	5.00	3.57
Clarification request	1	2	1.67	7.14
Commenting	6	4	10.00	14.29
Concluding	6	4	10.00	14.29
Confirmation request	6	1	10.00	3.57
Correcting mistake	0	1	--	3.57
Engaging	1	2	1.67	7.14
Exclamation	6	4	10.00	14.29
Informing	9	3	15.00	10.71
Interrupting	4	2	6.67	7.14
Marker	7	3	11.67	10.71
Supplying vocabulary	2	0	3.33	--
Totals/Average per examiner	60/7.5	28/2.8		

Table 7 shows that, though the examiners were engaged in a similar discourse context as in Task 1, especially when they asked the elaboration questions, some of the non-eliciting follow-up moves made by them in Task 1, such as back-channeling, challenging, confirmation requests, engaging, and interrupting, did not take place in Task 3. Also, the NNS examiners were eliciting less in this part of Task 3, and they employed more discourse features to do so, as well. They shared the same discursual behaviors such as concluding, informing and commenting with the native speakers, but they came up with more managing discourse features again such as markers and confirmation requests and supplying vocabulary, while the NS examiners still tended to comment, conclude, and agree with the examinees.

Table 7. Non-Eliciting Follow-up Moves by NS and NNS Examiners in Task 3

Types of Non-Eliciting Follow-up Moves: Task 3	No. of Occurrence		% of All Occurrences	
	NNS	NS	NNS	NS
Agreeing	0	3	--	16.67
Answering question	3	0	9.38	--
Clarification request	6	0	18.75	--
Commenting	2	2	6.25	11.11
Concluding	4	5	12.50	27.78
Correcting mistake	1	0	3.13	--
Elaborated answer	2	0	6.25	--
Exclamation	1	1	3.13	5.56
Informing	5	5	15.63	27.78
Marker	2	0	6.25	--
Supplying vocabulary	4	1	12.50	5.56
Totals/Average per examiner	31/3.44	18/1.8		

It is obvious that, while conducting Task 2 (see Table 8), if the examiners asked questions, gave information without having been asked, or supplied answers with elaboration, the examinees would not have as many opportunities as they could to encourage the examiners to provide the information needed to make a decision. The NNS examiners' commenting and informing and the NS examiners' concluding seem to have stopped the test takers from initiating in the discourse. Furthermore, the average number of non-eliciting moves per examiner for the NNS group decreased.

Table 8. Non-Eliciting Responding Moves by NS and NNS Examiners in Task 2

Types of Non-Eliciting Responding Moves: Task 2	No. of Occurrences		% of All Occurrences	
	NNS	NS	NNS	NS
Asking question	4	4	18.18	25.00
Clarification request	2	0	9.10	--
Commenting	4	1	18.18	6.25
Concluding	0	4	--	25.00
Confirmation request	1	0	4.55	--
Engaging	1	1	4.55	6.25
Elaborated answer	4	5	22.73	31.25
Informing	5	1	21.74	6.25
Marker	1	0	4.55	--
Totals/Average per examiner	22/2.76	16/1.6		

From the totals and averages per examiner for Task 3 (Table 9), we can see that the NNS examiners were for the first time not producing more non-eliciting moves than their NS counterparts. Concluding seems to have stopped candidates most effectively from elaborating their reasons for choosing something by both groups. Noticeably, it is still the NS group that used commenting for this effect.

Table 9. Non-Eliciting Elaboration of Choice Moves by NS and NNS Examiners in Task 3

Types of Non-Eliciting Moves for Elaboration of Choice: Task 3	No. of Occurrences		% of All Occurrences	
	NNS	NS	NNS	NS
Changing topic	1	1	16.67	16.67
Commenting	0	2	--	33.33
Concluding	3	2	50.00	33.33
Confirmation request	0	1	--	16.67
Engaging	1	0	16.67	--
Informing	1	0	16.67	--
Marker	0	1	--	16.67
Totals/Averages per examiner	6/0.67	7/0.7		

Interestingly, the amount of non-eliciting moves for elaboration of non-choice increased compared to the ones for elaborating choice. Table 10 shows that changing the topic by both groups and agreeing with the examinees by the NS examiners are the main causes. The task-specific discourse feature, Changing Topic, means that the examiner terminated the discourse for eliciting candidate's justification for not choosing something, limiting the effectiveness of the task. The total number of occurrences for this discourse feature is eight, which implies that eight out of nineteen examinees were not assessed for this test feature.

Table 10. Non-Eliciting Elaboration of Non-Choice Moves by NS and NNS Examiners in Task 3

Types of Non-Eliciting Moves for Elaboration of Non-Choice: Task 3	No. of Occurrences		% of All Occurrences	
	NNS	NS	NNS	NS
Agreeing	0	6	--	50.00
Back-channeling	0	1	--	8.33
Changing topic	5	3	45.45	25.00
Challenging	3	0	27.27	--
Concluding	1	0	9.10	--
Commenting	0	1	--	8.33
Interrupting	2	1	18.18	8.33
Totals/Average per examiner	12/1.22	11/1.1		

Discussion

The Effect of Discourse Variation on Oral Examiners' Discoursal Performance

In general, the NNS examiners are less facilitative of the examinees' elaboration in replying to their initiation in discourse and initiative in seeking information. The amount of eliciting discourse features that they used is sometimes half of that used by the NS examiners, while the amount of the non-eliciting ones made by the NNS examiners is often twice that made by their NS counterparts. Nevertheless, this situation changed by different degrees when they were engaged in different discourse contexts generated by Tasks 2 and 3. There was a decrease in the number of non-eliciting moves by both NNS and NS examiners, especially in Task 3. As a result, there was not a substantial difference between the NNS and NS figures. This may indicate an effect of discourse variation on the examiners' discoursal performance. Examiners tend to discourage the candidates from prolonging their replies and expanding on the topic initiated by them, but the examiners did not do this so much when they were eliciting initiatives and elaboration on the choice and non-choice.

If the effect of discourse variation is justifiable, the possibility that the NNS examiners are less eliciting in discourse needs to be reconsidered. They may only be unable to elicit elaborated replies successfully. In other discourse patterns where their roles are different, they may be more ready to facilitate elaboration and initiative. Therefore, NNS examiners of oral tests seem to be in need of training or standardization for the question-answer or interview discourse pattern. However, it might also be the consequence of the NNS examiners' strong sense of goal-orientation in the discourse. They were more focused on the completion of the

tasks, and therefore, tended to shift topics more frequently and prevent examinees from necessary elaboration on the topics.

The Effect of the Diverting Discourse Features by NNS and NS Examiners

A number of the task-specific discourse features included in the specific DA models of this research are considered deviations from the general guidelines and task design by the ELI-UM. They are interrupting, correcting mistakes, supplying vocabulary, asking questions, and supplying elaborated answers in Task 2, answering questions in Task 3, and changing topic to elicit elaboration of non-choice in Task 3. These occurrences again show that training and standardizing examiners is a challenging undertaking to test developers of speaking tests.

The fact that all these diverting features, except elaborating answers, did not appear in the eliciting discourse features shows that they are genuine deviations that have negative effects. The deviations can be seen as frequent (nine elaborated answers in Task 2, nine interrupting moves in Tasks 1 and 3, and six supplying vocabulary moves by the examiners involved), since this research involves 19 examiners. And most importantly, these incidents of deviation are not idiosyncratic. They scatter over the examiners in the above-mentioned parts of the exam, which may suggest again the unpredictability of examiners' behavior in conducting oral tests, so that training and coaching the examiners to follow the test structure, task procedure, and guidelines for discursial performance are vital before administration in order to prevent them from misrepresenting the construct and affecting the validity of the test.

The Effect of the NNS Examiners' Management Agenda and the Discoursally More Involved NS Examiners

The results show that, apart from the diverting discourse behaviors, both groups also tended to use certain discourse features that stopped the test takers from elaborating or initiating. However, the most frequent discourse features produced by the two groups vary. In Figure 3, comparisons of the average numbers per examiner of the most frequent discourse features in each group are presented to reveal the tendencies.

The NNS group produced significantly more back-channeling, clarification and confirmation requests (C. requests in Figure 3), informing and markers. Although those discourse elements, apart from back-channeling, are all follow-up moves in an exchange originally, they become initiations after spoken and demand a response from the other party in the conversation. According to Hoey (1991), they can be regarded as the follow-up moves that are treated as initiation to represent the disruption in spoken discourse. Obviously, they terminate the previous topic, which the examinees may have intended to elaborate. For example, to the clarification request "Phone a friend. Do you mean I can phone a friend to help?" the other party has to give a reply "Yes" or "No" to comply with the rule for a natural conversation, which cannot be considered a sufficient language sample for an oral examiner to use to make a rating.

**Comparison of the Most Frequent Non-eliciting Moves by NNS and NS Examiners
(Diverting features not included)**

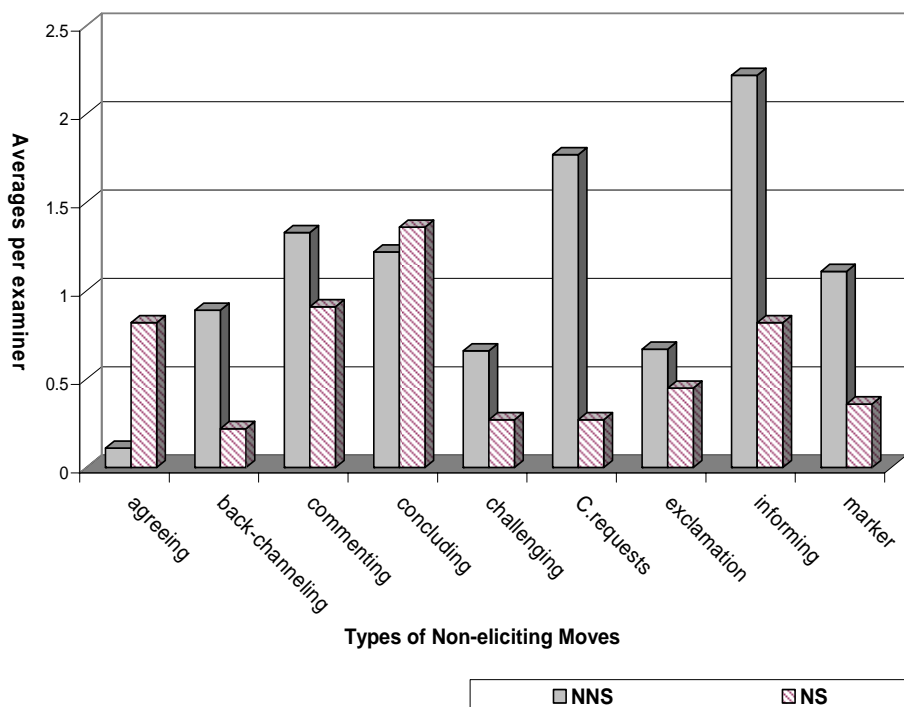


Figure 3. Comparison of the Most Frequent Non-Eliciting Moves by NNS and NS Examiners.

In contrast, the NS group’s most frequent discourse features that function as non-eliciting moves do not necessarily demand a response but may require a follow-up move to acknowledge or accept. This is self-evident because, to someone who has just expressed agreement, we are not obliged to respond. We can express appreciation verbally or use paralinguistic features to complete the episode of the oral communication. The same is also true with commenting and concluding, which the NS group produced comparatively more often than they did the other features. These features seem to have acted as the genuine follow-up moves in response to the examinees’ replies which show interest in what the examinee had said. They suggest a higher degree of involvement in the discourse by the NS examiners, though they operated as non-eliciting to the candidates’ expected discorsal performance. In fact, the non-eliciting moves as shown in Figure 3 all appeared as eliciting moves. However, only the NS examiners’ commenting was one of the top three that effectively initiated test takers’ prolonged speech.

The specific non-eliciting discourse features produced more by the NNS examiners, namely challenging, C. requests, informing, and markers, function not only as initiating moves as discussed earlier, but also operate as what Bygate (1987) describes as “agenda management” (p. 36) speaking skill features. They deal with starting, maintaining, directing, or ending a topic, which correspond to the NNS examiners’ marker (starting or ending), informing (starting or maintaining), challenging (directing or maintaining), and C. requests (directing or maintaining). As a result, it seems that the NNS examiners attempted to take

more control in the oral interaction than the NS examiners. These efforts seem to have partly operated negatively against the expected test taker's discursal performance. This finding matches the results by Berwick and Ross (1996).

Conclusion and Implications

This research has applied a discourse analytic approach to investigate the non-eliciting effect of the ECCE oral examiners' discourse behavior on the construct of the test and the expected examinee discursal performance. It was found that, on the one hand, in general the ECCE Speaking Test examiners, regardless their linguistic and cultural backgrounds, have followed the test-developer's guidelines and presented the instrument to the test takers in order to assess the targeted discursal performance: elaboration and initiative in spoken interaction. On the other hand, there were deviations from the requirements by the ELI-UM and the task requirements by both NNS and NS examiners. As a result, Tasks 2 and 3 could have been conducted more effectively to assess the ability to take initiative and support decisions. In comparison, the NNS examiners performed less eliciting behavior and more non-eliciting behavior than the NS examiners. Their non-eliciting discursal performance varies with the discourse variation borne by the task types. Therefore, their discursal performance and adherence to the examiners' guidelines are similar to that of their NS counterparts in initiating elaboration of choices, but not in other discourse contexts, such as being an information-provider or in the follow-up move after examinees' replies to their initiation. Furthermore, there did seem to be a cultural/pragmatic relativity caused by preferences of the specific discourse features by the NNS and NS examiner, the effect of which may be non-eliciting. It was noted that the NNS examiners in this study tend to take control and be goal-oriented in the follow-up move, thus depriving the test takers of chances to elaborate, while the NS examiners seemed to have been more involved in the oral interaction, paying attention to the content of what was being said by the candidates.

The findings of this research may imply that the institutional nature of face-to-face OPTs with role-based activities cannot be neglected. Though examiners sometimes subconsciously have the tendency of treating the oral interaction as natural conversation, the goals of OPTs determine that the spoken discourse involved is limited in terms of naturalness, interactiveness, and range of discourse behavior that can be appropriate for the intended constructs. This seems to indicate that interlocutor frames could be necessary for standardization of oral examiners' task and discursal performance. It is possible that the NNS examiners' OPT conversational styles are influenced by their first languages and cultures. However, the results indicate that it is the NNS examiners' inclination to control discourse that had the effect of not initiating the expected examinee discursal performance.

In summary, this study is a small-scale investigation with data randomly selected which might not be representative of the examiners or candidature. Therefore, any conclusion and implications drawn from the study have to be considered cautiously, and further study with more data and a wider range of participants will be needed to confirm the results of the present research to make generalizations about the effect of examiners' non-eliciting discourse behavior on the reliability and validity of oral assessment.

Acknowledgments

I would like to express my deeply felt gratitude to the Spaan Fellowship Committee for accepting my research proposal to make this study a reality. I would also like to thank Ms. Mary Spaan's practical guidance on the project and the efficiency of all involved staff at the ELI-UM.

References

- Berwick, R., & Ross, S. (1996). Cross-cultural pragmatics in oral proficiency interview strategies. In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition, and assessment: Selected papers from the 15th language testing research colloquium, Cambridge and Arnhem* (pp. 34–54). Cambridge, UK: Cambridge University Press.
- Brown, A. (1995). The effect of rater variables in the development of an occupation specific language performance test. *Language Testing*, 12(1), 1–15.
- Brown, A. (1998). *Interviewer style and candidate performance in the IELTS oral interview*. In S. Woods (Ed.), *Research Reports 1997, Vol. 1* (pp. 173–191). Sydney, Australia: ELICOS.
- Brown, A., & Hill, K. (1998). Interviewer style and candidate performance in the IELTS oral interview. *IELTS Research Reports 1998, 1*, 1–19.
- Brown, A., & Lumley, T. (1997). Interviewer variability in specific-purpose language performance tests. In A. Huhta, V. Kohonen, L. Kurki-Suonio, & S. Luoma (Eds.), *Current developments and alternatives in language assessment* (pp. 137–150). Jyväskylä, Finland: University of Jyväskylä.
- Burton, D. (1981). Analyzing spoken discourse. In M. Coulthard & M. Montgomery (Eds.), *Studies in discourse analysis* (pp. 61–83). London: Routledge.
- Bygate, M. (1987). *Speaking*. Oxford, UK: Oxford University Press.
- Carroll, B. J. (1980). *Testing communicative performance*. Oxford, UK: Pergamon.
- Coulthard, J., & Montgomery, M. (1981). *Studies in discourse analysis*. London: Routledge.
- Eggins, S., & Slade, D. (1997). *Analyzing casual conversation*. London: Cassell.
- English Language Institute. (2004). *ECCE oral examiner's manual, May-June 2004*. Ann Arbor, MI: English Language Institute, University of Michigan.
- Fayer, J. M., & Karshiski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37, 313–326.
- Francis, G., & Huston, S. (1992). Analyzing everyday conversation. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 123–161). London: Routledge.
- Fulcher, G. (1987). Tests of oral performance: The need for data-based criteria. *ELT Journal*, 41, 287–291.
- Hasselgren, A. (1997). Oral test subskill scores: What they tell us about raters and pupils. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC, 96* (pp. 226–241). Jyväskylä, Finland: University of Jyväskylä.
- 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. 1–26). Philadelphia: John Benjamins.

- Hoey, M. (1991). Some properties of spoken discourse. In R. Bowers & D. Brumfit (Eds.), *Applied linguistics and English language teaching* (pp. 65–84). Basingstoke, UK: Macmillan.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.
- Lazaraton, A. (1992). The structural organization of a language interview: A conversation analytic perspective. *System*, 20, 373–386.
- Lazaraton, A. (1996a). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13(2), 151–172.
- Lazaraton, A. (1996b). A qualitative approach to monitoring examiner conduct in the Cambridge assessment of spoken English (CASE). In M. Milanovic & N. Saville (Eds.), *Performance testing, cognition, and assessment: Selected papers from the 15th Language Testing Research Colloquium, Cambridge and Arnhem* (pp. 18–33). Cambridge, UK: Cambridge University Press.
- Lazaraton, A. (2002). *A qualitative approach to the validation of oral language tests*. Cambridge, UK: Cambridge University Press.
- Lazaraton, A., & Saville, N. (1994, February). *Process and outcomes in oral assessment*. Paper presented at the 16th Language Testing Research Colloquium, Washington, DC.
- Lazaraton, A., & Wagner, S. (1996). *The revised test of spoken English (TSE): Analysis of native speaker and nonnative speaker data*. (TOEFL Monograph Series MS-7). Princeton, NJ: Education Testing Service.
- Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.
- Lu, Y. (2003, July). *Test-takers' first languages and their discursal performance in oral proficiency tests*. Paper presented at the 25th Language Testing Research Colloquium, Reading, UK.
- Lu, Y. (2005). *A discourse analytic validation study of discourse competence in oral proficiency tests*. Unpublished doctoral dissertation, University of Reading, Reading, UK.
- McCarthy, M., & Carter, R. (1994). *Language as discourse: Perspectives for language teaching*. London: Longman.
- O'Loughlin, K. (1997). *The comparability of direct and semi-direct speaking tests: A case study*. Unpublished doctoral dissertation, University of Melbourne, Melbourne, Australia.
- O'Sullivan, B., & Lu, Y. (2004). *An empirical study on examiner deviation from the set interlocutor frame in the IELTS speaking paper*. Unpublished manuscript.
- Reed, D., & Cohen, A. (2001). Revisiting raters and ratings in oral language assessment. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 82–96). Cambridge, UK: Cambridge University Press.
- Reed, D.J., & Halleck, G.B. (1997). Probing above the ceiling in oral interviews: What's up there? In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.), *Current developments and alternatives in language assessment: Proceedings of LTRC 96* (pp. 225–238). Jyväskylä, Finland: University of Jyväskylä.
- Ross, S., & Berwick, R. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 159–176.

- Sheorey, R. (1985, April). *Goof gravity in ESL: Native vs. nonnative perceptions*. Paper presented at the 19th Annual TESOL Convention, New York.
- Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11*, 99–123.
- Sinclair, J., & Coulthard, M. (1975). *Toward an analysis of discourse*. Oxford, UK: Oxford University Press.
- Sinclair, J., & Coulthard, M. (1992). Toward an analysis of discourse. In M. Coulthard (Ed.), *Advances in spoken discourse analysis* (pp. 1–34). London: Routledge.
- Van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly, 23*, 487–508.
- Young, R. (1995). Conversational styles in language proficiency interviews. *Language Learning, 45*, 3–42.
- Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition, 14*, 403–424.

