

Such a Nice Gesture: Paired Spanish Interaction in Oral Test Discourse

Ana Maria Ducasse
RMIT University, Australia

Abstract—The problem faced by researchers regarding the construct of ‘interaction’ in paired peer-peer tasks in oral language proficiency testing is addressed by analysing verbal protocols of raters watching candidates in a videoed paired speaking task in a test in Spanish as a foreign language context. The aim is to investigate any correspondence between observable nonverbal features in the candidates’ videoed speech sample that raters claim to attend to while observing paired interaction and those features of body language that candidates are aware of while performing the test. This study is different to previous studies because the task is for paired beginner foreign language students of Spanish. The data is drawn from 17 transcribed paired speech samples, transcribed Verbal Protocols collected from 12 Spanish teacher/raters and 25 candidates Stimulated Verbal Recalls. The analysis is guided by non verbal communication performance features already identified in two previous studies: a previous rater cognition study where raters comment on their observation of candidate performance through Verbal Protocols (Ducasse and Brown, 2009) and a candidate Stimulated Verbal Recall where candidates comment on their interaction in a test performance (Ducasse, 2007). Findings indicate that of the non verbal communication features gaze and gesture are equally important to candidates during performance and to raters during rating process. These findings support findings in Conversation Analysis and Gesture studies and have implications for rating peer-peer language proficiency and for the ‘interaction’ construct.

Index Terms—paired speaking tests, rating criteria, Spanish L2

I. INTRODUCTION

Following wide range of research into speaking test interviews from within applied linguistics, and in particular language testing, paired tasks were introduced in oral proficiency testing to add another dimension to speaking tests. The new dimension was peer-peer communication. The advent of this additional task sparked new research. Currently, the field is investigating the effect on language output, of pairing peers. This language output of pair performance is then judged by raters with scales and rating criteria that take into account the known reality of the paired ‘construct’ as in what it is we think we test when we see two candidates performing together.

The majority of research into pairs concentrates on the interaction of variables between the pair of candidates such as familiarity, intra- and extro- vertedness, background culture, level of proficiency, and the effect of such variables on language output. Researchers look for causes of variability in the output which can make test results unreliable. The output however, is always considered to be the ‘audio band’ with non verbal communication ignored. The test outcomes should as a result be based ostensibly on that ‘audio’ output.

There is a lack of studies from the perspective of either candidates or raters during the ‘test event’ to ascertain whether the non verbal communication, beyond the audio output, has any effect on candidates speaking to each other or on raters’ view of the performance. A test event comprises a combination of two processes: a candidate performance which produces discourse as output and a rater performance which produces ratings as outcome. These two processes are simultaneous: while candidates speak to each other, raters rate.

By concentrating on rater performance the research identifies and demonstrates how describing the non verbal communication features that raters attend to can be a practical solution to reduce variability. By using this approach some of the obstacles to variability are removed by suggesting features of interaction preciously unincorporated into language tests might be included in the test event.

Context of the study

This paper focuses on the speech sample of peer-peer candidate pairs performing a 10 minute speaking task for an in house university test in Spanish as a foreign language. The task is for paired beginner LOTE students introducing, maintaining and changing between three different topics each from a task card.

Statement of the problem

If any nonverbal communication is taken into account by raters as part of paired interaction this broadens what is so far known about the construct. If the construct changes then rating scales need to be adapted to reflect it. At present non verbal communication is not included as a feature with rating criteria in the rating scales. The possibility that raters notice nonverbal communication as a factor affecting peer-peer performance is problematic if it is not present in the marking system for such a test.

Aim of the study

The aim is to demonstrate if and how raters and candidates view non-verbal communication in 10 minute clips of candidates performing a paired speaking task. The results challenge researchers to consider whether they should be incorporating nonverbal communication as part of rating oral proficiency.

Scope and focus

To realistically achieve the aim it is necessary to narrow the focus to the non-verbal performance features already identified. Using the results from these two previous discourse studies the central reason for focusing on non verbal communication is that both sets of protocols both the candidates and the raters raise the issue. Although paired speaking is necessarily made up of the audio channel i.e. candidates speaking and listening during the interaction, the presence of the video channel i.e. the non-verbal communication, cannot be ignored. This calls for further investigation in to nonverbal communication in peer-peer interaction to inform rating scale development, rater moderation or candidate test preparation.

It begs the questions: Do raters notice it? It is important to know the features raters focus on while rating because of the implications this has for include these features in rating scale development (Politt & Murray, 1996). Do candidates notice it? It is interesting for testers to know how aware candidates are of nonverbal communication because it has been studied in relation to second language listening comprehension (Sueyoshi & Hardison, 2005). If it is a helpful strategy it could become part of peer-peer test preparation.

Language teachers and language testers will need to consider their own testing context in order to decide whether gesture may also be likely to cause variability if it is overlooked.

Rating pairs and groups

Recent studies for rating pairs or groups all investigate the difficulty for scales developers and raters, the scale users, to adapt to this context: Nunn (2000) tackles the problem by designing rating scales for small group interaction during classroom activities; Orr (2002) analyses verbal reports given by raters during the decision making process for the First Certificate of English (including a pair/group task) and Bonk and Ockey (2003) in a Rasch analysis of a group oral discussion task, find that “rater and scale reliability were achievable under real testing conditions”. However, despite this finding in a more recent validity study on a university group oral test found the most frequent cause of variability to be “the more intangible interpersonal factors in the way group members react to each other” (Van Moere, 2006, p. 436).

The ‘intangible’ remains unexplained in the peer-peer test situation. What is missing and remains unexplained is information about the construct, the manner in which raters construe paired interaction when they are rating it and whether non verbal communication is attended to by raters or candidates.

Discourse studies

Discourse studies in language testing where rater Verbal Protocols have been used such as Politt and Murray (1996), Meiron (1998) and Brown (2000) indicate ways in which using rater protocols could broaden the so far ‘intangible’ peer-peer interaction construct. Brown Iwashita and McNamara (2005, p. 6) did not provide raters with guidance for the types of comments to make when giving verbal protocols in a simulated rating situation. Through this means it is possible to “ensure that the content of the scales is relevant to the context and meaningful to the raters” which is important for validating the scale development process. These empirically successful studies had delved into the rating process through qualitative methods so the researcher undertook a discourse study to investigate the manner in which raters construed peer-peer interaction.

Language specialists’ orientation to paired interaction

By conducting such a Verbal Protocol study (see Green 1998 for use of this methodology in language testing) raters verbalised the manner in which they construed interaction. In that study, there were 12 language specialists commenting on three pairs each taken randomly from 17 candidates, self selected into the study, performing 10 minute paired tests. The instructions for the verbal protocols were to watch 3 different candidate pairs and comment into a tape recorder about the features noticed about the peer-peer interaction.

A content analysis of the protocols uncovered the key features that reflected the view of the majority of the participants. It suggested that raters observing peer-peer interaction oriented to three main features: interactive listening, nonverbal communication and speaker engagement (Ducasse, 2009). The manner in which raters construe paired interaction which had remained previously unexplained was put forward for a particular context with paired beginners. Language testers for the first time were provided with an overview of the language specialists’ perspective on peer-peer interaction on a paired task.

This has implications for future scale development. When a scale is developed that incorporates features from the language specialists view of interaction, it will answer a call from the literature for including “in a scale what raters attend to” (Politt and Murray, 1996).

To gain greater insight, each of the three focal areas: interactive listening, nonverbal communication and speaker engagement requires further investigation. Only nonverbal communication falls within the scope of this paper but nevertheless a part of the knowledge gap regarding the ‘intangible’ in peer-peer interaction has been filled with the details on raters’ view of nonverbal communication providing deeper understanding of how they construct paired interaction.

Literature on non verbal communication

Relevant literature to set up a context for this study would be from the field of Gesture studies, the most influential being Mc Neill (1992) whose work has been used as a basis for other studies. Space prohibits a discussion on nonverbal communication which is defined as the meaning conveyed by the body, head arm, face, gaze, posture and inter personal distance (Kellerman 1992). It is a broad field studied in its own right beyond Applied Linguistics and this paper, and necessarily the scope is narrowed here from non verbal communication in Second Language Acquisition (SLA) to its presence in research for testing speaking in a foreign language context.

One studies each, from SLA speaking, listening and language testing, is included below. The speaking study provides an example of Second Language (L2) speakers using gesture interactionally, the listening study explains how gestures aid comprehension and the testing study, which is on peer-peer discourse, demonstrates candidates' perception of nonverbal communication in test performance.

In the first study, 16 Dutch learners speaking French at intermediate level Gullberg (2006) investigates whether nonverbal interaction, in particular gesture, is purely interactional or if it is linked to developmental SLA stages in a story retelling from pictures activity. In the findings when speech is ambiguous and gesture is unambiguous then they combine for meaning. Also when gestures occur with ambiguous expressions addressees take gestures into account. Thus the gestures are multifunctional serving both interactionally and to compensate for any internal processing problems.

In the second study on gesture and facial cues and L2 listening (Sueyoshi & Hardison, 2005) there is a survey of learners finding helpfulness in hand and arm gestures by lecturers while completing a listening comprehension task i.e. a lecture. A total of 42 lower intermediate and advanced learners were randomly exposed to a lecture under three conditions: audio only, AV face only and AV with gesture and face. The lower group preferred the face and gesture the higher group the face only. It confirms previous research findings that gesture helps L2 learners to comprehend speech (Cassell, McNeill & McCullough 1999).

Gesture in oral proficiency testing

In the third and only study located that reports on the attention paid by raters and candidates to nonverbal communication in a peer-peer testing context is the Coniam (2005) study on wearing face masks during the Sars outbreak in Hong Kong. In the test there was a role play and a four candidate discussion task. Half the raters (7 out of 15) thought the masks had not adversely affected performance because students had compensated by speaking louder and making greater eye contact. In the same study most of the 145 candidates perceived they had performed at a lower level. To compensate approximately two thirds reported speaking louder and slower they also reported facial expression disappearing behind the mask. This confirmed findings of previous studies in SLA that had also shown that gestures as well as facial cues could facilitate face to face interaction with second language learners in the form of comprehensible input and feedback (Gass 1997, Long 1996 and Pica 1994).

In short there is a knowledge gap in the area of nonverbal behaviour during performance of candidates in language tests. Non verbal communication in testing has as yet to be studied but the call for greater attention from within SLA has already been voiced in a recent discourse study (Lazaraton, 2004) where an ESL teacher's gestures are micro analysed. It concludes that gestures are significant to L2 learners and that gaze is also a significant factor that merits serious attention and asks why so few videos of performance have been analysed in SLA research. It must be a practical or technical problem hindering research which will change with technological advances?

Summary

In SLA and in language testing both from the speaking and listening perspective gesture, gaze and facial features have been found to aid communication. Nonverbal communication is a component of what makes peer-peer interaction, in pairs and group, difficult to rate. Discourse studies have perspicaciously drawn out raters' perception of the rating process and their perception of interaction for the paired construct included non verbal communication. This could be part of the 'intangible' searched for by van Moere (2006) and which warrants further investigation. Research questions were thus set out as follows:

What nonverbal communication is attended to by raters of peer-peer interaction?

What nonverbal communication is commented on by candidates after the peer-peer test experience?

II. METHOD

Verbal protocols are widely used to probe cognitive processes for learners. They have been used for raters in first and second language writing and for rating speaking on EAP tasks. To this point verbal protocols have not been used to investigate rating paired orals for L2 Spanish. This methodology allows the researcher to uncover key features that reflect the view of the majority of the participant raters while they are simulating being 'on the job'.

As was mentioned above, the data for this study come from a larger study for developing rating scales for paired interaction in a large Australian University. Data for the protocols from raters were collected in 2005 and the data for the protocols for the candidates were collected in 2003 soon after candidates performed their videoed 10 minute paired orals on which the study is built. For the focus of this study only the data coded with Nvivo as nonverbal communication were considered.

Participants

The group of 17 pairs of candidates were all acquainted, self selected pairs from their cohort who volunteered for the project in exchange for personalised individual feedback on their performance. The pairs were representative of the types that have previously been researched of same or mixed: proficiency level, gender, nationality and personality type.

The group of 12 raters that were recruited to participate were all experienced male and female teachers from the university language program where the students were taught in the target language with an emphasis on oral skills at beginner level. The raters were native and non native speakers of Spanish and represented varieties from Spain and Latin America. Their role as lecturers, teaching assistants and casual staff meant they were experienced in conducting the Paired Oral Test (POT) and had been previously trained and moderated to rate with the existing criteria. In essence, they were all experienced with the task and candidate level they were requested to observe.

Data collection

The paired orals were regularly audio taped for double marking and rater moderation. In this instance all the students in the cohort 145 were videoed. Informed and written consent was obtained before taping and subsequently the students, 25 in all, who were interested in watching them participated in the remainder of the study.

The Stimulated Verbal Recall took place as soon as possible in the week following the oral test. Students watched their performance and commented without guidance. The Rater Verbal Protocol took place the following year after the exam period. The twelve raters commented on the interaction on the performance without guidance.

Data transcription and analysis

The 17 pairs of candidates in the oral, the 12 raters commenting on the paired interaction and the 25 candidates commenting on their own performance were all transcribed as part of the larger study.

The protocols were coded and the inter coder reliability for both sets of protocols reached more than 80% inter-rater agreement. The preliminary coding grid from the patterns that had emerged could categorize 80% of the examples with a good fit. The remaining 20% were discussed by the coders and this helped tighten the definitions on the grid and clarify the coding for the remainder.

To limit the scope of the study those features that emerged from the verbal protocols from raters and candidates as a result of content analysis were used to guide the analysis below. Nonverbal communication was one of the features of interaction that raters and candidates commented on.

Second analysis

Using Nvivo to search for previously coded protocols produced results that included references made by candidates and by raters to non verbal communication. The approach taken generally worked well because the system was already set up for such a search.

The findings that emerged were divided as follows: general body language, body position, movement, gaze, lips, hands, gestures, appearance and head.

The data was divided into candidates' and raters' content analysis of the verbal protocols. In fact as the two columns in the tables below for each of the features, show that the candidates commented little on their body language. On the other hand the raters comment repeatedly on the same features, having had no guidance or instruction to do so, and they are very explicit about the presence or lack of non verbal communication and its connection to interaction.

III. RESULTS

Without guidance candidates and raters made reference to non verbal behaviour when asked to comment on paired interaction. Raters comments were divided into four categories for discussion which ranged from general to very specific: general body language; the body: position, appearance and movement; hands and gestures; and gaze. Candidates commented on all of the above but the general body language.

1 Body language

General comments on 'body language' as a category covering all aspects of nonverbal communication includes comments from the raters on the use of body language for "encouraging listening or "to help communication". Comments on the absence of body language are equally important because it suggests that it is equally noticeable in its absences or presence.

TABLE 1
BODY LANGUAGE

| |
|---|
| Raters comments |
| Body language is encouraging listening |
| The stronger one uses body language to help communication |
| Neither of them use body language |

2 Body: position, appearance and movement

This next category includes body position, appearance and movement which are grouped together because they are not hands or gaze, the remaining two and larger categories.

In nearly all the cases where the raters comment on the body position they elaborate and comment on the effect such a position has on the observer. Comments in table 2 below for example: "it doesn't help him at all"; "it is very telling"; "I would rate then as poor"; "it makes him know less than he does" or "a negative posture, he is defensive" all show

how the raters are very involved with the performance. This very personal reaction demonstrates the effect body position has on the rater that is observing.

The candidates' comments are few in this category so little can be surmised. It is interesting that they both refer to opposing levels of comfort in the test "upright" compared with "comfortable" both of which would be visible to the rater.

Appearance

After body position, appearance at least that you want to interact is important for the raters

'Smile every so often and recognise and gesture that you understand. You don't have to laugh right through but I would rate this as very poor interaction'. The perception this rater has of paired interaction is embodied in a single comment. It is expected that candidates acknowledge each other, signal comprehension and be cooperative by smiling. Comments made by candidates on appearance acknowledge their nerves which are as visible on video as they would be to raters.

Movement

Movement is a small category, with the candidates explaining why they are using nonverbal communication: embarrassment and the end of the exam drawing near. From the raters' perspective the comment "her whole body appears to say yes or no" shows a level of speaker engagement with the other candidate that the raters have observed. The nerves also are observed in the comment "his legs are shaking". Raters are expressing where their eyes are taken to as they observe the pairs performing and they seem to vary from the whole body perspective to the smaller nonverbal communication feature such as where candidates look which is the next category.

TABLE 2
BODY POSITION, APPEARANCE AND MOVEMENT

| Candidate comments body position | Raters comments body position |
|---|---|
| I look so upright | His position and the visor don't help him at all. |
| I put myself in a comfortable position | She has her legs crossed away in terms of physical interaction it is very telling |
| | They don't move their body and don't communicate physically that they understand. I don't know but in general terms I would rate them as very poor |
| | Body posture make him look as if he knows less than he does |
| | He has a negative posture he is the defensive |
| | Their chairs are close |
| Candidate comments appearance | Raters comments appearance |
| I am nervous and prepared she is the same | They relax and connect |
| I was nervous and self conscious | Her being relaxed adds to the interaction |
| Big smile it is over | Facial expression is good |
| I am biting them (lips) under pressure | Smile every so often and recognise and gesture that you understand you don't have to laugh right through but I would rate this as very poor interaction |
| Candidate comments movement | Raters comments movement |
| I knew it was coming to an end so I fidgeted | Her whole body appears to say yes or no |
| I am embarrassed and fiddling with my fingers | His legs are shaking |
| | Head was tilted inwards and really listening |
| | She nods to indicate she has understood |
| | She nods while the other is speaking to give her confidence |

3 Gaze

Gaze is the term used in Conversation Analysis to indicate where participants look during a conversation and it is linked interactionally with speech. There are a significant number of comments on comprehension and engagement that provide insight into the process of peer-peer interaction during the test from both the candidates and the raters.

First, the comments on engagement which divide into two types: comments that indicate that raters have observed signs of engagement or disengagement such as can be seen in table 4. Candidates' looking at each other is valued and an appraisal is added to the observation by the raters for example: "The visual support generally makes good communication" or "The way they look at each other is good". The impression is that both candidates are looking at each other from the comments and this is the way raters like it. Conversely the comment "Interaction is difficult because the situation is tense and it is hard to focus because they are not looking at each other" expresses how raters observe the difficulty in performing the task without looking at each other. Other comments that indicate that the partners are not engaged in the interaction such as "He looks at the paper rather than follow the conversation"; "He never looks up or at his partner" or "He is more interested in his shoe than his partner" are examples of how raters notice where one of the candidates is not helping the performance. This is not particular to peer-peer interaction. If a person is being interviewed and fails to look up then the feeling of rejection or disengagement would heighten for the interviewer. Raters can empathise with the person not being looked at and verbalise that rules of interaction are being broken by lack of inclusion by candidates averting gaze.

TABLE 4
GAZE

| Candidate comments | Raters comments |
|---|--|
| Gaze | Gaze |
| He was losing me you could see blankness in his eyes | She does not look at her partner very often |
| We locked eyes I worked out he was thinking “I don’t know what you’re saying” but then he got it” | The stare is unnatural and un spontaneous. She is drinking in the words with a fear of not understanding |
| it is a comfort thing to look down when you are speaking so you are in no-where land | She is looking up but not looking him in the eye as she answers |
| | He never looks up or at his partner He is more interested in his shoe than his partner |
| | They look at each other most of the time and make the most of it |
| | He does not look at her though she does look at him right through. At one point she speaks French and communication is broken completely then he is obliged to looks up and as she corrects herself communication is re established by the look. |
| | The visual support generally makes good communication |
| | The way they look at each other is good |
| | Interaction is difficult because the situation is tense and it is hard to focus because they are not looking at each other |
| | He looks at the paper rather than follow the conversation |

The second group of comments on comprehension come from both raters and candidates. Where previously looking at your partner is appraised positively here it is the contrary: “The stare is unnatural and not spontaneous. She is drinking in the words with a fear of not understanding”. The rater observes that it is not an interactive kind of ‘look’ but akin to ‘a rabbit in front of the headlights’ stare.

From the candidates’ perspective the comment “We locked eyes. I worked out he was thinking “I don’t know what you are saying” but then he got it” shows how intensely they are looking at each other and how communicative the expression is to convey so much meaning to the partner; equally expressed in the comment “He was losing me. You could see blankness in his eyes”. These comments demonstrate the importance of looking at your partner because of what a look is capable of conveying.

Looking at another person can be uncomfortable in an already difficult situation such as a test and the comment “it is a comfort thing to look down when you are speaking so you are in no-where land” highlights this. Raters notice if they look at each other in the eye or not as in the example “She is looking up but not looking him in the eye as she answers.” which shows that raters are constructing a norm within the peer-peer test: a preference for candidates to look at one another.

In summary, it can be seen in table 4 that raters observe if candidates gaze enhancing is hindering or enhancing interaction. Looking or gaze could be summarized as dividing up into + or – gaze = + or – engagement. The most important comment “He does not look at her though she does look at him right through. At one point she speaks French and communication is broken completely then he is obliged to looks up and as she corrects herself communication is re established by the look” shows how looking has a function here that the raters have noticed. It has repaired the broken interaction caused by her speaking French.

4 Hands and gestures

The comments on gesturing can be grouped into gestures that support the interaction, show comprehension or convey meaning.

Gestures that directly convey meaning are illustrated by “The stronger one mimes eating to the weaker one”. This support between candidates is similar to an expert/novice situation such as student /teacher or mother/ child. Mime is the only resource to ensure comprehension.

The situation is different in the comment “He uses his hands a lot when he can’t find the word he is looking for” because while the candidate is retrieving the word, he is holding the floor by gesturing in a way that can try to explain his meaning.

As in the comments mad on gaze above, candidates show engagement in the interaction also by using gesture “He uses hand gestures to draw his partner into the conversation” or “She makes supportive gestures”. Raters are expressing their awareness of support for the other person and when this is lacking they comment on its absence: “These two are not interested in communicating. She has her hands under her legs” which is interesting to see the manner in which raters judge so much on non verbal communication.

Finally, awareness by the raters of gestures made between candidates to signal comprehension such as “She gestures to indicate she is following the conversation” or “When she repeats it her partner give her a thumbs up to show she has understood” show a recognition that peer-peer interaction is two-way so it is essential that there be comprehension between the pair. It is important to signal it and one way is to gesture.

TABLE 6
HANDS AND GESTURES

| Candidate comments hands | Raters comments hands |
|--|--|
| I didn't know how to say that with my hands. | He uses his hands a lot when he can't find the word he is looking for |
| What is with those hands? | She is talking a lot with her hands |
| | He uses hand gestures to draw his partner into the conversation |
| | Great to see other cultures(referring to asian students) using their hands like us not like the anglo Saxons. |
| | When she repeats it her partner give her a thumbs up to show she has understood |
| | She nervously grabbed her head with her hands |
| | The stronger one mimes eating to the weaker one |
| | I put a lot of emphasis on hand movements. He has a successful interaction by using his hands a lot |
| | these two are not interested in communicating she has her hands under her legs |
| | They use their hands |
| | She makes supportive gestures |
| | She gestures to indicate she is following the conversation |

In sum, the findings from the content analysis for nonverbal communication from the two verbal protocols were reported in the tables above. They report on the amount of comment on nonverbal behaviour by candidates and raters. It can be claimed that raters are highly conscious of nonverbal communication as part of paired interaction. The comments on body language and body: movement, position and appearance all provide an insight into generalizations that raters make based on appearances. The comments on gesture are also enlightening in that they spell out that presence or absence of gesture is noticed in peer-peer interaction. Finally, gaze is also an important part of non verbal communication showing engagement and comprehension,

Candidates on the other hand although commenting under the same conditions as raters, with no guidance, give only a little insight into why they carried out certain actions.

IV. DISCUSSION

This qualitative research has shown language testers that from a rater's perspective nonverbal behaviour is a fundamental aspect of L2 performance and is a means of observable 'comprehensible input' between two candidates. Although "the role of gesture as primarily intrapersonal, interpersonal or both is being debated in the gesture studies literature": (Lazaraton, 2004. p. 109) there is strong evidence that in peer-peer interaction the interpersonal is what is observed and valued by raters.

On the issue of gesture being multifunctional and serving both interactionally and to compensate for any internal processing problems (Gullberg, 1998), there is evidence that raters observe beginner L2 learners using gesture as a communication strategy to compensate for verbal difficulties which is born out by comments on mime.

In the original coding, body language covered all the separate categories above so it was difficult to clearly picture the key features that made up nonverbal communication for raters. The range of comments was evenly spread from the different raters but the predominant ones were gaze and hand combined with gesture. These support previous findings of other SLA studies reported in the Literature review in the introduction.

A possible weakness that could effect the results is that this division into secondary level of coding was not checked for inter rater reliability with another coder. However, the coding is based on concrete words that refer to parts of the body so it leaves the definition very clear and would not cause confusion for another researcher to comb the data and find similar results.

The data in this study point to the significance of nonverbal behaviour as noticed by candidates and raters as part of paired oral interaction. Even without analysing the verbal input of the candidate or the gestures as appearing on the video the comments from the protocols are evidence that nonverbal behaviour mattered to the raters and to a lesser extent to candidates. The data reinforce the claim that there is a fundamental relationship between non verbal communication and 'spoken interaction' in face to face peer-to peer paired discourse.

These protocols shed light on what raters notices and an attempt is made to ascertain why the candidates behaved the way they did. They seem to be explaining in their comments.. in this type of frame work that fails it is not answered very fully because the candidates were not lead to comment in a particular direction as the raters. This is a shortcoming of the methodology

We can make claims that raters notice the effectiveness of gesturing or looking to keep the conversation going. This is interesting to see raters are aware of candidates using gesture to support a covert strategy to secure ongoing L2 speech (Swain, 2000).

The comments on gaze and gesture are the most enlightening on the importance to the pair in communicating in the test. The example of the pair regaining communication after one speaks French and the other looks up would not be a regular occurrence but claims can be made based on that clear cut example that had they not looked there might have been a communication breakdown. It is harder to prove if they keep looking and it is successful or they don't look and it is not successful as in both cases there could be other factors.

V. CONCLUSIONS

It was put forward that the construct for peer-peer interaction was not fully described making it a problem for rating pairs or groups. A verbal report study showed non verbal communication to be part of rater construct of peer-peer interaction. This study describes rater and candidate perceptions of nonverbal communication including gaze, gesture and other body language, including posture and movement. It could be argued that nonverbal communication might be part of that 'intangible' element that Van moere indicated was affecting group rating. It has been described in this study which has implications for the construct of peer-peer interaction and the consequences that broadening the construct has on rating scales for peer-peer oral proficiency testing.

VI. FURTHER RESEARCH

Some potential areas for further study include

- To examine candidate discourse and gesture and to connect gesture to comprehension or engagement in next turn.
- A micro-analysis with gesture and gaze, the two main features identified by candidates and raters, would provide a unique insight into paired test talk

REFERENCES

- [1] Bonk, W. J. & G. J. Ockey. (2003). A many facet Rasch analysis of the second language group oral discussion task. *Language Testing* 20.1, 89-110.
- [2] Brown, A. (ed) (2000). An investigation of rater's orientation in awarding scores in the IELTS interview. Canberra, Australia.
- [3] Brown A, Iwashita N & McNamara, T. (2005). An Examination of Rater orientations and Test-taker Performance on English-for-Academic Purposes Speaking Tasks. Educational Testing Service: Princeton
- [4] Cassell, J. McNeill, D. & McCullough, K. E. (1999). Speech gesture mismatches: Evidence for one underlying representation of linguistic and non linguistic information. *Pragmatics and Cognition* 7, 1-33.
- [5] Conium, D. (2005). The Impact of Wearing a Face Mask in a high stakes Oral examination: an exploratory Post Sars Study in Hong Kong. *Language Assessment Quarterly* 2.4, 235-261.
- [6] Ducasse, A. M. (2009) 'An empirically based rating scale for interaction in a Paired Test'. In Brown, A. and Hill, K. (eds.) *Tasks and Criteria in Performance Assessment: Proceedings of the 28th Annual Language Testing Research Colloquium*. Frankfurt: Peter Lang 1-22.
- [7] Ducasse, A. M. (2007). How do candidates view interaction in a paired oral. In Gitsaki C (ed), *Language and languages*. In: Global and Local tensions Newcastle: Cambridge Scholars Publishing, 184-200.
- [8] Ducasse, A. M. & Brown, A. (2009). Assessing paired orals: Raters' orientation to interaction. In: L. Taylor & G Wigglesworth (eds) *Pair work in learning and assessment contexts*. *Language Testing*, 26.3, 423-443.
- [9] Green, A. (1998). Verbal Protocol analysis in language testing research: A handbook. Cambridge, England: Cambridge University Press.
- [10] Gullberg, M. (2006). Handling discourse: Gestures, Reference Tracking and Communication Strategies in early L2. *Language Learning* 56.1, 155-196.
- [11] Kellerman, S. (1992). I see what you mean: The role of kinesic behaviour in listening and implications for foreign and second language learning. *Applied Linguistics* 13, 239 - 281.
- [12] Lazaraton, A. (2004). Gesture and Speech in the Vocabulary Explanations of One ESL Teacher: A Microanalytic Inquiry. *Language Learning* 54.1, 79-117.
- [13] McNeill, D. (1992). Hand and mind: What the hands reveal about thought. Chicago: University of Chicago Press.
- [14] Meiron, B. E. (1998). Rating oral proficiency tests: A triangulated study of rater thought processes. Unpublished master's thesis. University of California Los Angeles.
- [15] Nunn, R. (2000). Designing rating tasks for small group interaction. *ELT Journal* 54.2, 169-178.
- [16] Orr, M. (2002). The FCE Speaking test: Using rater Reports to help Interpret Test Scores. *System* 30.2, 143-152.
- [17] Politt, A. and Murray, N. L. (1960). What raters really pay attention to? In M. Milanovic and N. Saville (eds.). *Performance testing, cognition and assessment: selected papers from the 15th Language Testing Research Colloquium*, Cambridge England: Cambridge University Press, 74-91.
- [18] Sueyoshi, A. & Hardison, D. M. (2005). The Role of Gesture and Facial Cues in Second Language Listening Comprehension. *Language Learning* 55.4, 661-699.
- [19] Swain, M. (2000). The output hypothesis and beyond: Mediating acquisition through collaborative dialogue. In J, P, Lantolf (ed.). *Sociocultural theory and the second language learner*, Oxford: Oxford University Press, 97-114.
- [20] Van Moere, A. (2006) .Validity evidence in a university group oral test. *Language Testing* 23.4, 411-440.



Ana Maria Ducasse was born in Bristol, and was educated in Australia.

B.A. Spanish Honours, Language and Literature Linguistics major, French minor, Monash University Victoria, Australia Graduate Diploma of Education, Method and Practice of Teaching Modern Languages: Spanish/French/Italian and ESL Monash University, Victoria, Australia

MA Applied Linguistics The University of Melbourne, Victoria, Australia PhD Applied Linguistics The University of Melbourne, Victoria, Australia

She is fluent in Spanish, French and Italian apart from English. She has worked as an educator and program director in various primary, secondary and tertiary institutions teaching Italian, ESL and currently

Spanish. Her current position is lecturer in Spanish at RMIT University, Melbourne and her research focus is oral assessment.

Her three most recent publications include:

Ducasse, A. M. & Brown, A. (2011) The role of interactive communication in IELTS speaking and its relationship to candidates' preparedness for study or training contexts. IELTS Research Reports Vol 12 www.ielts.org/researchers

Ducasse, A.M. (2011). Spanish writing: what do raters look for? In Candel Torres, I. et al (Eds.) ICERI 2011 Proceedings CD. IATED: Madrid www.iated.org ISBN: 978-84-615-3324-4

Ducasse, A.M. (2010). Interaction in Paired Oral Proficiency Assessment in Spanish: Rater and Candidate Input into Evidence Based Scale Development and Construct Definition, Peter Lang, Frankfurt. A specialist series: "Language Testing and Evaluation"

Dr Ducasse is a member of ILTA, HERDSA and ALTAANZ professional societies. She is on the CELE/AP scientific committee involved in selecting papers and organising the biennial Asia Pacific Espanol Lengua Extranjera conference. She is also on the editorial committee of PLTA: Papers in Language Testing and Assessment.

Finally, she has been awarded prizes for tertiary teaching: in 2011 the Vice Chancellor's award for Teaching Excellence at La Trobe University where she was team leader for the Spanish program application; in 2010 Outstanding Teacher in Learning and teaching and was a La Trobe University nomination for ALTC citation and in 2009 was awarded the La Trobe University Citation for Outstanding Contribution to Student Learning.