

## Validity and Reliability Study for the Turkish Version of Number Sense Screener for 60-71 Months Old Children

Gozde Inal Kızıltepe

Correspondence: Gozde Inal Kızıltepe, Faculty of Education, Aydın Adnan Menderes University, 0910 Aydın, Turkey.

Received: December 2, 2018

Accepted: December 11, 2018

Online Published: December 17, 2018

doi:10.11114/jets.v7i2.3888

URL: <https://doi.org/10.11114/jets.v7i2.3888>

### Abstract

In this study, it was aimed to carry out the validity and reliability study of Number Sense Screener (NSS)'s Turkish version for 60-71 year-old children. The universe of the study in which general survey model was used consists of children at 60-71 years of age, who were continuing their education during the spring term of academic year 2017-2018 in public and private kindergartens under the Ministry of National Education within the central district of Aydın/Turkey. As for the sample of the study, 658 children continuing their education in 12 different preschools pre-determined according to socioeconomic status (low-middle-high) were included through systematic sampling method. Validity and reliability studies were carried out using data from the NSS application. Drawing from the analyses, it was determined that item difficulty values of the assessment tool ranged between 0.08 and 0.96, item discrimination indices ranged between 0.17 and 0.53 and item infit and outfit values ranged between 0.50-1.50. Having used Guttman's lambda-2, Alpha coefficient, Feldt-Gilmer coefficient, Feldt-Brennan coefficient and Raju's beta coefficient, it was determined that reliability coefficient calculated for the entire assessment tool varied between .826 and .837. Following DIF analysis, it was concluded that assessment tool items were not biased towards gender. According to the results of the study, it was concluded that the Turkish version of NSS is a valid and reliable assessment tool for assessing the number sense of children at 60-71 months of age.

**Keywords:** number sense, number sense screener, preschool, validity and reliability

### 1. Introduction

The fact that it's proven by many studies that mathematical concepts and skills acquired during preschool period are the strongest predictor for later achievement in mathematics led to more importance given to the acquisition of mathematical concepts and skills in both national and international education programs (Mazzocco & Thompson, 2005; Jordan, Kaplan, Olah & Locuniak, 2006; Duncan et al., 2007; Lopez, Gallimore, Garnier & Reese, 2007; Romano, Babchishin, Pagani & Kohen, 2010).

The most important mathematical concept emphasized in preschool period is that of number (Charlesworth & Lind, 2010; Onkol, 2012). The concept of number is a cornerstone for mathematics programs (NAEYC, 2010) and is a prerequisite for teaching advanced mathematics skills (Charlesworth, 2012; Nguyen et al., 2016). The National Council of Teachers of Mathematics-NCTM (2000) emphasizes the importance of developing number skills in mathematics programs during early childhood and highlights that the longest duration of preschool education should be allocated for number skills (as cited in Yılmaz, 2012). Number knowledge in preschool period includes many interrelated skills. These skills are often referred to as 'number sense' in international literature (Gersten, Jordan, & Flojo, 2005; Jordan, Gluttin & Dyson, 2012; Cekirdekci, Sengul & Dogan, 2016).

Number sense has become one of the focal points of mathematics education at international level in the last two decades. When the studies are examined, however, it is seen that many different definitions about number sense are available (Gersten et al., 2005; Sengul & Gulbagcı Dede, 2013). Berch (2005) suggests that the number sense is a sense about the meanings of numbers. The National Council of Teachers of Mathematics-NCTM (1989) defines number sense as the ability to understand the meaning of numbers, describe different relationships among numbers, identify the relative size of numbers, use referents for measuring objects and events, and think in a flexible way with numbers (as cited in Lago & DiPerna, 2010). Baroody ve Wilkins (1999) describe number sense as a concrete understanding of numerical relationships. In the light of these definitions, the term generally refers to one's understanding of numbers, ways of representing them, and relationships among numbers (Reys et al., 1999; Gersten et al., 2005). When definitions of

number sense are examined, it is seen that two common points are emphasized. First, numbers and the relationships among numbers or in other words, operations are mentioned. The second common point is about numbers and flexibility while using operations (Sengul & Gulbağcı Dede, 2013). In this respect, it was concluded that the special skills defined as part of the number sense include verbal and object count, quantity comparison, numerical definition and basic calculation (Howell & Kemp, 2009; Lago & DiPerna, 2010). Jordan et al. (2006) suggest that the basic components of number sense are counting, number knowledge, number transformation, estimation and number patterns.

In many studies, number sense is defined as predictive of later mathematics achievement. For example, Mazzocco & Thompson (2005) found that learning disorders in mathematics in later years are due to the deprivation of number sense in preschool period. Jordan, Kaplan, Locuniak & Ramineni (2007) found that number sense in preschool period is quite effective in predicting achievement in mathematics in primary school. Additionally, it is predicted that difficulties with number sense faced by children during preschool period are very likely to grow further and continue in later stages of education (Aunio, Hautamäki & Van Luit; 2005). As can be seen, number sense is one of the most crucial concepts to be developed in early mathematics and is a key predictor of subsequent mathematical achievement both in short (Aunio & Niemivirta, 2010) and long term. Number sense was also considered to be one of the content standards of mathematics education by National Council for Teachers of Mathematics-NCTM (2000). According to the standards put forth for mathematics education by NCTM, young children should develop a foundational number sense by the end of second grade (NCTM, 2000).

Number sense gradually develops from infancy and matures with the experience and knowledge acquired in the preschool period as a result of these children's' interaction with their informal or formal environment (Reys & Yang, 1998; Ginsburg Lee & Boyd, 2008; Jordan et al., 2012a). Factors such as the diversity of mathematical concepts used in line with the characteristics of familial environment (socioeconomic status of the family, parents' level of education and so forth) where the children are born into, attitude towards mathematics, playing number-related games at home and being able to make use of house chores for educational purposes result in children having different levels of number sense competency (Howse, Lange, Farran & Boyles, 2003; Penner & Paret 2008; Ivrendi, 2011; Ramani, Siegler, & Hitti, 2012; Laski & Siegler, 2014, Gulec & İvrendi, 2017). Garon-Carrier et al. (2018) also suggest that children acquire different levels of skills about number sense in their home setting. Pittalis, Pitta-Pantazi and Christou (2016), on the other hand, emphasize that the number sense is highly personal. It is highly important to evaluate the number sense of preschool children due to individual differences and its impacts on primary school period and to develop early intervention programs especially for children in the risk group (who have low number sense) (Aunio et al., 2005; Jordan et al., 2006). Because of the importance of number sense, there is a need for valid and reliable assessment tools for assessing preschool children (Gersten & Chard, 1999). Lago and DiPerna (2010) put forth that a variety of tests have been developed to measure number sense, yet emphasize that a very few of the assessment tools examine reliability and validity. Number Sense Screener (NSS) is one of the assessment tools used in many different studies in order to measure number sense of preschool children and its validity and reliability are proven in line with longitudinal studies (Jordan et al., 2006; Lago & DiPerna; 2010; Jordan et al., 2012a; Dyson, Jordan, Beliakoff, Hassinger-Das, 2015; Starr, DeWind & Brannon, 2017). Number Sense Screener is an important assessment tool for being a psychoeducational assessment instrument to monitor number development, predict later mathematics achievement, and identifies children at risk and to prepare effective intervention programs accordingly (Jordan et al., 2012a).

When examining the literature related to the subject in Turkey, there is an increase in the body of research dedicated to the number skills of children in preschool period in parallel with the greater importance given to the concept of number in the international literature in recent years and there has been a development of various assessment tools for assessing the number skills of children preschool in this direction or their versions adapted to Turkish language. (e.g., Onkol, 2012; Olkun, Fidan, Babacan Ozer, 2013; Pekince & Daglıoğlu, 2017; Yılmaz & Inal Kiziltepe, 2017). In an attempt to evaluate the number sense of preschool children, there is another assessment tool called 'Assessing Number Sense Instrument' developed by Wakefield and Ivrendi (2008) (as cited in Ivrendi, 2011). From this point of view, this study aims to conduct the validity reliability study of the Number Sense Screener (NSS), which is frequently used in international literature to evaluate the number sense of preschool children, for children at 60-71 months of age. It is believed that the validity and reliability study of the Turkish version of NSS is crucial in terms of providing a valid and reliable standard to determine the number sense levels of preschool children in this group of age, identify children with lower number sense, prepare early intervention programs, measure the success of applied training programs and determine the factors affecting number sense.

## 2. Method

### 2.1 Research Model

The research model falls within general survey since the study is to determine the current situation with analysis of the data obtained. In general survey, within a universe consisting of a large number of elements, a general survey of the universe is carried out on the entirety of the universe or on a group, sample or sample taken from it (Karasar, 2010).

### 2.2 Participants

The universe of the study consisted of children at 60-71 months of age attending public and private kindergartens under the Ministry of National Education in Efeler district center of Aydın province in the spring term of the academic year 2017-2018. The sample of the study consisted of 658 children attending 12 preschool education institutions which were determined according to socioeconomic level (low-middle-high income) in Aydın, Turkey. The mean age of children is 6.8 and the standard deviation is 5.7. 342 of the children were girls (52.0%) and 316 were boys (48.0%). The study was carried out in the preschool education institutions which volunteered to participate in the study in accordance with the legal permits obtained from the Provincial Directorate of National Education. Moreover, permission was obtained from the parents of the children included in the sample through the form developed by the researcher.

Sampling of children was determined by systematic sampling method. The participants of the study were not selected according to any specific characteristic they may possess accordingly, and hence the validity of the research was enhanced. Systematic sampling is a method whereby a starting point ( $X^{\text{th}}$ ) is taken from a unit and selecting every subsequent 'x' unit for the sample (Cıngı, 1994). To this end, a list of the classes in which 60-71 month-old-children were enrolled, was obtained in the schools included in the sampling. In line with this list, the assessment tool (second, fourth, sixth, eighth...) was applied to the children determined by the number of 'two' selected as the initial unit. In the absence of parental leave form of children selected through systematic sampling method, the assessment tool was applied to the next child in the list.

### 2.3 Data Collection Tool

The data were collected through "Number Sense Screener (NSS), which consists of six subtests.

**Number Sense Screener (NSS):** Assessment tool, developed for the purpose of assessing the early numerical competencies of the first grade children of preschool and primary school (e.g., Jordan et al., 2006; Jordan et al., 2007) is a short name for the research-based assessment tool named 'Number Sense Brief (33 items)'. Assessment tool is of six subtests and its duration of application varies between 15 and 20 minutes. In this study, the validity and reliability analyses of the NSS form consisting of 29 items were performed (Jordan et al., 2012a). Information on the contents and number of items of the Number Sense Screener's subtests is given below:

**Counting Skills:** This subtest consists of three items. The subtest includes items containing counting principles (one-to-one correspondence, cardinality, and stable order) and rhythmic counting (counting rhythmically until a predetermined number) (Jordan et al., 2007; Jordan et al., 2012a).

**Number Recognition:** In the number recognition subtest consisting of four items, children are asked to name the numbers shown to them (such as 13, 37) (Jordan et al., 2009; Jordan et al., 2012a).

**Number Comparisons:** Number comparison subtest adapted from Griffin (2002) consists of seven items. In this segment of the test, children are asked to know what number comes after a pre-given number or what number comes two numbers after that pre-given number (e.g., 4) and also given two numbers (e.g., 4 and 3), children were expected to tell which number was bigger or smaller. Furthermore, children were shown a series of numbers, each of which was placed in a corner of an equilateral triangle (e.g., 7, 3, and 6). Children were then asked to identify which number was closer (e.g., 6) to the number in the top corner of the triangle (Jordan, et al., 2012a; Dyson, Jordan & Glutting, 2013).

**Nonverbal Calculation:** In this subtest, there are a total of four items: three additions and one subtraction. The nonverbal calculation task was adapted from Levine et al. 1992. Items of this section are asked to the children, using a white mat and a box (the cover of the box has an opening cut into the side through which dots can be pushed into the box) and 10 black dots of the same size. For example, three buttons are placed on the box, which can be seen by the children for the collection process and the examiner says "See? Here are three dots". Letting children observe the buttons, buttons are then taken back into the box and the box cover is put back on. Following that, another black dot is placed on the mat and the examiner says "Here is another black dot" and asks children to watch attentively and black dot is slid with box lid and put back into the box. After that, pages related to the item in the assessment tool are opened and the children are asked to point to the option showing the total number of black dots inside the box. Because this is a nonverbal task, the child is not penalized if he or she points to the correct rectangle with dots but says the wrong word (Jordan, et al., 2012a; Jordan, et al., 2012b).

**Story Problems:** Story problems subtest consists of a total of five items including three additions and two subtractions. In this part, the children are informed that to find the right answers, they can use their fingers to ask for help or use the number list (given with the assessment tool) or use paper and pencil to that end. The addition problems are phrased simply, following this basic format: “Susan has  $m$  pennies. Jim gives her  $n$  more pennies. How many pennies does Susan have now?” Similarly, the subtraction problems are phrased, “Susan has  $m$  pennies. Jim takes away  $n$  of her pennies. How many pennies does Susan have now?” (Jordan, Glutting, Ramineni; Watkins, 2010; Jordan et al., 2012a).

**Number Combinations:** Number combinations section consists of six items including four additions and two subtractions and verbally asked as follows: “How much is  $m$  and  $n$ ” and “How much is  $n$  take away  $m$ ?” In this section, as in story problems test, children are informed that to find the right answer, they can use fingers to ask for help or use number list or use paper and pencil (Locuniak & Jordan, 2008; Jordan et al., 2012a).

The items included in the subtests are evaluated as true, false and no response; one point for correct answers, zero score for no response and false answers. For each test, total score is calculated through summing the correct answers given by children for all the six tests. Total number sense score is obtained by the summation of the scores of all subtests (Jordan et al., 2007; Jordan et al., 2012a). Accordingly, maximum score that be achieved for this assessment tool is 29.

The Turkish version of NSS was applied individually to the children included in the group under study by a researcher or examiners trained about the assessment tool, in a quieter setting other than children’s usual classrooms provided by their preschool institutions, complying with the rules and directives in the application booklet. Six examiners selected from among Preschool Education Department’s undergraduate and graduate students were informed in detail about the assessment tool and how to apply it. Afterwards, how each examiner applied the assessment tool to a child (between 60-71 months old) not included in the study group was observed and examiners were given necessary feedbacks according to the notes taken during observation. Furthermore, examiners were divided into dual groups. One of the practitioners in the dual group did not apply the assessment tool to a different child between 60-71 months old, which was not included in the study group, and then calculated the score of the child obtained from the assessment tool in accordance with the sign-up form. The other examiner in the dual group calculated the score obtained by the child by filling in the answers given by the child into his/her own sign-up form. Following that, whether each two examiner scored the tasks the same way was checked.

#### 2.4 Adaptation of NSS into Turkish

The adaptation of the NSS into Turkish has been started with the translation of the assessment tool into Turkish with the original instruction booklet in English language. First, forms (instructions booklet and assessment tool) were translated into Turkish by three language experts and then translated back into English through back-translation technique. Forms that had been translated into Turkish were then translated back into English using back-translation technique. Forms were reviewed by a language expert who had a good command of both Turkish and English and it was concluded that forms were in accord with one another in terms of expression and cohesion. Forms translated into Turkish were reviewed by a Turkish language expert and the necessary changes were made and then they were finalized.

Expert opinion was used to determine the qualitative and quantitative efficacy of the assessment tool items in assessing and encompassing (content validity) a trait which is aimed to be assessed (McGartland et al., 2003). To this end, forms were sent out to seven field experts (preschool education, class education, mathematics education) and they were asked to evaluate the suitability of instructions and test items of NSS in terms of expression, Turkish culture, evaluation the number sense of children at 60-71 months of age, the fields that include subtest items.

With the consensus of the experts, the items they found appropriate were taken as they were into the Turkish version of the assessment tool and the items proposed to be corrected were modified accordingly. In this respect, the writing of the number four was changed (4 instead of 4) in such a way that it is used more frequently in Turkish language. In an effort to have an idea about comprehensibility of the assessment tool by children and the approximate duration of application, the Turkish version of the NSS was applied on a small group ( $n = 10$ ) and it was observed that the scale items were understood by the children.

#### 2.5 Data Analysis

Content validity of Number Sense Screener was conducted through Lawshe technique in line with expert opinions and content validity ratios (CVR) and content validity indices (CVI) were calculated. In the reliability and validity analyses of Number Sense Screener, Rasch analysis through jMetrik 4.111 software which is based CMLE=Conditional Maximum Likelihood Estimation and PROX=Normal Approximation methods was used.

Table 1 shows the acceptable ranges for the measures to be used for unweighted and standardized goodness-of-fit statistics performed by Linacre (2002) in the evaluation of the results obtained from the software.

Table 1. Interpretation of parameter-level mean-square fit statistics

Criterion	Interval	Comment
WMS and UMS	> .20	It distorts subject matter and lowers assessment quality.
	1.5 – 2.0	Unproductive for assessment process yet not unwieldy either.
	0.5 – 1.5	Highly suitable for assessment.
	≤ 50	Less productive for assessment process yet not that ineffective. Can produce high reliability and discrimination values.
Std.WMS and Std.UMS	≤ 3	Data do not fit the model. A larger sample may be needed.
	2.0 – 2.9	Data cannot be predicted clearly.
	-1.9 – 1.9	Data have a plausible predictability.
	≤ -2.0	Data are highly predictable. Other dimensions may limit the answer patterns.

When Table 1 is examined, it is determined that the ideal range for WMS and UMS values is between 0.50 and 1.50. When Table 1 is examined, similarly, it is determined that the ideal range for standardized WMS and UMS values is between -1.90 and 1.90. The columns WMS and Std.WMS show the weighted and standardized mean square outfit statistics, respectively. Following these, the columns UMS and Std.UMS show the weighted and standardized mean square deviation statistics. (Guzeller, Eser & Aksu, 2018).

The reliability coefficients for the assessment tool subtests and for the entire assessment tool were calculated by using Guttman's lambda-2, Alpha coefficient, Feldt-Gilmer coefficient, Feldt-Brennan coefficient and Raju's beta coefficient methods. Moreover, differential item functioning (DIF) analysis was carried out so as to determine whether assessment tool items are biased towards gender. The Mantel-Haenszel contingency table approach was used to identify items displaying DIF.

### 3. Results

As part of the validity and reliability studies of Number Sense Screener (NSS) for 60-71 months old children, this section includes content validity indices difficulty and discrimination values of items, standard deviation values, reliability analyses for the entirety of assessment tool and its items, and differential item functioning (DIF) results to determine whether the item statistics and assessment tool items obtained through Rasch analysis are biased toward gender.

In Table 2, content validity indices for the sum of subtests and NSS are presented.

Table 2. Content validity indices for the total of NSS and its subtests

NSS	Items Numbers	CVI
Counting Skills	3	1,00
Number Recognition	4	1,00
Number Comparisons	7	1,00
Nonverbal Calculation	4	1,00
Story Problems	5	1,00
Number Combinations	6	1,00
Total	29	1,00

The technique developed by Lawshe requires at least 5 and at most 40 expert opinions. (akt: Yurdugul, 2005). Seven experts were consulted in this study.

Content validity ratios were obtained through a collection of different expert opinions on the measurement items. Veneziano and Hooper (1997) suggest that the minimum content validity ratio for the seven experts should be 0.99. CVI is calculated based on the total CVR averages of items that are significant at the level of  $p < 0.05$  and that are to be taken into the final form. (akt: Yurdugul, 2005).

According to the Lawshe technique, the CVI values for each subtest of the assessment tool were obtained as 1.00. Accordingly, it was decided that each item should stay in the measure and accepted that the test had content validity.

Descriptive statistics of difficulty, discrimination and standard deviation values of NSS items are shown in Table 3.

When Table 3 is examined, it is seen that the difficulty values of NSS items vary between 0.08 and 0.96, discrimination indices vary between 0.17 and 0.53. According to this result, it was determined that all the items in the assessment tool except for the 7th were at acceptable levels, 7th item was found to have low discrimination value and be highly difficult.

Table 3. Descriptive statistics for NSS items

Item	Item difficulty	Standard deviation	Item difficulty indices
I1	0,9666	0,1799	0,1747
I2	0,8951	0,3066	0,3255
I3	0,9255	0,2627	0,1748
I4	0,6733	0,4694	0,4186
I5	0,3055	0,4610	0,4055
I6	0,1945	0,3961	0,4302
<b>I7</b>	<b>0,0866</b>	<b>0,2815</b>	<b>0,2292</b>
I8	0,7736	0,4188	0,2729
I9	0,4088	0,4920	0,2651
I10	0,7234	0,4477	0,2332
I11	0,6155	0,4868	0,3714
I12	0,6626	0,4732	0,3402
I13	0,7234	0,4477	0,3127
I14	0,6368	0,4813	0,3001
I15	0,8708	0,3357	0,1985
I16	0,7264	0,4461	0,2658
I17	0,5198	0,5000	0,2304
I18	0,8495	0,3578	0,2288
I19	0,7492	0,4338	0,3692
I20	0,3967	0,4896	0,4052
I21	0,5334	0,4993	0,3384
I22	0,4103	0,4923	0,3707
I23	0,4818	0,5000	0,3465
I24	0,7386	0,4397	0,3507
I25	0,5426	0,4986	0,5129
I26	0,4392	0,4967	0,5382
I27	0,5061	0,5003	0,4431
I28	0,3647	0,4817	0,4511
I29	0,2933	0,4556	0,4496

Reliability values for NSS items obtained through different methods are given in Table 4.

Table 4. Reliability values for NSS items

Item	Guttman's lambda-2	Alpha Coefficient	Feldt-Gilmer coefficient	Feldt-Brennan coefficient	Raju's Beta coefficient
I1	0,8363	0,8256	0,8294	0,8288	0,8256
I2	0,8336	0,8219	0,8266	0,8259	0,8219
I3	0,8365	0,8255	0,8296	0,8290	0,8255
I4	0,8302	0,8179	0,8230	0,8223	0,8179
I5	0,8305	0,8185	0,8234	0,8227	0,8185
I6	0,8301	0,8181	0,8229	0,8222	0,8181
I7	0,8354	0,8243	0,8286	0,8280	0,8243
I8	0,8349	0,8233	0,8282	0,8275	0,8233
I9	0,8359	0,8241	0,8290	0,8283	0,8241
I10	0,8364	0,8249	0,8296	0,8289	0,8249
I11	0,8320	0,8198	0,8248	0,8240	0,8198
I12	0,8330	0,8210	0,8260	0,8253	0,8210
I13	0,8340	0,8220	0,8268	0,8261	0,8220
I14	0,8346	0,8226	0,8274	0,8268	0,8226
I15	0,8365	0,8252	0,8294	0,8289	0,8252
I16	0,8354	0,8237	0,8286	0,8279	0,8237
I17	0,8373	0,8256	0,8304	0,8297	0,8256
I18	0,8359	0,8244	0,8289	0,8283	0,8244
I19	0,8319	0,8199	0,8244	0,8238	0,8199
I20	0,8303	0,8184	0,8228	0,8222	0,8184
I21	0,8326	0,8212	0,8254	0,8249	0,8212
I22	0,8312	0,8198	0,8241	0,8236	0,8198
I23	0,8323	0,8208	0,8251	0,8246	0,8208
I24	0,8321	0,8206	0,8249	0,8244	0,8206
I25	0,8249	0,8139	0,8177	0,8173	0,8139
I26	0,8236	0,8128	0,8165	0,8161	0,8128
I27	0,8282	0,8168	0,8209	0,8204	0,8168
I28	0,8277	0,8166	0,8207	0,8202	0,8166
I29	0,8265	0,8152	0,8194	0,8189	0,8250

In Table 4, it is seen that the reliability coefficient calculated according to five different methods is above the value of .80 value which is considered as critical for the assessment tools. At the same time, it was determined that the 7th item which has lower difficulty value than other items and found difficult by children has a significantly high reliability coefficient. According to this result, it was determined that the NSS materials meet the assumption of reliability in terms of internal consistency.

The reliability coefficients for the sum of the assessment tool and the 95% confidence interval and standard error values for this value are shown in Table 5.

Table 5. Reliability Analysis Results for NSS

Method	Coefficient	Confidence interval lower limit	Confidence interval upper limit	SEM
Guttman's lambda-2	0,8373	0,8189	0,8547	2,1240
Alpha Coefficient	0,8261	0,8064	0,8447	2,1962
Feldt-Gilmer coefficient	0,8304	0,8112	0,8485	2,1686
Feldt-Brennan coefficient	0,8299	0,8106	0,8480	2,1722
Raju's Beta coefficient	0,8261	0,8064	0,8447	2,1962

When Table 5 is examined, it is seen that the reliability coefficients determined by different methods for the total of 29 items vary between .826 and .837. According to this result, the results obtained from the sum of the assessment tool are considered to be reliable as well.

In terms of validity and reliability, it is not sufficient to examine only item difficulty and discrimination indices when determining of substances that should be included in an assessment tool. For this reason, goodness-of-fit indices with converted discrimination indices were calculated via Rasch analysis being one of the 1-parameter logistic models. Table 6 shows the item statistics obtained through Rasch analysis according to the following parameters: maximal number of iterations = 150, convergence criteria= 0.005 and extreme score adjustment= 0.3.

Table 6. Rasch Analysis Results for NSS Items

Item	Difficulty	Std. Error	WMS	Std. WMS	UMS	Std. UMS
I1	-3,24	0,21	0,89	-0,58	0,62	-1,00
I2	-1,97	0,13	0,87	-1,46	0,72	-1,47
I3	-2,37	0,15	0,92	-0,69	1,79	2,67
I4	-0,35	0,09	0,93	-1,92	0,83	-2,14
I5	1,54	0,10	0,94	-1,33	1,00	-0,01
I6	2,28	0,11	0,88	-1,98	0,72	-2,69
I7	3,42	0,15	1,01	0,12	1,09	0,46
I8	-0,95	0,10	1,02	0,34	1,04	0,36
<b>I9</b>	<b>0,98</b>	<b>0,09</b>	<b>1,15</b>	<b>3,85</b>	<b>1,19</b>	<b>3,05</b>
I10	-0,64	0,10	1,11	2,41	1,12	1,22
I11	-0,05	0,09	1,00	-0,08	0,94	-0,86
I12	-0,30	0,09	1,02	0,54	0,95	-0,55
I13	-0,64	0,10	1,03	0,66	0,95	-0,51
I14	-0,16	0,09	1,07	1,90	1,02	0,36
I15	-1,71	0,12	1,00	0,00	1,56	2,78
I16	-0,66	0,10	1,06	1,24	1,28	2,64
<b>I17</b>	<b>0,42</b>	<b>0,09</b>	<b>1,17</b>	<b>4,82</b>	<b>1,28</b>	<b>4,51</b>
I18	-1,51	0,12	1,01	0,18	1,03	0,27
I19	-0,79	0,10	0,94	-1,29	1,08	0,77
I20	1,04	0,09	0,99	-0,37	0,97	-0,40
I21	0,36	0,09	1,05	1,44	1,06	0,99
I22	0,97	0,09	1,00	0,11	1,03	0,54
I23	0,61	0,09	1,04	1,22	1,04	0,65
I24	-0,73	0,10	0,97	-0,56	0,88	-1,16
I25	0,31	0,09	0,85	-4,68	0,81	-3,43
<b>I26</b>	<b>0,82</b>	<b>0,09</b>	<b>0,82</b>	<b>-5,25</b>	<b>0,75</b>	<b>-4,72</b>
I27	0,49	0,09	0,93	-2,13	0,94	-1,06
I28	1,21	0,09	0,91	-2,24	0,85	-2,52
<b>I29</b>	<b>1,61</b>	<b>0,10</b>	<b>0,90</b>	<b>-2,26</b>	<b>0,83</b>	<b>-2,27</b>

Unweighted Mean Square (UMS) and Weighted Mean Square (WMS) fit statistics are the goodness-of-fit statistics for the items. Within these values, WMS is considered to be the infit measure, while UMS stands for outfit. As can be seen

in the Table, the assessment tool consisting of 29 items is suitable for the assessment process to be conducted according to WMS and UMS values (Linacre, 2002). As the infit and outfit values of the items were in the range of 0.50-1.50, the-goodness-of-fit statistics of the substances were found to be quite satisfactory. According to standardized infit and outfit values, it was determined that the 9th, 17th, 26th and 29th items of the assessment tool did not provide a model-data compatibility.

DIF analysis was carried out to find out whether NSS items were biased towards gender. The results obtained based on the common odds ratio with Mantel-Haenszel method are shown in Table 7.

Table 7. DIF Analysis Results

Item	Chi-Square	p-value	Valid N	Effect size	Class
I1	0,37	0,54	332	0,76	A
I2	0,07	0,79	463	0,93	A
I3	0,06	0,80	495	1,09	A
I4	3,07	0,08	549	0,72	A
I5	0,04	0,84	615	0,96	A
I6	0,18	0,67	456	0,91	A
I7	0,33	0,57	495	1,18	A
I8	3,37	0,07	590	0,68	A
I9	0,21	0,65	632	0,92	A
<b>I10</b>	<b>6,39</b>	<b>0,01</b>	<b>611</b>	<b>1,59</b>	<b>B-</b>
I11	0,86	0,35	616	1,18	A
I12	0,27	0,61	594	1,10	A
I13	3,96	0,05	564	0,68	A
I14	0,33	0,57	593	0,90	A
I15	0,37	0,54	604	0,86	A
I16	0,48	0,49	622	0,88	A
I17	1,02	0,31	638	0,85	A
I18	0,33	0,56	574	0,87	A
I19	1,46	0,23	570	1,28	A
I20	2,88	0,09	592	1,36	A
I21	3,54	0,06	615	1,38	A
I22	0,01	0,93	642	1,02	A
I23	0,36	0,55	634	1,11	A
I24	3,47	0,06	562	0,69	A
I25	0,32	0,57	612	0,90	A
I26	0,00	0,97	563	1,01	A
I27	0,04	0,84	638	0,96	A
I28	1,35	0,25	578	1,26	A
I29	1,93	0,17	584	1,33	A

As can be seen in Table 7, it is determined that all items except for the 10th item have a negligible level of item function. The Item Characteristic Curve of the 10th item, determined to be biased at an intermediate (B-) level, is shown in Figure 1.

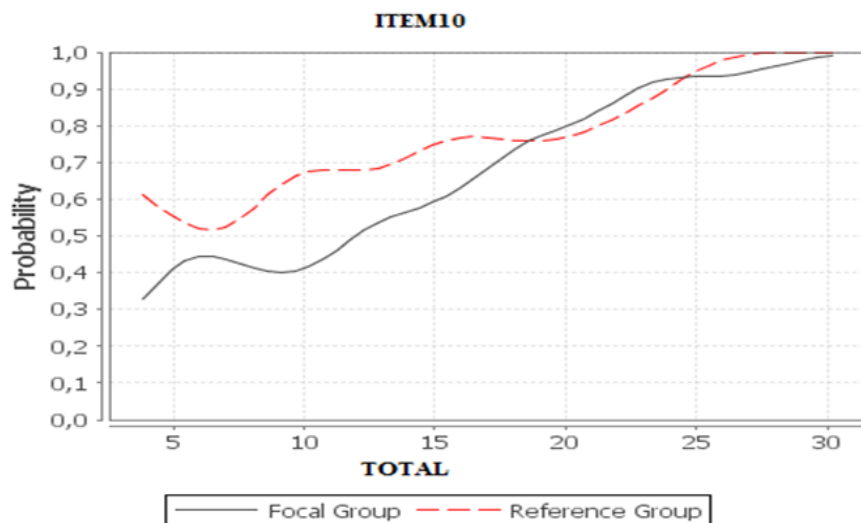


Figure 1. Item Characteristic Curve for item 10



When Figure 1 is examined, it is seen that girls who were specified as the reference group scored higher than boys when the total of points were in the range of 0-20. However, boys scored higher when the total points were in the range of 20-25 and the case changes again as the range goes above 25. This result validates that 10th item had a negligible level of DIF. On the other hand, it was concluded that the results obtained for the entirety of 29th item were not biased such that it would leave any group at an advantage or disadvantage since an item has to have a level C DIF to be determined as biased (Koyuncu, Aksu & Kelecioğlu, 2018).

#### 4. Discussion and Conclusions

The validity and reliability study of the Number Sense Screener (NSS), which was used frequently for preschool period in abroad, was performed in order to introduce a valid and reliable assessment tool into literature and to assess the number sense of 60-71-month-old Turkish children. As part of the validity study, content validity of the assessment tool was examined and evaluated in accordance with seven field experts through Lawshe Technique. According to the Lawshe technique, the CVI values for each subtest of NSS were obtained as 1.00. Accordingly, it was decided that each item should stay in the assessment tool and accepted that the assessment tool had content validity.

Results of the analyses show that the difficulty values of NSS vary between 0.08 and 0.96, discrimination indices vary between 0.17 and 0.53. According to this result, it was determined that all the items in the assessment tool except for the 7th were at acceptable levels, 7th item was found to have low discrimination value and be a difficult question for the children at 60-71 months of age. In addition to this, reliability coefficient of the 7th item varied between .824 and .835 and in the light of expert opinion, it was not to be removed from the assessment tool. In this item being a part of number recognition subtest, children are asked to name the number '124'. According to the Ministry of Education's Preschool Education Program which is currently implemented in Turkey and last updated in 2013, a study associated with numbers should first include numbers from 1-20. In other words, studies with only single and two-digit numbers are found necessary in accordance with the education program in preschool education institutions. Therefore, children included in the sample are thought to have difficulty naming the number '124' as it is a three-digit number.

For the items included in the assessment tool, it was found that the item reliability coefficients calculated by using Guttman's lambda-2, Alpha coefficient, Feldt-Gilmer coefficient, Feldt-Brennan coefficient and Raju's beta coefficient methods were found to be above .80 which is accepted to be the critical threshold. In addition, it was determined that the reliability coefficients calculated using the same methods for the entire assessment tool varied between .826 and .837. High reliability coefficients obtained indicate that the assessment tool is reliable. The original reliability and validity study of Number Sense Screener was conducted on children in the same sampling group, in different periods of time. Assessment tool was applied to children included in the sample in three different periods of time as follows: fall term of kindergarten (mean age=5,8 and SD=3.7), spring term of kindergarten (mean age=6,2 and SD=3.7) and fall term of first grade (mean age=6,8 and SD=3.7). As a result of the analyzes, it was determined that the reliability coefficient for the entirety of assessment tool was .82 for the fall term of kindergarten, .86 for the spring term of kindergarten, .87 for the fall term of first grade and .85 for average (Jordan et al., 2012a). In the reliability and validity study conducted by Jordan et al. (2006), Cronbach's alpha ( $\alpha$ ) coefficient, calculated for the entirety of the assessment tool, was also found to be above the critical value of .80. These results show similarity with the reliability coefficients obtained from the Turkish version.

The infit and outfit values of the Turkish version of the NSS were found to be in the range of 0.50-1.50. In the literature, it is put forth that the ideal range for infit and outfit values is between 0.50 and 1.50 (Guzeller et al., 2018). In this respect, it can be said that the goodness-of-fit statistics of the assessment tool are in the ideal range. In the results of the Rasch analysis of the original assessment tool using the Winsteps software, similarly, it was determined that the infit and outfit values were in the range of 0.50-1.50 for the first and third items (item 1 outfit = 0.43, item outfit = 0.42) (Jordan et al., 2012a).

As a result of the DIF analysis conducted to determine whether the NSS items were biased towards gender, only the 10th item was found to be intermediate-level (B-) biased. In order for an item to be specified as biased towards a particular group, it must show a minimum C level of DIF (Koyuncu et al., 2018). Therefore, it can be concluded that the 29 items in the Turkish version of the assessment tool are not biased in a way that would leave a certain group advantageous or disadvantageous in terms of gender. It was examined with the Mantel-Haenszel method whether the assessment tool items in the original form of the NSS were biased towards gender and it was decided that the assessment tool was not biased towards gender since only a negligible level of item function was determined in one item (Jordan et al., 2012a).

It was determined that the results of validity and reliability analysis carried out for 72-84 month-old children also indicate that reliability coefficients for the entirety of the assessment tool range between .876 and .884, infit and outfit values are in the ideal range (0.50-1.50) as considered in the literature (Uyanık- Aktulun, 2019).

In the light of the findings obtained from the study, it was concluded that the Turkish version of NSS is a valid and reliable assessment tool for assessing the number sense of children between 60-71 months of age. One of the strengths of this study is that it evaluates the validity and reliability of the data obtained from NSS using different methods. On the other hand, there are certain limitations to this study. The findings obtained from this study are limited to 658 children at 60-71 months old, who still continue their education in preschool institutions in a province of Aydın/Turkey. In this respect, the validity and reliability of the assessment tool can be re-tested through expanding the group under study and new studies to be conducted. Moreover, cross-cultural studies can be carried out to compare and evaluate the number sense of children in preschool period, with different language versions of NSS. Additionally, it is considered that examining the relationship between the scales related to number skills adapted to Turkish version of NSS and 60-71 months old children is crucial in terms of introducing multiple assessment tools to the field to evaluate the number sense and number skills of children in this group of age.

## References

- Aunio, P., & Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learning and individual differences*, 20(5), 427-435. <https://doi.org/10.1016/j.lindif.2010.06.003>
- Aunio, P., Hautamäki, J., & Van Luit, J. E. (2005). Mathematical thinking intervention programmes for preschool children with normal and low number sense. *European Journal of Special Needs Education*, 20(2), 131-146. <https://doi.org/10.1080/08856250500055578>
- Baroody, A. J., & Wilkins, J. (1999). The development of informal counting, number, and arithmetic skills and concepts. In J. Copley (Ed). *Mathematics in the early years* (pp. 48–65). Reston, VA: National Council of Teachers of Mathematics.
- Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities*, 38(4), 333–339. <https://doi.org/10.1177/00222194050380040901>.
- Cekirdekçi, S., Sengül, S., & Doğan, M. C. (2016). Examining the relationship between number sense and mathematics achievement of the 4th grade students. *Qualitative Studies (NWSAQs)*, 11(4), 48-66.
- Charlesworth, R. (2012). *Experience in math four young children*, 6th edition. Belmont, CA: Wadsworth Cengage.
- Charlesworth, R., & Lind, K. K. (2010). *Math and science for young children* (6. Edition).USA:Wadsworth, Cengage Learning.
- Cingi, H. (1994). *Sampling theory*. Ankara: Hacettepe University Faculty of Science Publications.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428-1446. <http://dx.doi.org/10.1037/0012-1649.43.6.1428>
- Dyson, N. I., Jordan, N. C., & Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *Journal of learning disabilities*, 46(2), 166-181. <https://doi.org/10.1177/0022219411410233>.
- Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A kindergarten number-sense intervention with contrasting practice conditions for low-achieving children. *Journal for research in mathematics education*, 46(3), 331-370. <https://doi.org/10.5951/jresmetheduc.46.3.0331>
- Garon-Carrier, G., Boivin, M., Lemelin, J. P., Kovas, Y., Parent, S., Séguin, J., & Dionne, G. (2018). Early developmental trajectories of number knowledge and math achievement from 4 to 10 years: Low-persistent profile and early-life predictors. *Journal of school psychology*, 68, 84-98. <https://doi.org/10.1016/j.jsp.2018.02.004>
- Gersten, R., & Chard, D. (1999). Number sense: rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education*, 33(1), 18-28. <https://doi.org/10.1177/002246699903300102>
- Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, 38(4), 293–304. <https://doi.org/10.1177/00222194050380040301>
- Ginsburg, H. P., Lee, J. S., & Boyd, J. S. (2008). Mathematics education for young children: What it is and how to promote it. *Social Policy Report*, 22(1), 3-22. <https://doi.org/10.1002/j.2379-3988.2008.tb00054.x>
- Gulec, N., & Ivrendi, A. (2017). Predicting 5-6 years old children's number concept skills in terms of parent and teacher variables. *HU Journal of Education*, 32(1), 81-98.
- Guzeller, C. O., Eser, M. T., & Aksu, G. (2018). *Measurement applications with JMetrik*. Ankara: Maya Academic.
- Howell, S., & Kemp, C. (2009). A participatory approach to the identification of measures of number sense in children

- prior to school entry. *International Journal of Early Years Education*, 17(1), 47-65. <https://doi.org/10.1080/09669760802699902>
- Howse, R. B., Lange, G., Farran, D. C., & Boyles, C. D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *Journal of Experimental Education*, 71(2), 151–174. <https://doi.org/10.1080/00220970309602061>
- Ivrendi, A. (2011). Influence of self-regulation on the development of children's number sense. *Early Childhood Education Journal*, 39(4), 239- 247. <https://doi.org/10.1007/s10643-011-0462-0>
- Jordan, N. C., Glutting, J. J., & Dyson, N. (2012a). *Number sense screener™ (NSS™) User's guide, k-1, research edition*. Baltimore: Brookes Publishing.
- Jordan, N. C., Glutting, J., Dyson, N., Hassinger-Das, B., & Irwin, C. (2012b). Building kindergartners' number sense: A randomized controlled study. *Journal of Educational Psychology*, 104(3), 647-660. <https://doi.org/10.1037/a0029018>
- Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, 39(2), 181-195.
- Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice*, 22(1), 36-46. <https://doi.org/10.1111/j.1540-5826.2007.00229.x>
- Jordan, N. C., Kaplan, D., Olah, L., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, 77, 153-175. <https://doi.org/10.1111/j.1467-8624.2006.00862.x>
- Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Developmental psychology*, 45(3), 850-867. <http://dx.doi.org/10.1037/a0014939>
- Karasar, N. (2010). *Scientific research methods* (21st Edition). Ankara: Nobel academic publishing.
- Koyuncu, I., Aksu, G., & Kelecioğlu, H. (2018). Comparison of Mantel-haenszel, logistic regression and likelihood ratio methods to evaluate differential item functioning by using different computer software. *Elementary Education Online*, 17(2), 909-925. <https://doi.org/10.17051/ilkonline.2018.419339>
- Lago, R. M., & DiPerna, J. C. (2010). Number sense in kindergarten: A factor-analytic study of the construct. *School Psychology Review*, 39(2), 164–180.
- Laski, E. V., & Siegler, R. S. (2014). Learning from number board games: You learn what you encode. *Developmental Psychology*, 50, 853–864. <http://dx.doi.org/10.1037/a0034321>
- Linacre, J. M. (2002). What do infit and outfit, Mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878.
- Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities*, 41(5), 451-459. <https://doi.org/10.1177/0022219408321126>
- Lopez, E. M., Gallimore, R., Garnier, H., & Reese, L. (2007). Preschool antecedents of mathematics achievement of Latinos: The influence of family resources, early literacy experiences, and preschool attendance. *Hispanic Journal of Behavioral Sciences*, 29(4), 456-471. <https://doi.org/10.1177/0739986307305910>
- Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice*, 22, 152-155. <https://doi.org/10.1111/j.1540-5826.2005.00129.x>
- McGartland, R. D., Berg-Weger, M., Tebb, S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27(2), 94-104. <https://doi.org/10.1093/swr/27.2.94>
- NAEYC- National Association for the Education of Young Children (2010). *Early childhood mathematics: Promoting good beginnings*. Available online at: <http://www.naeyc.org/files/naeyc/file/positions/psmath.pdf>. Retrieved on November 06, 2018
- NCTM- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM. Retrieved November 06, 2018, from <http://standards.nctm.org/>
- Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., & Spitler, M. E. (2016). Which

- preschool mathematics competencies are most predictive of fifth grade achievement?. *Early childhood research quarterly*, 36, 550-560. <https://doi.org/10.1016/j.ecresq.2016.02.003>
- Olkun, S., Fidan, E., & Babacan Ozer, A. (2013). The development of number concept and the use of counting in problem solving of 5-7 Year olds. *Education and Science*, 38(169), 235-248.
- Onkol, F. L. (2012). *The adaptation of the Early Numeracy Test and investigation of the effects of the Early Number Development Programme on six years old children*. Unpublished unpublished doctoral thesis, Marmara University, Institute of Educational Sciences, Istanbul.
- Pekince, P., & Daglıoğlu, H. E. (2017). Validity and reliability study of counting principles scale. *Elementary Education Online*, 16(2), 765-778. <https://doi.org/10.17051/ilkonline.2017.304733>
- Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37, 239-253. <https://doi.org/10.1016/j.ssresearch.2007.06.012>
- Pittalis, M., Pitta-Pantazi, D., & Christou, C. (2015). The development of student's early number sense. In *CERME 9-Ninth Congress of the European Society for Research in Mathematics Education* (pp. 446-452). <https://hal.archives-ouvertes.fr/hal-01286931/document> Retrieved on November 06, 2018
- Ramani, G. B., Siegler, R. S., & Hitti, A. (2012). Taking it to the classroom: Number board games as a small group learning activity. *Journal of Educational Psychology*, 104, 661-672. <https://doi.org/10.1037/a0028995>
- Reys, R. E., & Yang, D. C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education*, 29(2), 225-237. <https://doi.org/10.2307/749900>
- Reys, R., Reys, B., McIntosh, A., Emanuelsson, G., Johansson, B., & Yang, D. C. (1999). Assessing number sense of students in Australia, Sweden, Taiwan, and the United States. *School Science and Mathematics*, 99(2), 61-70. <https://doi.org/10.1111/j.1949-8594.1999.tb17449.x>
- Romano, E., Babchishin, L., Pagani, L. S., & Kohen, D. (2010). School readiness and later achievement: Replication and extension using a nationwide Canadian survey. *Developmental Psychology*, 46(5), 995-1007. <https://doi.org/10.1037/a0018880>
- Sengul, S., & Gulbagcı, D. H. (2013). An investigation of classification of number sense components. *The Journal of Academic Social Science Studies*, 6(8), 645-645.
- Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition*, 168, 222-233. <https://doi.org/10.1016/j.cognition.2017.07.004>
- Uyanık- Aktulun, Ö. (in press). Validity and reliability study for the Turkish version of number sense screener for 72-84 months old children. *Journal of Education and Training Studies*.
- Yılmaz, Z. (2012). Let's open the math box of our childhood together [Çocukluğumuzun matematik kutusunu beraber açalım]. *Monthly Children, Mother, Father and Educator Journal [Çocuk Çocuk Aylık Anne Baba Eğitimci Dergisi]*, 94, 8-10.
- Yılmaz, B., & Inal, K. G. (2017). Validity and reliability study of early numeracy assessment for 48-60 month children. *Adnan Menderes University Education Faculty Journal of Educational Sciences*, 8(2), 81-91.
- Yurdugul, H. (2005). Use of content validity index for content validity in scale development studies [Ölçek geliştirme çalışmalarında kapsam geçerlilik için kapsam geçerlik indekslerinin kullanılması]. XIV. National Educational Sciences Congress Pamukkale University Faculty of Education 28-30 September. Available online at: <http://yunus.hacettepe.edu.tr/~yurdugul/3/indir/PamukkaleBildiri.pdf>. Retrieved on: November 12, 2018.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.