EXAMINERS’ AND TEST TAKERS’ ATTITUDES TOWARD INTEGRATION OF INTERACTIVE COMPONENTS OF COMMUNICATION INTO ORAL EXAMINATIONS

Abstract

Speaking skill in EFL context can be assessed through various examinations. Among these examinations, commonly used interviews seem to lack the necessary setting to assess all features of speaking skill. Although spoken production is assessed by means of exam questions in an interview format, spoken interaction cannot be assessed due to the absence of an interactive format. On the other hand, incorporation of a section into examinations, where test takers, interacting with another participant, can be assessed in their interactional skills, may help the assessment of both features of speaking skill. Therefore, our aim in this study is to find out the attitudes of examiners and test takers toward integration of interactive components of communication into oral examinations. The participants of this study were the teachers and students in a language school. The data were collected by means of checklist to teachers and interviews with teachers and students.

Keywords: Speaking skill; interaction; oral exams; paired testing.

1 This study constitutes some part of the MA thesis titled “Both Examiners’ and Test Takers’ Attitudes toward Integration of Interactive Components of Communication into Oral Examinations” conducted by Havayıoğlu (2015).
SÝNAV YAPAN ÖĞRETMENLER VE SINAVA GÝREN ÖĞRENCÝLERÝN SÝZLÝ SINAVLARÁ INTERAKTİF İLETÝÞİM BÝLEÞENLERÝNÝN KATÝLMASÝNA YÖNELÝK TUTÝMLARÝ

Öz

İngilizcenin yabancı dil olarak öğretiminde konuşma becerileri farklı sözlü sınavlar ile ölçülebilmektedir. Bu sınavlar arasında yaygın bir kullanıma sahip olan mülakatların konuşma becerilerinin tüm özelliklerini ölçebilecek bir ortam oluşturmadığı ekste olduğu görülmektedir; çünkü sınav giren öğrencilerin sözlü üretim becerileri mülakatlarında yöneltilen sorular ile ölçülen karşılıklı konuşma becerilerini ölçübilecek bir format bu sınav türüne bulunmamaktadır. Öte yandan, sözlü sınavlara giren öğrencilerin başka bir katılımcıyla iletişim kurarken değerlendirilebileceği bir bölüm eklenmesi, konuşma becerilerinin gerektirdiği hem sözlü hem de karşılıklı konuşma özelliklerinin ölçülmesine yardımcı olabilmektedir. Dolayısıyla bu çalışmadaki amacı, sınav yapan öğretmenlerin ve sınav alan öğrencilerin, interaktif iletişim bileşenlerinin sözlü sınavlara eklenmesi yönündeki tutumlarını belirlemektir. Çalışmanın katılımcıları bir dil okulunda bulunan öğretmenler ve öğrencilerden oluşmaktadır. Veriler, öğretmenlerle kontrol listesi ve öğretmen ve öğrencilerle sözlü görüşmeler sayesinde toplanmıştır.

Anahtar kelimeler: Konuşma becerisi, karşılıklı konuşma, sözlü sınavlar, ikili sınavlar, interaktif iletişim bileşenlerinin katılımı.

1. Introduction

Oral proficiency testing has become preferred assessment method for oral language ability since the emergence of communicative language teaching as a part of communicative language teaching (CLT) (Morrow, 2012). Until the 1980s, Oral Proficiency Interviews (OPIs) were dominantly utilized for the assessment of oral language ability. However, pair or group tasks involving peer-to-peer interaction have been integrated since the late 1980s as a result of the changing views on the nature of what speaking tests should involve (Ducasse and Brown, 2009).

The transition from interview to peer-to-peer testing is claimed to have positive washback on the classroom (Egyud and Glover 2001). Peer-to-peer assessment also reduces the assessment workload as students are tested together in pairs or groups, which results in time savings compared to an individual interview format (Ducasse and Brown, 2009). In addition, van Lier (1989, p.501) states that individual interviews with student focus on the elicitation of language use whereas “successful conversation” is not represented during this assessment method.

The key component of a successful conversation is interaction which was defined by Ellis and Fotos (1999, p.3) as “the interpersonal activity that arises during face-to-face communication” and “the intrapersonal activity involved in mental processing”. Kramsch (1986, p.367) points out that:

Whether it is a face-to-face interaction between two or several speakers, or the interaction between a reader and a written text, successful interaction presupposes not only a shared knowledge of the world, the reference to a common external context of communication, but also the construction of a shared internal context or "sphere of inter-subjectivity" that is built through the collaborative efforts of the interactional partners.
Young (2000, p.5) compares the theory of interactional competence with the theory of “communicative competence” (Canale and Swain, 1980). He explains that communicative competence focuses on “an individual language user in a social context” whereas interactional competence focuses on “the co-construction of discursive practices by all participants involved rather than on a single person.” The acceptance of ‘interactional competence’ in language testing has led to many empirical studies on the design of tests to validate the construct as an effective component for assessment.

The studies investigating the characteristics of interview discourse between examiner and examinee have shown that there is a limited amount of interaction due to the power differential (Hughes, 2003) which leads to asymmetrical conversation (Lier, 1986; Salaberry, 2000). However, the studies examining the paired/group oral tests indicate that test takers interacting with each other produced more interactional features (Brooks, 2009), collaborated for the co-construction of interaction (Lazaraton and Davis 2008; Gan, 2010) and showed symmetrical results (Galaczi, 2008).

As the above mentioned studies reveal, since the integration of peer-to-peer interaction into oral interviews contributes to the assessment of dynamic process of communication, it is also worth exploring the examiners’ and test takers’ attitudes on the implementation of paired tests. To do so, a language school, providing English language courses for over eight years in Adana, Turkey, was chosen to implement paired tests. The school adopts CEFR-based language teaching approach and utilizes the scales of CEFR for assessment.

Testing is done through only oral examinations in the school. After the completion of each level, students are individually interviewed by an examiner to qualify for the next level or to be certified with a level of language competency. Interviews, depending on the availability of students, are conducted either immediately after or within a period of one week at the end of levels. Interview questions, prepared in accordance with syllabus subjects, are comprised of both closed (yes/no, choosing among possible options, identifying a certain piece of information) and open (recall and process, rhetorical, funneling) questions.

Furthermore, examinations assess the components of grammatical resource, lexical resource and discourse management. However, interactive components of communication are not assessed through these oral examinations even though the philosophy underlying the educational approach of the school is teaching communication.

“In interactive activities, the language user acts alternately as speaker and listener with one or more interlocutors so as to construct conjointly, through the negotiation of meaning following the co-operative principle, conversational discourse” (CEFR, 2011). Based on this principle, standardized examinations such as KET, PET, FCE, CAE and CPE employ strategies to assess interactive communication skills of test-takers. Furthermore, the Ministry of National Education in Turkey although it has not fully achieved its goal in students’ attainment of speaking skill yet, puts a great emphasis on the spoken interaction as a part of speaking skill and claims the need to integrate interactional activities such as dialogues, discussions, debates, improvisation and role-plays into the curricula in government schools (Ortaöğretim Kurumları İngilizce Dersi Öğretim Programı, 2011).

1.1. Oral Proficiency Interview versus Paired Testing
In the 1970s and 1980s, communicative language testing emerged as a “part of the broader movement to develop ‘communicative’ language teaching (CLT)” (Morrow, 2012, p.140). The purpose of communicative tests, as defined by Morrow, is “to find out what a learner can ‘do’ with the language, rather than to establish how much of the grammatical/lexical/phonological resources of the language he/she knows”. In this respect, tests measuring language use of test takers in more natural settings e.g. daily life, were designed and implemented. Among these tests assessing oral proficiency of test takers, Oral Proficiency Interviews (OPIs) utilized by Foreign Service Institute (FSI) in 1950s and later developed by American Council on the Teaching of Foreign Languages (ACTFL) had been taken by many test takers who were required for educational and work purposes or simply needed an official evaluation of their oral language abilities. These tests, although paired format has been incorporated recently (ACTFL Proficiency Guidelines, 2012), had long been used in one-to-one interview format.

Since the 1980s, OPI has received criticism for validity issues (Csépes, 2009). Savignon (as cited in Halleck 1992) states that OPI focuses on the elicitation of grammatical features and lacks the construct of communicative competence. Hughes (2003) also notes that due to the match of candidates with more proficient language users (examiners/interlocutors), they are limited to produce only one speech style in the interview format and do not allow space for the elicitation of other functions e.g. “asking for information”. Salaberry (2000, p.296) points out that ACTFL tester training manual does not specify a clear descriptor for the evaluation of “various factors” which results in “inconsistent ratings between and among raters”. Salaberry also argues OPI interviews do not share the essential characteristics of a conversation, do not reflect symmetrical conversation and need to be revised to include more interactional features. Kramsch (1986, p.367) indicates that the goals of ACTFL and ETS diverge from interactional goals since “they focus on behavioral functions rather than on conceptual notional development… have a static rather than a dynamic view of content… emphasize accuracy to the detriment of discourse aptitude”. Similarly, Lier (1989) states that the OPIs’ format provides an efficient interview setting, but not a conversational one and this leads to asymmetrical results. Therefore, he suggests that asymmetry can be eliminated by integrating paired format in which candidates are engaged in task-based activities to demonstrate their oral proficiency. In the following figure, Lier (1989) illustrates four patterns of social interaction in terms of contingency. The first two patterns are associated with the characteristics of oral proficiency interviews while the latter two are associated with the features of a successful conversation.
Figure 2. Classes of social interaction in terms of contingency

In this figure, “A and B represent interactants; R represents a reaction by one interactant to the other; → and ⇢ represent primary and minimal (or no) contingency, respectively; pseudocontingency describes speech events such as acting in a play or rituals (e.g., greetings); asymmetrical contingency describes the interaction of (traditional) teaching and interviewing; reactive contingency describes events such as a rambling conversation; and mutual contingency describes negotiations, serious discussions and the like.” (Lier, 1989, p.497)

Parallel to the development of OPIs, paired format has gained popularity as an assessment tool for oral proficiency and has been extensively used by many high stakes of Cambridge ESOL Examinations. The history of paired test, as Saville and Hargreaves (1999) state, dates back to 1980s when it was first used in FCE and CPE as an optional format. After the introduction of CAE in 1991, the paired format became an obligatory feature of one of the main suite tests in its history. It was also adopted by KET in 1993, revised PET in 1995 and revised FCE in 1996. Fairness in language assessment could be a serious concern of examiners and test takers, thus procedures and measures used in assessment should be valid and reliable (Birjandi & Bagherkazemi, 2011). Therefore, the paired format was adopted by Cambridge ESOL examinations after various alternative formats and feedback collected from oral examiners, teachers, students and candidates taking the tests were evaluated.
Foot (1999) asserts that paired testing might negatively contribute to the existing problems in the assessment of speaking. He holds the view that paired testing consumes time due to standardization and validity issues and leads to a less natural setting with the participation of two examiners. Besides, pairs may not maintain a conversation if there is a mismatch, and time allocated in paired testing is not sufficient for the assessment of language proficiency of test takers.

However, in a study conducted by Egyud and Glover (2001), which was also reflected as a reaction to Foot’s criticism on the dearth of research favoring paired testing, oral tests were implemented at a vocational secondary school in Hungary and students were questioned on their views of the use of pairs in oral tests. Students were reported to find paired format more satisfactory than one-to-one format. Egyud and Glover (2001) commented on the findings as the following:

- Students like pairings,
- Pairings give students a better opportunity to produce their best,
- Pairings help to produce better English than one-to-one format,
- Pairings support good teaching,
- Whatever the format of the exam, it will work better with through preparation and training (p.70).

One of the studies comparing the performance of test takers between the individual and paired format in oral proficiency tests was conducted by Brooks (2009). In her study, she investigated the rate of the interaction between test taker - interviewer and test taker - test taker for which the data were obtained from a larger study (Brooks, in preparation as cited in Brooks, 2009). The findings of the study showed that students performed better in paired tests compared to the individual format. In addition to higher exam scores in paired tests, pairs demonstrated a more complicated performance involving more co-constructed interaction, negotiation of meaning and interlocutor’s attentiveness to his/her partner.

A study on the effectiveness of co-constructed interaction was conducted by Gan (2010) in a secondary school in Hong Kong. Eight students in two groups were assessed and the data were collected from a case study of these two groups. The students in the higher-scoring group were awarded so, as they "engaged constructively and contingently with each other's ideas, demonstrating a range of speech functions such as suggestions, agreement or disagreement, explanations, and challenges". In the lower-scoring group, although students did not sustain a topical talk, they did collaborate to negotiate meaning. Since it is a small scale study, Gan (2010) tentatively concludes that testing students in pairs demonstrates their interactional skills in real life and their progression socially and linguistically.

1.2. Interlocutor Effect

Aside from the effectiveness of paired format on the elicitation of interactional features, candidates also contribute to the quality of this format. Norton (2005) states that pairing affects the linguistic performance, amount of talk produced and qualitative participation of each test taker if partners share uncommon features of linguistic ability, familiarity and gender.

In a study on the discourse of interlocutor effect, Lazaraton and Davis (2008) examined the effect of identity on the proficiency of paired candidates. The purpose of the research was to investigate how test takers in two-large scale speaking tests (Cambridge ESOL’s FCE and PET)
represented themselves as proficient users of English. The discourse produced by four groups of pairs was analyzed and categorized as doing (or being) proficient, doing interactive, doing supportive and doing assertive respectively. The results of the study showed that pairs whose proficiency levels were similar in the first two categories achieved high scores by being collaboratively, individually, and interactively proficient whereas the candidates who had varying proficiency levels in the last two groups were awarded high scores as they either scaffolded their partner or acted assertively. Lazaraton and Davis (2008) concluded that though the qualities of a good pairing such as similar proficiency levels, collaborative interaction, support and confidence, work for the benefit of the test taker, they should not be regarded as the only criteria for the interlocutor effect.

Davis (2009) also examined the effect of interlocutor proficiency in a paired format and he states that a difference between the proficiency levels of candidates played a minimal role on the scores awarded for each participant. Based on his findings, he suggests that assessment of paired candidates in oral exams might give possibly fair results even at varying proficiency levels.

Brooks (2009), in her study on co-construction for a better performance, found that the interlocutor effect plays a significant role on the rate of candidates' interaction. The test takers who interacted with each other in oral tests illustrated more features of interaction (e.g. “seeking confirmation, asking a question, asking for agreement, clarification request, incorporating words, prompting elaboration, finishing sentences [...]”, P.352) than when they interacted with an interviewer. The results of the study also showed that in individual format turn takings between the test taker and the interlocutor resembled one-way information transfer not only because of the interlocutor’s role as the interrogator but also test taker’s preference to further the conversation individually. On the other hand, the conversation ran more interactively in the paired format as test takers displayed more empathetic behaviors due to their shared features of interaction.

Familiarity effect is another issue investigated by O’Sullivan (2002). In the study, a group of 32 Japanese university students (24 women and 8 men) were paired with an acquainted and a strange participant. Although there seemed to be no significant superiority of one gender to another, students were awarded higher scores when they were engaged in tasks with someone they knew. Similarly, Ying (2009) carried out a research in which 31 university students in groups of three attended two examinations. In one of the examinations, two participants already knew each other with a stranger whereas, in the other one, all the participants were stranger to each other. The overall result shows that familiarity with the interlocutor had a minor effect on the performance of test takers although there seemed to be a correlation between the challenge students faced and familiarity. When the students were acquainted with their partners, they found the assessment less difficult and the other way round.

Nakatsuhara’s (2011) study also showed that the number and characteristics (extraversion / introversion level) of test-takers have an effect on the co-constructed discourse. 269 Japanese upper-secondary school students participated in an exam in either groups of three or groups of four. Students who were assessed in groups of three worked more collaboratively and involvement of an introverted participant, if there was any, was easy whereas introverted students in the latter group avoided talking which was not difficult due to the number of participants. Turn-takings also appeared to be asymmetric in the latter group compared to the first group. She con-
cludes that unless students in groups of three are less proficient than groups of four, the second group seems to benefit negatively in terms of size and personality factors.

2. Methodology

The research was conducted in the form of a descriptive study. As the data gathered during the study were qualitative, both examination criteria checklist and interviews were subjected to content analysis. “Content analysis is a research method for making replicable and valid inferences from data to their context, with the purpose of providing knowledge, new insights, a representation of facts and a practical guide to action (Krippendorff, 1980 as cited in Elo, S., & Kyngäs, H., 2008).” Weber (as cited in Hsieh, H. F., & Shannon, S. E., 2005, p.1277) states that “qualitative content analysis goes beyond merely counting words to examining language intensely for the purpose of classifying large amounts of text into an efficient number of categories that represent similar meanings”.

2.1. Participants

Convenience random sampling was used to select the participants of the study. The participants comprised of two parties of people. To elaborate, totally 26 participants, that is five EFL examiners and 21 test-takers in a language school in Adana, Turkey took part in this study. Being a full-time teacher at the institution was the only criterion for the selection of the examiners. That is, variables as experience, age, gender and the like were not taken into consideration. Still, we can say that of the totally five ELT graduate participants, three were females while two were males and their teaching experiences varied from four to ten years. Twenty one EFL students having taken at least one oral examination at previous levels and attending the classes of B1 and B2 levels, were selected as participants. The ages, genders and educational backgrounds of the participating students were not considered as a variable; similar to the selection criteria used for teachers, the only criterion to be selected as a participant was that, as stated above, they had to have experience on the previous exam and were about to take the new version at above mentioned levels. Still, we can say that of the twenty-one participants, sixteen were females while five were males and their ages ranged from eighteen to forty five.

2.2. Instrumentation

In order to bring more insight to the study and realize triangulation in this study, exam criteria checklist and interviews were used.

2.2.1. Examination Criteria Checklist

As the first step of the study, teachers were given a checklist to evaluate the extent current examinations correspond to instructional criteria of the school. Examination criteria checklist is comprised of 70 assessment questions used at Intro, A1, A2, B1, B2, C1 and C2 levels. Teachers were asked to check the questions in line with the productive and interactive components of spoken skill they correspond to.

2.2.2. Interviews

After the checklist was evaluated, teachers were interviewed so as to elicit more thorough understanding of checklist responses. Following the implementation of new examinations, teachers were re-interviewed to find out their evaluation of the changes on the format of examinations.
In the same vein, students were interviewed on their views on examinations conducted in the school before and after the implementation of the new examination system.

3. Findings and Discussion

The purpose of this study was to investigate the attitudes of examiners and test takers toward the incorporation of interactive components of communication into oral examinations in a language school. The study was also concerned with the effectiveness of the implementation and the possible backwash effects on the classroom instruction in the teachers’ opinion. Thus, the study attempted to find out answers to the following research questions:

1- Is there a conflict between the principles of instruction and the nature of oral examinations conducted in the language school?
2- What is the nature of examiners’ attitudes toward oral examinations conducted in the language school?
3- What is the nature of test-takers’ attitudes toward oral examinations conducted in the language school?
4- What are the teachers’ attitudes toward the integration of interactive communication strategies into oral examinations?
5- What are the test-takers’ attitudes toward the integration of interactive communication strategies into oral examinations?
6- In what ways, if any, does this implementation have backwash effects on the classroom instruction in the school as teachers perceive?

In this respect, five data collection tools were administered; an exam criteria checklist and four interviews including preliminary and follow-up interviews with examining teachers and test takers. In the following pages, the research questions will be discussed in line with the findings obtained from the data.

Research Questions 1: Is there a conflict between the principles of instruction and the nature of oral examinations conducted in the language school?

The syllabi of the language school reveal that the principles of instruction are in line with the teaching objectives of CEFR, that is, students are engaged in tasks and activities in listening, speaking, reading and writing skills with a specific emphasis on speaking skill. However, the assessment system of the language school does not fully conform to the assessment requirements of CEFR. Since the language school adopts a communicative approach in language teaching, the testing system is designed to assess students’ proficiency orally. However, the assessment questions correspond to only assessment of spoken production criteria of speaking skill described in CEFR which also suggests the assessment of spoken interaction components as well as spoken production. In this respect, the principles of instruction and the oral examinations conducted in the language school do not wholly cohere with each other.

Implications: In similar occasions, in order to eliminate the gap between the instruction and the assessment, the testing system of those institutions may need to be redesigned to involve the components of both spoken production and spoken interaction, as is the case in our study. Such an exam, if designed well, is likely to create a setting where students are able to interact with an interlocutor in addition to the interview sessions, allowing examining teachers to assess stu-
students’ proficiency more thoroughly, that is, making the assessment system well-suited to the standards of CEFR.

**Research Question 2:** What is the nature of examiners’ attitudes toward oral examinations conducted in the language school?

The results of the exam criteria checklist indicate that the previous assessment system in the language school lacks questions assessing interactive components of speaking skill and all the oral assessment questions are associated with the assessment of spoken production skill by the participating teachers. The results obtained from the preliminary interview held with the examining teachers support these findings. A gap between instruction and assessment is the examiners’ first reaction. This finding orients us to think that the implementation may possibly fill in the gap between instruction and assessment. Egyud & Glover (2001, P.75) states that “we expect examination practices to reflect good practice, and to encourage it. Examinations and tests must connect with the classroom, because they exist as a part of a longer and more important process: the process of learning and teaching”. Moreover inconvenient setting for the involvement of students and teachers, elicitation of only the subjects taught in classrooms, language school’s deprecation of the implementation, appropriateness of the exams for a quick assessment process and business logic of the school are the other negative reactions of teachers toward the absence of interactional questions in the previously conducted oral exams.

The interview results held with the teachers also reveal several reactions. Involvement of teachers and variation in questions for a more natural conversational environment, increasing comprehensiveness of the examinations and assessment of spoken interaction skills of students are regarded as critical factors in previous oral exams.

It is agreed by all the participating teachers that the previous testing system does not fully conform to the assessment criteria of CEFR. The vast majority of the participants also indicate that the examinations conducted before the implementation of the new component do not conform to the teaching philosophy of the language school.

**Implications:** The above findings inherently entail a revision in the exam format. Designing a new format conforming to the teaching philosophy of the language school, thereupon the assessment criteria of CEFR, necessitates an appropriate setting which facilitates the elicitation of interactional features of student discourse authentically. At this point, assessment of two-way interaction between pairs seems to be rational. However, the choice of pairing requires a careful thought since it might lead to one-way information transfer if the interlocutor is an examiner. Hughes (2003, p.119) notes that “[t]he relationship between the tester and the candidate is usually such that the candidate speaks as to a superior and is unwilling to take initiative. As a result, only one style of speech is elicited, and many functions (such as asking for information) are not represented in the candidate’s performance.” On the other hand, Brooks (2009, p.341) indicates “[w]hen test-takers interacted with other students… paired testing format resulted in more interaction, negotiation of meaning, consideration of the interlocutor and more complex output.” Therefore, by considering the concerns of the language school, utilizing paired oral tests assessing examinee-examinee interaction appears to be a more appropriate option.

**Research Question 3:** What is the nature of test-takers’ attitudes toward oral examinations conducted in the language school?
The results of the preliminary interview held with the participating students indicate that the overwhelming majority of the participants have positive attitudes toward the previously conducted oral exams. The function of language learning for communicative purposes, assessment of students’ real performance, more efficient assessment of students’ proficiency compared to written exams and the comfort are the reasons provided by the students for their positive attitudes.

Still, those exams cannot fully be said to be viewed positively. Utilization of both written and oral examinations for a better assessment system is supported by the participating students. Similar to Nakatsuha’s study (2011) which illustrates the involvement of introverted students into the conversation during group oral exams, peer/group examination is viewed to contribute to a more relaxing atmosphere. Furthermore, the subjectivity of the examiners, variation of students’ performance in oral examinations, integration of components of writing skills into syllabi and examinations, utilization of written exams in lessons to assess vocabulary development of students, and a gap between the instruction and the assessment are the critical reactions that need to be reconsidered for a better assessment system, as the findings of this study suggest.

Besides, a small number of participants find the previously conducted examinations ineffective due to the lack of accuracy in terms of assessment of students’ proficiency resulting in variation of proficiency levels among students in the classrooms, and difficulty and anxiety levels in oral examinations.

Implications: As the students of the study are aware that speaking takes precedence over the other skills in language instruction of the school, most of them seem to feel quite content with the exams, yet the concerns regarding the exams still require attention. Written exams for the assessment of either lexical items or writing skills could be integrated into the assessment procedure, and a certain proportion of the scores gained on these exams might be added up to the overall scores of test takers to provide a more satisfactory assessment system for the students while still upholding the assessment principles of the school. For the improvement of oral exams, paired testing with the participation of two examiners seems to be a remedy as it helps test takers “feel relaxed” (Foot, 1999) and lead to fairer assessment by making them “feel reassured that their mark does not just depend on one person” (Saville and Hargreaves, 1999).

Research Question 4: What are the examiners’ attitudes toward the integration of interactive communication strategies into oral examinations?

Before the implementation of the study, the majority of the participating teachers viewed the incorporation of interactive components into oral exams as a contributory factor to a more efficient assessment system. In this regard, considering the concerns and suggestions put forward by the teachers, a new assessment system, assessing students’ oral proficiency in terms of both spoken production and spoken interaction criteria, was designed.

After the implementation of the new components, that is, the inclusion of the spoken interaction criteria into oral examinations, the majority of the teacher participants find the new examinations more efficient than the previous ones. The dialogues constructed in the exam are found to be more authentic leading to higher number of spoken features such as “seeking confirmation, asking a question, asking for agreement, clarification request, incorporating words, prompting elaboration, finishing sentences […]” (Brooks, 2009, P.352) which, in return, facilitates the assessment of student’s performance. Interactive assessment, real life like dialogues,
comprehensiveness of assessment questions, comfort students feel and use of visual aids are the other leading factors contributing to the efficiency of the new examinations as perceived by the participants. Over all, the majority of the participating teachers indicate that the study served its purpose in terms of involvement of interactive components in oral exams.

The participants also pointed out that the implementation conforms with:

- The needs of the school as it coheres with the teaching principles of the language school and helps the school stand out among its competitors.
- The requirements of CEFR as it contributes positively to classroom instruction, by itself corresponds to a criterion in ELP and is in line with the assessment descriptors of CEFR.
- The needs of students and teachers as paired format offers a more comfortable and natural atmosphere, makes interactive assessment available for the teachers and prepares students for real life conversational settings.

Notwithstanding the positive attitudes of participants, they also indicate that the new exams are not as time efficient as expected, which was also of a concern to Foot (1999), might complicate the decision-making process of the examiners due to the fewer number of questions and need to be improved. The new exam is also regarded as inappropriate to be conducted at B1 level which is the threshold level as described in CEFR. Thus, it can be concluded that Intro, A1, A2 and B1 levels are not perceived to be suitable for the assessment of spoken interaction but appropriate for B2, C1 and C2.

Implications: As seen in the findings above, the incorporation of interactive components seem to increase the efficiency of oral exams in terms of the quality of assessment with a more authentic setting. The implementation also seems to cohere with the speaking standards of CEFR and benefit the school, its students and teachers. However, it is not impeccable. Below are some recommendations for the implementation of the oral exams.

Paired format might take up equal time or longer when it is compared to the interview format. However, sharing the work load of each assessment session with the involvement of more examiners or appointing pairs/groups according to the availability of examiners might eliminate the concerns of teachers for this implementation.

As this implementation is the first attempt, examining teachers might have possibly directed fewer numbers of questions to compensate for the time or extended the duration of the exams to direct more questions. To level this imparity off, all the examining teachers could reach a compromise on the interrelation between time and number of questions to standardize the tests.

It is evident that more proficient students can interact with other students more easily since they have more control over the language. However, this does not necessarily mean that students with lower language competencies should not be assessed for their interactional skills. The number of interactive activities in lower levels can be increased so that students would practice more and gain familiarity with the format before they take the exams.

The critical views reflected in this study need to be handled carefully. After all, the above mentioned issues might be reflected in interview format as well, thus they do not only pertain to paired testing.
Research Question 5: What are the test-takers’ attitudes toward the integration of interactive communication strategies into oral examinations?

According to the preliminary student interviews, although the majority of the participants had positive views toward the previously conducted exams in the language school, they favored the inclusion of interactive speaking skills into assessment procedure with the reasons that they could possibly contribute positively to the assessment system, test takers, examiners and the classroom instruction. Minority of participants also expressed their concerns regarding the mismatch between student pairs which contradicts with the findings of Lazaraton & Davis (2008) and Davis (2009) on the efficiency of the paired format even at varying proficiency levels. Anxiety and topic choice are the remaining concerns put forward by the student participants. The difficulty level of the exams, the stress added on the exams and inadequate proficiency level of students were the probable drawbacks toward the new implementation brought by the participating students. The last aspect of the findings is supported by Norton (2005) who states that pairing affects the linguistic performance, amount of talk produced and qualitative participation of each test taker if partners share uncommon features of linguistic ability, familiarity and gender.

After the implementation of the study, the vast majority of the participants indicate that the new exams assess overall speaking skill more efficiently. Effective interactive assessment, friendly environment, more opportunities to speak, confidence gained with the participation of two examiners and test takers, authentic setting, and coherence with the classroom instruction and assessment are the reasons given for the efficiency by the student participants.

Apart from the positive views, they raise concerns over the design, appropriateness to the test takers and comprehensibility of the new examinations. In addition, time allocated for the assessment of test takers individually and in pairs, and effect of peer on student’s performance are also regarded as critical reactions. Finally, perceived higher anxiety and inappropriateness of the exams were among the negative factors.

Implications: The findings illustrated above indicate that students seem to benefit more from paired testing. This type of implementation is likely to raise students’ awareness on the importance of interactional skill and they might show an increased willingness while participating in pair work activities in the classroom. In addition, students will likely feel less stressed in the exams as they perceive the new format more friendly and assuring.

If the exams, as mentioned earlier, are standardized, the concerns regarding the design, appropriateness, comprehensibility and duration will possibly be eliminated. Besides, except for special conditions such as disadvantageous match due to power differential, if students are allowed to choose their partners before the exam, they might feel more comfortable (Ying, 2009) and probably demonstrate a better performance. After all, test takers are examined with a partner from the same classroom.

Research Question 6: In what ways, if any, does this implementation have backwash effects on the classroom instruction in the school as teachers perceive?

The data collected from the teacher interviews reveal some positive backwash effects on the classroom instruction. The findings show that if the paired format is adopted by the institution, it is likely to change students’ perception toward the speaking activities, in specific interactional tasks in classroom which supports the idea that “pairings support good teaching” (Egyud
Moreover, the new exams may possibly spur interactional talks among students and even the participation of bashful students into activities. Finally, students will likely be more conscious of their learning as the implementation is perceived to increase students’ language portfolio use in classes.

**Implications:** Despite the fact that there are few, the findings are hope giving in terms of bettering classroom instruction, motivating students to involve themselves in any conversational settings more enthusiastically and contributing to a more coherent educational system.

In conclusion, despite the negative reactions, paired testing appears to be a more efficient format and offers a more comprehensive and thorough oral proficiency assessment in the field of communicative language testing.

**REFERENCES**

American Council on the Teaching of Foreign Languages (ACTFL), (2012). *ACTFL proficiency guidelines*. Alexandria, VA: ACTFL.


