

# The Effects of Test Type, Pronunciation, and Proficiency Level on EFL Learners' Speaking Exam Scores

Kardelen Kilinc<sup>1</sup> & Ozgur Yildirim<sup>2,\*</sup>

<sup>1</sup>Eskisehir Technical University, Turkey

<sup>2</sup>Faculty of Education, Anadolu University, Turkey

\*Correspondence: Faculty of Education, Anadolu University, Eskisehir, Turkey. E-mail: oyildirim@anadolu.edu.tr

Received: April 11, 2020

Accepted: June 6, 2020

Online Published: June 20, 2020

doi:10.5430/wje.v10n3p95

URL: <https://doi.org/10.5430/wje.v10n3p95>

## Abstract

The present study aims to reveal the effects of test type, pronunciation and proficiency levels of the students on speaking test scores. A total of 147 Turkish EFL students consisting of 38 beginner, 36 elementary, 37 pre-intermediate and 36 intermediate levels participated in the study. Presentation as planned, and paired speaking test as unplanned one were used as instruments to figure out the effects of two different test types on the test scores. Being prepared for their performances, the students did their presentations lasting between 5 and 8 minutes in front of two raters. The same participants, at the end of each level, were invited to the paired-speaking tests designed for the students to show their spontaneous performances. The performances in both tests were scored through a scale with five criteria, and the criteria along with overall scores were examined through Paired Samples t-test to reveal the effects of test type, and through bivariate regression to see the proportion of pronunciation aspect on overall scores. The results showed that in beginner level, although no differences were found in the overall scores, the test type induced differences in pronunciation, vocabulary and relevance aspect of their speaking performances. Similar results were found in elementary level besides the difference found in accuracy aspect, which resulted in a significant difference in overall scores. However, in pre-intermediate level, the only significant difference was found in pronunciation. On the other hand, in intermediate level all the aspects along with the overall scores were found to be affected by the test type except for fluency and pronunciation. The bivariate regression revealed that the effect of pronunciation sub-score on overall scores is significant in each level and test type.

**Keywords:** speaking tests, language testing, pronunciation sub-score

## 1. Background to the Study

Language testing has a great value in ELT not only due to it being used to assess students' performance, but also to reflect the quality of teaching. Therefore, preparing and administering valid and reliable tests are of high value in terms of revealing students' competence and performance accurately. That's why, language testing has gained more and more importance recently (Güneş, 2010).

When compared to testing the areas of grammar, vocabulary, and pronunciation; and testing the receptive skills of listening and reading; testing productive skills is more challenging for teachers since they include a lot of aspects to be assessed. Despite the fact that speaking and writing, as productive ones, have common points; the former differs from the latter in many respects. Speaking is not an act of producing the written pieces of works orally. What makes speaking highly demanding for students and more challenging to assess for teachers is the higher number of components it has (Güllüoğlu, 2004). Being able to speak in a foreign language necessitates mastering a number of subskills concurrently. Basically, the subskills of speaking to be tested analytically are range, accuracy, fluency, interaction and coherence (Luoma, 2004). In a speaking test, these components affect the scores the students achieve naturally. Hence, it is crucial to be aware of these components to understand the underlying reasons behind the test takers' low and high scores.

In-depth, assessment serves as a diagnostic tool revealing the problematic issues the learners have and the extent to which each criterion constitutes the whole scores. As a sub-skill of speaking, the effect of pronunciation on speaking scores has been neglected for years. Kang and Ginther (2018) put forward that pronunciation testing has been a

research topic since 2005, yet empirical studies on this area can hardly be found. As Luoma (2004) states, in spite of having a central role in speaking assessment, hardly any of the studies in testing speaking skill are specifically concerned with pronunciation. This statement is also related to the idea that the specific patterns of pronunciation affect the overall assessment of speaking (Celce-Murcia, Brinton, & Goodwin, 2010).

The intelligibility of speaking is a prominent factor of comprehensibility in communication since communicative effectiveness and comprehensible pronunciation are two facets that are interwoven (Luoma, 2004). However, no consensus has been reached regarding the role of pronunciation aspects in general speaking ratings (Kang, 2013; Ma, 2015). These facts lead us to examine the extent to which pronunciation affects speaking assessment so that the ways to improve speaking skills can be discovered.

Ma (2015) attempted to find out the weight of pronunciation in overall speaking proficiency test among the students at an IEP (Intensive English Program). She studied with 226 English as a Second Language students to reveal the weights of the segmental and suprasegmental features of pronunciation in overall speaking ratings. The results proved that pronunciation, especially sentence stress, intonation, and rhythm, could constitute approximately 42% of the speaking proficiency ratings, which also showed that suprasegmentals had a larger effect than segmentals and among all the aspects of it, sentence stress had the biggest impact on the scores.

In another study, Higgs and Clifford (1982) revealed the weights of five speaking subskills, vocabulary, grammar, pronunciation, fluency and sociolinguistic in speaking test ratings according to different levels. Fifty teachers rated the test performances based on five subskills, and the results proved that pronunciation played a larger role at the lower levels. In other words, the result of this study showed that at certain proficiency levels, pronunciation might have a more critical effect on the overall speaking test while the effect might decrease as the proficiency level increases.

In a similar study with the previously mentioned one, De Jong and Van Ginkel (1992) attempted to figure out the role of pronunciation in different levels. A total of 25 Dutch learners of French as a foreign language were given different speaking tasks and their performances were evaluated through a 4-point comprehensibility scale. Additionally, the impressions of the raters on the performances were also gathered. In the end, the findings showed that pronunciation component had the biggest role in comparison to the other components in the low levels. In the higher levels, however, all facets were found to have an equal role.

Apart from the subskills of a speaking test, another major indicator of the scores is the test type; that is, the way proficiency is tested is of great importance. According to Underhill (1987), there are many kinds of speaking tests in literature such as sentence completion, sentence construction, translation and interpreting, reading a blank dialogue, oral presentation, verbal essay, using pictures and picture story, giving instructions, explanations, and descriptions, information-gap, oral interview, controlled interview, free interview, role-playing, simulation and discussion. To be able to design a valid speaking test aiming at revealing students' actual oral performances, both students' perceptions and the effects of test types on their speaking scores should be taken into consideration.

There are a few studies conducted to reveal the learners' perceptions about the test types assessing their oral skills and the effects of the test type on their scores. Brooks (2009) examined the students' interactions in paired and individual tests. The data was collected at Academic Preparation Program at the Canadian University with eight pairs. Data analysis showed that the students performed better in paired-tests than individual ones, during which they were required to discuss specific issues with the examiner. The paired test format enabled the students to produce more complex language with the help of co-construction.

In Turkish context, Öncel (2016) carried out a study to see the students' preferences about the two test types; paired and individual, and the reasons behind their preferences. Based on the results gathered from the semi-structured interviews, the students favored paired speaking tests more than the individual type; however, the researcher defends that the proportion of the students who opt for individual test type rather than paired one cannot be ignored and should be analyzed, too. The reasons behind these perceptions were interpreted that the ones who preferred paired test thought that their partner enhanced their performance, during the test while the other group indicated that the same situation hindered their performance and affected them negatively. The researcher emphasizes the importance of individual differences based on the results of the study and the importance of variety on test types to assess speaking. The study also aimed at revealing the differences between the Turkish and international students' preferences on speaking test types and the results were surprising at this point. While the Turkish students favored the paired tests, the international students' results showed equality between two choices. These results also supported the idea that individual differences are affected by different background knowledge, intelligence types, learning styles, cultural values and proficiency levels.

Since the two common test types of speaking are paired and individual, the studies carried out focused on these two formats but still, the studies conducted in this area are very limited. Even though all assessment types have their characteristic features which determine whether the students prefer to be tested in one way or another, there is one more distinction determining the test performance of students; whether the students plan their speech in advance or not. Oral communication takes place in planned or unplanned occasions. It should be designed authentically not only the activities to teach speaking, but also the tasks to test. That's why speaking can be tested through the tasks requiring either preparation in advance or spontaneous speech.

Unplanned speech that is mostly tested in individual or paired formats occurs through interaction with other speakers. In planned speech such as lectures, conference presentations, and expert discussions; speakers prepare and rehearse their performances beforehand. Speakers, in unplanned speech, utter shorter idea units compared to the speech in planned ones where the presenters use longer and complex sentences (Luoma, 2004).

This idea may lead us to the opinion that planned speech showing enhanced performance of the test taker results in a more valid test since it prevents the external factors affecting students' performances, such as anxiety emerging from the nature of spontaneity of unplanned tests. On the other hand, Thornbury (2012) puts forward that unplanned speech shows us the test takers' performance under the real operating conditions which were described as urgency, unpredictability, and spontaneity. This reminds us of the authenticity component a task should include, which makes it valid compared to the planned speech. As a conclusion, both test types are valuable in terms of testing the students' oral performances. Hence, the question to be investigated is the effects of these two different test types on their performance and on the scores they achieve from the speaking tasks.

The studies in this area focused on the effect of planning speech on the aspects accuracy, complexity, discourse-level and fluency so far. First of all, Ellis (1987) aimed to see the effect of planning on intermediate level students' accuracy and designed a study with regard to past tense morphemes through a narrative discourse task. The findings of this study showed that accuracy improved through planning in speaking tasks. However, in two years, another study found different results. Crookes (1989) aimed to reveal the effect of planning factor on both accuracy and complexity among intermediate and advanced level students. The statistical results showed that students improved their performances in terms of complexity through planning their speech, yet their accuracy did not improve as expected. To prove this, Williams (1992) found out that complexity and accuracy improve through planning in speaking tasks; however, planning does not induce higher scores on accuracy.

Foster and Skehan (1996, 1997) carried out two different studies with pre-intermediate level students on the effect of planning factor on speech performances. In the first study (1996), they achieved the result that planning induced better performances in terms of complexity and discourse level, but they did not observe any improvement in accuracy. In 1997, they came up with different results and stated that accuracy, complexity, and fluency improved as a result of planning. To sum up, there has not been a consensus on the effect of planning on accuracy improvement. On the other hand, in two recent studies, Tajima (2003) and Martinez (2004) carried out researches to see the results of planning on fluency aspect of speaking, and both found that fluency improved as a result of planning speech. Just as the facets of speaking skill investigated in these studies, the effect of planning on pronunciation facet should also be examined.

Another factor determining the speaking scores is the level of the students. As one would expect, the speaking scores, along with all the components, are assumed to get higher as the level increases. However, it is hard to say that all the sub-skills in oral skills improve at a similar speed.

Considering all the aforementioned arguments, the current study aims at revealing the effects of two test types and pronunciation sub-score on overall speaking exam scores in relation to the different proficiency levels. With this motivation, this study aims to find answers to the following research questions:

- What is the effect of test type on the scores in unplanned and planned speaking exam tasks according to different EFL proficiency levels?
- What is the effect of pronunciation sub-scores on overall scores in unplanned and planned speaking exam tasks according to different EFL proficiency levels?

## 2. Method

A total number of 147 Turkish EFL students studying at a private English language teaching institution in Turkey participated in this study. Participants constituted of 38 beginner, 36 elementary, 37 pre-intermediate, and 36 intermediate level adult learners; they were placed in their levels by Cambridge English Placement Test.

'Presentation as the planned task' and 'paired-speaking test as the unplanned task' were used as the test types in this study. The scoring scale used in these tests had five components as accuracy, fluency, use of vocabulary, pronunciation, and relevance; each of which was rated on a twenty-point scale.

For the presentation, as the planned task, the students were asked to prepare a presentation lasting between 5 and 8 minutes in the last week of their term. They were given the topics pertinent to their level to give a presentation two weeks before the presentation day. Within a week, the students sent the drafts they prepared for the speech along with their PowerPoint presentation file. The instructors provided feedback on the major issues to lead the students to revise their speech. Then, on the presentation day, two different raters watched their performance and rated them by using the scoring scale.

For the paired-speaking test, as the unplanned task, the students were invited to a classroom in pairs to perform in front of two raters. The performances lasted between ten and fifteen minutes and the students were asked to pick a card on the table and read the question to their partner to be answered and the procedure continued by follow-up questions. Depending on their performances, the raters asked them to pick 3 to 5 cards. Their performances were recorded to be listened to and rated based on the same scoring scale used for the planned presentations.

To obtain the results of the present study, the students' scores in two types of tasks were analyzed. The first research question focused on the effects of two different test types on the scores. To answer that, paired-samples t-test was run for overall scores and sub-scores in unplanned and planned tasks. The second research question was about the effect of pronunciation component on overall scores in both test types. Bivariate regression was run to reveal the effect of pronunciation on overall scores for each test type and level.

### 3. Results and Discussion

#### 3.1 Recruitment the Effect of Test Type on the Scores of Beginner Level Students

As Table 1 represents, the results of the paired samples t-test indicate that there is not a statistically significant difference ( $t(37) = 0,488$ ;  $p > .05$ ) between the mean scores of the overall planned task ( $M = 71,63$ ;  $SD = 10,114$ ) and unplanned task ( $M = 70,66$ ;  $SD = 11,400$ ) in beginner level. However, the results also show that there is a statistically significant difference ( $t(37) = 2,125$ ;  $p < .05$ ) between the mean scores of the vocabulary sub-scores in planned task ( $M = 14,24$ ;  $SD = 2,889$ ) and in unplanned task ( $M = 13,03$ ;  $SD = 2,746$ ). Similarly, the results revealed a statistically significant difference ( $t(37) = -7,363$ ;  $p < .05$ ) between the mean scores of the pronunciation sub-score in planned task ( $M = 12,71$ ;  $SD = 3,328$ ) and in unplanned task ( $M = 15,84$ ;  $SD = 2,118$ ). Lastly, a statistically significant difference was also found ( $t(37) = 2,497$ ;  $p < .05$ ) between the mean scores of relevance sub-score in planned task ( $M = 19,00$ ;  $SD = 2,731$ ) and in unplanned task ( $M = 17,24$ ;  $SD = 2,889$ ).

**Table 1.** The Effect of Test Type on the Scores of Beginner Level Students

	Planned		Unplanned		t	df	p
	M	SD	M	SD			
Overall Score	71,63	10,114	70,66	11,400	,488	37	,629
Accuracy	12,53	3,073	11,71	3,376	1,347	37	,186
Fluency	13,16	3,098	12,84	3,468	,494	37	,624
Vocabulary	14,24	2,889	13,03	2,746	2,125	37	,040
Pronunciation	12,71	3,328	15,84	2,118	-7,36	37	,000
Relevance	19,00	2,731	17,24	2,889	2,497	37	,017

Considering the overall scores, the findings can be interpreted that beginner level students could not get higher scores in planned tasks despite their preparation. When each component is analyzed, accuracy and fluency are the only components which did not show a significant difference. That means the participants did not perform better in presentations in terms of accuracy and fluency. In lower levels, even though the students internalize grammar rules and can get high scores in paper-pen exams, they have difficulty in oral production since they still need time to practice more for automatization of the features (Lightbown & Spada, 2013). Similarly, fluency also develops in time and it is not likely to be observed in a few weeks even though they have the chance for preparation in advance. Consequently, the development of accuracy and fluency through preparation is unlikely to be observed in low-level students.

### 3.2 The Effect of Test Type on the Scores of Elementary Level Students

As Table 2 represents, the results of the paired samples t-test indicate that there is a statistically significant difference ( $t(35)=2,678$ ;  $p<.05$ ) between the mean scores of the overall planned task ( $M=71,33$ ;  $SD=8,281$ ) and unplanned task ( $M=66,97$ ;  $SD=11,305$ ) in elementary level. Contrary to the results in beginner level, there is a significant difference ( $t(35)=4,202$ ;  $p<.05$ ) between the mean scores of the accuracy component in planned task ( $M=12,42$ ;  $SD=3,225$ ) and in unplanned task ( $M=10,00$ ;  $SD=3,406$ ). The results related to the other components revealed that there is a significant difference ( $t(35)=3,404$ ;  $p<.05$ ) between the mean scores of the vocabulary component in planned task ( $M=14,06$ ;  $SD=2,414$ ) and in unplanned task ( $M=12,39$ ;  $SD=2,476$ ). Besides, a significant difference ( $t(35)=-4,033$ ;  $p<.05$ ) was found between the mean scores of the pronunciation component in planned task ( $M=12,25$ ;  $SD=2,922$ ) and in unplanned task ( $M=14,11$ ;  $SD=3,310$ ). In the same degree, the findings showed a significant difference ( $t(35)=5,759$ ;  $p<.05$ ) between the mean scores of the relevance component in planned task ( $M=20,00$ ;  $SD=0,00$ ) and in unplanned task ( $M=18,11$ ;  $SD=1,968$ ). On the other hand, as in beginner level results, no significant difference ( $t(35)=0,358$ ;  $p>.05$ ) was found between the mean scores of the fluency component in planned task ( $M=12,61$ ;  $SD=3,174$ ) and in unplanned task ( $M=12,36$ ;  $SD=3,83$ ).

**Table 2.** The Effect of Test Type on the Scores of Elementary Level Students

	Planned		Unplanned		t	df	p
	M	SD	M	SD			
Overall Score	71,33	8,281	66,97	11,305	2,678	35	,011
Accuracy	12,42	3,255	10,00	3,406	4,202	35	,000
Fluency	12,61	3,174	12,36	3,833	,358	35	,723
Vocabulary	14,06	2,414	12,39	2,476	3,404	35	,002
Pronunciation	12,25	2,922	14,11	3,310	-4,03	35	,000
Relevance	20,00	,000	18,11	1,968	5,759	35	,000

Even though the results were very similar to the ones found in beginner level, they show that preparation induces accuracy achievement in planned tasks as the level increases. However, fluency still does not make any changes either in planned or in spontaneous performance, which can be explained through the necessity of time for automatization. As another low-level, elementary still requires the students to spend more time to practice to be automatized. To sum up, it can be noted that it is still too early to observe a difference in fluency through preparation in elementary level; though, accuracy begins to take a role in overall scores in planned tasks. It shows us that planned tasks enable the learners to be more accurate in oral performances.

### 3.3 The Effect of Test Type on the Scores of Pre-Intermediate Level Students

The paired samples t-test results related to the pre-intermediate level group were found to be surprising as represented in Table 3. Although there was not any significant difference ( $t(35)=-,506$ ;  $p>.05$ ) between the mean scores of the overall planned task ( $M=65,56$ ;  $SD=8,647$ ) and the overall unplanned task ( $M=66,47$ ;  $SD=9,632$ ), the findings indicated a significant difference ( $t(35)=-5,079$ ) only between the mean scores of the pronunciation component in planned task ( $M=12,03$ ;  $SD=2,249$ ) and in unplanned task ( $M=14,39$ ;  $SD=2,811$ ). As opposed to the pronunciation component, the results indicated that there is not any significant difference ( $t(35)=-,052$ ) between the mean scores of the accuracy component in the planned task ( $M=9,89$ ;  $SD=3,151$ ) and in the unplanned task ( $M=9,92$ ;  $SD=3,074$ ). Secondly, there is not a significant difference ( $t(35)=0,602$ ;  $p>.05$ ) between the mean scores of the fluency component in planned task ( $M=12,39$ ;  $SD=3,548$ ) and in unplanned task ( $M=11,97$ ;  $SD=2,558$ ). Thirdly, there is not any significant difference ( $t(35)=1,768$ ;  $p>.05$ ) between the mean scores of the vocabulary component in planned task ( $M=12,39$ ;  $SD=3,548$ ) and in unplanned task ( $M=11,97$ ;  $SD=2,558$ ). Similarly, the results of the component vocabulary indicated no significant difference ( $t(35)=1,768$ ;  $p>.05$ ) between the mean scores in planned ( $M=12,78$ ;  $SD=2,085$ ) and in unplanned task ( $M=12,03$ ;  $SD=2,396$ ). Lastly, the results of the component relevance indicated no significant difference ( $t(35)=0,497$ ;  $p>.05$ ) between the mean scores in planned ( $M=18,47$ ;  $SD=2,793$ ) and in unplanned tasks ( $M=18,17$ ;  $SD=2,287$ ).

**Table 3.** The Effect of Test Type on the Scores of Pre-intermediate Level Students

	Planned		Unplanned		t	df	p
	M	SD	M	SD			
Overall Score	65,56	8,647	66,47	9,632	-,506	35	,616
Accuracy	9,89	3,151	9,92	3,074	-,052	35	,959
Fluency	12,39	3,548	11,97	2,558	,602	35	,551
Vocabulary	12,78	2,085	12,03	2,396	1,768	35	,086
Pronunciation	12,03	2,249	14,39	2,811	-5,07	35	,000
Relevance	18,47	2,793	18,17	2,287	,497	35	,622

Considering the place of the pre-intermediate level among the framework in CEFR, it can be noticeable that pre-intermediate level has different requirements from the students enrolled in that level. A pre-intermediate level student is assumed to be an independent user of the language as opposed to the basic users of that in A1 and A2. Being placed between the low levels and high ones causes the pre-intermediate level to be a Threshold level which is more challenging for the students. That's why the requirements of this level are higher than the basic levels since they are expected to produce the input they acquired in previous levels as output. When the output is in the case, it is known that the learners need practice; otherwise, preparation does not make sense, as the results indicate in the present study. However, the pronunciation component is the only one which still has a significant difference between two tasks, and this shows that even when the other components of oral skills are not affected by the test type, pronunciation is. Namely, test type brings about a difference in the sub-scores of the pronunciation aspect. In addition, it is also noticeable that accuracy facet does not improve by planning. Before this one, Foster and Skehan (1996, 1997) conducted two studies on this matter and found improvement in only one. The findings of the study they conducted in 1996 are in parallel with these results.

### 3.4 The Effect of Test Type on the Scores of Intermediate Level Students

Showing different aspects of the spoken language affected by the test type, in intermediate level, the test results indicated that there is a significant difference ( $t(35)=4,089$ ;  $p<.05$ ) between the overall mean scores in planned ( $M=76,22$ ;  $SD=10,139$ ) and unplanned tasks ( $M=70,56$ ;  $SD=10,697$ ). The findings related to accuracy component indicated a significant difference ( $t(35)=2,404$ ;  $p<.05$ ) between the mean scores in planned ( $M=11,61$ ;  $SD=3,442$ ) and in unplanned task ( $M=10,22$ ;  $SD=4,050$ ). Similar results were obtained for vocabulary and relevance, as well. A significant difference was found ( $t(35)=4,271$ ;  $p<.05$ ) between the overall mean scores of the vocabulary component in planned ( $M=14,94$ ;  $SD=2,574$ ) and unplanned tasks ( $M=12,61$ ;  $SD=12,911$ ). As it was in the other levels, relevance was found to have a significant difference ( $t(35)=4,877$ ;  $p<.05$ ) between the overall mean scores in planned ( $M=19,69$ ;  $SD=1,091$ ) and unplanned tasks ( $M=17,39$ ;  $SD=2,589$ ). The findings are represented in Table 4.

**Table 4.** The Effect of Test Type on the Scores of Intermediate Level Students

	Planned		Unplanned		t	df	p
	M	SD	M	SD			
Overall Score	76,22	10,139	70,56	10,697	4,089	35	,000
Accuracy	11,61	3,442	10,22	4,050	2,404	35	,022
Fluency	14,83	3,745	15,22	3,735	-,545	35	,590
Vocabulary	14,94	2,574	12,61	2,911	4,271	35	,000
Pronunciation	15,14	3,314	15,11	2,482	,050	35	,960
Relevance	19,69	1,091	17,39	2,589	4,877	35	,000

In intermediate level, the effect of test type can be interpreted as having an impact on overall scores, in favor of planned task. As seen, the students performed better when they were prepared in terms of accuracy, vocabulary, and relevance. The finding related to accuracy has been found in parallel with the previous study conducted by Ellis (1987). The improvement by planning in intermediate level was also observed in that research. However, fluency and pronunciation were not affected by the test type in contrary to the previous levels.

The overall effect of test type was found to be significant in elementary and intermediate levels. This shows us that preparation induces higher scores in oral exams in these levels. When we analyze the components to reveal the factors bringing about this difference; accuracy, vocabulary and relevance were the common aspects of their performance that resulted in a significant difference in overall scores. It can be inferred that the students can use the

language accurately, more enriched with vocabulary and depending on the topic they talk about with the condition of planning their speech. On the other hand, simultaneous speech is more likely to cause grammar mistakes, speaking out of the topic and giving a performance with a limited range of vocabulary.

As far as beginner and pre-intermediate levels are concerned, planning the speech in advance does not make a big difference in the scores. Bearing in mind the common facets of these two levels which were not found to be significant (accuracy and fluency), the following arguments can be explanatory. The common point of these two levels is being the term in which the students have more difficulty in producing the target language. In other words, in beginner level, the students meet the language for the first time and have difficulty in the grammar features. This challenge leads them to accuracy problems which bring low fluency scores since the students need to be accurate to be able to speak without many hesitations in this level. Integrating a spiral curriculum, the coursebook used in this institution enables the students to encounter a grammar feature repetitively in beginner and elementary level, thus being familiar with the language forms, the students can progress in communicative skills faster at the end of the elementary level. However, this is not the case in pre-intermediate level for they face new grammar units for the first time and they need more time to internalize them, which resulted in low accuracy and fluency sub-scores in this study. Consequently, the results of the first research question show that there is a significant impact of the test type on oral exams when the level of the students is not beginner or the Threshold in which the students have more challenges. In other words, they perform better when they had a chance to be exposed to the target language features repetitively in their level. This leads them to perform better when they are assessed through planned tasks.

Apart from all these, in three levels, beginner, elementary and pre-intermediate, all vocabulary sub-scores are higher in planned tasks than unplanned ones while it is just the opposite in pronunciation. An explanation of this outcome can be provided through the nature of planned tasks. In view of the time they have to prepare their speech out of the school, they attempt to use new vocabulary that they have not learned before. Their focus is on the meaning they wish to convey rather than the pronunciation. Hence, during their performance, they come up with a wide variety of vocabulary; however, their pronunciation errors increase in the same line with the number of the new words they use. Consequently, their score on pronunciation gets low, while that of vocabulary gets high.

### 3.5 The Effect of Pronunciation on Overall Scores

The results of bivariate regression reveal the effect of pronunciation sub-scores on overall scores for each level in Table 5.

**Table 5.** The Effect of Pronunciation on Speaking Scores

Level	Planned				Unplanned			
	R <sup>2</sup>	df	F	p	R <sup>2</sup>	df	F	p
Beginner	,448	1,36	29,262	,000	,433	1,36	27,458	,000
Elementary	,259	1,34	11,889	,002	,459	1,34	28,812	,000
Pre-Intermediate	,355	1,34	18,709	,000	,431	1,34	25,794	,000
Intermediate	,420	1,34	24,598	,000	,303	1,34	14,808	,000

In planned tasks, beginner level pronunciation sub-score significantly explains 44.8% of the variance in overall scores ( $F(1,36)=29,262$ ;  $p<.05$ ). Likewise, in unplanned tasks, it significantly explains 43.3% of the variance in overall scores ( $F(1,36)=27,458$ ;  $p<.05$ ). Evenly, in intermediate level planned tasks, it significantly explains 42% of the variance in overall scores ( $F(1,34)=24,598$ ;  $p<.05$ ); in unplanned tasks, it is 30.3% of the variance in overall scores ( $F(1,34)=14,808$ ;  $p<.05$ ). These results show us, in beginner and intermediate level, the effect of pronunciation sub-score on overall scores is more in planned tasks than unplanned ones. In beginner level, the students' pronunciation sub-scores on unplanned tasks were significantly higher than the ones in planned tasks. Not being autonomous, the students in this level find it challenging to learn the correct pronunciation of the words, and when they have the chance to prepare their speech and investigate new words to use in a presentation, they give more importance to their meaning and ignore the correct pronunciation of them. Thus, the pronunciation sub-score in the planned task has more impact on overall scores, yet in a negative way corresponding to this reason. As a result, low scores in pronunciation aspect are achieved; however, in unplanned tasks in which they can only use the words they were taught in the lessons, they pronounce those more accurately since the correct pronunciation of the words are emphasized by their teachers and they learn the words accurately in each aspect of it.

When the intermediate level is the case, considering the result that there was not a significant difference in pronunciation sub-score in two types of test; however, contrasting the other levels, the scores were higher in the

planned tasks. Being the highest level, intermediate level students are assumed to be more autonomous and more aware of the components of vocabulary. They are conscious of the fact that pronouncing the target words correctly is of a great importance in terms of intelligibility. When they have the chance for preparation for their speech, they take into account the pronunciation of the words they have just searched and learned; and they can pronounce them correctly during presentations. Thus, they take advantage of preparation and acquire higher scores on that aspect, which results in a high impact on overall scores in this study. To sum up, the results of the second research question also support that pronunciation sub-score plays a bigger role in overall scores in planned tasks than unplanned ones.

When we look at the results in elementary and pre-intermediate levels, it is obvious that the effect of pronunciation sub-score is higher in unplanned tasks than planned tasks. In other words, in elementary level planned tasks, pronunciation sub-scores significantly explain 25.9% of the variance in overall scores ( $F(1,34)=11,889$ ;  $p<.05$ ). Similarly, in pre-intermediate level planned tasks, pronunciation scores significantly explain 35.5% of the variance in overall scores ( $F(1,34)=18,709$ ;  $p<.05$ ). The results of these two levels are in parallel with the results of the first research question. We found that pronunciation had a significant difference in both levels but still biased in favor of unplanned tasks. In these levels, the participants achieved higher scores in paired speaking tests than they achieved in presentations. Similar results were obtained through the second research question that, in these two levels, pronunciation had a bigger impact on paired tasks than presentations. Bearing in mind the problem with the lack of knowledge about how to learn new words, the pronunciation errors induced these findings. However, as opposed to beginner level, in these two levels, a gradual increase was observed between the results in planned tasks (25,9%; 35,5%; 42%) and a decrease between the results in unplanned tasks (45,9%; 43,1%; 30,3%). This change reveals that the role of pronunciation in the scores increases from unplanned to planned tasks as the students' level increases. In other words, the students achieve a higher level of knowledge about how to acquire vocabulary with all its aspects and how to learn the pronunciation of them when their level increases and this may be related to autonomy.

Besides the comparisons, it is obvious that pronunciation sub-score has a significant role in each level and each test type. The results indicate that the role of pronunciation in speaking tasks cannot be ignored and should be taken into account in terms of teaching and testing oral skills. The results were also found consistent with the previous studies. Ma (2015) found that pronunciation component explained 42% variance in overall scores, which was very similar to the findings of the present study. Besides, the findings proved that pronunciation had the biggest role in beginner level on planned, and in elementary level on unplanned speaking exam tasks. Considering the fact that both levels are known as low levels, the results are also in parallel with the study findings of Higgs and Clifford (1982) and De Jong and Van Ginkel (1992) since it was indicated that the effect of pronunciation is more in low levels.

#### 4. Conclusion and Implications

The present study aimed to examine the effects of test type, pronunciation and proficiency level on speaking exam scores. Depending on the results gathered from the subjects, the findings indicated that in elementary and intermediate levels, the students achieve higher scores in planned tasks. This shows that the students take advantage of preparation for their performance in these levels, and it enables them to achieve higher scores. This shows us that testing oral skills through presentations is similar to the process-writing, in which the students have time to get ready, prepare drafts and get feedback before being assessed for their writing skills. In this test type, the students go through a similar process to be assessed for their oral skills.

A developmental measure can be used to assess the potential of learners (Yakışık, 2012). Instead of testing their actual performance immediately, involving the students in such a process that they have the opportunity to be oriented and mediated is likely to show their potential performance based on Vygotsky's Sociocultural Theory (1962). During planning their tasks, they can get help from others such as their instructors or peers; and some aids such as dictionaries or technology, and show their potential performance to achieve higher scores, as proven in the present study.

This kind of test has another advantage that the students can also monitor their process and feel more confident due to higher scores they get in the tasks. Knowing that the self-efficacy factor affects the students' performance as well, teachers can exploit the advantage of planned tasks to increase the students' motivation and self-confidence in speaking skills. Besides, involving in a process and monitoring their performance result in more autonomous learners in a long-term period.

Even though the positive effects of planned tasks were found in elementary and intermediate levels, the students in beginner and pre-intermediate levels did not achieve significantly higher scores in planned tasks. The common point of these two levels is being more challenging for the students in comparison to the other two. The results show that

the students' preparation did not induce higher scores on the tasks in these levels. Examining the sub-scores to reveal where the reasons lie, we see that pronunciation sub-scores are higher in unplanned tasks. Although it was expected that they could achieve higher scores with regards to this when they were prepared in advance; they, surprisingly, achieved lower scores. On the other hand, in both levels, the vocabulary scores increased while the pronunciation scores decreased. In unplanned tasks, they could only use the vocabulary they had been taught in the classes. However, due to the opportunity to prepare their speech, they must have investigated new words to convey their message during the performance beforehand. The problem arising at this point is their lack of knowledge about how to learn new vocabulary without the assistance of a teacher. Even though they can learn the meaning and the use of new vocabulary correctly, they fail in pronouncing them accurately. That's why, the scores of pronunciation component in planned tasks are lower than those in unplanned tasks, unlike the other components. Because of this problem, the students in specific levels such as at the beginning of language learning or at the Threshold level cannot use the advantages of planned tasks in speaking tests. To deal with this, teaching them how to learn the pronunciation of new words should be focused to lead the students to be more autonomous.

First and foremost, teaching the students how to use the dictionaries should be elaborated through electronic devices. Students should benefit from the audio materials in the dictionaries. Modeling the use of online dictionaries or applications of reliable sources can be helpful. However, the students without sufficient ear-training or the ones in low-levels have problems in imitating the sounds they hear. Furthermore, only listening to the correct pronunciation is not also enough for the others. Especially the students with visual intelligence (Gardner, 2011) suffer in the lessons without visual aids, and audio recorders are never enough for them to acquire language features. Thanks to International Phonetic Alphabet (hereafter, IPA), a more concrete material through the symbols of each sound constituting the words can be presented. Consolidating the IPA training with audio materials can improve learning since students can get benefit from the phonemic transcriptions on the dictionary entries as they need to learn how to pronounce a word in a situation with no chance to listen.

In addition, increasing the amount of time allocated for pronunciation teaching in class through the implementation of several different techniques can be helpful to help students overcome pronunciation problems. For example, 'conceptualization' can be integrated into curriculum (Hişmanoğlu, 2004). In conceptualization, through contrastive analysis, learning the sound system in stages, then the IPA and segmental structures, structuring the cognitive mind map in learners' minds and focusing on the stages of the sound system are the steps to be followed (Geylanoğlu, 2016).

As another technique, Audio Articulation Method presented by Demirezen (2008) and based on the steps indicated by Celce-Murcia, Brinton and Goodwin (2010) can be used to undo the pronunciation errors of specific sounds within a lesson hour. The steps in AAM are as following: prescribing a model, listening and imitating, using minimal pairs, visual aids, tongue twisters and developmental approximation drills, reading aloud, recitation, and recordings of learners' productions, respectively. In other words, the steps can be summarized as presentation with modeling, visual aids to show the place of articulation of each sound in a chart and drills; and the process is followed by recording students' voices so that they can evaluate their performance and monitor their progress. Through this technique, it is possible to teach phonemes when problems arise within a short time in the lessons.

In conclusion, the present study provides the effects of oral test types through the scores the students achieve and the underlying reasons behind by analyzing each component. The findings proved that pronunciation has a great value in speaking tests with a high proportion of overall scores. The effect of it revealed the necessity of more focus on teaching pronunciation. In addition, assessing the students' oral skills through planned tasks has been found to be effective in terms of showing their potential performance through a developmental process.

## References

- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341-366. <https://doi.org/10.1177/0265532209104666>
- Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (2010). *Teaching pronunciation hardback with audio CDs (2): A course book and reference guide*. Cambridge: Cambridge University Press.
- Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, 11, 367-383. <https://doi.org/10.1017/S0272263100008391>
- De Jong, J. H., & Van Ginkel, L. W. (1992). 15 Dimensions in oral foreign language proficiency. In L. Verhoeven & JHAL de Jong (Eds.), *The construct of language proficiency: Applications of psychological models to language*

- assessment, p. 187. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/z.62.19jon>
- Demirezen, M. (2008). The /æ/ and / ʌ / phonemes as fossilized pronunciation errors for Turkish English language teachers and students: Undoing the fossilized pronunciation error. *Journal of Language and Linguistic Studies*, 4(2), 1-82.
- Ellis, R. (1987). Interlanguage variability in narrative discourse: Style shifting in the use of the past tense. *Studies in Second Language Acquisition*, 9(1), 1-19. <https://doi.org/10.1017/S0272263100006483>
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18(3), 299-323. <https://doi.org/10.1017/S0272263100015047>
- Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Geylanioglu, A. S. (2016). *Developing English language learners' pronunciation through conceptualization* (M.A Thesis). Retrieved from Yükseköğretim Kurulu Başkanlığı. (Accession No. 448446).
- Güllüoğlu, Ö. (2004). *Attitudes towards Testing Speaking at Gazi University Preparatory School of English and Suggested Speaking Tests* (MA Thesis). Retrieved from Yükseköğretim Kurulu Başkanlığı. (Accession No: 191801).
- Güneş, A. L. (2010). *Differences and challenges involved in the assessment of speaking skill: the case of three universities in Ankara*. (MA Thesis). Retrieved from Yükseköğretim Kurulu Başkanlığı. (Accession No: 263018)
- Higgs, T. V., & Clifford, R. (1982). The push toward communication. In T. V. Higgs (Ed.), *Curriculum, competence and the foreign language teacher* (pp. 51-79). Lincolnwood, IL: National Textbook.
- Hişmanoğlu, M. (2004). *Problematic English consonants and vowels for Turks in relation to ELT: Text development and problem solutions* (Unpublished Doctoral Dissertation). Ankara: Hacettepe University, Institute of Social Sciences.
- Kang, O. (2013). Relative impact of pronunciation features on ratings of non-native speakers' oral proficiency. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (Aug. 2012. pp. 10-15). Ames, IA: Iowa State University.
- Kang, O., & Ginther, A. (Ed.). (2018). *Assessment in second language pronunciation*. Routledge: New York. <https://doi.org/10.4324/9781315170756>
- Lightbown, P. M., & Spada, N. (2013). *How Languages are Learned 4th edition-Oxford Handbooks for Language Teachers*. Oxford: Oxford University Press.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511733017>
- Ma, R. (2015). *The role of pronunciation in speaking test ratings* (MA Thesis). Brigham Young University: Provo: 4426. Retrieved from: <https://scholarsarchive.byu.edu/etd/4426>
- Martinez, A. A. P. (2004). *Fluency and the planned and unplanned speech of Spanish-English bilinguals. Doctoral dissertation*. Puerto Rico: University of Puerto Rico, Graduate School of Educational Sciences. Available from ProQuest Dissertations & Theses Global database. (Accession Order No. AAT 1422010).
- Öncel, A. (2016). *Learner preferences for individual and paired speaking tests in a multicultural EFL context* (M.A. Thesis). Retrieved from Yükseköğretim Kurulu Başkanlığı. (Accession: 429303).
- Skehan, P., & Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language teaching research*, 1(3), 185-211. <https://doi.org/10.1177/136216889700100302>
- Tajima, M. (2003). *The effects of planning on oral performance of Japanese as a foreign language* (Unpublished Dissertation). West Lafayette: Purdue University, Graduate School of Languages and Cultures. Available from ProQuest Dissertations & Theses Global database. (Accession Order No. AAT 3113880).
- Thornbury, S. (2012). Speaking instruction. In J.C. Richards & A. Burns (Eds.), *Cambridge guide to pedagogy and practice in second language teaching* (pp. 198-207). Cambridge: Cambridge University Press.
- Underhill, N. (1987). *Testing spoken language: A handbook of oral testing techniques*. Cambridge: Cambridge University Press.
- Vygotsky, L. S. (1962). *Language and thought*. Ontario: Massachusetts Institute of Technology Press. <https://doi.org/10.1037/11193-000>

- Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistants. *TESOL Quarterly*, 26, 693-711. <https://doi.org/10.2307/3586869>
- Yakışık, B. Y. (2012). *Dynamic Assessment of ELT Students' Speaking Skills* (M. A. Thesis). Retrieved from Yükseköğretim Kurulu Başkanlığı. (Accession: 317052).