

An Integrative Multimodal Analysis of the Examiner-Examinee Interaction in the IELTS Speaking Examination

*Hanaa Youssef Shaarawy**

Introduction

Any oral communication is characterized by an interplay among verbal and non-verbal means of interaction. These different modes could be “embodied” (i.e. language, gesture, gaze, head movement, posture, proxemics) or “disembodied” (e.g. layout that includes furniture, print) (Norris, 2004). Among the different oral interactions is the examiner-examinee exchange that is common in contexts that require a score for a job, admission in a university, summative assessment or even travel for academic or immigration purposes. This study aims at investigating the examiner-examinee interaction from a multimodal perspective with special reference to the IELTS Speaking examination. IELTS Speaking is one of four modules that is to be completed alongside with Writing, Reading and Listening by anyone who wants to be admitted in some private universities in Egypt or to be qualified for travelling abroad for study or immigration purposes. The test comprises three parts. According to the different IELTS recognizable sites, Part One is labelled “interview”, Part Two is called “a long-turn” and Part Three is given the term “discussion”. The focus of Part One is to communicate on familiar topics; Part Two tests the ability of the test-taker to talk on a given topic for two minutes; and Part Three examines the ability of the candidate to discuss, justify and analyze issues related to Part Two topic. (For more information on IELTS, see “IELTS”, n.d.).

Literature review

Multimodality has been the interest of many researchers. Kress and Van Leeuwen (1996) on the one hand and O’Toole (1995) on the other discussed two major approaches for multimodality: the former applied theoretic generalizations to different texts, and the latter paid attention to specific texts and derived generalizations out of these texts to modify the theory (as cited in O’Halloran & Smith, 2012). Multimodality has been used to analyze digital texts by many researchers (e.g. Alfonso & Giralt, 2013; Jewitt, 2013a; Jewitt, 2013b; O’Halloran, Tan, Smith & Podlasov, 2011; O’Halloran, Marissa, Podlasov &

* Assistant Professor of Linguistics, Faculty of Languages, MSA University.

Tan, 2013). Multimodality has also been the framework used by much research in the field of education (e.g. Bezemer & Jewitt, 2010; Jewitt, 2003; Jewitt, 2007; Jewitt, 2008; Rowe, 2012). Zollo (2013) integrated multimodality with critical discourse analysis when investigating the Council of Europe's (COE) persuasive communication campaigns.

The IELTS Speaking examination has also been subject to a plethora of research. Some studies discussed the marking of the exam (e.g. Brown, 2003; Nakatsuhara, Inoue & Taylor, 2017; Seedhouse, Harris, Naeb & Üstünel, 2014). Others compared the examinee and examiner performance features through two different modes of delivery: face to face and video-conferencing (e.g. Nakatsuhara, Inoue, Berry & Galaczi, 2016). Researchers also examined the relationship between Listening proficiency and Speaking performance (e.g. Nakatsuhara, 2009). Nevertheless, no research on IELTS has tackled the examiner-examinee interaction, the focus of the present study.

Theoretical background and framework

Challenging the concept that language is the superordinate means of communication, Norris (2004) offered a framework of multimodal interaction analysis where she integrated multimodality (taking into consideration such semiotic features like gestures, print and layout), "interactional sociolinguistics" that focuses on real time interaction, and "mediated discourse" whose emphasis is on the mediated action (p. 10).

Norris (2004) chose "action" to be the "unit of analysis", and she differentiated between "higher-level" actions and "lower-level" actions. Higher-level actions are marked by a beginning and an end and they are made up of multiple chains of lower-level actions (the smallest unit of analysis). An example given by Norris to illustrate this concept is a meeting among three friends. The meeting is a higher-level action that has a beginning and an end, and it is made up of multiple chains of gestures, postural shifts, gaze shifts and intonation units. Higher-level actions can also encompass other higher-level actions. In the same example of the meeting among three, there could be a conversation between two of the members (Norris, 2004, pp. 13-14).

Norris (2004) categorized modes into two different types: embodied modes and disembodied modes that she also called "frozen actions". She defined embodied modes as the modes individuals use to express their feelings, thoughts and perceptions. These modes could be spoken language, gestures or even a print mode (i.e. written texts). On the other hand, disembodied modes are material objects developed by someone prior to an interaction. These objects might include a magazine on a table, furniture or pictures on the wall. For these objects to be present, chains of lower-level actions are performed. For a magazine to be

present, someone must have bought it and placed it on the table. Perceiving a chain of lower-level actions for the presence of the magazine, furniture or any other material object, necessitates the presence of higher-level disembodied actions that are labelled “frozen actions” (p. 14). Norris added that one mode can function as an embodied mode or as a disembodied mode depending on context. For example, the print mode can function as an embodied mode if the individual uses it to express his feelings, thoughts or perceptions, or it can function as a disembodied mode when s/he reacts to the print performed by others (pp. 41-44). Within this framework, Norris discussed several communicative modes: “spoken language”, “proxemics”, “posture”, “gesture”, “head movement”, “gaze”, “music”, “print” and “layout”, and she affirmed their interdependence on one another. (Norris, 2004, pp. 15-57).

Closely associated with the above framework is the concept of “modal density”. Norris (2004, 2011) discussed modal density in terms of “modal intensity” and “modal complexity”. Modal intensity has to do with the weight and importance of a mode which depends on the context of interaction. Within a given context, if a mode changes the higher-level action radically if it is discontinued, this mode has high intensity; if a mode changes the higher-level action slightly if it is discontinued, this mode has a mid-intensity; and if the mode does not change the higher-level action if it is discontinued, this mode is of low intensity. On the other hand, modal complexity has to do with the multiplicity of different modes and it is achieved when there is an interplay among several communicative modes. (For a detailed discussion and examples of modal density, see Norris, 2004, pp. 79-91 and Norris, 2011, pp. 93-110).

In order to understand the relativity of modal density in interaction, Norris (2004, 2011) introduced the notions of “attention” and “foreground-background continuum”. Adopting Van Leeuwen’s (1996) notion of dividing any heard sounds into three groups and then “hierarchizing” these groups, treating some as more important than others (as cited in Norris, 2011), Norris (2011), divided higher-level actions (including disembodied/frozen actions) that are achieved through the employment of different communicative modes into groups, hierarchized them according to which higher-level action is attended/reacted to and which higher-level action is present, but “disattended” and treated as something listeners should not react to. Norris (2004) emphasized that

The higher-level action that a participant highly attends to and/or highly reacts to, and/or highly acts upon, is in the *foreground* of their *attention/awareness* ...

The higher-level action that a participant attends to in some degree and/or reacts to in some degree, and/or acts upon in some degree, is in the *mid-ground* of their *attention/awareness* ...

The higher-level action that a participant is only decreasingly aware of, disattends, and/or does not react to, and/or does not act upon, is in the *background* of their *attention/awareness*. (p. 97, italics in original)*

All that has been presented in the above theoretical background constitute the framework upon which the present study is based.

Research questions

The purpose of this study, as mentioned above, is to investigate the different modes of communication used in the examiner-examinee interaction in the three parts of the IELTS Speaking examination. The following research questions aid in this investigation:

1. What are the different modes of communication utilized in each part of the IELTS Speaking examination?
2. How do these modes unite to construct higher-level actions that suit the conventional features of each part of the examination?
3. What is the relationship between higher-level actions on the one hand and modal density with the foreground-background continuum of attention on the other?

To answer these questions, a qualitative analysis is carried out.

Method

Materials

The IELTS Speaking examination constitutes the data for the present study. All the IELTS Speaking examinations available on www.youtube.com from recognized IELTS sites (e.g. www.ielts.org and Cambridge English Language Assessment) were watched and twenty random samples representing the three parts of the exam were selected. I used the videos available on the internet, since I had no access to live IELTS to make my own recordings. The method of transcribing data was adopted from Norris (2004). Images were incorporated in the transcription to enrich the analysis by giving it details that cannot be captured by just transcribing spoken language. Albeit Norris labelled the video captures “Plates” (with a capital P), they are termed “figures” in the present study so as

* (For a detailed discussion of the foreground-background continuum, see Norris, 2004, pp. 95-111 and Norris, 2011, pp. 47-50).

not to confuse them with other meanings of “plates”. All other terminology present in the analysis: “higher-level” actions, “lower-level” actions, “embodied” and “disembodied” modes, “frozen actions”, “modal density” and “foreground-background continuum of attention” are adopted from Norris. (See theoretical background above).

Data Transcription, Analysis and Discussion

Before transcribing and analyzing the data, it is important to note that the whole IELTS Speaking examination is a higher-level action that has an opening and a closing. Within this higher-level action, there are three other higher-level actions instantiated in the three parts of the exam. Each part has a beginning and an end. In what follows, each of the examination parts is transcribed, analyzed and discussed to arrive at the multimodal interaction features that characterize this particular part.

Since the main aim of this study is to analyze the examiner-examinee interaction that characterizes the IELTS Speaking examination in general and not to analyze one individual examination, only the formulaic language of the examiner and the examinee is transcribed below, since the topics the examiner chooses from vary, and the answers of examinees vary. Figures presented in the analysis were selected to represent the shared modes found in all the IELTS Speaking tests. They are not idiosyncratic features of a specific interaction in a specific test, therefore spoken language transcribed is also deemed representative of spoken language found in all the IELTS Speaking tests not the language of a specific examiner, a specific examinee or a specific test. Analyzing the language of specific participants in an IELTS Speaking examination does not lie within the scope of the present study, since the study is not dealing with the different linguistic features that aid in getting high scores in the examination. In order not to confuse the transcription of the spoken language with the analysis, the mode of “spoken language” is transcribed in italics.

Part One (4-5 minutes): “Interview”

After analyzing all the data of the present study, it was found that all the “interviews” of Part One have a formulaic template that is made up of four sections. The first section is a morning or an afternoon greeting by the examiner and the examinee with an introduction of oneself to break the ice, the second section is a set of questions on either the place where the candidate lives or what s/he does in life, the third section is on a general topic, and the fourth section is on another general topic. The third and fourth sections are made up of a list of familiar topics from which examiners choose.

Section 1 template: greetings and introducing oneself.

Examiner: Good afternoon. My name is Michael Austin. Can you tell me your full name please?

Examinee: [Answers vary]

Examiner: Where you are from?

Examinee: [Answers vary]

Examiner: Can I see your identification please?

Examinee: Yes, sure. [hands on the ID]

Examiner [Checking that the person in the ID is the same as the person sitting]: Thank you. That's fine. [Returns the ID]



Figure 1. The modes of gaze, head movement and print

An afternoon greeting from the examiner and the examinee marks the launch of the whole examination, the commencement of Part One and the beginning of section one at the same time. The examiner, then, introduces himself and asks for the examinee's full name, nationality and ID. The examiner ended section one and marked the beginning of section two by saying *"Now in this first part"* and denotes that he is going to ask the examinee some questions about himself/herself *"I'd like to ask you some questions about yourself"*.

This section is a higher-level action, since it has a beginning and an end, and it is made of chains of lower-level actions. These chain are composed of different modes: "spoken language" discerned in a series of adjacency pairs between the examiner and the examinee articulated in friendly tones from the examiner's part, "gestures" recognized in the examinee handing on the ID and the examiner giving it back, "print" mode (Figure 1a) "gaze shifts", and "head movements" (Figure 1b) illustrated by the examiner when checking the identity of the examinee against the papers on the table, "proxemics" which is the distance between the examiner and the examinee, as they are sitting across each other revealing the formality of the relationship (Figure 2) and the "layout" mode illustrated in the physical environment of the examination: a table, two chairs across each other, papers in front of the examiner, a tape recorder (placed in a circle to the left of the examiner in the middle of the table), a timer that calculates

the time of each part of the examination (placed in a circle in front of the examiner to her right) and the examinee identification card placed in front of her (Figure 2).



Figure 2. Proxemics and layout

It is worth noting that the “proxemics” and the “layout” modes characterize the authenticity of the three parts of the examination. Different ‘proxemics’ and/or different “layout” would create a different atmosphere that does not belong to IELTS in any sense. In other words, these two modes are core elements in the IELTS Speaking examination setting. It is also worth noting that the mode of “spoken language” in all the data of the study is characterized by a friendly tone articulated slowly in a clear voice with conventional intonation units that accompany any interaction: falling intonation at the end of sentences and rising intonation when asking questions, pauses after meaningful units together with a medium steady pitch. These conventional features of the mode of “spoken language” are concomitant with the IELTS Speaking examination where international non-native English candidates from all over the world sit for this examination. The Speaking proficiency level is not known to the examiners. So, all the instructions, questions and prompts are standardized to address all levels, which secures the validity and reliability of the test.

Section 2 template: Candidate’s way of living or place of living.

Examiner: Now, in this first part, I’d like to ask you some questions about yourself. Let’s talk about what you do. Do you work or are you a student?

Examinee: I am a student.

Examiner: What subject are you studying?

Examinee: [Answers vary].

Examiner: Why did you choose this subject?

Examinee: [Answers vary]

Examiner: What do you hope to do when you finish your studies?

Examinee: [Answers vary]



Figure 3. The modes of gaze, head movement and print

This section is also a higher-level action whose beginning is marked by “*Now, in this first part*” and it comes to an end when the examiner uses one of the topic shifts: “*Let’s talk about ...*”, “*Let’s go on to talk about ...*”, “*Tell me about ...*” or “*Now, let’s talk about ...*”. This higher-level action is made up of multiple chains of lower-level actions presented in different modes: “spoken language” by means of a chain of adjacency pairs between the examiner and the examinee followed sometimes by “*why/why not?*” to prompt the examinee to elaborate on her talk, “gaze shifts” from reading the questions (Figure 3a) to looking at the examinee together with an upward sagittal “head movement” (Figure 3b) to follow with the examinee and show attention and interest or in nodding to show interaction and to encourage the examinee to go on talking, “print” mode that is also utilized by the examiner as an essential mode in the interaction and the modes of “proxemics” and “layout” that are constant throughout the whole examination.

Section 3 template: a general topic.

Examiner: Now, let’s talk about weekends. What do you usually do at the weekend?

Examinee: [Answers vary]

Examiner: What do you think you do next weekend? Why?

Examinee: [Answers vary]

Examiner: Do you enjoy weekends now more than you did when you were a child?

Examinee: [Answers vary]

Examiner: How important it is for you to relax at the end of the week?

Examinee: [Answers vary]



Figure 4. The modes of gaze, head movement and print

In section three, the examiner has a list of optional general topics to choose from. This section constitutes a third higher-level action in Part One. Its beginning is marked by one of the topic shifts presented in section two, and its end is also marked by one of the topic shifts. This topic shift is either the same as that of its beginning or different, but it is nothing other than the topic shifts presented above. Similarly, as section two, this section is made of chains of lower-level actions discerned in the modes of “spoken language”, “gaze shifts”, “print” mode, “head movements”, “proxemics” and “layout”. Figure 4a shows the “print” and “gaze” modes while reading the questions and Figure 4b illustrates the “gaze” mode accompanied by a sagittal “head movement” and a smile while looking at the examinee to denote friendliness and interest in what the examinee is saying.

Section 4 template: a general topic

Examiner: Let's talk about music. What sort of music do you usually enjoy listening to?

Examinee: [Answers vary].

Examiner: Has the kind of music you like changed over the years?

Examinee: [Answers vary]

Examiner: Do you prefer listening to live music or recorded music? Why?

Examinee: [Answers vary]

Examiner: Do you think listening to music helps you study?

Examinee: [Answers vary]



Figure 5. The modes of gaze, head movement and print

This section is the fourth higher-level action in Part One. Its beginning is marked by one of the topic shifts exactly as section three, but its end is marked by the beginning of Part Two of the whole examination. Other than that, it is a carbon copy of section three in the usage of the same modes that constitute chains of lower-level actions as apparent in the “gaze” and lowering the “head” while reading the question in Figure 5a and moving it up with a sagittal movement while looking at the examiner in Figure 5b.

All that has been presented delineates that Part One is made of four higher-level actions, and that each action comprises chains of lower-level actions instantiated in “spoken language”, “gaze shifts”, “print”, “head movements”, “proxemics” and “layout” together with “gestures” that are utilized only in section one. To the examiner, all the modes including the mode of “print” and the mode of “layout” are embodied modes, since they are either deployed to determine the topics that are to be communicated (“spoken language”, “gaze shifts”, “print”, “head movements”, “proxemics”) or they play a vital role in the interaction as setting the time, tape recording the examination, and arranging chairs show formality (the mode of “layout”). Thus, all modes are integrated into higher level actions to communicate formulaic messages, which results in having all these higher-level actions placed on the foreground of the examiner’s attention. On the other hand, only higher-level actions that include “spoken language” are placed in the foreground of the examinee’s attention. The modes of “print” and “layout” are frozen actions to the examinee, since s/e does not attend to them, and they are placed in her background attention. The examinee is only responsible for aligning her language to international speaking standards. She is not responsible for tape recording the examination or for taking care of the time. These tasks are the examiner’s along with assessing the examinee’s talk.

Consequently, some of the modes used in Part One show higher modal density than others. To the examiner, “spoken language” and “print” are of paramount importance, and the higher-level action will be drastically affected if

they are discontinued, and that is why they are never discontinued in any of the videos selected for the analysis. Thence, high modal density is achieved by the examiner through the intensity of “spoken language” and “print”, which complies with the international identity of the IELTS examination. Rigorous spoken language that is read from a script guarantees the test validity and reliability all over the world irrespective of the place of its administration; otherwise the test will lose its standardization and recognition throughout the whole world. On the other hand, “spoken language” is the sole essential mode for the examinee in this Part, and if it is discontinued, the higher-level action will be drastically affected. So, high modal density is achieved by the examinee through the intensity of “spoken language”, which is normal, since the main purpose of the examination is to test the Speaking proficiency of the candidate. Concerning the other modes used, “gestures” affect the higher-level action of section one only, since if the examinee refuses to hand on his/her ID, s/he will be in trouble as s/he has to prove his/her identity. So, gestures have high modal density only in section one. “Gaze shifts” and “head movements” aid in presenting the performance in a friendly courteous encouraging natural manner, but they do not affect the content in any way. Therefore, these modes are of mid modal density. They do not perform low modal density, since their absence will create tough, unfriendly atmosphere which might affect the performance of the examinee.

All that has been presented above responds to the three research questions raised in this study concerning Part One, and it shows the formulaic interaction between the examiner and the examinee in Part One of the IELTS Speaking examination.

Part Two (3-4 minutes): “A long talk”.

After analyzing all the data of the present study, it was found that Part Two is also presented in a formulaic rigid pattern. All the examiner’s modes of communication are standardized. In this part, the examiner has a number of prescribed topics from which s/he chooses. S/He gives the examinee “a topic card” that includes question prompts to a book the examinee has read, a place s/he has visited, a film s/he has watched, a person s/he will never forget, a gift s/he has recently presented or any topic that allows the examinee to talk about something that has happened in his/her life. The examinee is given one minute to prepare his/her talk, and then s/he talks for one to two minutes. (For more information about Part Two, see “IELTS Liz,” 2017). Part Two is made up of six sections with six formulaic templates. Section one is a non-verbal shift between Part One and Part Two, section two is a verbal shift from Part One to Part Two, section three is the introduction of the topic cue, section four is the

one minute given to the examinee to prepare notes, section five is the two minutes talk by the examinee and section six is the end of Part Two.

Section one: the non-verbal shift from Part One to Part Two. The shift from Part One to Part Two is achieved through a strict interplay among the mode of print, the mode of head movement, the mode of gesture and the mode of gaze with a friendly smile on the face that delivers support and comfort as shown in Figure 6.



Figure 6. The shift from part 1 to part 2 through turning the page and gazing at it

The examiner uses the papers in front of her (the print mode) and her fingers (the gesture mode) (Figure 6a) to turn the page (Figure 6b) with a deictic head movement (the head movement) and gaze at the new page to change the focus from Part One to Part Two (Figure 6c). All modes used in Figure 6 are embodied modes, since they merely mark the beginning of Part Two, and these modes in collaboration construct a higher-level action that begins in Figure 6a and ends in Figure 6c. This higher-level action lies in the foreground of the examiner's attention, but in the background of the examinee's, since the examinee does not react or respond to it as apparent in all the videos selected for analysis. The examinee attends to non-verbal communicative modes if they are accompanied by the mode of "spoken language" as shown below, which is ordinary in an international Speaking test.

Section two: the verbal shift from Part One to Part Two. To align to the norms of international Speaking examinations, the shift between Parts One and Two is signaled verbally using the mode of "spoken language" together with the mode of "head movement", the mode of "gaze" and the mode of "print" as apparent in Figure 7.

Examiner: Now I'm going to give you a topic, and I'd like you to talk about it for one to two minutes. Before you talk, you will have one minute to think about what you are going to say and you can make some notes if you wish. Do you understand?

Examinee: Yes.



Figure 7. The modes of spoken language, print, head movement and gaze at the beginning of Part Two

The examiner combines the modes of spoken language, print, head movement and gaze at the beginning of Part Two when he uses recurrent up and down sagittal head beats with lateral slants first to read and say what is written on the papers in front of him (Figure 7a) and then to look at the examinee (Figure 7b), and he uses the modes of spoken language with the mode of gaze when asking the examinee “*Do you understand?*” This examiner, as apparent in Figure 7, is clasping his hands; however, this gesture can be deemed irrelevant and idiosyncratic as it varies from one examiner to another.

All the modes in Figure 7 are embodied modes that form the second higher-level action in Part Two. This higher-level action starts with the examiner saying, “*Now I’m going to give you a topic*” and ends with the examinee answering “*Yes*”. It is placed in the foreground attention of the examiner and the examinee, since the latter must respond to the instructions. Moreover, modal density is achieved in this section through the intensity of the modes of “spoken language” and “print” from the examiner’s part and through the intensity of the mode “spoken language” only from the examinee’s part. The modes of “head movement” and “gaze” only aid the interaction by showing interest, encouragement and follow up.

Section three: introduction of the topic card. The examiner, then, integrates formulaic language with formulaic modes to present the topic which the examinee is to talk about (Figure 8).

Examiner: Here is some paper and a pencil for making notes, and here is your topic. Please don’t write anything on the booklet. I’d like you to describe a well-known person you like or admire.

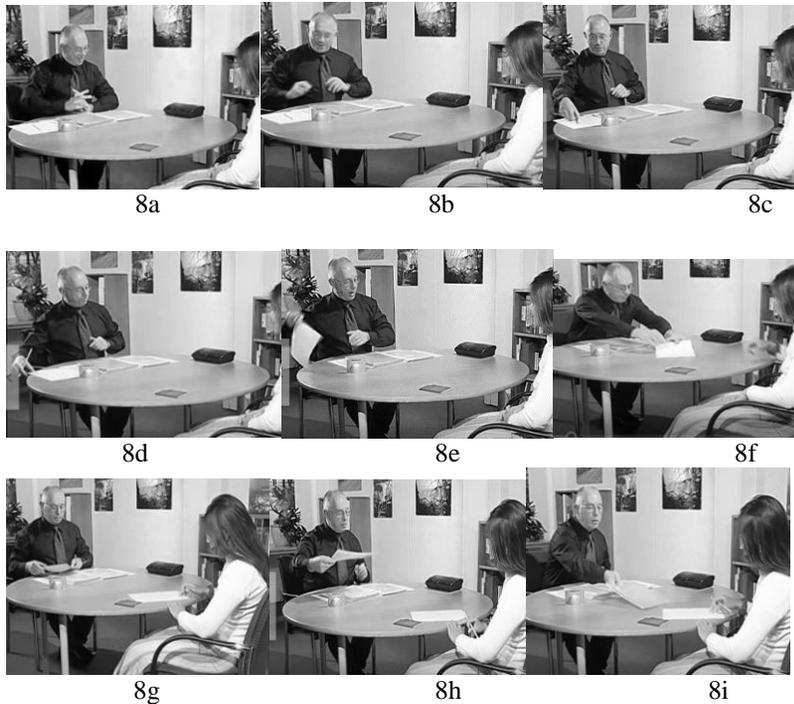


Figure 8. Integration of different modes

In Figure 8, the examiner has not only used the mode of “spoken language”. He has also integrated other modes of communication to complement “spoken language”. These modes are the mode of “gesture”, “gaze”, “posture” “layout” and “print”. The examiner leans slightly backwards (posture mode) (Figure 8a) unfolds his fists (gesture preparation stage), looks (the gaze mode) right at the pencil on the desk (the layout mode) (Figure 8b), holds it (gesture stroke stage) and starts the instructions (the beginning of the spoken language mode) (Figure 8c). While holding the pencil, the examiner opens his thumb and index finger (gesture preparation stage) (Figure 8d) to hold the paper (gesture stroke stage) (Figure 8e) and leans forward (the mode of posture) to put the paper and pencil (the mode of print) in front of the examinee (Figure 8f). After putting the pencil and paper in front of the examinee, he uses his right hand, takes the booklet (gesture preparation and print), which includes the topic cue (Figure 8g, 8h), and lays it down (the stroke of the gesture) in front of the examinee with a forward lean (posture mode) (Figure 8i). Figures 8c to 8i are accompanied by “*Here is some paper and a pencil for making notes, and here is your topic. Please don’t write anything on the booklet.*” After that, the examiner reads (the mode of gaze)

the topic he chooses for the examinee “*I’d like you to describe a well-known person you like or admire*”.

All lower-level actions represented in the different modes deployed in this section unite to construct the third higher-level action in Part Two that starts with “*Here*” and ends with mentioning the topic about which the examinee will talk. This higher-level action is placed in the foreground attention of both interlocutors, since it introduces the topic that assesses the Speaking proficiency of the examinee. Modal density is achieved in this section through complexity by integrating the modes of spoken language, print, gaze, gesture, posture and layout, and it is also achieved through the intensity of all these modes.

Section four: one-minute note-preparation. The candidate is given one minute to write notes on what she is going to describe (Figure 9).



Figure 9. embodied print and embodied layout

The one minute given to the examinee to write notes represents the fourth “higher-level” action within Part Two, since it is marked by a beginning and an end. The beginning commences with the start of the one minute and terminates with the end of it when the examiner says, “*All right.*” (See below). During this minute, the modes of print (booklet) and the mode of layout (the pencil, paper, desk and chair) are utilized by the examinee (Figure 9). These two modes are embodied, since the examinee is using them to write down her thoughts, perceptions and feelings. For the first time in the examination, this one-minute higher-level action that does not include any spoken language is placed in the foreground attention of the examinee. Moreover, modal density is achieved through the intensity of the modes of print and layout because if these two modes are absent, this higher-level action will be radically affected, which complies with the international norms of the IELTS Speaking examination that require this one-minute preparation period.

Section five: Two-minutes talk by the examinee. In this section, the examinee is given two minutes to talk about the prompt written on the card.

Examiner: All right. Remember you have one to two minutes for this. So, don't worry if I stop you. I'll tell you when the time is up. Can you start speaking now please?

Examinee: [Talks for one to two minutes].

Examiner: Thank you. Would you like to meet this person?

Examinee: Yeaas.

After the one-minute preparation ends (section four), the examinee, in this section, utilizes what she has written on paper to help her in the talk. By doing so, the examinee is using the mode of spoken language, the print mode to recall her thoughts and the gaze mode to read what has been written (Figure 10).



Figure 10. Spoken language, print and gaze

(It has been noticed in two of the videos selected for analysis that if the examinee talks for less than two minutes, the examiner helps by saying, “*Can you tell me more about that?*” accompanying her mode of “spoken language” by the modes of “gaze” and “gesture” as shown in Figure 11)



Figure 11. The mode of spoken language, gaze and gesture

As seen in Figure 11, the examiner uses the mode of gesture while saying “*Can you tell me more about that?*” She raises her hand from the still position (Figure 11a) directing it to the topic cue in front of the examinee (preparation) (Figure 11b), and in the third image, she looks (gaze mode) and points at the question she wants the examinee to elaborate on (stroke) with a forward body posture to

aid the deictic gesture with the index finger (Figure 11c). Then, there is usually a retraction to the rest position (although the retraction is not apparent in this segment).

Examinee: [Answers vary].

Examiner: *Thank you.*

It has also been noticed in the videos selected for the analysis that the examiner may follow-up the examinee's talk with a question or two if the time allocated for Part Two has not ended up.

This two-minute talk is the fifth higher-level action in Part Two since it has a beginning marked by the examiner saying, "*All right.*" and an end that is also marked by the examiner when he says, "*Can I take the booklet and the paper and pencil back please?*" (See below). In this higher-level action, the modes of spoken language, print and gaze are intertwined by the examinee to respond to the prompt. All these modes are embodied modes that express the examinee's thoughts and perceptions. Modal density is achieved through the intensity of the three modes, and the higher-level action is placed in the foreground attention of the examiner and the examinee.

Section six: the end of Part Two. The examiner, then, ends Part two by saying, *Examiner: Thank you. Can I take the booklet and the paper and pencil back please?*

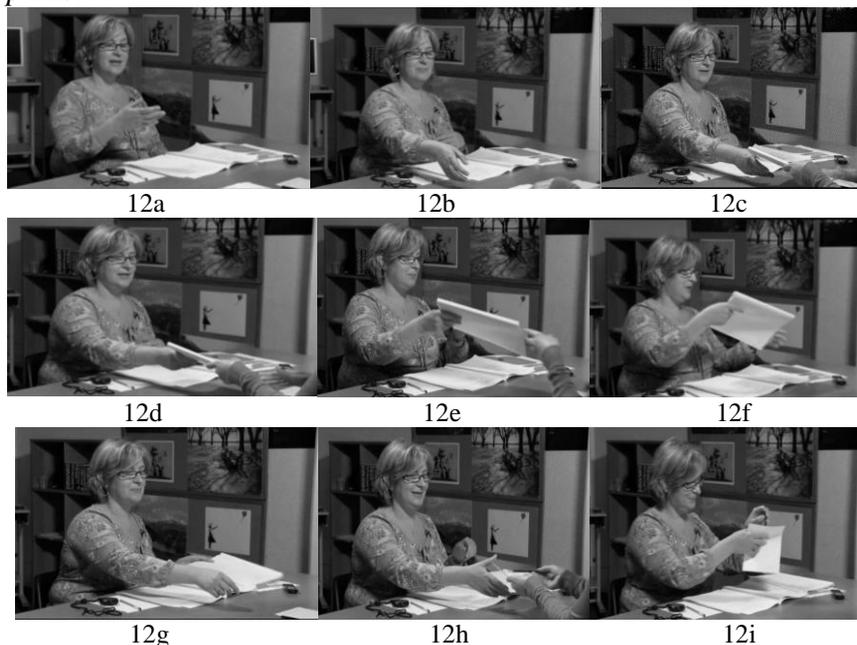


Figure 12. Gesture, gaze, spoken language and print

In Figure 12, the examiner uses the mode of gesture, gaze and spoken language to take back the print from the examinee. She starts by saying “*Can I take the booklet and the paper and pencil back please?*” and raising her hand from the rest position, moving it forward (Figure 12a) and downward (gesture preparation) (Figure 12b) to take the booklet (gesture stroke) (Figures 12c-12d). She, then, takes the booklet and puts it aside next to the papers found in front of her (gesture post stroke) (Figures 12e-12f), and repeats the same procedures to take the paper and pencil. She moves her hand forward and downward (gesture preparation) (Figure 12g); then, she holds the paper and pencil (gesture stroke) (Figure 12h) and she puts them aside (gesture post stroke) (Figure 12i). While doing so, she keeps looking (the mode of gaze) at the booklet, paper and pencil (print) to make sure she is having them all.

All the modes in Figure 12 are embodied modes that construct the sixth higher-level action in Part two. It has a beginning and an end. It starts with the examiner saying, “*Can I take the booklet and the paper and pencil back please?*” and ends with Figure 12i when the examiner puts the paper and pencil aside. This higher-level action is made of chains of lower-level actions instantiated in one utterance (spoken language) and a series of gestures, gaze and deployment of print. Modal density is created in this higher-level action through complexity illustrated in the multiplicity of the modes created and through intensity demonstrated in the importance of each mode. The high modal density of this higher-level action places it in the foreground of the examiner’s attention. If any of the modes is discontinued, Part two is not concluded and the whole examination stops.

The examiner taking the booklet and the paper and pencil back is not an idiosyncratic feature of the examiner presented in Figure 12. It is an observed behavior among all the examiners present in the videos selected for analysis, and it is achieved through “spoken language”, “gesture”, “gaze” and “print” in all the videos of the analysis. Thus, it is a ritualized action in concordance with the international standards of the IELTS examination. Everything is provided to the candidate at the beginning of the examination and taken from him/her at the end. The examinee should not take the booklet or the paper s/he wrote notes on with him/her outside the examination room for the sake of the privacy of the test. Moreover, the same booklet and pencil are re-used by other candidates, and that is why the examiner asks the examinee not to write anything on the booklet at the start of Part two “*Please don’t write anything on the booklet.*” (See above). The examiner, on the other hand, makes sure that everything is returned as one of his/her responsibilities in administering the examination.

All that has been presented in Part Two shows that this part is made of six higher-level actions that are created from unified chains of lower-level actions.

These lower-level actions are instantiated in different embodied modes of spoken language, print, layout, gesture, posture, gaze and head movement. All these higher-level actions are placed in the foreground attention of the examiner, and they are either placed in the foreground or the background attention of the examinee according to the moment of interaction. Modal density is achieved through the complexity and amalgamation of the different modes, and it is achieved through the intensity of the modes of spoken language, print, layout, gestures and postures. All this responds to the three research questions upon which the study is based and foreshadows the formulaic nature of the interaction that has to be so rigid to warrant the validity and reliability of the IELTS Speaking examination across the whole world.

Part Three (4-5 minutes): “Discussion”

The examiner taking the booklet, paper and pencil back in Part Two marks the end of Part two and the start of Part Three. Part Three starts with a formulaic sentence:

Examiner: We have been talking about [the topic of Part Two] hobbies, and I'd like to discuss with you one or two more general questions related to this.

[The examiner reads very little and discusses the topic with the examinee]



Figure 13 The modes of gaze, head movement and print

Part Three is an extension to Part Two. The same topic the examinee talks about in Part Two is elaborated on in Part Three through a discussion between the examiner and the examinee. Part Three is a higher-level action that starts with “*We have been talking about hobbies*” and ends with “*That is the end of the Speaking test*”. The modes of “gaze”, “head movement” and “print” accompany the mode of “spoken language” as apparent in Figure 13. The examiner reads from the papers in front of her (Figure 13a), and then she moves her head up while looking at the examinee (Figure 13b). Part Three appears to be of a semi-formulaic nature unlike Parts One and Two. To clarify, in Parts One and Two, questions asked to the examinee and instructions given by the examiner all have the same wordings in all the examinations that constitute the

data of the current study. Part Three, on the other hand, follows a rigid framework but different wordings that change among examiners even when tackling the same topic. For example, when the topic to be discussed is “*Hobbies*”, one examiner, after saying the formulaic sentence mentioned above, says, “*Let’s think about current and future hobbies. What are the most popular types of hobbies in your country?*”, another examiner says, “*Let’s think about hobbies now and hobbies in the future? What do most people like to do in your country?*”. So, both examiners discuss hobbies at present and in the future and ask the candidate to talk about the well-known hobbies in his/her country, but they use different wordings, unlike Parts One and Two where the language is identical among all examiners. This may be due to two factors: the incremental level of difficulty of the examination and the speaking proficiency level of the candidate.

Regarding the incremental level of difficulty, the examinee responds to a couple of questions on a familiar topic in Part One. This response is in one or two utterances just to convey the message, but the response lasts for two minutes long in Part Two on a prepared topic showing an increase in the level of speaking proficiency. In Part Three, the examinee is not given any time for preparation, like Part Two, and s/he has to discuss and analyze abstract issues by responding to the examiner’s questions fluently, which shows a higher level of speaking proficiency.

Concerning the speaking proficiency level of the examinee, it is a well-known fact that any non-native English speaker from any part of the world can sit for the IELTS Speaking examination. This non-native speaker may be of a low or a high speaking proficiency level. Examiners phrase the questions accordingly; i.e. if the candidate’s speaking level is advanced, the examiner uses advanced language in Part Three, and if the candidate is of a beginner or intermediate level, the examiner uses a matching language to prompt the candidate to talk as much as possible while retaining the framework of the prompt, which shows the great responsibility and attention of a certified IELTS Speaking examiner. At the end of Part Three, the examiner thanks the examinee and ends the test by saying, *Examiner: Thank you very much. That is the end of the Speaking test.*

Part Three is a higher-level action that has a beginning and an end. It is made of chains of lower-level actions illustrated in the modes of spoken language, gaze, head movement and print to the examiner and in the mode of spoken language only to the examinee. This higher-level action is intensely placed in the foreground attention of both participants, and modal density is primarily achieved through the high intensity of spoken language, print, gaze and head movement from the examiner’s part and through spoken language from the examinee’s parts. Other modes (e.g. gestures) may occur by examiners, but they

are all idiosyncratic features that do not affect the international identity of the examination in Part Three.

All that has been presented in Part Three responds to the three research questions upon which the current study is based and reflects the rigid international characteristics of the IELTS Speaking examination.

Conclusion

The purpose of this study was to explore the different communicative modes in the examiner-examinee interaction with a special focus on the IELTS Speaking examination. Three research questions were posed to investigate the different modes utilized in the interaction, to examine how these modes unite to construct higher-level actions that correspond to the conventional nature of each part of the examination, to construe the relationship between higher-level actions, modal density and the foreground-background continuum of the participants' attention. Findings indicate that there is a rigorous systematized interplay among different modes of interaction, and although "spoken language" is not the only superordinate mode deployed in all parts of the examination, it, together with "print", "layout" and "proxemics" verifies the identity of the IELTS Speaking examination. Modal density through complexity and intensity is highly affected by the conventional features that characterize each higher-level action. Higher-level actions are foregrounded or backgrounded in the participants' attention depending on the purpose of these actions in the interaction.

The present study is limited to the fact that the data used are ready-made videos available on the internet and not videos recorded by the researcher. Consequently, no interviews with the same examiners or the examinees were accessible to find out about how they regarded the use of different modes, which would validate the study. No observations were practiced exploring any other elements not taken by the camera in the available internet videos.

For future research, it is recommended to hold some interviews specially with examinees to know how they regard the embodied and the disembodied modes, whether the frozen actions affect their concentration and, consequently, their scores, and to discern how they regard the whole setting of the exam. Interviews with examiners are also needed to recognize some tips, other than those available on the internet, that can help test-takers score high bands with the aid of different communicative modes.

References

- Alfonso, R. & Giralt, M. (2013). Multimodal discourse and oral language skills in digital portfolios for learning foreign languages. *Linguistics and Literature Studies* 1(2), 111-118. DOI: 10.13189/lis.2013.010208
- Bezemer, J. & C. Jewitt (2010). Multimodal Analysis: Key issues. In L. Litosseliti (ed), *Research Methods in Linguistics*. London: Continuum. pp. 180-197. Retrieved from http://eprints.ncrm.ac.uk/2886/1/17_Jewitt_SAGE_Handbook_.pdf
- Brown, A. (2003). An examination of the rating process in the revised IELTS Speaking Test. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_rr_volume06_report2.ashx
- IELTS. (n.d.). Retrieved from <http://www.ielts.org/>
- IELTS Liz: Free IELTS Videos, Lessons & Tips. (2017). Retrieved from <http://ieltsliz.com/>
- Jewitt, C. (2003). Orchestrating debate: A multimodal analysis of classroom interaction. DOI: 10.1111/1467-9345.3702004
- Jewitt, C. (2007). A multimodal perspective on textuality and contexts. *Pedagogy, culture and Society*, 15(3), 275-789. DOI: 10.1080/14681360701601937
- Jewitt, C. (2008). Multimodal discourses across the curriculum. *Encyclopedia of Language and Education*, 1098-1108. DOI: 10.1007/978-0-387-30424-3_84
- Jewitt, C. (2013a). Multimodal methods for researching digital technologies. Retrieved from https://multimodalmethodologies.files.wordpress.com/2013/03/chapter_17.pdf
- Jewitt, C. (2013b). *Learning and communication in digital multimodal landscapes*. London: Institute of Education Press. Retrieved from <http://web.b.ebscohost.com/>
- Nakatsuhara, F. (2009). The relationship between test-takers' listening proficiency and their performance on the IELTS Speaking Test. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_rr_volume12_report4.ashx
- Nakatsuhara, F., Inoue, C., Berry, V., & Galaczi, E. (2016). Exploring performance across two delivery modes for the same L2 speaking test: Face-to-face and video-conferencing delivery A preliminary comparison of test-taker and examiner behavior. Retrieved from <https://www.ielts.org/-/media/research-reports/ielts-partnership-research-paper-1.ashx>
- Nakatsuhara, F., Inoue, C., & Taylor, L. (2017). An investigation into double-marking methods: comparing live, audio and video rating on performance on

- the IELTS Speaking test. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_online_rr_2017-1.ashx
- Norris, S. (2004). *Analyzing multimodal interaction: A methodological framework*. New York: Routledge.
- Norris, S. (2011). Identity in (inter)action: Introducing multimodal (inter)action analysis. USA: De Gruyter Mouton. Retrieved from <http://web.a.ebscohost.com/>
- O'Halloran, K. L., Marissa, K. L. E., Podlasov, A. & Tan, S. (2013). Multimodal digital semiotics: the interaction of language with other sources. *Text and Talk*, 33(4-5), 665-690. DOI: 10.1515/text-2013-0030
- O'Halloran, K. L. & Smith, B. A. (2012). Multimodal text analysis. Retrieved from <https://www.slideshare.net/theskyblue/01-multimodal-textanalysisohalloranandsmith>
- O'Halloran, K. L., Tan, S., Smith, B. A. & Polasov, A. (2011). Multimodal analysis within an interactive software environment: critical discourse perspectives. *Critical Discourse Studies*, 8(2), 109-125. DOI: 10.1080/17405904.2011.558687
- Rowe, D. W. (2012). The Affordances of Multimodal Interaction Analysis for Studying the Beginnings of Literacy. Retrieved from <https://pdfs.semanticscholar.org/>
- Seedhouse, P., Harris, A., Naeb, R., & Üstünel E. (2014). The relationship between speaking features and band descriptors: A mixed methods study. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_online_rr_2014-2.ashx
- Zollo, S. A. (2013). *Promotion, Popularisation and pedagogy: An analysis of the verbal and visual strategies in the COE's human rights campaigns*. Newcastle upon Tyne: Cambridge Scholars Publishing. Retrieved from <http://web.b.ebscohost.com/>