DO DIFFERENT SPEECH INTERACTIONS IN AN ORAL PROFICIENCY TEST YIELD DIFFERENT KINDS OF LANGUAGE?

PAVLOS PAVLOU

ABSTRACT

The aim of this paper is to provide evidence for or against the hypothesis that different speech interactions in a test battery of oral proficiency elicit different kinds of language (vocabulary, grammatical structures etc.). The objective of this study has been carried out through the analysis of speech samples taken from the COAST (Cyprus Oral Academic Skills Test). The test battery consisted of four different speech interactions; a group discussion, an oral report, two role plays and an oral proficiency interview (OPI). The four speech interactions are described, analyzed and compared to each other in order to detect the lines along which they differ. These structural differences are expected to result into the differentiated linguistic make up of the four speech interactions.

INTRODUCTION

In the testing of oral proficiency in a second/foreign language the OPI has been the most common elicitation technique. This has been recognized, among many scholars in the field, by Madsen (1980) and Underhill (1987). This fact can be attributed to the advantages of the OPI which, among others, are: a) the easiness of comparing the performance among various candidates since there is a predetermined set of questions, either written or in the mind of the examiner and b) the ability of the examiner to assist the candidate (through probing) to demonstrate his/her higher level of proficiency since the examiner can control the content of the interaction. The OPI has also many disadvantages, the major one being the fact that it is falsely thought to be representative of the whole spectrum of oral interaction (Shohamy et al., 1986; Raffaldini, 1988).

This disadvantage was the main reason of including four different speech interactions in the COAST. This decision was taken based on the following reasons. First, the four different speech interactions represent the oral trait much better than the OPI alone does, and therefore the picture of the candidates’ oral proficiency is more accurate. In addition to that, the inclusion of four different speech interactions enhances the fairness of the test.
since different students believe that different speech interactions reflect their true knowledge in speaking English differently (Nevo & Shohamy, 1986) and therefore each candidate should be given the chance to be tested in an interaction in which s/he feels comfortable. In the administration of the COAST the students had to complete all the tasks.

METHODOLOGY AND DATA COLLECTION

In order to test the above mentioned hypothesis the COAST was developed (see appendix 1). It consisted of a group discussion, an oral report, two role plays (one aiming at formal and one aiming at informal register) and an OPI. The COAST was administered to 60 high school students in Cyprus by the same examiner. The students were informed that the test was for research purposes. In order to avoid a sequencing effect the students took the four tests in different orders. The samples from the sixty students were audio- and videotaped and then transcribed. After that they were rated by three different raters. The rating scales used for the evaluation of the samples were the ones developed by Bachman and Palmer (1983). The scales which were sensitive to various aspects of communicative competence (linguistic, pragmatic and sociolinguistic) were slightly modified.

STRUCTURAL DIFFERENCES AMONG THE FOUR INTERACTIONS

Before looking at the results of the study we should discuss a fundamental question, namely, "Why should the performance of the subjects vary across these four speech interactions?" The claim is that these four speech interactions are different. But the crucial question is "HOW are they different?" The answer to this question is the ultimate guideline to what different aspects of language use we should expect the subject to exhibit in the four different speech interactions.

In order to discover the actual differences between the four speech interactions, we need to describe them abstractly, i.e. independently of possible content or specific topic addressed in a real test situation. The content and the topic are not inherent components of these interactions, rather they are additional definitions that need to be specified in a concrete test situation. In other words, we need to describe, for example, what "discussion" means rather than what "discussion on this topic with these persons and goal in mind", and try to develop a typology of spoken "genres".

The remainder of this section is devoted to the description of the similarities and differences among the four speech interactions. Not all possible comparisons among the four interactions are furnished because the
following discussions show that some interactions (e.g. oral interview and role play) are more similar than different. Where evidence for this is provided, one of the two interactions (functioning as a representative of both interactions) will be compared with a third interaction. The role play and oral interview have been collapsed only for juxtaposition with the group discussion because of the number of interlocutors, and not, for example, because of the similarity of the status of the interlocutors in the two tasks.

The aspects of discourse that will be compared across the four interactions are the number of interlocutors in each interaction and the status of the interlocutors. In a multi-participant interaction the relative status of the participants may have an impact on the control of the interaction something that is also going to be looked at in this section. It is hypothesized that the different combinations of these two factors constrain and delineate the linguistic outcome in each interaction.

**Oral Interview vs. Role Play**

In many respects these two interactions are similar, e.g. the number of participants is identical. However, they differ in respect to the relative status of the participants depending on what roles we assigned to the two interlocutors by the task description of the role play. There are three possibilities:

(a) examinee has higher status than the examiner  
(b) examinee has lower status than the examiner  
(c) examinee and examiner have equal status  

The first possibility is rare and in a test situation hard to achieve. Even if the instructions indicate that the examinee has higher status than the examiner, the examinee may be reluctant to express his/her putative higher status because of the realization of the real life role constellation; after all, the examiner is of higher status and frequently older; therefore, it is hard for the examinee to play the role of someone with higher status.

The second possibility is the most common and is clearly replicated in the oral interview. The third possibility, though less common, is not so difficult for the examinee to act out.

The recognition of the different status of the interlocutors will guide the examinee to choose the appropriate register and exhibit his knowledge of register differences in English.

The major difference among the two interactions is the capacity of the role play to detect sensitivity to register. Since the two interactions are so similar, the oral interview is chosen to represent both interactions in the subsequent comparisons.
Oral Interview vs. Group Discussion

These two interactions are similar in that they both contain the element of conversation. However, they exhibit a series of differences. These differences are:

(a) the number of interlocutors
(b) the relative status of interlocutors
(c) the control of the interaction (flow of information)

The Number of Interlocutors

The number of interlocutors is unlikely to cause changes in the use of the linguistic repertoire for two reasons. First, a multi-participant discussion can end up being a sum of multiple dyadic interactions, in the sense that each interlocutor addresses only one person at a time as in a free interview. More likely, however, speakers in discussions address all other participants at a time and sometimes address a specific member of the group. Second, because of the identical forms of singular and plural second person pronouns and verbs in English, the fact that a speaker addresses one or more persons does not put demands on the speaker's linguistic repertoire because s/he does not have to use different forms. Maybe some exceptions exist and are associated with very fluent native-like speech where we may see some marking of the multiple addressees as in the following examples:

(a) What do you guys think about that?
(b) What do you all think about that?

The Relative Status of Interlocutors

In a speech interaction the relative status of the interlocutors could differ. A speaker's status in relation to his/her interlocutor could be inferior, superior or equal and this may result in linguistic differences. Some of these linguistic differences are associated with certain registers. A discussion on register was provided in the comparison between oral interview and role play above.

The Control of the Interaction

In the group discussion, since the interlocutors are of equal status, there is no single participant controlling the flow of information as in the oral interview (in which the examiner asks and the examinee answers). Therefore, theoretically, each participant can say whatever s/he wants and speak whenever s/he wants. However, there are some sociolinguistic rules of conversation management, which help organize such interactions. Some sociolinguistic rules of conversational management are turn taking, turn allocating, topic introduction, and topic shift. It is the command of such
interactional skills that we should attempt to assess in an interaction of this type in a test of oral proficiency, because such skills form an important part of a speaker's competence that can not be assessed by more "traditional" tests of oral proficiency.

**Oral Interview vs. Oral Report**

The oral report is probably the most different among all four interactions since it involves only one person. However, the fact that there is only one person involved does not mean that there is no audience for this interaction. The audience (a group of American high school students) is "mentally" present and actually influences the candidate's linguistic production including the choice of register and the amount of background information provided. If the speaker assesses that there is not much background knowledge shared between him and the audience, he may choose to organize his report differently, providing the missing background information at the beginning.

Monologic (as in an oral report) and conversational speech interactions (as in an interview or group discussion) exhibit textuality characteristics such as cohesion and coherence. These aspects of a text are constructed through linguistic (cohesive devices, prosodic features) and paralinguistic features (gestures, kinesics) and it is the responsibility of the interlocutors to create these aspects of textuality.

In a multiparticipant interaction, it is more difficult to locate the source of lack of cohesion and coherence in the text. In other words, it is more difficult to say which of the participants, and to what extent, is responsible for the inadequate presence of these two elements in a text.

On the other hand, in a monologue, the sole speaker carries the entire responsibility for the lack of cohesion and coherence in the text. Since the speaker is not interrupted, there is more responsibility on his side to produce a unit of language in the form of a spoken text applying the principles of textual organization – cohesion and coherence (also referred to as rhetorical organization). Some important elements of rhetorical organization are the inclusion of an introduction or topic sentence, primary and secondary supporting sentences, transition signals, and conclusion.

This responsibility of the speaker does not end with the text production. The speaker must consider what information is necessary for the listener to interpret the text and try to include this information in his production. Considerations of this type include providing the assumed background information needed by the reader to comprehend the content of the text and explaining terms that the listener may not be familiar with.

The analysis offered above showed in what ways the four interactions differed. In the following sections we will examine whether the language of
the subjects tested by the coast was perceived to be significantly different and in what ways.

**Varied Performance Across Different Speech Interactions: What Does This Mean?**

The above analysis allows us to expect that various kinds of language will be elicited. But the claim, namely that a speaker’s oral proficiency varies across different speech interactions is not well defined. Another step that we need to take before actually looking at the results is to clarify what “varies” refers to. Does it refer to the actual proficiency or to the perceived proficiency? If we are talking about “perceived proficiency”, we are interested in whether the scores indicate if the actual proficiency of the candidate varies across tasks or not.

Before attempting to clarify “performance across the four speech interactions varies”, it is necessary to explain that the discussion of the stability of the scores presupposes that all other possible sources of error have been considered and thought not to cause any variability in the outcome. There are numerous such factors but here I am only going to mention the most important: learning effect of one speech interaction on the other, compatible content across interactions, and stable testing environment.

There are four possible answers to the question “What does varied mean?” Figure 1 shows schematically the possible answers. The first possibility is that the subject’s proficiency is indeed stable and this is either reflected or not reflected in the scores. The second possibility is that the subject’s proficiency is indeed unstable (varied) and this is again reflected or not reflected in the scores.

What we are interested in is the case where a subject’s proficiency is indeed varied and is reflected by the scores. The discussion of the other cases is beyond the scope of this paper. The hypothesis that a subject’s proficiency is indeed varied and this is reflected in the scores can be supported by the following assumptions. We hypothesize that each speech interaction produces different kinds of language. This hypothesis leads to the following question: Should the level of proficiency vary across these speech interactions because of the distinctive language in each speech interaction? One could support that this is indeed the case for the following reason. One way to talk about language proficiency is to associate proficiency with the ability to perform certain language functions such as narrating in the present, giving directions, complaining, apologizing etc. These functions are classified in various levels of proficiency. For example, giving directions is considered an intermediate level function whereas apologizing is considered an advanced functions (CAL, 1991). All functions are carried out through language and therefore we can conclude that some linguistic structures and features that are used to carry out higher level functions are a component of a higher
proficiency in a given language. In other words, a speaker’s command of the various structures and features associated with the different interactions and the functions that are represented in these interactions may vary. All four possibilities of language proficiency across different speech interactions depend on two factors: the raters and the rating scale they are using, as well as any interaction between the two factors. It may be the case that due to a lack of training, a specific group of raters cannot differentiate between varied performances of a subject even though a scale has been shown to be sensitive to such varied performance. Along the same lines, it can be the case that the scales are not suitable to differentiating among the varied performances of a subject. This can be true even though the scales are used by raters who have proven that they are able to differentiate provided they have the right tools, i.e. the right scale.

The inadequate training of raters can be surmounted by rigorous training (provided that the scale is valid) and, therefore, it will not be discussed here. Evidence for the positive effect of the training of the raters on the accuracy of rating is cited in Warmke and Billings (1979).

**Figure 1.** Perception and Possible Sources of the Stability of a Speaker’s Proficiency Across Different Speech Interactions
STUDENTS’ PERFORMANCE ACROSS THE FOUR DIFFERENT SPEECH INTERACTIONS

After having established that there is reason to expect different kinds of language across the four tasks of the test battery and after having somehow operationally defined the term “varied” we can turn to the results. The reader should be reminded that the aim of this study is to test whether or not the students’ performance varies according to the individual speech interaction tested. Evidence for or against this hypothesis is provided in two steps. First, by looking at the interaction on which the students achieved the best performance. By doing so we may be able to detect whether the subjects did substantially better on any given speech interaction. This will be accomplished by comparing the subjects’ best performance and worst performance. The second way to investigate whether the students’ performance varies across the four interactions is by looking at the correlations among the tasks and the correlations among the scores of each scale component across the five tasks.

COMPARISON BETWEEN STUDENTS’ BEST AND WORST PERFORMANCES

A comparison between the subjects’ best and worst performance could provide some indication as to whether or not the students’ performance varied according to the individual speech interaction tested. To the extent that there is a significant difference between the subjects’ best and worst scores across the four elicitation tasks, we can infer that the speech interactions measure different aspects of the speaker’s oral proficiency or that some students can perform better under certain conditions.

Table 1 indicates the total number of subjects per interaction in which the subjects received their highest score, i.e. their performance on that score was rated the best. For the purpose of this discussion, Role Play 1 and Role Play 2 will be considered as one speech interaction.

<table>
<thead>
<tr>
<th>INTERACTION</th>
<th># OF SUBJECTS’ BEST PERFORMANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Discussion</td>
<td>12</td>
</tr>
<tr>
<td>Oral Report</td>
<td>24</td>
</tr>
<tr>
<td>Oral Interview</td>
<td>14</td>
</tr>
<tr>
<td>Role Play 1/2</td>
<td>15</td>
</tr>
</tbody>
</table>
As we can see from Table 1, 24 out of 60 students did best on the oral report, 14 on the oral interview, 12 on the group discussion, and 15 on one of the two role plays. Some possible explanations for the results are offered in the following paragraphs.

The fact that 24 out of 60 students did best on the oral report can be traced back to factors pertinent to the test taker and is associated with the relative control s/he had over the task. Since the test taker is the only participant in this task s/he can have a choice of what to say and also monitor the pace of his delivery. These two parameters are not present in the other speech interactions of the test battery. A second possible factor relates to the raters and has to do with the number of participants in each interaction. The presence of three test takers makes it difficult for the rater to be consistent because it is difficult for him/her to pay attention to the language of three different candidates.

If the four different speech interactions were really assessing different aspects of the subject’s language, we would expect that the difference among scores on the four interactions would be significant.

However, the subjects’ scores on all speech interactions present the following picture. The difference between best-worst performance (DBWP) was:

For 33 students - less than 3 points (out of a total of 27 points)
For 52 students - less than 4 points
For 58 students - less than 5 points

For only eight students was the DBWP more than 4 points. (Note: Four points out of the 27 possible points represent approximately a 15% difference in the scores.) The two highest DBWP were 6.33 points for Student 5 and 7.17 points for Student 32. Fifty-two students did not perform more than 15% differently on any speech interaction. Therefore, the best performance test cannot provide us with any kind of evidence supporting the hypothesis that students’ performance varies across different interactions.

The test of best/worse performance showed that the four speech interactions did not offer any differential contributions. After examining the information pertaining to the interaction on which the subject did best, the next step is to look at whether the correlations among the different speech

---

1The number of students who scored higher in the role play interaction was calculated in two ways. First, I counted the number of students who did best either in RF1 or RF2. The result from this calculation was 15 students. Second, I calculated the average of the two role plays and then counted the students whose role play average score was their best performance. Again, the result was 15 students.

2If we add the above numbers 12 + 24 + 14 + 15, we get 65 instead of 60. The reason for this discrepancy is that there were cases in which a subject received the same high score on two or even three interactions.
interactions provide us with any information in regard to whether the subject's performance varies across different speech interactions.

Correlations Among the Tasks

The following table, Table 2 represents correlations among the task scores for each and from each task to the total. The values in the last row represent the correlations of each task component to the total after they have been corrected for part/whole overlap. The correlations were estimated in order to see whether the subjects' performances were stable or varied significantly across the four interactions. If the correlation coefficients are high, this is an indication that the subjects performed at a very similar level of proficiency on all four interactions. On the other hand, if the correlations are low, we can assume that each speech interaction measured a different aspect of the candidates' oral proficiency.

Table 2
Correlations Among the Tasks

<table>
<thead>
<tr>
<th></th>
<th>GD</th>
<th>OR</th>
<th>OI</th>
<th>RP1</th>
<th>RP2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Dis.</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Report</td>
<td>.892</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview</td>
<td>.884</td>
<td>.944</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Role P. 1</td>
<td>.876</td>
<td>.907</td>
<td>.944</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Role P. 2</td>
<td>.869</td>
<td>.917</td>
<td>.950</td>
<td>.973</td>
<td>1.00</td>
</tr>
<tr>
<td>Total</td>
<td>.930</td>
<td>.963</td>
<td>.979</td>
<td>.977</td>
<td>.979</td>
</tr>
<tr>
<td>F/W Cor.</td>
<td>.922</td>
<td>.951</td>
<td>.967</td>
<td>.963</td>
<td>.964</td>
</tr>
</tbody>
</table>

The results show that the subjects' rated proficiency was very stable across all five speech interactions. The results seem to be consistent with the results of the best/worst performance test.

It is worth mentioning that the above table shows that, though not statistically significant, the correlations of the Group Discussion are consistently lower than those involving the other three interactions.

Empirical Evidence for Distinctive Features

It has been claimed earlier in this paper that the four different interactions are structurally different and, for this reason, one would expect to see different kinds of language in them. Juxtaposition of the various speech interactions suggested some structural differences. The question that arises is: "What features of language index these differences?" This section lists some possible linguistic features that are unique or relatively more frequent in each speech
interaction.

It was claimed that a crucial characteristic of a group discussion is that, in order to be successful, the student should command the rules of conversational management. Such rules include: topic introduction, inviting contributions, agreeing and disagreeing, topic shift, acknowledging the contributions of other interlocutors, etc. Some of these rules are often associated with specific forms but some are not. For example, negotiating the meaning is often accomplished through questions, but interrupting can be accomplished in many ways.

Elements that are found frequently in a group discussion are:

(a) Phatic signals such as "Hm ... hm" (acknowledging someone's contribution).
(b) Questions either in order to negotiate meaning (What do you mean?) or to invite a contribution (What do you think about this?)

Here are some examples from the data that show elements of conversational management.

Excerpt 1:

GROUP 3

1 A: First of all, I think we have to take under consideration the ...
2
3 B: Hm hm ... But of course we should look after ...
4 the the accommodation we should say that there are
5 many luxuries hotels in Cyprus
6
7 A: Yes, I think it's a good idea but we have to point
8 out that Cyprus has a lot of interesting things
9 that you cannot find in other countries
10 B: Of course and the remains of the Greece politism
11 ... civilization is very very great (A: Hm hm)
12 so we should show them some history resorts
13 [A: Yea] like ... Kourio or the ...
14 A: Salamina theater
15 B: Salamina
16 A: Yes
17 B: [But of course Salamina theater is at the other
18 side
19 A: [Yes but we can tell them about it
20 B: We can show the the divide line, we can take them
21 to Nicosia and show them the division line so they
22 should know ... they could know about what happen
23 in Cyprus (A: Hm, hm, Nicosia is ...) about the
24 Turkish invasion
25 A: Yes
26 C: Yes
27 C: We could also point out the real problem we have in
28 Cyprus so that they can take it to the outside
29 world
30 [and they can say more about it
31 B: [Yes
32 C: [Of course Probably they can help us in giving
A: I think that we have to suggest that they should visit some of the museums.
B: Of course.
C: And learn about the civilization and...
A: I think we have to suggest that they should visit the old villages on Troodos mountains.
B: Certainly.
C: Yes.
A: Because in that way they will be able to have a better look of Cyprus traditional life.

As we see in the above excerpt, the students' contributions on lines 3, 6, 10, 11, and 22 are phatic communication devices which are used to show to the interlocutor that they agree with what s/he is saying and that the speaker can continue with his contribution. In many instances (e.g. in lines 11-12, 16-18, 33-34), there is a lot of overlapping which is another characteristic of a genuine conversation or discussion.

The students' contributions on lines 6 and 11, 37, 41, and 42 also function as approvals to the topic and utterances of one of the interlocutors. Again, all the mentioned features can be found only in a discussion and conversation but not in monologic speech such as an oral report.

In excerpt 2, we will look at other characteristics of a discussion, namely turn allocation techniques, topic initiation, and requests for clarification.

**Excerpt 2:**

GROUP 4

1. A: Cyprus is a tourist country ... There are some problems that Cypriots face.
2. B: I agree with you and I think the visitors are 2 millions we have 2 millions tourist and we have many problems to discuss.
3. C: Yes, I believe that we have good things from tourism. We have a lot of money in Cyprus we have a lot of...
4. B: What do you think about this season? We have problems with the war...
5. C: Yes we have problem this season with our tourism because we believe that we have only one and a half ...
6. B: Kelvin you think you can find a solution? Uh to that apparent problem? Just one answer. What do you think ours people can do?
7. A: First we have to find the reason why tourists don't visit Cyprus as they did before and if we find the reason we'll find the solution to the problem.
8. B: But we are ... we have many inhabitants.
9. C: MANY?
11. C: I believe with Kelvin that ... I agree I agree with Kelvin and I believe ...
12. B: But what is the reasons?
The reasons I believe that ... it's uh ... the
problems of East Mediterranean like the war in
Kuwait last year I believe ...
Yes but the poor tourists they don't give us
nothing, they give us they don't give us anything
because they come to Cyprus with only 10 pounds
that's only
You think tourists come here and work they give us
no exchange? ...
No I don't believe that the temperature is a reason
What's your opinion?
I think that we must say about the problems the...
I think the Chrisso region a place we can ...
Do you think that ... I think that Cyprus is not
divided to two you think ...
WHAT?
To two you think that we have more tourists

We see that a given speaker asks questions in order to invite another
speaker's contributions and, in this way, he allows his peer to take control
(lines 9, 14, 25, 33, 41, 51). The phrases which are in bold and capitalized
(lines 21 and 53) are requests for clarification which are only possible in a
non-monologic speech interaction. They are further encouraged in cases in
which the interlocutors are of equal status, as in a group discussion.

As far as the characteristics of a one participant speech event are
concerned, we saw earlier that in a monologue (such as an oral report) the
speaker has a greater responsibility to obey the rules of textual organization
since he is not interrupted by anyone else. Some elements of rhetorical
organization are an introduction or topic sentence, transition signals,
concluding paragraph and comment, etc. The following contain examples of
these features.

In Excerpt 3, we can see an example of a concluding paragraph.

Excerpt 3: GROUP 1 - STUDENT A

In my team the Americans students are visiting our country we decided that it should
be best to take them to visit Nicosia the capital of Cyprus which has a history, it is full of
museum a great...

This is about the things I think so it would be best to show them to show the American students
our traditional way of living our civilization where we are the most important thing.

In the above excerpt, the last paragraph functions as a concluding
paragraph and contributes to the creation of a text. It also includes a
summary signal ("This is about..."), indicating that the student is about to
summarize his suggestions.
Do Different Speech Interactions Yield Different Kinds of Language?

Excerpt 4: GROUP 3 - STUDENT B

Although Cyprus has been a English colony it is well well developed. There are many restaurants, hotels, nightspots, resorts that would attract the tourist ... There is a great variety of plants and animals although Cyprus is a very small place. Uhh In conclusion I would like to say that Cyprus would have been a heaven on earth if wasn't divided.

In the above excerpt, the student uses the transition signal “in conclusion” which is the signal par excellence of a concluding paragraph.

Excerpt 5: GROUP 9 - STUDENT C: REPORT

Cyprus is a small beautiful island in the Mediterranean sea. Tourists can visit it, ... So but tourist tourism ... is effect Cypriots and it has positive and negative effects. Tourists bring new ideas so Cypriots can open their horizons. Also tourism is bringing foreign exchange and it can be very useful. Also Cypriots can learn foreign languages more easily, more easily and they became fluent speakers. But ... tourism is having a lot of bad ... negative effects. Is spoiling the young people but also prices are rising because especially at tourist area ... hm and these effects the Cypriots and their cost of life ... It is going up continuously. Also when tourists visit Cyprus can learn a lot about our problem and Turkish troops is occupying our North part of Cyprus and [...] Also Cypriots can a lot about ... from tourist. They can develop their characters [...] I think it’s not right and also Cyprus can offer offer much who want to visit the island and also is very beautiful ... I’m not going to tell more. If you want to to know if you want to learn more about Cyprus come and visit us.

In this excerpt, we see a series of transitional markers (“so”, “but”, “also”) which are a good characteristic of a well structured text (when they are not overused). The transition marker (also) is overused. Ideally, there should be a variety of markers with similar functions (in this case, additives and conjunctions), e.g. “in addition”, “furthermore”, “besides”, etc. The last sentence functions as a concluding remark that is often a part of a concluding paragraph.

Excerpt 6: GROUP 11 - STUDENT A

Parents in Cyprus are different ... I believe that they are different than in other European countries. I believe because Cyprus is small country the relationship between the parents and the children are good and we all face many problems with our parents but there are some of course ... On the other hand children want to make their lives ... go with their friends and make up their minds by their selves so we have fights sometimes ... and ... At last I believe that parents must be less protective because they must understand that they are children are becoming older and some a day they will be at their place and they must be able to make a family and bring up their own children.

In the above excerpts, we see a clear topic sentence (“Parents in Cyprus are different”), a transitional marker (“On the other hand”), and a concluding paragraph which is signaled by the transition marker “At last”.

All the examples provided above present some empirical evidence for the hypothesized differences among the speech interactions. The speech samples
from the 60 subjects are filled with such examples which due to space limitations can not be presented in this paper.

In conclusion, two things need to be mentioned at this point. First, these distinctive characteristics are easier to observe in the use of language by native speakers and advanced non-native speakers rather than in the language of less proficient non-native speakers. Testers may choose to administer different speech interactions according to the assumed level of the testee. For example, if the testee is "lower intermediate" and the tester wishes to assess the test-taker's command of past tense then the OPI, which give the tester the opportunity to outline his agenda may be a more effective and illuminating than an oral report. Second, these characteristics can be divided into two groups. The first group contains elements that are categorical, i.e., they exist only in one speech event and do not exist in another. For example, turn-taking techniques exist only in a group discussion in OPI and in role plays but cannot exist in an oral report. The second group consists of elements that differentiate between speech events because of their relative frequency in these speech events. For example, asking for clarification is a characteristic of group discussion rather than an interview. In an oral interview where the examiner primarily asks questions and the test-takers predominantly answer questions, the test-taker has little chance and is inhibited from asking questions. On the other hand, in a group discussion there is no such restriction. This restriction is a consequence of the structure of the speech event.

CONCLUSION

In this paper an effort was made to provide evidence for or against the hypothesis that different speech interactions in test battery of oral proficiency yield different kinds of language. A structural description and comparison of the four speech interactions provided justification for the above hypothesis. Moreover the term 'varied' interaction was investigated and clarified.

The ratings of the 60 subjects on the four different speech interactions were subjected to various statistical tests and the results showed that the difference among the ratings on the four interactions were not significant. This finding is not congruous with the findings of the structural analysis of the four speech interactions; an analysis whose results supported a differentiated linguistic make-up for each speech interaction. However, a look at the data shows that there are some linguistic features that can be found in one interaction and not in another or can be found in different frequencies.

These results force us to look for certain explanations. As suggested by Pavlou (1995) the fact that the different kinds of language in the four speech
interactions were not reflected in the results of the study may be because all four interactions were assessed against the same scale. In order for the linguistic differences of the four interactions to be detected the speech samples should be assessed by means of scales that are sensitive to the individual characteristics of each speech interaction. Of course, it is still possible that even with the new interaction-specific scales the assessment of the four interaction would not vary. Further research in this area would probably provide a more definite answer to the central question of this study.

"The author would like to express his appreciation for the reviewers' comments.

REFERENCES


APPENDIX 1

COAST (CYPRUS ORAL ACADEMIC SKILLS TEST)

Components of COAST:

Group Discussion

Topic 1: Relationship between young people and their parents in Cyprus - One handout for each of the three participants.

Topic 2: Tourism - One handout for each of the three participants.

Oral Report

Based on the Group Discussion. Instructions are given on previous handouts.

Oral Interview

Follow procedures of FSI/OPI.

Role Plays

Role Play 1: Asking for Information

Handout A: Informal Register
Handout B: Formal Register

Role Play 2: Giving Directions

Handout A: Informal Register
Handout B: Formal Register
Map

Role Play 3: Describing a House

Handout A: Informal Register
Handout B: Formal Register

Role Play 4: Making an Appointment

Handout A: Informal Register
Handout B: Formal Register
Schedule