ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
2020; 11(2); 199-218

Four-Skill Assessment of Turkish Language: Results from a Pilot Project

Emine EROĞLU * H. Eren SUNA ** Hande TANBERKAN ***
Amine CANIDEMIR **** Umare ALTUN *****
Mahmut ÖZER ******

Abstract
This study analyses the results of the 'Four-Skill Test in Turkish Language' (FSTTL) project conducted by the Ministry of National Education to assess the language skills of students as a pilot project and investigates the effects of various variables on language skills. Relationships between language scores and school type, gender, preschool participation, parents' level of education, and course grades are investigated in this descriptive study. The sample is consisted of 1932 students in seventh grade who participated in the pilot study. Test battery, consisted of reading, listening, writing, and speaking subtests, is used to assess the language skills of students within the scope of the FSTTL. Findings show that students in imam-hatip middle schools and middle schools performed at a similar level in all subtests. Female students performed significantly higher than male students in all subtests. Students participated in pre-school education performed significantly higher than those who did not participate in reading, writing, and listening subtests. Findings also show that the increase in parents' level of education leads to an increase in students' subtest scores. The effect of parents' level of education on subtest scores is comparatively higher than the effects of other factors in focus. Significant correlations have been obtained between the four-skill scores and student's Turkish course, social sciences, mathematics, and science course grades. It is suggested that FSTTL must be developed based on the experiences of the pilot project as a standardized test in accordance with the international standards and actively used to improve educational processes.

Keywords: Language Skills, Four-Skill Test in Turkish Language, Language Teaching, Assessment

INTRODUCTION

Language is a living entity that provides communication between people, is dynamic, has its own specific rules, a system of secret treaties that it is not known when it was formed, and a social structure consisting of sounds (Ergin, 1998, p. 2). Language is the basic tool for people to engage with the environment and to express their thoughts and feelings. Any verbal and written reaction of the individual who perceives the events and actions in his environment is directly related to his language skills. In this respect, language skills are among the most basic skills expected for the individual to be able to adapt to daily life, to interact with his environment as an individual and to be a part of social life (Jing, 2006).

* Ministry of National Education, Ankara-Turkey, emineeroglu34@gmail.com, ORCID ID: 0000-0001-6611-3313
** PhD., Ministry of National Education, Ankara- Turkey, herensuna@gmail.com, ORCID ID: 0000-0002-6874-7472
*** PhD., Ministry of National Education, Ankara- Turkey, handetanberkan@gmail.com, ORCID ID: 0000-0001-7142-5397
**** Ministry of National Education, Ankara- Turkey, aminecanidemir@hotmail.com, ORCID ID: 0000-0003-0705-1604
***** Ministry of National Education, Ankara- Turkey, umarealtun@gmail.com, ORCID ID: 0000-0001-8995-0763
****** Prof. PhD., Ministry of National Education, Ankara- Turkey, mahmutozer2002@yahoo.com, ORCID ID: 0000-0001-8722-8670

To cite this article:
Eroğlu, E., Suna, H.E.,Tanberkan, Canıdemir A., Altun, U. & Özer, M.(2020). Four-skill assessment of Turkish language: Results from a pilot project. *Journal of Measurement and Evaluation in Education and Psycholog, 11*(2), 199-218. doi: 10.21031/epod.687758

Geliş Tarihi: 11.02.2020
Kabul Tarihi: 26.03.2020

Individuals use their acquired language skills for social and academic purposes. It is aimed to improve the language skills of individuals both in the social context they will use in daily life and the academic context that they will use throughout their education. Therefore, education systems are structured in such a way that individuals can improve their language skills in both social and academic contexts. Thus, it is aimed to raise individuals who can actively participate in the society, express their feelings and thoughts as they wish, and have literacy skills (Bayyurt, 2013; Cook, 1999). It is also important to note that individuals who have higher levels of language skills also have a significant advantage in employment in diverse sectors (Budria, Colino & Matinez de Ibarreta, 2019; Gazzola & Mazzacani, 2019). Recently, since widespread automation in the labor market is supported by artificial intelligence technologies (Perc, Ozer & Hojnik, 2019), language skills become a much more important factor for adaptability in new circumstances. In this manner, language skills still have a crucial role in communication between people towards the demands of the labor market.

Gaining language skills, which are the basic means for individuals to express their feelings and thoughts, plays an important role in the language to live and to be delivered to the next generations in a proper way. In order for a language to be properly learned, individuals must have gained reading, writing, and listening skills as well as speaking (MoNE, 2019a). Therefore, the acquisition of language skills requires the development of four basic language skills simultaneously (Gautam, 2019; Manaj-Sadiku, 2015). Verbal speech on any subject, texts read to learn, news listened to in daily life, or texts written in order to express their opinions provide individuals to meet their different needs. Four basic language skills, reading, writing, listening and speaking, have a natural relationship with each other, and the development of one skill positively affects the development of other language skills (Brown, 2001; Chengyu, 2018; Gautam, 2019).

Each of the four basic language skills ensures that different functions of the language are performed. Children learn their native language primarily through listening. This learning is also the basis of the individual's ability to learn the native language. Meaning and sounds come to the fore in reading skills. Some symbols need to be analyzed and interpreted to improve reading skills. Writing skill refers to the transformation of emotions, thoughts, opinions, and dreams into text. In writing skill, it is important for the individual to express what they saw, heard, thought, and lived in text. One of the general objectives of the Turkish curriculum was expressed as "to provide students with the ability and habit of describing what they see, watch, listen, read, examine and think, design with words or writing correctly and in accordance with the purpose". The accurate written communication depends on the fulfilment of the external structure, internal structure (narration), spelling, and punctuation dimensions (Deniz, 2000; Kantemir, 1997; Özkırımlı, 1994). Speaking skills can be explained as a set of skills that enable the individual to communicate in the target language (Barın, 1997). Speaking skill is considered as one of the most frequently used language skills of the individual to communicate in a social and academic context (Boonkit, 2010). As can be seen, four basic language skills are considered as components of the language skill of the individual. Models used in the development of language skills and assessment methods of language skills in Turkey is described respectively.

Four Basic Language Skill Approaches and Global Trends in Measuring Language Skill

The development of language skills has been one of the most important issues in education. Many different methods have been developed for the development of language skills, which is a basic communication requirement, and two of these methods are frequently used (Gautam, 2019; Widdowson, 1978). The first of these methods is the behavioural model (major skills model) that divides language skills into subskills such as reading, listening, speaking, and writing and focusing on the development of these subskills separately. In this approach, it is accepted that there is a natural link between basic skills, but each component is developed within itself (Akram & Malik, 2010; Hinkel, 2010). In the integrated model, language skill is seen as a whole with all subskills, and subskills are tried to be developed with the same methods (Xue-Ping, 1997). Both models have their own advantages and limitations, and the approaches used in education systems differ.

Although they are in an organic relationship with each other, the benefits of addressing these skills separately for the development of four basic language skills have been demonstrated by linguists based on data (Hinkel, 2010). Addressing basic skills separately in language teaching enables different methods to be used in developing these skills. In addition, individuals' gains, strengths, and aspects that are open to development can be examined separately according to their language skills (Hinkel, 2002; Stern, 1983). For example, a personalized development plan can be presented to an individual who has sufficient listening and speaking skills, but not sufficient writing and listening skills.

Linguistic scientists express that with the development of language skills separately, students can understand different layers of language faster and use different skills more effectively (Canale & Swain, 1980; Mitchell & Vidal, 2001). Developing language skills separately can shorten learning time and speed up the use of language skills. However, language skills must be used together for advanced applications in language teaching. For this reason, it is recommended to integrate the skills that are handled separately for the development of language skills after a certain level of competence, and to configure the language teaching accordingly at a later level (Halliday, 1978; Nunan, 1989; Widdowson, 1978).

Structuring the language teaching by grouping it according to the skills has led to a similar approach in the assessment of language skills. In order to assess the gains based on reading, listening, writing, and speaking, many tests that measure language skills are structured to consist of subtests that measure four basic skills separately. In tests designed in this way, each basic skill is accepted as a component of the language, and a score is calculated for each component as a result of the assessment (Bachman & Palmer, 1996).

In tests that evaluate four basic skills separately, test development processes specific to subtests for each skill can be followed; therefore, the approach of separating the skills according to subtests is frequently preferred. The use of skill-specific subtests has been used since the 1960s as it facilitates test development and implementation processes (Hinkel, 2010). Today, each basic language skill is measured through separate subtests within Test of English as a Foreign Language (TOEFL), Test of English for International Communication (TOEIC), International English Language Testing System (IELTS) and Pearson Test of English Academic (PTE ACADEMIC) that are used internationally to determine proficiency in various languages.

Assessment of Language Skills in Turkey

The main purpose of teaching Turkish is to make students proficient in the skill areas of their native language. It was stated in the program that language skills are related to daily life and that the development of the individual in every field is a prerequisite (MoNE, 2019a). When it is examined in detail, it is seen that the education and teaching of the Turkish language are structured on four basic language skills, which are reading, writing, listening and speaking, and grammar.

Understanding, one of the two most important aspects of the native language education and training process, is composed of listening and reading skills. Narration consists of speaking and writing skills (Kavcar, Oğuzkan & Sever, 1999). Listening and speaking skills are the skills that individuals acquire from the moment they are born and are learned before other skills. For this reason, it is aimed to support these skills in school-age children and to gain additional reading and writing skills. Unless the four basic skills are used together at a certain level, it is not possible to learn Turkish with all its functions (Doğan 2009).

Although it has an important place in the Turkish curriculum, there is no standard assessment method and assessment tool for students' four language skills in Turkish. Although there are learning outputs based on basic language skills at each grade level in the Turkish curriculum, assessment of these skills has been limited to in-class practices. In addition, no monitoring studies are conducted to assess the extent to which students have these basic skills. Language skills assessed in centralized interstage

transition examinations and periodic monitoring studies remain limited (MoNE, 2018; ÖSYM, 2018). There are subtests that assess the language skills of students in the central examinations which are applied within the scope of the High School Transition System (LGS) and Higher Education Institutions Exam (YKS), but these subtests focus only on reading skills (MoNE, 2018; ÖSYM, 2018). Turkish-Mathematics-Science Student Achievement Monitoring Study (TMF-ÖBA), which was implemented for the first time in 2019, and the Academic Skills Monitoring and Evaluation (ABİDE) focused on only the reading skill of students (MoNE, 2019b, MoNE, 2016). Additionally, the reading skill of students are assessed in international studies such as Programme for International Student Assessment (PISA) and Progress in International Reading Literacy Study (PIRLS). The results of these studies provide more important insights about students' achievements if the results are investigated in detail (Ozer, 2020).

Central examinations for assessment of basic four language skills are carried out for individuals who learn Turkish as a second language or live abroad. The Turkish Proficiency Exam (TYS) developed by Yunus Emre Institute, and the level determination and diploma exams developed by Turkish and Foreign Language Research and Application Centres (TÖMER) also assess four basic language skills. However, the target group of the examinations is individuals who learn Turkish as a foreign language. In order to assess students' basic four language skills in Turkish with standard measurement tools by overcoming this limitation, "Project for Determining and Assessing Turkish Language Proficiencies in Four Skills" was initiated by the Ministry of National Education (MoNE).

It is aimed to measure the language skills of the students within the framework of the competencies determined by the Project for Determining and Measuring Turkish Language Proficiencies in Four Skills. The results to be obtained will provide the important insights about the students competencies in language skills, language teaching and provide feedback on the effectiveness of the teaching process. Within the scope of this project, Four Skills Turkish Language Exam developed under the coordination of MoNE General Directorate of Measurement, Assessment and Examination Services. It is the first large-scale application to assess students' skills in the native language within the common assessment framework and in accordance with international assessment standards (MoNE, 2020). Language laboratories have been established in 15 provinces in order to perform the testing process at international standards. These language laboratories are equipped with headphones in which listening and recording can be performed and test cabinets that isolate external sounds.

The first step taken within the scope of the project is to develop an assessment framework to determine the scope of Turkish basic language skills. During the development of the framework, workshops were organized by the MoNE, and academics from Turkish education, experts from Turkish teaching, and measurement and evaluation specialists studied together in these workshops. Within the assessment framework developed, it was determined which behaviours to be observed in each of the basic skills, and concrete behavioural responses of language skills were developed.

The development of the assessment framework is one of the initial studies in which student behaviours to be observed within the scope of Turkish four basic language skills are determined. Although widely accepted assessment frameworks have been developed in many foreign languages, there is no framework reflecting the common view of experts in Turkish before this study. The item and task development process was carried out after the completion of the assessment framework. Each item and task developed was harmonized with the assessment framework. A pilot study of the Turkish Language Exam in Four Skills by the MoNE was conducted on 24-26 April 2019 in language laboratories with the participation of 1932 7th grade students in 15 provinces including Adıyaman, Ankara, Antalya, Aydın, Bursa, Denizli, Erzurum, Gaziantep, İstanbul, Konya, Kütahya, Muğla, Samsun, Şanlıurfa, and Trabzon. Within the scope of the test, all subtests related to four basic language skills were applied in the computerized environment.

Due to the fact that existing test applications focus only on reading skills, developed for students who learn Turkish as a second language or do not conducted as large-scale application, it is not possible to have valid and reliable data reflecting the language competencies of the students in Turkey. The number of studies focusing on determining the variables that affect the development of language skills is also

very limited for the same reason (Erkek, Batur, Kaplan & Ercan, 2017; Lüle Mert, 2013, 2014). Turkish Language Exam for Four Skills is an important step taken in order to overcome this deficiency, and the pilot study has been successfully carried out in accordance with international assessment standards. The results obtained will make it possible to implement data-based studies to develop these skills and to meet the needs of our education system by making them sustainable practices. The project outputs will provide important feedback in determining the improvements to be made in the curriculum and the development of Turkish language teaching. It will also make it possible to develop four skill tests with international standards on different levels of Turkish proficiency.

The psychometric analysis made with the data obtained from the pilot study is important in terms of ensuring that the test will be more qualifying in the initial application. Similarly, the analysis results for student characteristics on pilot study data will provide important information about the role of student characteristics in language skills. In this context, it is considered that the first results presented by the Four Skills Turkish Language Test regarding the quality of the pilot study implementation data and the relationship between student characteristics and language skills are important.

In this study, the pilot study results of the Four Skills Turkish Language Test conducted under the coordination of MoNE General Directorate of Measurement, Assessment, and Examination Services were examined, and it was aimed to determine the change of language skill performances in terms of various students, parents and school characteristics.

For this purpose, this study is conducted to answer the following research questions:

1. Is there any significant difference in students' reading, listening, writing and speaking subtest mean scores

    1 a. according to the type of school?

    1.b. according to the participation in pre-school education?

    1.c. according to the gender groups?

    1.d. according to the education levels of the parents?

2. Is there a significant relationship between students' language skills scores and 7th grade scores in Turkish, social sciences, mathematics, and science courses?

METHOD

Research Model

In the study, the current situation of the participants regarding language skills was assessed, and the relationship between language skills and various variables was examined. The descriptive correlational model was used in the design of the research. In descriptive models, phenomenon or condition in focus is examined as it is, and the current situation is described in detail (Karasar, 1999). In the descriptive correlational model, which is one of the submodels of the descriptive model, the relationships between variables are examined in detail without any external intervention.

Population and Sample

The research population is composed of the students in the seventh grade in Turkey during the academic year 2018-2019. In the sample of the study, there are 1932 seventh grade students in 15 provinces. In the sampling process, two-stage convenience sampling method was used. In this sampling type, it is possible to describe and compare the characteristics of various subgroups that are considered to be suitable according to various criteria (Büyüköztürk et al., 2016). Schools were selected according to their distance to language laboratories, type (secondary school and imam hatip secondary school), and

gender distribution criteria. After the schools were selected according to these criteria, the seventh grade classes in the school were included in the sample. In other words, all students in selected branches were applied, and after selecting the school, cluster sampling was carried out in the selection of students. In Table 1, the distribution of the study sample according to the student characteristics within the scope of the research aim is given.

Table 1. Demographic Characteristics of Students in the Study Sample

Variable	Sub Group	Frequency (f)	Ratio (%)
Gender	Female	1027	53.2
	Male	905	46.8
School Type	Secondary School	1302	67.4
	İmam Hatip Sec. School	630	32.6
Preschool Education Status	Participated	1498	77.5
	Not Participated	434	22.5
Mother's Education Level	Primary School	563	29.1
	Secondary School	287	14.9
	High School	485	25.1
	Higher Education	424	21.9
	Not Available Data	173	9.0
Father's Education Level	Primary School	304	15.7
	Secondary School	243	12.6
	High School	551	28.5
	Higher Education	663	34.3
	Not Available Data	171	8.9

As seen in Table 1, the gender distribution of the students in the study sample is quite balanced. 67.4% of the students are in secondary school, and 32.6% of them are in imam hatip secondary school. The majority of the students in the sample (77.5%) participated in pre-school education. It is determined that the ratio of students whose mothers are educated at high school or higher education level is 47%. The ratio of students whose father is educated at high school or higher education level is 62.8%.

Data Collection Tools

The data used in the study were obtained through the test battery developed for the Four Skills Turkish Language Test. Before the test battery was developed, a well-attended workshop was organized to determine the Turkish language skills to be measured, and an assessment framework was developed. Following the developing of the assessment framework, the most appropriate item and task formats were decided to assess the four language skills. A specialist group consists of Turkish linguistic experts, senior teachers in Turkish teaching practices, and measurement and evaluation experts evaluate the assessment framework and educational outputs which they expected. They agreed on item formats considering four-skill language assessment practices around the world. In this manner, it is decided to develop items related to reading and listening skills in multiple-choice format. Additionally, it is determined to develop tasks for speaking and writing skills, which enable students to structure their responses with a broader extent (MoNE, 2020). Accordingly, test-blue prints are prepared for each of the four-skills in Turkish for students in 7th grade. To decide on the cognitive levels of educational outcomes and related items, diverse taxonomies are considered, and four-level taxonomy is selected by the specialist group. In Table 2, four-level taxonomy, which is used in the pilot project, is given.

Table 2. Four-Level Taxonomy of Four Skills Turkish Language Exam

Level 1	Level 2	Level 3	Level 4
Remembering, Recognizing and Selection	Understanding and Inference (Comprehend explicitly stated information)	Inference and Interpretation (Comprehend explicitly not stated information)	Evaluation and Reflection

Educational outputs and items which are considered within the scope of the pilot project are mapped with cognitive levels in Table 2. In listening and reading sections, which consist of multiple-choice items, items are mapped with cognitive levels between level 1 and level 3 due to the limitations of item format. In the pilot project, two online test booklets for listening and reading subtests, and five online booklets for writing and speaking are developed as parallel tests. All items are developed by senior item writers in Turkish language and Turkish linguistics, and item revisions are conducted by measurement and evaluation experts. Concurrently, rubrics for open-ended tasks are developed by the specialist group, and rubrics are evaluated externally by academics from Turkish language education. Lastly, all approved items were clustered to online test booklets considering the balance of educational outcomes and item difficulties.

In the test battery, students were subjected to reading, listening, speaking, and writing subtests, respectively. The questions, tasks, and times for response according to the subtests are given in Table 3.

Table 3. Structure of Subtests in Four Skills Turkish Language Test

Subtest	Item and Task Type	Item or Task Number	Time
Reading	Multiple-Choice Item	20	30
Listening	Multiple-Choice Item	20	30
Speaking	Structured Task	2	10
Writing	Structured Task	4	60

As seen in Table 3, each of the reading and listening subtests consists of twenty multiple-choice items. In these subtests, students were given thirty minutes of response time. In the speaking subtest, students were given two tasks that were asked to explain themselves and the other to explain the steps of a process or the situation presented with the visual. Students complete this subtest in about ten minutes. In the writing subtest, students were given four tasks, including preparing a short text consisting of sentences, paragraphs, and a text including several paragraphs. The response time given to students to complete the four tasks is 60 minutes.

The subtests in the developed test battery differ structurally. Reliability analyses for reading and listening subtests consisting of multiple-choice items were performed with the Kuder-Richardson 20 coefficient frequently used in this item type. The Kuder-Richardson 20 (KR-20) coefficient is a coefficient used to calculate the internal consistency of items scored in two categories as correct and incorrect (Cronbach, 1951; Kuder & Ricardson, 1937). Kuder-Richardson 20 coefficients calculated for A and B forms of reading and listening subtests are given in Table 4.

Table 4. Internal Consistency Coefficients in Reading and Listening Subtests

Subtest	Form	Item Number	KR-20 Coefficient
Reading	A Form	20	0.720
Reading	B Form	20	0.771
Listening	A Form	20	0.768
Listening	B Form	20	0.779

As seen in Table 4, the KR-20 coefficients calculated for both forms of reading and listening subtests ranged from 0.720 to 0.779. The reliability coefficients calculated at 0.70 and above for measurement

tools used in education and psychology are considered as acceptable (Cronbach, 1951; Kuder & Richardson, 1937; Tavakol & Dennick, 2011).

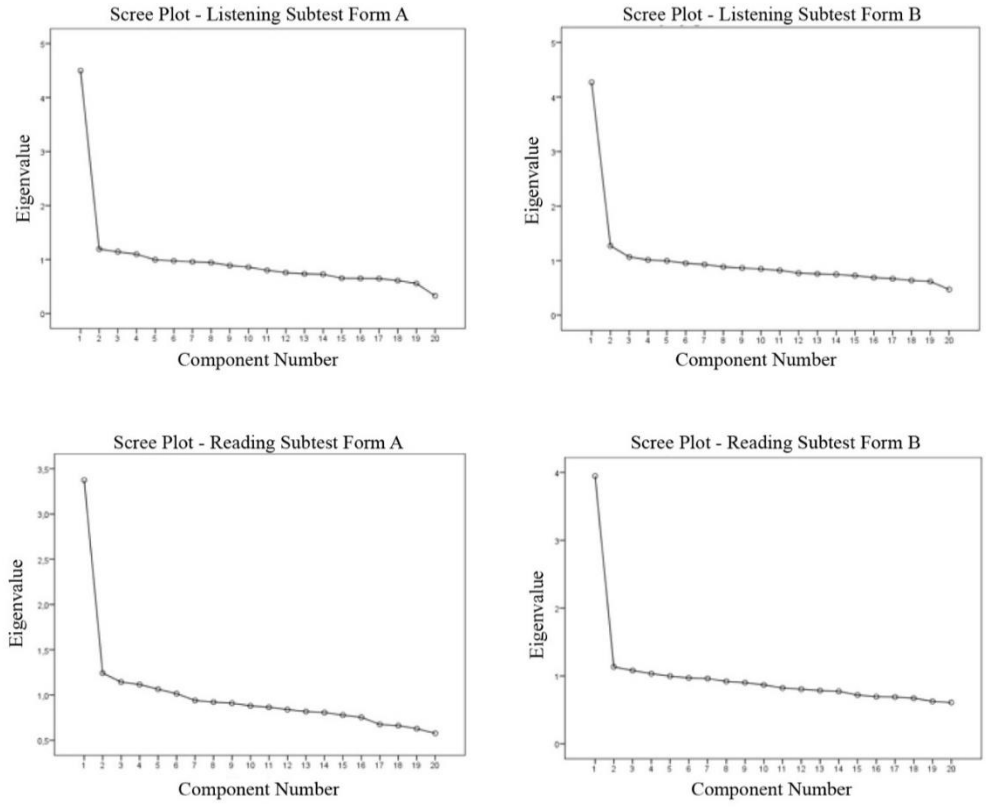
In order to provide information about the validity of the results obtained from reading and listening subtests, exploratory factor analysis (AFA) was performed to reveal the structural dimensions of both subtests. Kaiser-Meyer-Olkin (KMO) values obtained in the forms of both subtests, factor numbers with eigenvalues greater than one, variance ratio explained by the dominant factor, and factor loadings of the items under the dominant factor are given in Table 5 and scree plots are given in Chart 1.

Table 5. Exploratory Factor Analysis Results of Listening and Reading Subtests

Subtest	Form	KMO Value	Number of Factors with Eigenvalue > 1	Variance Explained by the Dominant Factor	Range of Factor Loadings
Reading	A Form	0.828	6	17%	0.047-0.532
Reading	B Form	0.888	4	20%	0.044-0.578
Listening	A Form	0.898	4	22%	0.091-0.774
Listening	B Form	0.901	4	21%	0.189-0.699

*Items below factor loading 0.30 are revised before taken into the test.

Chart 1. Scree Plots for Reading and Listening Subtests



The KMO values given in Table 5 show that the items in the forms related to the two subtests can be resolved by factor analysis. Although there are possible factors with eigenvalues greater than one in all forms, there is clearly a sharp decrease in the scree plots in Chart 1. This indicates that the items in the forms for both subtests are grouped under a single and dominant factor. The factor loadings of two of the items in each subtest form are below 0.32. The relevant items need to be strengthened in the initial

application. However, they were included in the analysis since they did not have a negative loading in this study.

In speaking and writing subtests, students are asked to answer open-ended tasks. In these subtests, students' performances are scored by assessment specialists through the answers they give in the tasks presented to them. For this purpose, assessors are trained about open-ended task assessment via rubrics, and all responses of students are assessed by assessors via a well-attended workshop conducted by MoNE. Each of the open-ended tasks is assessed by two assessors with a blinded approach, and consistency between two assessors is considered. When the score difference between assessments is significant, the final score is determined by a high-level assessor, who is a senior assessor in Turkish language education. Evidence for reliability in these types of tests is mostly provided by the interrater reliability method. In this method, the consistency between the scores given by the raters for the answers of students to the tasks is examined (Crocker & Algina, 1986). Cramer's V coefficient (Cramer's V) and contingency coefficient were used to obtain evidence of inter-rater reliability, and the coefficients with regard to speaking subtest were given in Table 6 based on forms.

Table 6. Consistency Coefficients between Raters in the Speaking Subtest *

Form	Task Type	Cramer's V (Mean)	Contingency Coefficient (Mean)
A Form	First Task	0.56	0.68
	Second Task	0.58	0.69
B Form	First Task	0.47	0.62
	Second Task	0.66	0.75
C Form	First Task	0.56	0.69
	Second Task	0.56	0.67
D Form	First Task	0.59	0.69
	Second Task	0.62	0.69
E Form	First Task	0.49	0.58
	Second Task	0.46	0.56

* *In the first task, students are asked to introduce themselves, in the second task to explain the steps of a process or the situation presented with the visual.*

As seen in Table 6, in the speaking subtest, the V coefficients calculated between the raters were between 0.46 and 0.66, and the contingency coefficients were between 0.56 and 0.75. There are no generally accepted standards as in the other types of reliability for the V coefficient and the contingency coefficient, whose values vary between 0 and 1. However, V coefficients greater than 0.25 are considered to provide information about the general agreement between the two variables (Akoğlu, 2018). It is seen that the V coefficients given in Table 4 are well above this criterion. The contingency coefficients calculated in this subtest are higher than the V coefficients and indicate that the consistency between raters is relatively high.

The reliability coefficients between the raters calculated in the writing subtest are given in Table 7.

Table 7. Interrater Consistency Coefficients Calculated in Writing Subtest *

Form	Task Type	Cramer's V (Mean)	Contingency Coefficient (Mean)
A Form	First Task	0.93	0.85
	Second Task	0.81	0.81
	Third Task	0.60	0.72
	Fourth Task	0.68	0.74
B Form	First Task	0.86	0.83
	Second Task	0.71	0.71
	Third Task	0.56	0.66
	Fourth Task	0.70	0.76
C Form	First Task	0.91	0.85
	Second Task	0.80	0.81
	Third Task	0.66	0.75
	Fourth Task	0.69	0.77
D Form	First Task	1.00	0.87
	Second Task	1.00	0.82
	Third Task	0.68	0.75
	Fourth Task	0.66	0.74
E Form	First Task	0.83	0.82
	Second Task	0.92	0.68
	Third Task	0.59	0.69
	Fourth Task	0.74	0.71

**Students are expected to write a sentence in the first and second tasks, a paragraph in the third task, and a text composed of several paragraphs in the fourth task.*

As seen in Table 7, the mean V coefficients calculated in different forms of the writing subtest ranged from 0.56 to 1, and the mean consistency coefficients ranged from 0.66 to 0.85. These coefficients indicate a high level of consistency among raters, as in the speaking subtest.

Evidence regarding the reliability and validity of the results obtained from the test battery shows that the data obtained from the pilot study is sufficient in terms of psychometric perspective. As it can be seen from Table 4, Table 5, and Table 6, it is possible to revise particular items and tasks in the test battery to be more qualified in the initial application, but in this study, all items and tasks are included in the analysis in their current form.

Ethics Committee Permission

The data of this research were used with the letter number of 42497731-605.99-E.6452557 dated 17.04.2020 of the General Directorate of MEB Measurement, Evaluation and Examination Services.

Data Analysis

In the study, t-test, single-factor variance analysis (ANOVA), effect size analysis, Pearson correlation analysis were used for the analysis of quantitative data obtained with the test battery. The t-test and single-factor ANOVA were used to examine the significance of the difference between the language skill mean scores of the groups, and the eta-square effect size was used to analyze the effect of the variables on the scores. The Pearson correlation analysis was used to determine the and significancy and strength of the relationship between the variables. Significant differences between the groups were interpreted by taking into account their effect size. Criteria for effect size (partial eta-square) are as follows: PES<0.02 is small, 0.02<PES<0.13 is medium, and PES<0.13 is a high level of effect (Miles & Shelvin, 2001).

Participation in pre-school education, gender, and parents' level of education are selected as possible effective variables on the language skill of students. It is shown that these demographic and educational

variables lead to significant changes on students' language development and skills (Bakken, Brown & Downing, 2015; Catts, Fey, Zhang & Tomblin, 2001; Reilly, Neuman & Andrews, 2019; Schermse et al., 2018, Storch & Whitehurst, 2002). The difference between school types is also examined to have insights about the possible effect of educational program differences on language skills of students.

RESULTS

In the findings section of the research, descriptive statistics, and the findings related to each research question are given, respectively.

The mean scores and other descriptive statistics obtained by students in subtests for language skills are given in Table 8.

Table 8. Descriptive Statistics of Language Skill Subtest Scores

Subtest	Possible Score Range	Lowest Score	Highest Score	$\bar{X}$	SD
Reading	0-20	0	20	10.63	3.63
Listening	0-20	0	20	11.70	2.98
Writing	0-36	1	36	16.82	8.09
Speaking	0-36	15	36	27.21	3.95

As seen in Table 8, the mean scores calculated in the reading and listening subtests, where the scores that can vary between 0 and 20, are quite close. In the writing and speaking subtests ranging from 0 to 36, the students perform quite differently. It can be seen in Table 7 that the students perform relatively high in the speaking subtest (X = 27.21, SS = 3.95) and that they showed relatively low performance in the writing subtest (X = 16.82, SS = 8.09).

Findings Related to the First Research Question

The findings of the t-test and effect size related to the research question 'is there any significant difference in students' reading, listening, writing and speaking subtest mean scores according to the type of school?' are shown in Table 9.

Table 9. t-Test Results of Language Skill Subtest Scores by School Type

Subtest	School Type	n	$\bar{X}$	SD	df	*t*	η^2
Reading	Secondary School	1238	10.56	3.63	1832	1.058	---
	İmam Hatip Sec. School	596	10.78	3.62			
Listening	Secondary School	1240	11.73	2.99	1836	0.555	---
	İmam Hatip Sec. School	598	11.65	2.99			
Writing	Secondary School	1024	16.64	8.05	1523	0.883	---
	İmam Hatip Sec. School	501	17.18	8.26			
Speaking	Secondary School	677	27.24	3.98	1019	0.276	---
	İmam Hatip Sec. School	344	27.15	3.83			

As can be seen from the t-test results in Table 9, the type of school has not a significant effect on the language skills of the students. In other words, the students who attend secondary school and imam hatip secondary school have a similar level of scores in reading, listening, writing, and speaking subtests. The effect sizes show that the effect of school type on students' language skills is negligible.

As seen from the t-test results, there is no significant difference between the reading subtest mean scores by school type ($t_{(1832)} = 1.058$, $p>0.05$). In the reading subtest of the students in imam hatip secondary

school, the mean score is calculated as 10.78, and the mean score of the students in secondary schools is 10.56. It is observed that the effect of school type on reading subtest scores is negligible.

It is observed that mean listening subtest scores given by school type are quite close to each other, and students in diverse secondary school types perform similarly in this subtest. There is no significant difference between the listening subtest scores of the students according to school type ($t_{(1836)} = 0.555$, $p>0.05$). The effect size analysis also showed that the school type does not have a significant effect on the listening subtest scores.

In the listening subtest, the mean score of the students who are in imam hatip secondary school is 17.18, and that of the students who are in secondary schools is 16.64. As can be seen from the t-test results, there is no significant difference between the mean scores of the students in both school types in writing subtest ($t_{(1523)} = 0.883$, $p>0.05$). As a result of the effect size analysis, it is shown that the school type does not have a significant effect on writing subtest scores.

The mean speaking subtest score of the students who are in imam hatip secondary school is calculated as 27.15. The mean subtest score of the students in secondary school is 27.24. According to the t-test test results, there is no significant difference between the mean scores of the speaking subtest by school type ($t_{(1019)} = 0.276$, $p>0.05$). The result of the effect size analysis shows that the school type does not have a significant effect on the speaking subtest scores.

The findings of the t-test to find the answer to the research question "Is there any significant difference in students' reading, listening, writing and speaking subtest mean scores according to the participation to pre-school education' are shown in Table 10 together with descriptive statistics.

Table 10. t-Test Results of Language Skill Subtest Scores According to Preschool Education Status

Subtest	Pre-School Education Status	n	$\underline{X}$	SD	df	*t*	η^2
Reading	Par. Pre-School Edu.	1419	10.87	3.58	1832	6.328*	.021
	Not Par. Pre-School Edu.	415	9.60	3.62			
Listening	Par. Pre-School Edu.	1422	11.91	2.92	1836	5.624*	.017
	Not Par. Pre-School Edu.	416	10.99	3.10			
Writing	Par. Pre-School Edu.	1171	17.38	8.24	1523	5.272*	.018
	Not Par. Pre-School Edu.	354	14.81	7.37			
Speaking	Par. Pre-School Edu.	832	27.22	3.94	1019	0.313	---
	Not Par. Pre-School Edu.	189	27.32	3.88			

*$p<0.05$.

The t-test results given in Table 10 show that participation in pre-school education leads to a significant difference in all subtest scores except speaking. Therefore, the reading, listening, and writing subtest scores of students who participate in preschool education are significantly higher. It is seen that participating preschool has its strongest effect on reading skill. As can be seen from the t-test results, there is a significant difference between the mean reading skills scores of students according to their preschool education status ($t_{(1832)} = 6.328$, $p<0.05$, $n^2= 0.021$). Students who participate in preschool education have a higher mean score in the reading subtest. Effect size analysis shows that preschool education has a significant effect on reading subtest scores, but this size of effect is small.

The mean listening score of students who did not participate in pre-school education is calculated as 10.99 in this subtest. The mean listening score of students receiving preschool education is 11.91. There is a significant difference between the mean listening subtest scores of the students according to their pre-school education status ($t_{(1836)} = 5.624$, $p<0.05$, $n^2= 0.017$). The mean listening subtest score of the students who participated in preschool education is higher. According to the results of the effect size, it was determined that the effect of participating in preschool education on listening subtest scores was low.

The mean writing subtest score of students who did not participate in preschool education is 14.81. The mean score of the students who participated in preschool education in the writing subtest is calculated as 17.38. The t-test results show that students who participated in preschool education have significantly higher scores in writing subtest than students who did not participate in preschool education ($t_{(1523)}$ = 5.272, p<0.05, n^2= 0.018). The effect size results show that preschool education has a low impact on students' writing subtest scores.

The mean score in the speaking subtest of students who did not participate in preschool education is calculated as 27.32. The mean score of the students who participated in preschool education is 27.22. The t-test results show that preschool education does not lead to a significant difference between the mean speaking scores ($t_{(1019)}$ = 0.313, p>0.05).

The findings of the t-test to answer the research question of 'is there any significant difference in students' reading, listening, writing and speaking subtest mean scores according to the gender groups?' are presented in Table 11 together with descriptive statistics.

Table 11. t-Test Results of Language Skill Subtest Scores by Gender

Sub Test	Gender	n	$\underline{X}$	SD	df	*t*	η^2
Reading	Female	971	10.92	3.57	1832	4.163*	.009
	Male	863	10.21	3.66			
Listening	Female	975	12.00	2.79	1836	4.512*	.011
	Male	863	11.37	3.17			
Writing	Female	812	18.32	8.12	1523	8.055*	.041
	Male	713	15.03	7.76			
Speaking	Female	561	27.96	3.79	1019	6.618*	.041
	Male	460	26.36	3.92			

**p<0.05*.

According to the results given in Table 11, the effect of gender on language skills leads to a significant difference in all subtests. It was determined that the mean scores of female students in all reading, listening, writing, and speaking subtests are significantly higher than male students. The effect size analysis shows that the difference between the mean scores of female and male students is even greater in writing and speaking subtests. According to the t-test results related to the reading subtest scores, there is a significant difference between the mean scores of male and female students ($t_{(1832)}$ = 4.163, p<0.05, n^2= 0.009). Female students' mean reading scores are higher than male students. According to the results of the effect size analysis, the effect of gender on the reading subtest scores is low.

In the listening subtest, the mean score of male students is calculated as 11.37, and the mean score of female students is 12. According to the t-test results related to the listening subtest scores, there is a significant difference between the mean scores of female and male students ($t_{(1836)}$ = 4.512, p<0.05, n^2= 0.011). Listening mean scores of female students are significantly higher than male students. In the effect size analysis, it is observed that the effect of gender on reading scores is low.

As can be seen from the t-test results, there is a significant difference between female students' mean writing score and male students' mean writing score ($t_{(1523)}$ = 8.055, p<0.05, n^2= 0.041). Female students' mean writing scores are higher than male students. It is determined that the effect of gender on writing scores is low.

The mean score of male students in the speaking subtest is calculated as 26.36 and female students as 27.96. The mean score of female students in the speaking subtest is significantly higher than the mean of the male students ($t_{(1019)}$ = 6.618, p<0.05, n^2= 0.041), but the effect of gender on the speaking subtest is found to be low.

The single-factor ANOVA findings to the research question "Is there any significant difference in students' reading, listening, writing, and speaking subtest mean scores according to the education levels of the mothers?' are shown in Table 12 together with descriptive statistics.

Table 12. ANOVA Results of Language Skill Scores According to Mother's Education Level

Subtest	Education Level	n	$\underline{X}$	SD	df	*F*	η^2
Reading	Primary Sch.	543	9.27	3.56	3	67.817*	.109
	Secondary Sch.	277	10.00	3.55			
	High School	461	11.88	3.21			
	Higher Edu.	386	12.86	3.41			
Listening	Primary Sch.	545	10.84	3.04	3	38.569*	.065
	Secondary Sch.	277	11.43	2.89			
	High School	461	11.88	2.70			
	Higher Edu.	388	12.87	2.91			
Writing	Primary Sch.	451	14.88	7.80	3	20.131*	.041
	Secondary Sch.	232	16.77	8.24			
	High School	382	16.47	7.58			
	Higher Edu.	324	19.37	8.36			
Speaking	Primary Sch.	275	26.03	4.14	3	17.957*	.056
	Secondary Sch.	144	26.92	3.80			
	High School	264	27.59	3.72			
	Higher Edu.	237	28.45	3.67			

*$p<0.05$.

As can be seen from the ANOVA results in Table 12, the education level of the mother leads to a significant difference in all subtest scores. In other words, students whose mothers graduated from higher education have significantly higher reading, listening, writing, and speaking scores. It is observed that the education level of the mother has its greatest impact on reading scores.

There is a significant difference between the mean reading scores of the students according to the education level of the mother ($F_{(3,1667)} = 67.817$, $p<0.05$, $n^2= 0.109$). As the education level of the mother increases, students' mean scores in the reading subtest increase. According to the results of the effect size analysis, the mother education level has a small effect on reading scores.

It is seen that the education level of the mothers leads to a significant difference in the listening scores ($F_{(3,1671)} = 38.569$, $p<0.05$, $n^2= 0.065$). It is determined that the mean score of the students whose mothers are graduates of higher education is 12.87 in the listening subtest, and the mean of the students whose mothers are primary school graduates is 10.84. As a result of the effect size analysis, it is determined that the effect of mother education level is low on the listening scores.

The mean of the writing subtest scores of the students whose mothers are higher education graduates is 19.37, and those whose mothers are graduated from primary school are calculated as 14.88. ANOVA results show that students whose mothers have higher education levels have significantly higher scores than other students' scores ($F_{(3,1389)} = 20.131$, $p<0.05$, $n^2= 0.041$). The effect of mother education level on students' writing subtest scores is examined, and it is showed that this effect is low.

In line with the ANOVA results, it was observed that the level of mother education leads to a significant difference between the students' speaking scores ($F_{(3,920)} = 17.957$, $p<0.05$, $n^2= 0.056$). As a result of the effect size analysis, it is determined that the effect of mother education level on students' speaking scores is low.

The findings of the single-factor ANOVA to find the answer to the research question 'is there any significant difference in students' reading, listening, writing, and speaking subtest mean scores according to the education levels of the fathers? are shown in Table 13, together with descriptive statistics.

Table 13. ANOVA Results of Language Skill Subtest Scores According to Father's Education Level

Subtest	Education Level	n	$\underline{X}$	SD	df	*F*	η^2
Reading	Primary Sch.	296	8.91	3.60	3	61.218*	.100
	Secondary Sch.	233	9.70	3.47			
	High School	531	10.38	3.44			
	Higher Edu.	611	11.94	3.35			
Listening	Primary Sch.	296	10.48	3.10	3	40.270*	.067
	Secondary Sch.	234	11.33	2.78			
	High School	532	11.54	2.90			
	Higher Edu.	613	12.61	2.98			
Writing	Primary Sch.	249	14.27	7.84	3	20.458*	.042
	Secondary Sch.	194	15.46	7.44			
	High School	431	16.48	7.87			
	Higher Edu.	514	18.69	8.20			
Speaking	Primary Sch.	139	26.25	4.12	3	11.019*	.035
	Secondary Sch.	126	26.43	3.86			
	High School	300	27.01	3.93			
	Higher Edu.	358	28.09	3.70			

**p<0.05*.

As seen in Table 13, the father's education level leads to a significant difference in all subtest scores. Therefore, students whose father graduated from a higher education level have higher reading, listening, writing, and speaking subtest scores. It is shown that the education level of the father has its strongest effect on reading scores.

According to ANOVA results, it is seen that the father's education level leads to a significant difference in mean reading scores of the students ($F_{(3.1671)} = 61.218$, $p<0.05$, $n^2= 0.100$). Effect size results showed that the father's education level has a small effect on students' reading scores.

It is shown that the father's education level also leads to a significant difference in the mean listening scores of students ($F_{(3.1675)} = 40.270$, $p<0.05$, $n^2= 0.067$). Effect size results showed that the father's education level has a small impact on students' listening scores.

According to the ANOVA results, there is a significant difference between the mean writing scores according to the father's education level ($F_{(3.1388)} = 20.458$, $p<0.05$, $n^2= 0.042$). The mean of the speaking scores of students whose father graduated from higher education is calculated as 18.69, and the mean of the speaking scores of the students whose father graduated from primary school is calculated as 14.27. The results of the effect size analysis showed that the level of father's education has a low impact on students' writing scores.

The mean speaking score of the students whose fathers are higher education graduates is calculated as 28.09, and the mean score of the students whose fathers are primary school graduates is calculated as 26.25. The results show that there is a significant difference between the students' mean speaking score according to the education level of the father ($F_{(3.923)} = 11.019$, $p<0.05$, $n^2= 0.035$). The results of the effect size analysis showed that the level of father's education has a low impact on students' speaking scores.

Findings for the Second Research Question

The relationship between the students' reading, listening, writing and speaking scores and their scores in Turkish, social sciences, mathematics and science courses was analyzed via Pearson correlation coefficient to answer the research question "*Is there a significant relationship between the students' language skills scores and the 7th grade scores in Turkish, social sciences, mathematics, and science?*" and the findings are shown in Table 14.

Table 14. The Relationship between Students' Reading, Listening, Writing, Speaking Scores and Their Scores in Turkish, Social Sciences, Mathematics and Science Courses in 7th Grade*

Subtest	Course Score	r
Reading	Turkish	0.66*
	Social Studies	0.64*
	Mathematics	0.61*
	Science	0.62*
Listening	Turkish	0.53*
	Social Studies	0.53*
	Mathematics	0.49*
	Science	0.50*
Writing	Turkish	0.38*
	Social Studies	0.35*
	Mathematics	0.35*
	Science	0.33*
Speaking	Turkish	0.29*
	Social Studies	0.26*
	Mathematics	0.28*
	Science	0.24*

* *p<0.05*

As can be seen in Table 14, there are significant relationships between all four subtests of Four Skills Turkish Language Test and scores of Turkish, social sciences, mathematics, and science courses in 7th grade. Correlation coefficients calculated at the subtest level are explained below.

According to the results of Pearson correlation analysis given in Table 14, there is positive, statistically significant relationships between the students' reading subtest scores and scores of Turkish ($r = 0.66$, $p < 0.05$), social sciences ($r = 0.64$, $p < 0.05$), mathematics ($r = 0.61$, $p < 0.05$) and science courses ($r = 0.62$, $p < 0.05$). These results show that the performances of the students in the reading subtest and their performances in all four courses are significantly related.

There is positive, statistically meaningful and medium level relationships between the students' listening scores and scores of Turkish ($r = 0.53$, $p < 0.05$), social sciences ($r = 0.53$, $p < 0.05$), mathematics ($r = 0.49$, $p < 0.05$) and science courses ($r = 0.50$, $p < 0.05$). These results show that the performances of the students in the listening subtest and their performances in all four courses are significantly related.

There is positive, significant and medium-level relationships between the students' writing scores and scores of Turkish ($r = 0.38$, $p < 0.05$), social sciences ($r = 0.35$, $p < 0.05$), mathematics ($r = 0.35$, $p < 0.05$) and science courses ($r = 0.33$, $p < 0.05$). According to these findings, the students' writing scores are significantly correlated with their performance in all four courses. The correlation between the subtest scores and the course scores is found to be higher in the writing subtest than in speaking subtest.

There is positive, significant and low-level relationships between the students' speaking scores and scores of Turkish ($r = 0.29$, $p < 0.05$), social sciences ($r = 0.26$, $p < 0.05$), mathematics ($r = 0.28$, $p < 0.05$) and science courses ($r = 0.24$, $p < 0.05$). These results show that the scores of the students in the speaking are significantly related to their performance in all four courses, but the level of relationship between them is low.

CONCLUSION AND DISCUSSION

Language skill is one of the basic skills that individuals must have in order to express themselves and be a part of the society. It has been shown in academic studies that the individual's competencies in the native language and many educational outcomes are related, especially academic achievement (Akbaşlı, Şahin & Yaykıran, 2016; Mahmud, 2014; Shali, 2017). Therefore, language skills have an important

role in the social and academic life of individuals. Because of this role, language skills are among the most important skills acquired through education.

The ways and methods in the development of language skills have also influenced the methods used to evaluate these skills. Approaches in which the four skills are assessed and scored separately in the assessment of native language and foreign language skills are the majority. Today, these basic skills are assessed separately in exams such as TEOFL, TOEIC, IELTS, PTE, which are often used for qualification.

In the Turkish language teaching program as the native language, there are educational outcomes to improve students' basic four language skills. However, there is no standard measurement method to assess the extent to which students have these basic skills, and no monitoring study is available on this subject. In interstage transition examinations such as LGS and YKS, and periodic monitoring studies as TMF-ÖBA and ABİDE focus only on reading skills. In this context, detailed data on students' listening, writing, and speaking skills are not available. In order to overcome this important deficiency, MoNE developed the Four Skills Turkish Language Test in 2019, and the pilot study is conducted under the coordination of the General Directorate of Measurement, Evaluation, and Examination Services.

The results indicate that students performed relatively high in the speaking subtest and relatively low in the writing subtest. It is observed that female students had higher mean scores than male students in all subtests. It can be stated that this result is consistent with the results of inter-stage examinations and monitoring studies in particular for the reading subtest (MoNE, 2018, MoNE2019b, ÖSYM, 2018). Findings are in coherence with the results of the monitoring study examined by Reilly, Neuman, and Andrews (2019). There are additional findings on internationally applied TOEFL and NAEP exams that female students are more successful, but the difference between gender groups is small (ETS, 2001; ETS, 2017). The fact that female students are performing better than male students in some language skills is seen in Dutch (van der Silk, van Hout & Schepens, 2015). Findings are also consistent with the PISA 2018 application that female students are performing better than male students in the reading field in the sample of Turkey (MoNE, 2019c). This finding shows that students from different gender groups may have diverse levels of linguistic skills.

It is determined that the mean scores of students attending secondary school and imam hatip secondary school did not show any significant difference in any subtest. In other words, the type of school the student attends does not have a significant effect on students' language skills. These findings are in coherence with that the graduates enrolled in imam hatip secondary school and other secondary schools according to the 2018 LGS central exam results (MoNE, 2018). Similarly, within the scope of 8th grade application of ABİDE 2016, it is determined that mean scores of imam hatip secondary school and secondary school students are quite close (MoNE, 2016). According to the results, students in two school types performed at the same level in their listening, writing, reading, and speaking skills.

It is determined that the students who participated in preschool education show higher performance than the students who did not participate in preschool education in all subtests except speaking. Considering the effect of preschool education on language development, these findings are seen to be consistent with the literature (Bakken, Brown & Downing, 2015; Schermse et al., 2018). It has been demonstrated in academic research that providing students with verbal skills through education in the preschool period positively contributes to language development and the psychological development of the individuals (Catts, Fey, Zhang, & Tomblin, 2001; Storch & Whitehurst, 2002). In speaking skills, why preschool education does not have a significant effect requires a more detailed investigation as a separate research subject.

Another finding is that the increase in parents' education level also increases the mean scores of the students in all subtests. The fact that parents from higher education levels use comparatively higher level of intellectual and complex language in home and read more with their children (Raikes et al, 2006; Rowe, Pan & Ayoub, 2005; Tamis-Lemonda & Rodriguez, 2009) is a possible reason for this significant difference between students. As parents' level of education is one of the components of students' social

background, and social background has a significant impact on students' academic achievement (Ozer & Perc, 2020; Schuetz, Ursprung & Woessmann, 2008), it is expected that students' language skills are positively correlated with parents' level of education. It is determined that the effect of mother and father education level on students' listening skills is higher than their gender and participation in preschool education. The findings are consistent with inter-stage examination results (MoNE, 2018) and academic studies abroad (Khodadady & Alaee, 2012; Richels, Johnson, Walden & Conture, 2013).

The study also examined the relationship between language skills and students' Turkish language, social sciences, mathematics, and science course scores. It has been determined that reading, listening, writing, and speaking skills have a significant relationship between the scores of four courses in levels ranging from low to medium (r=0.24 - r=0.66). This finding, which is important information about the validity of the study, also revealed an important implementation regarding the assessment of language skills in classrooms. The relationship between the reading scores and scores of the four courses is quite higher than the other language skills (between r = 0.61 - r = 0.66). One possible reason for this is that reading skills are used intensively in classroom assessments.

Findings obtained within the scope of the pilot study show that the test battery will make important contributions to the assessment of students' basic language skills. The findings support the validity of the pilot study, provide sufficient psychometric evidence, and the findings are supported by language development and assessment literature. The results of the future initial study will provide important feedback for native language teaching. Findings of the pilot study of the 'Four Skills Turkish Language Test' conducted by MoNE for the first time show that the assessment framework and data provide valid and reliable findings as a whole. Based on the data from the pilot study, it will be possible to develop certain levels of exams at the same standards with international qualifications in four skills of Turkish, both to strengthen native language education in schools and to make a more detailed analysis and to enhance educational processes.

REFERENCES

Akbaşlı, S., Şahin, M., & Yaykıran, Z. (2016). The effect of reading comprehension on the performance in science and mathematics. *Journal of Education and Practice, 7*(6), 108-121.

Akoğlu, H. (2018). User's guide to correlation coefficients. *Turk J Emerg Med,18*(3), 91–93.

Akram, A., & Malik, A. (2010). Integration of language learning skills in second language acquisition. *International Journal of Arts and Sciences, 3*(14), 231 – 240.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* USA: Oxford University Press.

Bakken, L., Brown, N., & Downing, B. (2015). Early childhood education: The long-term benefits. *Journal of Research in Childhood Education, 31*(2), 255-269.

Barın, M. (2000). Yabancı dil öğretiminde konuşma becerisinin önemi. *Atatürk Üniversitesi Sosyal Bilimler Dergisi*, 26, 123-127.

Bayyurt, Y. (2013). Current perspectives on sociolinguistics and English language education. *The Journal of Language Teaching and Learning, 1*, 69-78.

Boonkit, K. (2010). Enhancing the development of speaking skills for non-native speakers of English. *Procedia Social and Behavioral Sciences, 2*, 1305–1309.

Brown, H. D. (2001). *Teaching by principles: An interactive approach to language pedagogy* (2nd ed.). New York: Pearson Education.

Budria, S., Colino, A., & Martined de Ibarreta, C. (2019). The impact of host language proficiency on employment outcomes among immigrants in Spain. *Empirica, 46*, 625–652.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2016). *Bilimsel araştırma yöntemleri.* Ankara: Pegem Akademi Yayınları.

Canale, M., & Merrill, S. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Catts, H. W., Fey, M.E., Zhang, X., & Tomblin, B. (2001). Estimating the risk of future reading difficulties in kindergarten children: A research-based model and its clinical implementation. *Language, Speech, and Hearing Services in Schools, 32*, 38–50.

Chengyu, N. (2018). Implications of interrelationship among four language skills for high school English teaching. *Journal of Language Teaching and Research, 9*(2), 418-423.

Cook, V. J. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33*(2), 185-209.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Deniz, K. (2000). *Yazılı anlatım becerileri bakımından köy ve kent beşinci sınıf öğrencilerinin durumu*. (Yayınlanmamış yüksek lisans tezi). Çanakkale On Sekiz Mart Üniversitesi Sosyal Bilimler Enstitüsü, Çanakkale.

Doğan, Y. (2009). Konuşma becerisinin geliştirilmesine yönelik etkinlik önerileri. *Türk Eğitim Bilimleri Dergisi*, 7(1), 185-204.

Ergin, M. (1998). *Türk dil bilgisi.* İstanbul: Bayrak Yayınları.

Erkek, G., Batur, Z., Kaplan, K., & Ercan, E. (2017). Türkçe eğitimi ve yabancı dil öğretiminde dört temel dil becerisinin edinimine ilişkin öğretmen görüşleri. *Avrasya Dil Eğitimi ve Araştırmaları Dergisi, 1*(1), 42-75.

Educational Testing Service (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work*. Policy Information Report. Princeton: Education Testing Service.

Educational Testing Service (2017). *Test and score data: Score and data summary for TOEFL iBT Test*. Princeton: Education Testing Service.

Gautam, P. (2019). Integrated and segregated teaching of language skills: An exploration. *Journal of NELTA Gandaki (JoNG), 1,* 100-107.

Gazzola, M., & Mazzacani, D. (2019). Foreign language skills and employment status of European natives: Evidence from Germany, Italy and Spain. *Empirica, 46,* 713–740.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* London: Edward Arnold.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical.* Mahwah: Lawrence Erlbaum.

Hinkel, E. (2010). Integrating the four skills: Current and historical perspectives. In Robert B. Kaplan. (Ed.), *The Oxford handbook of applied linguistics* (2nd ed.) (pp. 110-125).

Jing, W.U. (2006). Integrating skills for teaching EFL: Activity design for the communicative classroom. *Sino-US English Teaching, 3*(12).

Kantemir, E. (1997). *Yazılı ve sözlü anlatım*. Ankara: Engin Yayınevi.

Karasar, N. (1999). *Bilimsel araştırma yöntemi.* Ankara: Nobel Yayınevi.

Kavcar C., Oğuzkan F., & Sever, S. (1995). *Türkçe öğretimi (Türkçe ve sınıf öğretmenleri için)*. Ankara: Engin Yayınevi.

Khodadady, E., & Alaee, F. F. (2012). Parent education and high school achievement in English as a foreign language. *Theory and Practice in Language Studies, 2*(9), 1811-1817.

Kuder, G. F., & Richardson, M. W. (1937). The theory of estimation of test reliability. *Psychometrika. 2,* 151-160.

Lüle Mert, E. (2013). Türkçe öğretmen adaylarının dört temel dil becerisine ilişkin algılarının metaforlar aracılığıyla analizi. *Uluslararası Sosyal Araştırmalar Dergisi, 6*(27), 357-372.

Lüle Mert, E. (2014). Türkçenin eğitimi ve öğretiminde dört temel dil becerisinin geliştirilmesi sürecinde kullanılabilecek etkinlik örnekleri. *Ana Dili Eğitimi Dergisi, 2*(1), 23-48.

Mahmud, M. M. (2014). Communication aptitude and academic success. *Procedia - Social and Behavioral Sciences, 134,* 125 – 133.

Manaj-Sadiku, L. (2015). The importance of four skills reading, speaking, writing, listening in a lesson hour. *European Journal of Language and Literature Studies, 1*(1), 29-31.

Ministry of National Education (2016). *Akademik becerilerin izlenmesi ve değerlendirilmesi: 8. sınıf raporu.* Ankara: MEB Yayınları.

Ministry of National Education (2018). *2018 Liselere geçiş sistemi: Merkezi sınavla yerleşen öğrencilerin performansı.* Eğitim Analiz ve Değerlendirme Raporları Serisi No:3. Ankara: MEB Yayınları.

Ministry of National Education (2019a). *Türkçe dersi öğretim programı*. Ankara: MEB Yayınları.

Ministry of National Education (2019b). *Türkçe-Matematik-Fen Bilimleri öğrenci başarı izleme araştırması (TMF-ÖBA)-I: 2019 4. sınıf seviyesi.* Eğitim Analiz ve Değerlendirme Raporları Serisi No:9. Ankara: MEB Yayınları.

Ministry of National Education (2019c). *PISA 2018 Türkiye ön raporu.* Eğitim Analiz ve Değerlendirme Raporları Serisi No:10. Ankara: MEB Yayınları.

Ministry of National Education (2020). *Dört beceride Türkçe dil sınavı: Pilot çalışma sonuçları.* Eğitim Analiz ve Değerlendirme Raporları Serisi No:11. Ankara: MEB Yayınları.

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers.* London: Sage Publishing.

Mitchell, C. B., & Vidal, K. E. (2001). Weighing the ways of the flow: Twentieth century language instruction. *Modern Language Journal, 85*, 26-38.

Nunan, D. (1989). Designing *tasks for the communicative classroom.* Cambridge: Cambridge University Press.

Ölçme, Seçme ve Yerleştirme Merkezi (2018). *2018 YKS değerlendirme raporu*. Değerlendirme Raporları Serisi No: 9. Ankara: ÖSYM Yayınları.

Ozer, M., & Perc, M. (2020). Dreams and realities of school tracking and vocational education. *Palgrave Communications, 6,* 34.

Ozer, M. (2020). What PISA tells us about performance of education systems?. *Bartın University Journal of Faculty of Education, 9*(2), 217-228.

Özkırımlı, A. (1994). *Dil ve anlatım*. Ankara: Ümit Yayınevi.

Perc, M., Ozer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications, 5,* 61.

Raikes, H., Pan, B. A., Luze, G., Tamis-LeMonda, C.S., Brooks-Gunn, J., Constantine, J., Tarullo, L. B., Raikes, H.A., & Rodriguez, E.T. (2006). Mother-child bookreading in low-income families: Correlates and outcomes during the first three years of life. *Child Development, 77*(4), 924-953.

Reilly, D., Neumann D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist, 74*(4), 445– 458.

Richels, C. G., Johnson, K. N., Walden, T. A., & Conture, E. G. (2013). Socioeconomic status, parental education, vocabulary and language skills of children who stutter. *J Commun Disord., 46*(4), 361-74.

Rowe, M. L., Pan, B. A., & Ayoub, C. (2005). Predictors of variation in maternal talk to children: A longitudinal study of low-income families. *Parenting: Science and Practice, 5*(3), 259-283.

Schmerse, D., Yvonne A., Flöter, M., Wieduwilt, N., Rossbach, H. G., & Tietze, W. (2018). Differential effects of home and preschool learning environments on early language development. *British Educational Research Journal, 44*(2), 338-357.

Schuetz, G., Ursprung, H. W., & Woessmann, L. (2008). Education policy and equality of opportunity. *Kyklos 61*(2), 279–308.

Shali, S. K. (2017). The power of listening ability and its effects on academic performance: An examination of college students. *Imperial Journal of Interdisciplinary Research, 3*(5), 1891-1896.

Stern, H. H. (1983). *Fundamental concepts of language teaching*. Oxford: Oxford University Press.

Storch, S.A, & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*, 934–947.

Tamis-LeMonda, C. S., & Rodriguez, E. T. (2009). *Parents' role in fostering young children's learning and language development.* Language Development and Literacy: Encyclopedia on Early Childhood Development. Retrieved from http://www.child-encyclopedia.com/sites/default/files/textes-experts/en/622/parents-role-in-fostering-young-childrens-learning-and-language-development.pdf.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education, 2*, 53-55.

Widdowson, H. G. (1978). *Teaching language as communication.* Oxford: Oxford University Press.

Xue-Ping, G. (1997). A scheme for the obtaining of language skills. *The Internet TESL Journal, 3*(6).

van der Slik, F. W. P., van Hout, Roeland, W. N. M., & Schepens, J. J. (2015). The gender gap in second language acquisition: Gender differences in the acquisition of Dutch among immigrants from 88 countries with 49 mother tongues. *PLoS One, 10*(11).