

The Usage of Meaning Identification Technique in Measuring Reading Comprehension Skills*

Received: 12 December 2018

Revised: 5 March 2019

Accepted: 6 March 2019

ISSN: 1307-9298

Copyright © IEJEE

www.iejee.com

Bariş Esmer^{a,**}, Ahmet Melih Güneş^b

DOI: 10.26822/iejee.2019450836

Abstract

In this study, it is aimed to determine at what level the Meaning Identification Technique (MIT) can measure the reading comprehension skills of the 4th grade primary school students and to test the validity and reliability of the MIT test with different types of texts. In the study carried out in the survey model, MIT and Sentence Verification Technique (SVT) tests, which were developed by the researcher and are peculiar to narrative and informative texts, were used. The readability of the texts was tested with Cloze Tests (CT) developed specifically for each type of text. The study was carried out with the participation of 87 primary school 4th grade students in two primary schools randomly selected among the public schools in the city of Ankara. The results of the study demonstrated that MIT tests measure reading comprehension skills as reliable as SVT tests. While there was a significant difference between comprehension and the numbers of reading in narrative texts, it was found that there was no significant difference with the numbers of reading in informative texts. A great majority of students (approximately 75%) felt the need to read two or more times while making sense of the text. On the other hand, MIT tests are a significant predictor of Turkish course achievements.

Keywords: Reading Comprehension, Sentence Verification Technique, Meaning Identification Technique, Cloze Test

Introduction

The assessment of reading comprehension skills has been one of the important milestones in the field of education for over 100 years. According to Marcotte, Rick and Wells (2018), reading comprehension skills play an important role in course achievements. It is particularly of importance for researchers, teachers, and material developers to have correct and purposeful measurements for reading comprehension skills.

Several assessment tools were developed for the identification and handling of reading comprehension skills (reading comprehension tests, cloze tests, sentence verification tests etc.). At the international level, literacy skills are determined by standardized tests through the Programme for International Student Assessment (PISA). These exams are composed of different types of questions and the majority of these questions are created with multiple choice questions. It is rather difficult for individuals who want to determine their reading comprehension skills to access, prepare, and evaluate PISA questions in this way (OECD, 2016). Despite the development of many assessment tools, it is seen that there are no reading evaluation tests standardized according to class level and available to be used at the national level in Turkey (Ulusoy & Çetinkaya, 2012).

Research and Development (RAND) Reading Study Group (2002) is designed for a good reading comprehension test. This study group was created for different text types to provide descriptive information on readers. It is also structured for different age ranges, to measure more than one of the components of comprehension altogether based on a theoretical structure and to be sensitive to social, cultural, and linguistic differences (Snow, 2002). Caccamise et al. (2008) emphasize that formative assessment in the classroom environment provides better results in determining students' competencies in a field. In this view, improving reading com-

prehension tests may change depending on many factors. the purpose of this study was to focus on the adaption of Meaning Identification Technique (MIT) tests with a particular attention to the Turkish language features rather than the standardization of the text.

The reading comprehension tests that are called traditional show only how the text can be better understood. These tests do not address how the cognitive and meta-cognitive processes are used, or where the challenges are faced in the process of interpretation (Klingner, 2004). On the contrary, the Sentence Verification Technique (SVT), which is developed by using texts, can identify both reading and listening comprehension skills. SVT is created for different purposes and to demonstrate what kind of mistakes have been made in the process of understanding. SVT was influenced by the idea that understanding is a structural process. SVT is based on the theory that the meaning is an image that is formed in our minds as a result of the interaction of the reader's prior knowledge with the information received from the text (Kintsch, 1994; Kintsch & Kintsch, 2005; Kintsch & Van Dijk, 1978; Royer & Cunningham, 1981; Royer et al., 1979). The researchers emphasize that the processing memory, the discernment of the text's structure, and the prior knowledge and scheme of the reader are influential in the process of reading and listening comprehension. This theory asserts that reading or listening comprehension can be measured by mental representations formed in the minds after reading or listening texts.

In the preparation of the SVT test (Royer, 2001; Royer et al. 1987), the texts are rearranged in a way that each consists of 12 sentences. Each sentence in the text is then rewritten in accordance with four types of sentences: "original, paraphrase, meaning change, and distractors." The original sentences are written exactly the same way written as in the text. On the other hand, in paraphrases, at least a few words should be changed from the sentence in the text, but the

* This study was presented at the 18th International Classroom Teaching Education Symposium as an oral presentation.

^{a**}Corresponding Author: Barış Esmer, Gazi University, Gazi Faculty of Education, Department of Primary Education, Turkey. E-mail: barisesmer@gazi.edu.tr

^bAhmet Melih Güneş, Balıkesir University, Necatibey Faculty of Education, Department of Primary Education, Turkey. E-mail: melih.gunes@balikesir.edu.tr

meaning should remain the same. In sentences that their meanings were changed, the meaning should be changed by replacing only one or two words from the sentence in the text. The new sentence and the sentence structure in the text should be rather similar to each other. As for the distractors, by sticking to the syntactic features and the main theme of the text, sentences are formed with a completely different meaning. By writing 4 sentences from each sentence type, a new text with 16 sentences is obtained.

The literature illustrates that the SVT tests is a useful technique for measuring reading comprehension in English language text. SVT tests are often used in measuring reading comprehension skills (Durwin & Sherman, 2008; Lastrapes, 2015; Marcotte & Hintze, 2009; McKeown et al., 2009; Pichette et al., 2014), in measuring listening comprehension skills (Greene et al., 1990; Kertoy & Goetz, 1995; Lehto & Anttila, 2003) in defining reading problems (Carlisle, 1989; Carlisle, 1999; Royer & Sinatra, 1994), as a descriptive study in determining students' reading levels (İlter, 2017; Royer et al., 1990), in measuring the language skills of students of English as a second language (Mestre & Royer, 1991; Royer & Carlo, 1991; Royer et al., 1991), in determining the levels of texts (Jones & Smith, 2014; Yazıcı & Kurudayıoğlu, 2017) and in predicting learning skills (Mooney et al., 2016; Mooney & Lastrapes, 2018; Royer et al. 1987; Royer et al., 1990).

The MIT is applied differently from the SVT. The purpose of the MIT is to eliminate the features that can reduce test reliability while evaluating the meaning of the text. The MIT tests can be applied to both narrative and informative texts. The texts selected for the MIT should have only following sections: problem, event, and conclusion. The MIT is based on the idea that the meaning is structurally constructed. Tests constructed for this technique consist of two types of questions which are "meaning change" and "paraphrase types. (Marchant et al., 1988). The MIT consists of 12 sentences which is written as 6 items from each of the two sentence types. Practitioners who respond to this test are expected to answer the questions by considering and evaluating their meaning in the text. Indeed et al. (1995) itemize the characteristics of good readers as interacting with the text through their experience to understand the text, text interpretation, synthesis, and evaluation. In this context, Yıldırım et al. (2010) alleged that the narrative texts are easier to understand with compared to informative texts. The fact that the informative texts are more difficult to understand, the following factors have effect on texts: the lack of sufficient knowledge on the text content and structure (Armbruster et al. 1987; Diakidoy et al., 2005; Kamberelis & Bovino, 1999), less interaction with the texts (Singer et al., 1997), and the need for prior knowledge and vocabulary (Afflerbach, 1986; Best et al., 2008; McNamara et al., 2011).

The studies on the SVT and the MIT reveal that there are a limited number of studies on the MIT tests (Marchant et al., 1988; Ushiro et al., 2013). although there are many studies claiming that the SVT tests can be used as valid and reliable tools (Lynch, 1982; Marchant et al., 1988; Marcotte et al., 2018; Royer et al., 1979; Royer et al., 1984), These studies, on the other hand, assert that the MIT can be used as a reliable tool for measuring both superficial and inferential comprehension (KR-20=.77). However, it is safe to say that previous studies on this subject did not consider and implement the MIT test in reading comprehension in Turkish language text. Hence, further investigations are required to determine students' level of reading comprehension in Turkish texts.

Reading comprehension tests are accepted as important indicators of both reading comprehension and learning performance. These tests are used as important tools in determining student achievement (Royer et al., 1990). Royer et al. (1990) found that SVT tests explained 36% of the variance affecting school success whereas they explained 30% of the

course achievement. The purpose of their study is to determine to what extent the Meaning Identification Technique (MIT) can measure the comprehension skills of the 4th grade primary school students in reading narrative and informative texts as well as testing the validity and reliability of the MIT test in different text types. In addition to that, the readability of the texts is determined by Cloze Tests. study aims to revive interest in measuring reading comprehension in Turkish by implementing the MIT tests.

The research questions of the study are as follows:

1. What are the students' reading scores from the MIT tests?
2. Whether and to what extent the students' reading scores from the MIT tests associated their genders?
3. Whether and to what extent the students' reading scores from the MIT tests are associated with their numbers of reading the text?
4. Whether and to what extent the SVT tests and the MIT tests are associated with Turkish course achievements?
5. To what extent does the students' SVT and MIT scores affect their Turkish course achievements?

Method

Survey model is the main method adopted by this study. a survey was constructed to see to what extent the MIT is able to measure students' reading comprehension skills. It also analyses the validity and reliability of the MIT test in different text types. The study followed the process figured below.

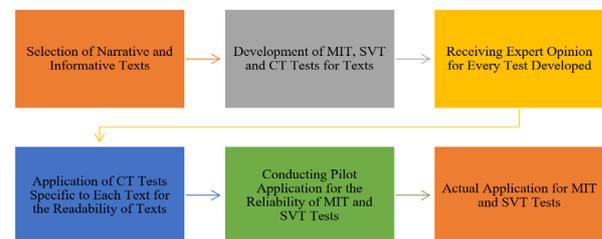


Figure 1. Data collection procedure

Subjects for the Study

The unit of analysis of the study was fourth graders. The participants of the study were

87 fourth grade students from 3 different classes of the public schools in Ankara, the capital city of Turkey. The data was collected between February 26th and February 28th in 2018. Participants have similar socioeconomic backgrounds and their ages vary between 9 and 10 years with a 9.86 mean age. Students were identified with convenience sampling method. This type of sample is conducted with people who are willing to participate in the study (Johnson & Christensen, 2004) and who can easily participate in the study (Patton, 2002). The consent form—which explained that the participation of respondents is on a voluntary base—was placed at the beginning of the survey. Consent form to participate in this study was obtained from school administration and class teachers ensured the students' participation. Detailed information on the purpose of the research was given to administration, class teachers, and student to assure the confidentiality and that no personal identifiable information was collected. The descriptive analysis of the study depicts that 57.5% of the participants was female while 42.5% of them was male.

Data Collection Tools

Three copyrighted tests, consisting of 53 items were used to construct a survey questionnaire. These scales are SVT, MIT, and CT tests.

SVT and MIT Tests. SVT and MIT tests, which are peculiar to narrative and informative texts, were adopted as the data collection tool. In the determination of the narrative and informative texts, primary school 4th grade Turkish textbooks published by the Ministry of National Education (MNE) were examined and the text of "Fairy Tales to Eat" was selected as a narrative (N) and the text of "A Tiny Box: The Seed" was selected as an informative text (I) from the textbook. The following procedure, which was determined by Royer et al. (1987), was used to form SVT-MIT tests:

1. The narrative and informative texts were rearranged in a way that each consists of 12 sentences.
2. For the SVT tests, a total of 16 sentences were written, composed of 4 sentences from each of the original, meaning change, paraphrase, and distractor types.
3. For the MIT tests, a total of 12 sentences were written, composed of 6 sentences from each of meaning change, and paraphrase types.
4. 'Yes' and 'No' options were added to each sentence, and students were asked to compare the real meaning of what they read and the meanings of the sentences in the SVT and MIT tests.
5. The answer 'Yes' means that the meanings in the sentence and in the text are the same while the answer 'No' means that the meaning in the sentence and meaning in the text are different from each other.
6. The texts used in the study and the SVT and MIT were prepared in a way to be on separate pages.
7. The sentences in the first half of the text were the first questions in the SVT and MIT tests. In this way, it was aimed to prevent students from answering test questions with short-term memory.
8. In the interpretation of the scores obtained from the SVT and MIT tests, 70% and below scores were evaluated as frustrating comprehension level, the scores between 71% and 79% were considered as the instructional level of comprehension, and the scores 80% and above were assessed as independent level Figure 2 below depicts some examples of developed test items.

Everyday, one or more fairy tales were disappearing, and families were having difficulty in finding fairy tales to tell their children (sentence in the text) * p. 86

•Everyday, one or more fairy tales were being added, and families were having difficulty in choosing fairy tales to tell their children. (Meaning Change)

If you split an apple in half right in the middle, you can easily see the pips that are the apple tree seeds (Sentence in the Text) * p. 119

•It is enough to split the apple in half in order to reach the pips of an apple (Paraphrase)

* Bozbey et al. (2016). Primary School Turkish Textbook and Workbook. Ankara: Özgün.

Figure 2. Examples to the MIT test items

Cloze Tests. Narrative texts and informative texts consisting of 136 and 188 words were used in this study. In order to determine the readability of the texts that are used in the SVT and MIT tests, Cloze Tests (CT) were prepared,

the sixth word of which was deleted. CT was developed by Taylor (1953) to determine readability. This test is also used as a valid and reliable tool in determining the readability of the Turkish texts (Keskin & Akilli, 2013; Tunçer & Erden, 2015; Ulusoy, 2009). The first and the last sentences of the texts are provided as a hint, when the test was being prepared. From the second sentence, every sixth word has been deleted. Four answer options were added to each gap, three of which were distractors while one was the correct answer. 33 students were employed to use CT tool. Students were asked to mark the correct answer and select one of the appropriate options for the deleted word. The application was completed in a 40-minute. Table 1 provided the readability scores of the texts.

Table 1. Readability values of texts

Type of Text	Percentage of Readability	Assessment Criteria
Fairy Tales to Eat (H)	59.37%	60% and above Independent reading level - Easy text 40% - 59% Instructional Level
A Tiny Box: The Seed (B)	47.26%	Below 40% Frustration-level- Difficult Text (Rankin & Culhane, 1969).

N: Narrative text, I: Informative Text

Table 1 show that both narrative and informative texts are at instructional level. After determining that the texts were appropriate for the students' levels, the opinions of two subject-matter expert faculty members were consulted for the scope validity of the SVT and MIT. For the reliability of the tests, a pilot study was conducted with 26 students studying in a public school in Ankara. Table 2 depicts the KR-20 values to determine the reliability of the tests.

Table 2. Results of the texts' reliability analysis

Type of Text	KR-20 Values		
	SVT	MIT	Assessment Criteria
Fairy Tales to Eat	.74	.72	> .70 (Fraenkel & Wallen, 2006).
A Tiny Box: The Seed	.73	.78	

Table 2 reveals that both SVT and MIT tests can give reliable results and they can be used in the actual application.

Turkish Course Success. In the determination of the students' success in the Turkish language course, Turkish course exams conducted by teachers were taken into consideration. Turkish course achievement scores were determined by calculating the average of students' scores for two exams.

Process

Prior to the actual application, students were informed about the SVT and MIT tests and were asked whether they are familiar with the test. Firstly, the papers with 12-sentence narrative text (Fairy Tales to Eat) were distributed to the students and they were allowed to read until they understand. No time limitation was applied for students so they can focus on understanding the text. Students noted number of readings on upper right corner of the paper. After that, the papers of the MIT and SVT tests were distributed to the students and they were asked to answer the test. The same procedure was repeated for the informative text as well (A Tiny Box: The Seed). The students were asked to answer four different tests in a total of 60 minutes.

Data Analysis

The data obtained as a result of the research was firstly transferred to the computer environment and was examined in terms of missing or incorrect value, outlier, and multivariate. 4 of the 91 observations in the data set were excluded because they were outliers. Thus, data obtained from 87 students were included in the study. Kolmogorov-Smirnov and Shapiro-Wilk Tests were conducted to see whether the data has normal distribution or not. The test results provided us that the data had normal distribution ($p \geq .05$, see Table 3). Students' levels of reading were determined by calculating the percentage, frequency, and means of the scores obtained from the SVT and MIT tests. In the tests, it was concluded that the data was distributed normally. Independent t-test was implemented to determine whether and to what extend the narrative and informative reading texts and the reading scores of the male and female students are associated. One-Way Analysis of Variance was conducted to determine the difference by the reading numbers. For assumptions (homogeneity of variance) of underlying one-way Anova test was used Informative MIT ($F_{(2,84)} = .179$; $p = .836$). Narrative MIT ($F_{(2,84)} = 5.09$; $p = .008$, for normality you also see Table 3) correlation analysis was utilized to test the relationships and the regression analysis was used for the prediction of Turkish course comprehension success.

Table 3. Normality Tests of Variables

Variables	Skewness	Kurtosis	Kolmogorov-Smirnov	Shapiro-Wilks	
N-MIT	Male	-.769	-.141	.001	***
	Female	-.831	.855	.000	***
N-NR	One	-.408	-.764	***	.63
	Two	-.608	-.107	***	.11
I-MIT	Three and more	-.293	-.445	***	.38
	Male	-.673	.013	.007	***
I-NR	Female	-.304	-.720	.009	***
	One	-.593	-.529	***	.024
	Two	-.341	-.508	***	.087
	Three and more	-.112	-1.342	***	.23

N-MIT: Narrative meaning identification test, N-NR: Narrative text number of reading, I-MIT: Informational Meaning Identification Test I-NR: Informative text number of reading, Acceptable Range: +1 <kurtosis and skewness <-1, Tabachnick, Fidell & Osterlind (2013)

Findings

The research findings were provided in order by taking into account of the research questions.

What are the students' reading scores from the MIT tests?

Table 4 and Table 5 indicate the values obtained by analyzing students' reading scores from the MIT tests.

Table 4. Percentages, means, and standard deviations regarding students' reading scores

Texts	Reading Scores			
	n	%	M	SD
Fairy Tales to Eat	87	73.46	8.81	14.62
A Tiny Box: The Seed	87	62.16	7.45	15.18

The results provided us that the students got higher scores from the narrative texts. In the light of the results, the students could be grouped as the ones who have weak (62%) comprehension skills from informative texts and those who have instructional comprehension skills (73%) from narrative texts, in accordance with Royer's (2001) SVT test evaluation criteria.

Table 5. Distribution of students' MIT reading scores by good, average, and poor understanding levels

Texts	Reading Scores							
	Good		Average		Poor		Total	
	f	%	f	%	f	%	f	%
Fairy Tales to Eat (H)	35	40.2	22	25.3	30	34.5	87	100
A Tiny Box: The Seed (B)	10	11.5	21	24.1	56	64.4	87	100

Table 5 reveals that the highest scores on the narrative text type are at the level of good understanding ($f = 35$) and the weak understanding seems to be the highest on the informative text type ($f = 56$).

Whether and to what extend the students' reading scores from the MIT tests associated with their genders?

In order to see whether and to what extend the students' reading scores from the MIT tests associated their genders, an independent t-test were conducted. Table 6 below depicts the results.

Table 6. Independent t-test results regarding MIT tests of male and female students

Text Type	Female			Male			df	t	p
	n	M	SD	n	M	SD			
Fairy Tales to Eat (H)	50	75.83	14.21	37	70.26	14.76	85	1,28	.202
A Tiny Box: The Seed (B)	50	62.16	16.16	37	62.15	13.97	85	.001	.999

It is seen that the MIT scores of the students are not significantly associated with their genders. The results provided that the average scores of male and female students are very close to each other.

Whether and to what extend the students' reading scores from the MIT tests are associated with their numbers of reading the text?

Levene F test results depict that variances are homogeneous ($F_{(2,84)} = 5.09$; $p = .008$). For this reason, LSD from post hoc tests were conducted to evaluate differences among the means. A one-way analysis of variance is implemented to check whether there is a significant difference between the students' numbers of reading and reading comprehension skills. Tables 7, 8, 9, and 10 reveal the results.

Table 7. Descriptive statistics of narrative text MIT test scores

Number of Reading	n	M	SD
One	24	66.66	18.22
Two	42	75.99	13.04
Three and more	21	76.19	10.63

Table 8. One-way Anova results for reading scores of narrative MIT test by the number of reading

Source of variance	Sum of squares	df	Mean square	F	p	Difference
Between groups	1533.12	2	766.56	3.82	.026	2-1, 3-1
Within group	16873.42	84	200.87			
Total	18406.53	86				

* $p < .05$

Table 8 indicates that there is a significant difference between the number of reading and reading comprehension of narrative texts ($F_{(2,84)} = 3.82$; $p < .05$). It is seen that this difference occurred in the LSD multiple comparison test in favor of the students who read the text three times ($M = 76.19$) and twice ($M = 75.99$). The percentage of the students who

stated that they read the narrative text at least twice to make sense of the text is 72%.

Table 9. Descriptive statistics of informative text MIT test scores

Number of Reading	n	M	SD
One	27	58.64	16.58
Two	37	65.08	14.68
Three and more	23	61.59	13.93

Table 10. One-way Anova results for the reading scores of informative MIT test with respect to the number of reading

Source of variance	Sum of squares	df	Mean square	F	p
Between groups	659.04	2	329.52	1.44	.242
Within group	19175.1	84	228.27		
Total	19834.15	86			

Levene F test scores show that variances are not homogeneous ($F_{(2,84)} = 5.09; p = .836$). Table 10 shows that there is no difference in reading comprehension scores obtained from the MIT test by the numbers of reading of the text ($F_{(2,84)} = 1.44$). In other words, further reading of the text does not indicate that better understanding will occur. The results illustrate that the students who read the text twice reached higher means ($M = 65.08$) than the other students. The percentage of the students who stated that they should read the text at least twice is 69%, while the percentage of the students who stated that they should read at least three times in order to make sense of it is 26%.

Whether and to what extent the SVT tests and MIT tests are associated with Turkish course achievements?

Assumption of underlying the significance tested and variables are normally (bivariately) distributed (see Table 3). Table 11 provided the means and standard deviation values of the study variables and the correlation coefficients between these variables.

Table 11. Correlation analysis results to determine the relationship between variables

Variables	M	SD	1	2	3	4	5
1. Turkish Course Achievement	86.25	9.75	1.00	.41**	.26*	.36**	.43**
2. Narrative MIT	73.46	14.63		1.00	.13	.51**	.35**
3. Informative MIT	62.16	15.19			1.00	.13	.54**
4. Narrative SVT	81.32	13.36				1.00	.21
5. Informative SVT	68.90	14.54					1.00

** $p < .01$; * $p < .05$

The figures provided by Table 11 reveals that the students became more successful with the tests prepared with narrative texts. The students obtained a higher mean score from the SVT tests ($M_{NSVT} = 81.32, M_{ISVT} = 68.9$). The correlation coefficients values between the variables indicated that there is a moderate and positive correlation between the students' Turkish course achievements and I-SVT ($r = .43, p < .01$) and N-MIT ($r = .41, p < .01$), whereas a positive and low-level correlation was found with N-SVT ($r = .36, p < .01$) and I-MIT ($r = .26, p < .05$). Accordingly, the more students' SVT and MIT scores increase, the more Turkish course achievements increase. In addition, there are positive, average-level correlations between SVT and MIT tests prepared with informative texts ($r = .54, p < .01$), as well as between SVT and MIT ($r = .51, p < .01$) tests prepared with narrative texts. In other words, as the students' scores

from the SVT tests prepared with both narrative and informative texts increase, their MIT scores also increase.

To what extent does the students' SVT and MIT scores affect their Turkish course achievements?

Table 12 and Table 13 postulate the results of multiple linear regression analysis for predicting Turkish course achievements. Assumptions of underlying significance test of multiple linear regression tested (MIT-VIF= 1.018, SVT-VIF= 1.047, MIT-Tolerance= .983, SVT Tolerance= .955, Condition Index (CI)= < 30) and Tolerance values approaching to zero, VIF greater than 10 and CI above 30 was not found.

Table 12. Multiple linear regression results of MIT scores' prediction of Turkish course achievements

Variable	B	SE	β	t	p
Invariant	59.05	5.83		10.13	.000
Narrative MIT	.26	.065	.39	3.95	.000
Informative MIT	.13	.063	.21	2.13	.035

$R = .46; R^2 = .21; \text{Adjusted } R^2 = .195; F_{(2,84)} = 11.4; p < .05$
VIF:1.018, Tolerance: .983, CI<30

The multiple regression analysis reveals that there is a positive, moderate, and significant relationship between the MIT scores and Turkish course achievements ($R = .46, R^2 = .21, \text{Adjusted } R^2 = .195, p < .05$). The narrative MIT ($\beta = .26, p < .05$) test and the informative MIT ($\beta = .13, p < .05$) test are positive significant predictors of the Turkish course achievements. Moreover, the findings postulated that approximately 21% of total variance of Turkish course achievements are explained by the MIT scores.

Table 13. Multiple linear regression results of SVT scores' prediction of Turkish course achievements

Variable	B	SE	β	t	p
Invariant	52.26	6.72		7.78	.000
Narrative SVT	.21	.072	.28	2.85	.005
Informative SVT	.25	.067	.37	3.75	.000

$R = .51; R^2 = .26; \text{Adjusted } R^2 = .241; F_{(2,80)} = 14.02; p < .05$
VIF:1.047, Tolerance: .955, CI<30

The results of the multiple regression analysis reveal that there is a positive, moderate, and significant relationship between SVT scores and Turkish course achievements ($R = .51, R^2 = .26, \text{Adjusted } R^2 = .241, p < .05$). Hence, it is safe to say that informative SVT ($\beta = .25, p < .05$) tests and narrative SVT ($\beta = .21, p < .05$) tests are positive significant predictors of Turkish course achievements. Moreover, the findings of the study indicate that SVT scores explain approximately 26% of the total variance of Turkish course achievements.

Conclusion and Discussion

The results of the study provided us that the MIT test is a reliable tool ($KR-20 = .72; .78$) to measure reading comprehension skills. The results of the cloze tests (959.37 - %47.26) revealed that both types of texts were instructional (not above or below the children's level). In the study, it was found that poor comprehension (62%) occurred with informative texts and average-level comprehension (73%) was reached with narrative texts. These results of the study are consistent with the studies revealing that narrative texts are more easily understood than informative texts (Best et al., 2008; Rasool & Royer, 2001; Şahin, 2013; Yıldırım et al., 2010).

The findings of the study show that comprehension skills in reading narrative and informative texts did not differ

significantly by their gender. The reading tests results reveal that the students' scores were very close to each other. On the other hand, some studies (Çiftçi & Temizyürek, 2008; Rasool & Royer, 2001; Şahin, 2013; Yıldırım et al., 2010) reveal significant differences in favor of female students. Considering that these two text types are also at instructional level, this situation can be explained by the facts that all students use similar strategies in the process of understanding (see Pressley & Afflerbach, 1995) and that they were allowed to read until they understand during the study process. These studies obtained similar findings. For instance, Şahin (2013) studied fourth and fifth graders and found no significance difference between the informative texts reading comprehension marks of fifth grade male and female students. The study of Çiftçi & Temizyürek (2008) reveal that there isn't a significant difference between reading comprehension levels with regards to male and female students. Students performed similar understandings in 56% of total learning outcome.

When comprehension skills were compared in relation to the numbers of reading the text, a significant difference was only found between comprehension of narrative texts and the numbers of reading. The findings of the study indicate that students who read the text more than once received higher scores than the students who read only once. On the other hand, 72% of the students mentioned that they read the narrative text at least twice to make sense of it. The results show that 69% of the students read the text at least two times and 26% of them read at least three times in order to make sense of it. According to Royer et al. (1984), the SVT is sensitive to the degree and level of difficulty of the text read when determining reading comprehension. In the study of Best et al. (2008), they concluded that word recognition is important in making sense of narrative texts whereas vocabulary and prior knowledge on the subject are significant in informative texts. . The data at hand provides us the fact that students felt the need to read the text more than once, although the texts were at instructional level could be explained with the factors such as being unfamiliar with MIT tests, the lack of prior knowledge about the subject and also the lack of detailed information about the text structures.

The findings of the study revealed that there was a positive significant relationship between the students' Turkish course achievements and the MIT and SVT tests. In addition, the results postulated that there are positive, moderate-level ($r=.51$, $p<.01$) and significant relationships between MIT and SVT ($r=.54$, $p<.01$), and between MIT and SVT tests. In other words, students' SVT test scores increase as their scores obtained from MIT tests prepared with both narrative and informative texts increase. On the other hand, both MIT and SVT are significant predictors of Turkish course achievements. The MIT tests accounted for 21% of the variance affecting Turkish course achievements while SVT tests explained 26% of them. These results show similarities with many studies in the literature (Mooney et al., 2016; Marchant et al., 1988; Royer et al., 1990). The study reached the conclusion that the variance of the MIT tests that affected the course achievements was in varying rates (7% - 30%).

Limitations and Future Directions

This study was designed to test whether the MIT test is a valid and reliable tool to measure comprehension skills in reading different text types. However, as in all studies, there are some limitations also in this study that need to be addressed in future researches. First of all, the sample of this study consisted of 4th grade students studying at a primary school of a neighborhood considered to be at middle-level in socio-economic terms. Therefore, the results of the current research may not reflect the general condition of the 4th grade students. Future researches should be repeated in a way to include a larger sample, covering different so-

cio-economic levels. In this way, the reorganization of MIT test's evaluation and assessment criteria peculiar to students in Turkey might be ensured.

Secondly, assessments were made over the given correct answers in the developed MIT and SVT test scores. Kintsch & Kintsch (2005) emphasize that comprehension requires inference and inference requires knowledge. Researchers indicate that reading comprehension skills will become more understandable by determining how the information is used in the reading process. In this context, by reviewing the answers given to MIT and SVT test items, the mistakes made by the students in the process of understanding can be revealed. In other words, it can be determined on which sentence type (original, meaning change, paraphrase, distractor) the students were more successful/unsuccessful. In conclusion, it is easy to prepare, apply and evaluate the MIT tests in different text types; it is thought that the MIT test can high availability be used in our schools. There is a need for the future researches to check its usability on other courses at primary schools and for different education levels as well as to measure students' listening comprehension skills.

References

- Afflerbach, P. (1986). The influence of prior knowledge on expert readers' importance assignment process. In J. A. Niles & R. V. Lalik (Eds.), *National reading conference yearbook, 35. Solving problems in literacy: Learners, teachers and researchers* (pp.30-40). Rochester, NY: National Reading Conference.
- Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text?. *Reading Research Quarterly, 22*(3), 331-346.
- Best, R. M., Floyd, R. G., & McNamara, D. S. (2008). Differential competencies contributing to children's comprehension of narrative and expository texts. *Reading Psychology, 29*(2), 137-164.
- Carlisle, J. F. (1989). Diagnosing comprehension deficits through listening and reading. *Annals of Dyslexia, 39*(1), 159-176.
- Carlisle, J. F. (1999). Free recall as a test of reading comprehension for students with learning disabilities. *Learning Disability Quarterly, 22*(1), 11-22.
- Çiftçi, Ö., & Temizyürek, F. (2008). İlköğretim 5. sınıf öğrencilerinin okuduğunu anlama becerilerinin ölçülmesi. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 5*(9), 109-129.
- Diakidoy, I. A. N., Stylianou, P., Karefillidou, C., & Papageorgiou, P. (2005). The relationship between listening and reading comprehension of different types of text at increasing grade levels. *Reading Psychology, 26*(1), 55-80.
- Durwin, C. C., & Sherman, W. M. (2008). Does choice of college textbook make a difference in students' comprehension?. *College teaching, 56*(1), 28-34.
- Greene, B. A., Royer, J. M., & Anzalone, S. (1990). A new technique for measuring listening and reading literacy in developing countries. *International Review of Education, 36*(1), 57-68.
- İlter, İ. (2017). Farklı okuma yeteneklerindeki öğrencilere kelime anlamlarının öğretimi: bağlam temelli öğretimin okuduğunu anlama üzerinde kontrollü bir değerlendirilmesi. *Eğitim ve Bilim, 42*(190), 437-463.

- Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches* (2nd edition). Pearson.
- Jones, M., & Smith, M. (2014). Traditional and alternative methods of measuring the understandability of accounting narratives. *Accounting, Auditing & Accountability Journal*, 27(1), 183-208.
- Kamberelis, G., & Bovino, T. D. (1999). Cultural artifacts as scaffolds for genre development. *Reading Research Quarterly*, 34(2), 138-170.
- Kertoy, M. K., & Goetz, K. M. (1995). The relationship between listening performance on the sentence verification technique and other measures of listening comprehension. *Contemporary Educational Psychology*, 20(3), 320-339.
- Kintsch, E., & Kintsch, W. (2005). *Comprehension in children's reading comprehension and assessment* (pp. 89-110). Routledge.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49(4), 294.
- Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363-394.
- Klingner, J. K. (2004). Assessing reading comprehension. *Assessment for effective intervention*, 29(4), 59-70.
- Lastrapes, R. E. E. (2015). Evaluating technical adequacy features of sentence verification technique as a general outcome measure of content knowledge. *lsu doctoral dissertations*. 1232. Retrieved from https://digitalcommons.lsu.edu/gradschool_dissertations/1232
- Lehto, J. E., & Anttila, M. (2003). Listening comprehension in primary level grades two, four and six. *Scandinavian Journal of Educational research*, 47(2), 133-143.
- Lynch, D. J. (1982). A construct validity study of the sentence verification technique as a method of measuring reading comprehension. *Masters Theses 1911 - February 2014*. 1745. Retrieved from <https://scholarworks.umass.edu/theses/1745>.
- Marchant, H. G., Royer, J. M., & Greene, B. A. (1988). Superior reliability and validity for a new form of the Sentence Verification Technique for measuring comprehension. *Educational and Psychological Measurement*, 48(3), 827-834.
- Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology*, 47(5), 315-335.
- Marcotte, A. M., Rick, F., & Wells, C. S. (2018). Investigating the Reliability of the Sentence Verification Technique. *International Journal of Testing*, 1-22.
- McKeown, M. G., Beck, I. L., & Blake, R. G. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44(3), 218-253.
- McNamara, D. S., Ozuru, Y., & Floyd, R. G. (2017). Comprehension challenges in the fourth grade: The roles of text cohesion, text genre, and readers' prior knowledge. *International Electronic Journal of Elementary Education*, 4(1), 229-257.
- Mestre, J. P., & Royer, J. M. (1991). Cultural and linguistic influences on Latino testing. (Ed Gary D. Keller, James R. Deneen, Rafael J. Magallan.), *Assessment and Access: Hispanics in Higher Education*, (pp. 39-66). State University of New York Press.
- Mooney, P., & Lastrapes, R. E. (2018). Replicating criterion validity in science content for the combination of critical content monitoring and Sentence Verification Technique. *Assessment for Effective Intervention*, 1-8. <https://doi.org/10.1177/1534508418758362>
- Mooney, P., Lastrapes, R. E., Marcotte, A. M., & Amy Matthews Matthews, B. S. (2016). Validity of two general outcome measures of science and social studies achievement. *Specialis Ugdymas/Special Education*, 1(34), 145-188.
- Organisation for Economic Co-operation and Development (OECD). (2016). *PISA 2015 results in focus*.
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. (3rd edition). Sage publications. California, Thousand Oaks
- Pichette, F., Béland, S., De Serres, L., & Lafontaine, M. (2014). Psychometric validation of the Sentence Verification Technique to assess L2 reading comprehension ability. *The Quantitative methods for Psychology*, 10(2), 95-106.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale NJ: Erlbaum
- Royer, J. (2001). Developing reading and listening comprehension tests based on the sentence verification technique (svt). *Journal of Adolescent & Adult Literacy*, 45(1), 30-41.
- Royer, J. M., & Carlo, M. S. (1991). Transfer of comprehension skills from native to second language. *Journal of Reading*, 34(6), 450-455.
- Royer, J. M., & Cunningham, D. J. (1981). On the theory and measurement of reading comprehension. *Contemporary Educational Psychology*, 6(3), 187-216.
- Royer, J. M., & Sinatra, G. M. (1994). A cognitive theoretical approach to reading diagnostics. *Educational Psychology Review*, 6(2), 81-113.
- Royer, J. M., Abranovic, W. A., & Sinatra, G. M. (1987). Using entering reading comprehension performance as a predictor of performance in college classes. *Journal of Educational Psychology*, 79(1), 19.
- Royer, J. M., Carlo, M. S., Carlisle, J. F., & Furman, G. A. (1991). A new procedure for assessing progress in transitional bilingual education programs. *Bilingual Review/La Revista Bilingüe*, 16(1), 3-14.
- Royer, J. M., Greene, B. A., & Sinatra, G. M. (1987). The sentence verification technique: A practical procedure for testing comprehension. *Journal of Reading*, 30(5), 414-422.
- Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Reading Behavior*, 11(4), 355-363.
- Royer, J. M., Lynch, D. J., Hambleton, R. K., & Bulgareli, C. (1984). Using the sentence verification technique to assess the comprehension of technical text as a function of subject matter expertise. *American Educational Research Journal*, 21(4), 839-869.

- Royer, J. M., Marchant, H. G., Sinatra, G. M., & Lovejoy, D. A. (1990). The prediction of college course performance from reading comprehension performance: Evidence for general and specific prediction factors. *American Educational Research Journal*, 27(1), 158-179.
- Singer, M., Harkness, D., & Stewart, S. T. (1997). Constructing inferences in expository text comprehension. *Discourse Processes*, 24(2-3), 199-228.
- Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. RAND Corporation.
- Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2013). *Using multivariate statistics* (6th ed.). Pearson.
- Tunçer, B. K., & Erden, G. (2015). Boşluk doldurma testlerinin ilkökul 4. sınıf öğrencilerinin okuduğunu anlama düzeylerini belirlemede kullanılabilirliği. *Bartın Üniversitesi Eğitim Fakültesi Dergisi*, 318-324.
- Ulusoy, M. (2009). Boşluk tamamlama testinin okuma düzeyini ve okunabilirliği ölçmede kullanılması. *Türk Eğitim Bilimleri Dergisi*, 7(1), 105-126.
- Ulusoy, M., & Çetinkaya, Ç. (2012). Cümle doğrulama tekniğinin okuma ve dinlemenin ölçülmesinde kullanılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 43(43), 460-471.
- Ushiro, Y., Takaki, S., Kobayashi, M., Hasegawa, Y., Nahatame, S., Hamada, A., & Kimura, Y. (2013). Measures of macroproposition construction in EFL reading: Summary writing task vs. the Meaning Identification Technique. *JLTA Journal*, 16, 185-204.
- Yazıcı, E., & Kurudayıoğlu, M. (2017). 5. sınıf Türkçe ders kitaplarındaki dinleme metinlerinin öğrencilerin seviyesine uygunluğunun incelenmesi. *Ana Dili Eğitimi Dergisi*, 5(4), 967-984.
- Yıldırım, K., Yıldız, M., Ateş, S., & Rasinski, T. (2010). İlköğretim beşinci sınıf Türk öğrencilerin metin türlerine göre okuduğunu ve dinlediğini anlama düzeyleri. *Elementary Education Online*, 9, 44-51.