

Validity and Reliability Study of Turkish Version of Number Sense Screener for Children Aged 72-83 Months

Özgün Uyanık Aktulun

Correspondence: Özgün Uyanık Aktulun, Faculty of Education, Afyon Kocatepe University, Afyonkarahisar, Turkey.

Received: December 5, 2018

Accepted: December 26, 2018

Online Published: January 3, 2019

doi:10.11114/jets.v7i2.3935

URL: <https://doi.org/10.11114/jets.v7i2.3935>

Abstract

The objective of this study is to perform a validity and reliability examination of Number Sense Screener (NSS) by applying it to Turkish children aged 72-83 months. The universe of this general survey model study was constituted by 72-83-month old children showing normal development and studying at primary schools under the authority of the Ministry of National Education in central Afyonkarahisar province, Turkey. In sample selection, primary schools with different socioeconomic levels (low, medium, and high) were determined by taking into account the representation capability of the universe in Afyonkarahisar province, Turkey, and 672 children were selected from these schools through systematic sampling method. Validity and reliability studies were carried out using data from the NSS application. In light of the performed analyses, the difficulty and discrimination indexes of the items of NSS have been determined to range between 0.38 and 0.97, and 0.16 and 0.64, respectively. The reliability coefficients calculated for the whole of NSS have been determined to be between .876 and .884. The infit and outfit statistics of the items of NSS have been concluded to be extremely good. In DIF analysis performed in order to determine whether the items of NSS are gender-biased, the scale items have been concluded to be unbiased.

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

1. Introduction

Children's success in the field of mathematics is among the primary goals of education. Gaining mathematical concepts and skills in up-to-date national and international educational programs is becoming more and more crucial. High mathematical success serves as a precondition especially to work in jobs requiring technological proficiency and sustain other daily activities qualitatively (Mazzocco & Thompson, 2005). National Science Board (2003) has stated that success in the fields of mathematics and science is a must to work in well-paid jobs. At this juncture, number sense is among the concepts used to define mathematical success since it constitutes the basis for earning mathematical concepts and skills (Gersten & Chard, 1999; Jordan, Glutting & Ramineni, 2010a). According to Reys (1994), number sense is a way of thinking and one of the crucial skills in teaching and learning mathematics.

In the body of literature, number skills within the scope of mathematics have been seen to be stated mainly as number sense (Gersten, Jordan & Flojo, 2005; Charlesworth & Lind, 2007). Numerous efforts to argue about and define the concept of number sense and to constitute its components have been made by educators of mathematics (teachers, program developers, and researchers) and cognitive psychologists (McIntosh, Reys & Reys, 1992; Berch, 2005; Gersten et al., 2005). Therefore, instead of a mutual classification of the components of number sense, there are several different classifications and definitions in the body of literature. Berch (1998), for example, defined number sense as "a child's fluidity and flexibility with numbers, the sense of what numbers mean, and an ability to perform mental mathematics and to look at the world and make comparisons" (as cited in Gersten & Chard, 1999, p. 19). Kalchman, Moss, and Case (2001) defined number sense as fluency in estimating and judging magnitude; the ability to recognize unreasonable results; flexibility when mentally computing; the ability to move among different representations and to use the most appropriate representation for a given situation; and the ability to represent the same number or function in multiple ways, depending on the context and purpose of this representation. Berch (2005) defines number sense as the possessed sense with regard to the meaning of numbers. According to Berch, number sense is classified into awareness, intuition, recognition, knowledge, skill, ability, feel, process, conceptual structure, and mental activities (Jordan, Glutting & Dyson, 2012). Although it is defined varyingly within the body of literature, number sense can generally be regarded as one's understanding of numbers, ways of representing numbers, and relationships among numbers. A well-developed number sense promotes skills such as fluency in estimation and magnitude comparison, greater ease and flexibility in computation, and the ability to recognize

unreasonable results (Kalchman et al., 2001). In addition, two dimensions of number sense are underlined in the body of literature; preverbal and symbolic number knowledge (Jordan & Levine, 2009). The preverbal component as the basis of numerical cognition includes “precise representation of small numbers (3 or less) and approximate representation of larger numerosities” (Feigenson, Deheane & Spelke, 2004, p. 310; Jordan & Levine, 2009, p. 61). For instance, children aged 3 or 4 can compare two small numbers and find out the larger and smaller one (Gersten & Chard 1999). The symbolic number component involves “verbal subitizing, counting, numerical magnitude comparisons, linear representations of number, and arithmetic operations”. This component is considered as a complex set of skills that more likely develop within instruction and experiences (Jordan & Levine, 2009, pp. 61–62).

Even though it shows a slow development in the first seven years of life, number sense starts to develop beginning from babyhood. Therefore, children develop some meanings regarding numbers even before learning to count (Deheane, 1997). The National Council of Teachers of Mathematics [NCTM] (2000) defined number sense as one of the content standards of education of mathematics. Number sense is a crucial factor in using mathematical information fluently, developing arithmetic calculations (Gersten & Chard, 1999), gaining basic mathematical knowledge, and solving simple and complex mathematical problems (Jordan et al., 2010a). It has been stated that children with number sense understand the meaning of numbers better, can develop multiple correlations between numbers, recognize the relative magnitudes of numbers, understand the effects of operations on numbers, and can establish comparison (reference) points for measuring surrounding objects (NCTM, 2000). A child with developed number sense skills can discover his/her own methods while performing numerical operations, express a number through several ways in accordance with a given situation, and switch between quantities in the real world and numbers in the mathematical world easily. Besides, he/she can make comparisons in terms of quantity by developing strategies to solve complex mathematical problems (Berch, 2005). McIntosh et al. (1992) expressed that a person with a developed number sense thinks about numbers, operations, and emerging results, and examines them thoroughly.

Children start to earn such skills during early childhood as a result of informal or formal interactions with their environment (Ginsburg, Lee & Boyd, 2008; National Mathematics Advisory Panel [NMAP], 2008). In this process, gaining number sense may be related to factors like the socioeconomic status of the child’s family (Jordan, Huttenlocher & Levine, 1994; Jordan, Kaplan, Olah & Locuniak, 2006), the motivation of the child, his/her family, and the society they live in towards mathematics (Reys et al., 1999; Howse, Lange, Farran & Boyles, 2003), the child’s gender (Penner & Paret, 2008; Ivrendi, 2011), the child’s age and the educational background of the mother (Ivrendi, 2011). Therefore, not every child starts school with the same levels of number sense skills (Gersten & Chard, 1999). National Mathematics Advisory Panel (NMAP, 2008) has emphasized the critical importance of gaining number sense early in terms of primary school mathematical skills. Students having problems with gaining mathematical skills at primary school can be said to have less developed number sense skills compared to their successful peers (Gersten & Chard, 1999; Berch, 2005). Upon examining related studies, we can see that the primary school mathematical success is linked to pre-school period number sense development (Jordan, Kaplan, Ramineni & Locuniak, 2009; Libertus, Feigenson & Halberda, 2011-2013), while secondary school mathematical success is connected to primary school period number sense development (Halberda, Mazocco & Feigenson, 2008). This can have negative effects on mathematical skills that children gain cumulatively throughout their lives (Gersten, et al., 2005). Therefore, studies to be carried out about number sense in early childhood and early intervention endorsements based on the results of such studies are of vital importance especially in terms of number sense developments of children at risk (Griffin, Case & Siegler, 1994; Bowman, Donovan & Burns, 2001; Denton & West, 2002; Griffin, 2004; Aunio, Hautamäki & Van Luit, 2005; Jordan et al., 2006).

There are various studies carried out with children aged varyingly and from different grades frequently emphasizing on the importance of number sense in early childhood in the international literature (Malofeeva, Day, Saco, Young & Ciancio, 2004; Gersten et al., 2005; Jordan, et al., 2006; Jordan et al., 2009; Jordan, 2010a). Judging from the studies performed in Turkey on number sense of children, it can be seen that the concept has started to be recognized only recently and the studies are limited (Harc, 2010; Kayhan Altay, 2010; Ivrendi, 2011; Sengul & Gulbagci, 2012; Inal Kızıltepe, 2018). Number sense skills should be assessed early by using various scale instruments in different environments in order to make regulations enabling the development of children’s number sense and increase success levels beginning from early childhood. Various scale instruments are seen to be used for assessing children’s number sense skills in the international literature (Okamoto & Case, 1996; Van de Rijt, Van Luit & Pennings, 1999; Jordan et al., 2006; Moomaw, 2008; Zhao, 2006). One such scale instrument is Number Sense Screener (NSS) (Jordan et al., 2012). Number Sense Screener is crucial in that it is a psycho-educative instrument providing researchers to catch at-risk students early, predict later mathematical achievement, start planning effective intervention programs, and monitor progress (Jordan et al., 2012). It can be seen in the body of literature that Number Sense Screener has been used in various studies, such as by Szkudlarek & Brannon (2018) to examine the effects of the arithmetic education program attended by pre-school children with low success levels on number and letter recognition and specific mathematical

skills; by Starr, DeWind & Brannon (2017) to analyze the contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement; by Jordan & Dyson (2016) to develop the numerical proficiencies of children and adults at risk with the Number Sense Intervention Project; by Dyson, Jordan, Beliakoff & Hassinger-Das (2015) to examine the effects of research-based number sense interventions applied to pre-school children with low success levels; and by Jordan, Glutting, Ramineni & Watkins (2010b) and by Jordan, Kaplan, Locuniak & Ramineni (2007) to analyze and assess pre-school and primary school period number sense success as the predictor of third-grade primary school number sense success. It has been concluded that in these studies, number sense is assessed during pre-school and primary school periods; number sense successes, especially in kindergarten and first-grade primary school, predict children's later mathematical achievements; and number sense-based educational programs implemented during these periods are effective.

In the Turkish body of literature, it can be seen that there are related scale instruments limited in number, such as Number Sense Test about Decimal Numbers (NSTDN) developed by Sengul & Gulbagcı (2012) adapted for fifth-grade primary school students to directly assess the number sense skills in children during early childhood; Assessing Number Sense Instrument developed by Wakefield & Ivrendi (2008) (as cited in Ivrendi, 2011); and Number Sense Screener adopted by Inal Kızıltepe (2018) for Turkish children aged 60-71 months. In this context, measuring number sense skills with valid and reliable instruments is thought to be beneficial to assess and support the number sense skills of children in the early period, and increase the predictability of children's mathematical successes beginning from the early period. Therefore, the purpose of this study is to introduce a scale instrument that can enable longitudinal studies to be performed to the body of literature by adopting Inal Kızıltepe's (2018) version of Number Sense Screener for Turkish children aged 72-83 months and performing a validity and reliability examination on it.

2. Method

This research is a general survey model study. Survey models are studies that determine the views or features such as interests, skills, abilities, attitudes etc. of participants regarding an issue or an incident, generally carried out over larger numbers of samples than other studies (Buyukozturk, Kılıc Cakmak, Akgun, Karadeniz & Demirel, 2012).

2.1 Research Sample

The universe of the study was constituted by children aged 72-83 months studying in public and private primary schools under the authority of the Ministry of National Education in central Afyonkarahisar province, Turkey, during the fall semester of 2017/2018 academic year. In sample selection, primary schools with different socioeconomic levels (low, medium, and high) were determined by taking into account the representation capability of the universe, and which students studying at these schools would be included in the sample was determined through *systematic sampling*. Thanks to systematic sampling, the selection of participants was not performed based on any feature, which enabled the validity of the study to increase. Systematic sampling is a method in which a starting point (x) from a population of elements is picked, from which onward every "x" element is selected (Cingi, 1994). To that end, a list of classes in which children aged 72-83 months studied was obtained from each of the sample schools. In accordance with these lists, the assessment instrument was applied to children picked based on number three (third, sixth, ninth...) which had been selected as the starting point. In the case that a child selected through systematic sampling method could not present a parent permission form, the scale instrument was applied to the next child on the list.

672 children studying at primary schools were selected as samples through the systematic sampling method. Among these children, 342 were female (50.9%), and 330 were male (49.1%). The study was performed at schools that voluntarily participated in accordance with legal permissions obtained from Afyonkarahisar Provincial Directorate of National Education. In addition, permissions from the participant children's parents were obtained through a form developed by the researcher.

2.2 Data Collection Tool

Data from the study were collected with "Number Sense Screener (NSS)" which is composed of six subtests.

2.2.1 Number Sense Screener (NSS)

Number Sense Screener is a shortened form version of the research-based Number Sense Brief scale instrument (composed of 33 items) developed in order to assess the numerical proficiencies of kindergarten and first-grade primary school students (e.g., Jordan et al., 2006; Jordan et al., 2007). NSS, which was analyzed in terms of validity and reliability within the scope of this study, is composed of six subtests and 29 items in total (Jordan et al, 2012). The duration of the test ranges between 20 and 25 minutes. Information on the contents of the subtests of Number Sense Screener has been given below:

Counting Skills: This subtest, containing counting principles (one-to-one correspondence, cardinality, and stable order) and rhythmic counting (counting rhythmically up to a given number), is composed of three items (Jordan et al., 2012).

Number Recognition: Children are asked to name the displayed numbers (such as 13 or 37). This subtest is composed of four items in total (Jordan et al., 2012).

Number Comparisons: This subtest has been adapted from Griffin (2002). Given a number (e.g., 7), children are asked what number comes right after or two numbers after that number. Given two numbers (e.g., 5 and 4), children are asked which number is the bigger or the smaller one. Children are also shown visual arrays of three numbers (e.g., 6, 2, and 5), each placed on the point of an equilateral triangle. They are then asked to identify which number is closer to the target number (e.g., 5) at the triangle's apex. The subtest is composed of seven items in total (Jordan et al., 2012).

Nonverbal Calculation: This subtest has been adapted from Levine et al. (1992). The items in this subtest are asked to children using a white mat, a cartoon 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 buttons of the same size. To give an example of addition, the examiner places two buttons on the mat in front of the children and says "See? Here are two buttons", and then puts the buttons in the box through the opening on its side after allowing the children to observe. After that, the examiner places another button on the mat and says "Here is another button here" asking the children to watch carefully, and puts it in the same box through the opening on its side. Then the examiner opens the page related to the item on the scale instrument and asks the children to point to the set that has the same number of items hiding under the box. There are four items in total, three for addition and one for subtraction, in the subtest (Jordan et al., 2012).

Story Problems: In the items in this subtest, children are optionally asked to use their fingers, the number list (given along with the scale instrument), or a pen and paper to find the answer. The addition problems are phrased simply following this basic format: "Sue has m pennies. Jim gives her n more pennies. How many pennies does Sue have now?" Similarly, the subtraction problems are phrased, "Sue has m pennies. Jim takes away n of her pennies. How many pennies does Sue have now?" The subtest is composed of five items in total, three for addition and two for subtraction (Jordan et al., 2012).

Number Combinations: In this subtest, children are likewise optionally asked to use their fingers, the number list, or a pen and paper to find the answer. The Number Combinations subtest is composed of six items containing four addition and two subtraction questions phrased as "How much is m and n " and "How much is n takeaway m ?" (Jordan et al., 2012).

2.3 Data Collection Process

The test was applied in accordance with the rules on the application manual, and to each child individually by the researcher or one of the four testers who had been trained by the researcher about the scale instrument. These four testers, selected among the undergraduate and graduate pre-school education students, were extensively trained by the researcher about the scale instrument and how it should be applied. Then, each tester was observed while applying the scale instrument to a child (72-83 months old) that was not a participant in the study, and the necessary feedback was given to the testers in accordance with the notes taken. In addition, the testers were divided into two groups; while one of the testers was applying the scale instrument, the other was asked to give grades to the items on the scale instrument based on the answers given by the child; whether both testers gave the same grade was checked. The test was performed in a quiet and comfortable place at the schools that the sample children studied during the application process, placing emphasis on establishing positive relationships with the children. The testers started the test with a sample question in order to enable students to feel relaxed and have an idea about the actual test. The answers on the subtests were assessed as true (1 point), false, or no answer (0 points for either). The total point in each subtest was calculated by summing up the total number of correct answers in that subtest, while the total number sense point was obtained by summing up the points of all the subtests (Jordan et al., 2012). Accordingly, the maximum point to be obtained on the scale instrument is 29. The children who did not want to answer the test were not included in the application. Firstly, a pre-application was carried out by applying the test to 6 children in total from different income groups (2 from low, 2 from middle, and 2 from high income levels) selected among the universe through random sampling in order to determine the capability of the scale instrument to properly scale a concept abstracted from the behavior context desired to be scaled. Then, the scale instrument was applied to 672 children included in the sample group in the fall semester of the academic year.

2.4 Data Analysis

The content validity of Number Sense Screener was determined through Lawshe's method in accordance with expert opinions, and its content validity ratios (CVR) and content validity indexes (CVI) were calculated accordingly. The reliability and validity analyses of the scale instrument used within the context of the study were performed using jMetrik 4.1.1 software. The acceptable ranges for the criteria to be used for unweighted and standardized fit statistics produced by Linacre (2002) in assessing the results from the aforementioned software are shown in Table 1.

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

Criteria	Range	Comments
WMS and UMS	> .20	Makes the scale distort the features of the subject matter and degrade the quality of the scale.
	1.5 – 2.0	Unproductive for construction of measurement, but not degrading.
	0.5 – 1.5	Productive for measurement.
Std WMS and Std UMS	<.50	Less productive for measurement, but not degrading. May produce misleadingly high reliability and separation coefficients.
	≥ 3	Data very unexpected if they fit the model (perfectly), so they probably do not. But, with large sample size, substantive misfit may be small.
	2.0 – 2.9	Data noticeably unpredictable.
	-1.9 – 1.9	Data have reasonable predictability.
	≤ -2.0	Data are too predictable. Other "dimensions" may be constraining the response patterns.

Judging from Table 1, the most ideal range for WMS and UMS values has been determined to be between 0.50 and 1.50. Likewise, it can be said that the most ideal range for standardized WMS and UMS values is between -1.90 and 1.90. *WMS* and *Std WMS* columns indicate weighted and standardized mean square misfit statistics respectively, followed by *UMS* and *Std UMS* which indicate unweighted and standardized mean square outfit statistics respectively (Guzeller, Eser & Aksu, 2018).

3. Results

In this section, information on the results of the validity and reliability study of Number Sense Screener (NSS) for Turkish children aged 72-83 months is presented.

In the first stage for content validity, NSS was translated into Turkish with the original instruction manual in English. The forms related to NSS (the instruction manual and the scale instrument) were translated into Turkish and then back to English through back translation method by three linguists. Then, the forms were examined by an expert having comprehensive knowledge of Turkish and English languages, who also checked the expression and content integrity between both forms. The forms translated into Turkish were analyzed by a Turkish linguistics expert and were finalized following the necessary corrections. For expert opinion, the scale instrument was sent to seven domain experts (from the departments of pre-school education, classroom instruction education, mathematics education, and educational sciences) performing their respective duties at various universities, who were then asked to assess the suitability of NSS in terms of the expression of its instruction and test items, Turkish culture, its capability to assess 72-83-month-old children's number sense, and whether its subtests are in accord with their respective fields.

The items that were deemed suitable unanimously by the experts were included in the Turkish version of the scale instrument as they were, while the items that were recommended to be corrected were rearranged. Accordingly, the number "four" was decided to be expressed as the way it is more frequently used in Turkish (not 4, but 4).

Data obtained from expert opinions were analyzed according to Lawshe's method in order to determine the capability of the scale instrument to stably represent the defined universe of the content it aimed to measure or its specific fields, and whether it was quantitatively and qualitatively efficient enough to measure the behavior (feature) aimed to be measured (Colton & Covert, 2007). The content validity indexes regarding the subtests and the whole of NSS are shown in Table 2.

Table 2. The content validity indexes regarding the subtests and the whole of NSS

Number Sense Screener	The Number of Items	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

CVR and CVI values of the subtests and the whole of the scale instrument according to Lawshe's method were calculated to be 1.00. Accordingly, it was decided that each item should be included in the scale, and the test was acknowledged to have content validity.

Descriptive statistics regarding the NSS items (item difficulty and discrimination values, and standard deviation values) are shown in Table 3.

Table 3. Descriptive statistics regarding the NSS items

Item No	Item Difficulty	Std Deviation	Discrimination Index
I1	0.97	0.15	0.17
I2	0.94	0.22	0.19
I3	0.93	0.25	0.16
I4	0.96	0.19	0.41
I5	0.83	0.37	0.59
I6	0.80	0.40	0.52
I7	0.38	0.48	0.35
I8	0.90	0.29	0.37
I9	0.65	0.47	0.25
I10	0.93	0.25	0.34
I11	0.90	0.29	0.53
I12	0.91	0.29	0.50
I13	0.91	0.27	0.44
I14	0.78	0.41	0.33
I15	0.93	0.25	0.25
I16	0.91	0.28	0.25
I17	0.78	0.41	0.31
I18	0.91	0.27	0.22
I19	0.88	0.31	0.47
I20	0.68	0.46	0.51
I21	0.77	0.41	0.44
I22	0.66	0.47	0.52
I23	0.78	0.41	0.45
I24	0.90	0.28	0.52
I25	0.83	0.37	0.57
I26	0.77	0.41	0.64
I27	0.80	0.39	0.47
I28	0.77	0.41	0.54
I29	0.76	0.42	0.59

From Table 3, it has been determined that the item difficulty values range between 0.38 and 0.97, while the discrimination indexes range between 0.16 and 0.64.

Reliability values obtained for each item on the scale instrument through various reliability determination methods are shown in Table 4.

Table 4. Reliability values for NSS items

Item No	Guttman's lambda-2	Alpha coefficient	Feldt-Gilmer coefficient	Feldt-Brennan coefficient	Raju's beta coefficient
I1	0.8843	0.8768	0.8829	0.8823	0.8768
I2	0.8844	0.8767	0.8831	0.8826	0.8767
I3	0.8851	0.8774	0.8837	0.8832	0.8774
I4	0.8820	0.8738	0.8806	0.8799	0.8738
I5	0.8766	0.8681	0.8751	0.8745	0.8681
I6	0.8784	0.8699	0.8770	0.8764	0.8699
I7	0.8841	0.8755	0.8826	0.8820	0.8755
I8	0.8822	0.8737	0.8808	0.8801	0.8737
I9	0.8869	0.8783	0.8855	0.8850	0.8783
I10	0.8825	0.8742	0.8810	0.8804	0.8742
I11	0.8793	0.8706	0.8777	0.8771	0.8706
I12	0.8799	0.8712	0.8784	0.8778	0.8712
I13	0.8809	0.8724	0.8794	0.8788	0.8724
I14	0.8837	0.8750	0.8823	0.8817	0.8750
I15	0.8839	0.8759	0.8826	0.8820	0.8759
I16	0.8841	0.8760	0.8827	0.8821	0.8760
I17	0.8841	0.8755	0.8828	0.8821	0.8755
I18	0.8845	0.8765	0.8833	0.8827	0.8765
I19	0.8802	0.8715	0.8787	0.8780	0.8715
I20	0.8787	0.8701	0.8770	0.8764	0.8701
I21	0.8808	0.8721	0.8793	0.8786	0.8721
I22	0.8780	0.8697	0.8764	0.8758	0.8697
I23	0.8803	0.8716	0.8787	0.8781	0.8716
I24	0.8794	0.8708	0.8779	0.8773	0.8708
I25	0.8771	0.8686	0.8755	0.8750	0.8686
I26	0.8745	0.8664	0.8729	0.8724	0.8664
I27	0.8797	0.8712	0.8782	0.8776	0.8712
I28	0.8774	0.8691	0.8759	0.8754	0.8691

It has been concluded from Table 4 that all of 29 items on the scale instrument have reliability coefficients above .80, which is regarded to be a critical point, according to 5 different methods of reliability determination. Accordingly, the items on the scale instrument have been acknowledged to meet the criteria of internal consistency reliability.

Reliability coefficients calculated for the whole of the scale and a 95% confidence interval along with standard error values regarding the reliability coefficients are shown in Table 5.

Table 5. Reliability analysis results regarding the whole of NSS

Method	Coefficient	Confidence interval lower bound	Confidence interval upper bound	SEM
Guttman's L2	0.8847	0.8718	0.8969	1.6639
Coefficient Alpha	0.8766	0.8627	0.8896	1.7216
Feldt-Gilmer	0.8833	0.8702	0.8956	1.6742
Feldt-Brennan	0.8827	0.8695	0.8951	1.6782
Raju's Beta	0.8766	0.8627	0.8896	1.7216

It has been concluded from Table 5 that the reliability coefficients determined through different methods for the whole of the scale instrument composed of 29 items range between .876 and .884. Accordingly, the results obtained from the scale instruments are acknowledged to be reliable.

In contemplation of the possibility that looking only at the item difficulty and discrimination indexes would not be sufficient while determining the items to be included in the scale instrument within the scope of validity and reliability analyses, converted discrimination indexes and fit indices regarding the items were calculated through Rasch Analysis, one of the one-parameter logistic models. The item statistics obtained based on the parameters determined as 150 for maximum iteration number, 0.005 for convergence criteria, and 0.3 for regulation of extreme values in Rasch Analysis are shown in Table 6.

Table 6. Rasch analysis results regarding the NSS items

Item No	Difficulty	Std Error	WMS	Std WMS	UMS	Std UMS
I1	-2.20	0.25	0.98	-0.04	1.16	0.46
I2	-1.35	0.18	1.08	0.66	1.64	1.75
I3	-1.02	0.16	1.20	1.65	1.75	2.31
I4	-1.73	0.21	0.81	-1.26	0.37	-1.99
I5	0.21	0.12	0.81	-2.87	0.60	-3.22
I6	0.53	0.11	0.93	-1.05	0.82	-1.53
I7	3.24	0.10	1.09	2.02	1.31	2.66
I8	-0.61	0.15	1.00	0.00	1.17	0.79
I9	1.54	0.10	1.35	6.83	1.39	4.77
I10	-1.02	0.16	0.99	-0.01	1.20	0.77
I11	-0.65	0.15	0.82	-1.90	0.46	-2.93
I12	-0.70	0.15	0.85	-1.57	0.66	-1.56
I13	-0.82	0.16	0.90	-0.98	0.68	-1.39
I14	0.62	0.11	1.20	3.19	1.24	1.97
I15	-1.04	0.17	1.10	0.87	1.26	0.96
I16	-0.74	0.15	1.16	1.57	1.21	0.90
I17	0.61	0.11	1.22	3.34	1.15	1.27
I18	-0.79	0.15	1.15	1.39	1.67	2.37
I19	-0.42	0.14	0.91	-1.05	1.09	0.48
I20	1.32	0.10	0.96	-0.79	0.93	-0.84
I21	0.69	0.11	1.05	0.84	1.09	0.82
I22	1.49	0.10	0.93	-1.50	0.90	-1.32
I23	0.64	0.11	1.02	0.38	1.11	0.96
I24	-0.66	0.15	0.83	-1.79	0.48	-2.75
I25	0.23	0.12	0.83	-2.56	0.82	-1.32
I26	0.67	0.11	0.77	-4.18	0.67	-3.35
I27	0.49	0.11	0.99	-0.12	0.90	-0.77
I28	0.66	0.11	0.90	-1.72	0.80	-1.92
I29	0.80	0.11	0.84	-2.96	0.78	-2.24

The unweighted mean square (UMS) and weighted mean square (WMS) values shown in Table 6 represent fit statistics for the items. While WMS is regarded as infit criteria, UMS is regarded as outfit criteria. It can be seen from Table 6 that the WMS and UMS values of the scale instrument composed of 29 items have acceptable ranges for measurement

(Linacre, 2002). Although the UMS values of the items no 2 and 3 are above the critical point of 1.50, the fit statistics of the items have been concluded to be extremely good due to 1.50 – 2.00 range being regarded as acceptable (Linacre, 2002). Based on standardized WMS values, it has been determined that the items no 5, 9, 14, 17 and 26 do not provide a model-data fit, while the other items do so. From standardized UMS values, it can be seen that among the items not providing a model-data fit, only no 5, 9 and 26 constitute a problem.

DIF analysis was performed to determine whether the items on the scale instrument were gender-biased. The results obtained based on common odds ratio values through the Mantel-Haenszel method are shown in Table 7.

Table 7. DIF analysis results

Item No	Chi-square	p-value	Valid N	Effect size	Class
I1	0.14	0.71	345	0.82	A
I2	0.03	0.86	541	1.07	A
I3	3.73	0.05	545	0.54	A
I4	0.12	0.73	196	1.19	A
I5	1.47	0.23	392	0.73	A
I6	12.43	0.00	567	0.43	B+
I7	16.44	0.00	522	0.44	B+
I8	3.58	0.06	564	1.84	A
I9	5.4	0.02	582	0.65	A
I10	3.57	0.06	571	1.92	A
I11	0.3	0.59	400	0.83	A
I12	1.47	0.23	397	1.52	A
I13	5.77	0.02	402	2.3	B-
I14	0.01	0.93	582	1.02	A
I15	0.39	0.53	568	1.23	A
I16	0.44	0.51	577	1.23	A
I17	4.24	0.04	582	1.55	B-
I18	1.47	0.23	553	1.47	A
I19	0.61	0.44	493	1.27	A
I20	1.09	0.30	558	0.8	A
I21	2.28	0.13	579	1.39	A
I22	0.01	0.93	553	1.02	A
I23	1.63	0.20	580	1.34	A
I24	4.18	0.04	360	0.51	B+
I25	0.01	0.93	560	0.98	A
I26	0.04	0.84	558	1.05	A
I27	1.08	0.30	584	0.78	A
I28	5.28	0.02	480	1.75	B-
I29	5.02	0.03	551	1.74	B-

From Table 7, it has been determined that differential item functioning in all the items except for 6, 7, 13, 17, 24, 28 and 29 are negligible. However, among the items determined to be moderately biased (B+/B-), while the items no 6, 7 and 24 function in favor of the focus group, the items no 13, 17, 28 and 29 function in favor of the reference group. Nevertheless, all 29 items on the scale instrument have been concluded to be unbiased since an item needs to represent at least C-level DIF values in order for it to be regarded as biased (Koyuncu, Aksu & Kelecioğlu, 2018).

4. Discussion

The objective of this study, carried out with a view to introducing a reliable scale instrument that can measure Turkish children's number sense to the body of literature, is to perform a validity and reliability analysis for Number Sense Screener, which is frequently used in the international literature, by adapting it for Turkish children aged 72-83 months. In the first place, for the content validity of the scale instrument translated into Turkish, the Turkish forms (instruction manual and the scale instrument) were sent to seven domain experts, whose opinions were assessed through Lawshe's method. Based on this method, the CVR and CVI values regarding the subtests and the whole of NSS were determined to be 1.00. Accordingly, each item was decided to be included in the scale instrument, and the scale instrument itself was acknowledged to have content validity.

Based on performed analyses, it has been determined that the difficulty values of the NSS items range between 0.38 and 0.97, while their discrimination indexes range between 0.16 and 0.64. Accordingly, each item on the scale instrument has been determined to be acceptable.

The item reliability coefficients for the items on the scale instrument, calculated through the methods of Guttman's lambda-2, Alpha coefficient, Feldt-Gilmer coefficient, Feldt-Brennan coefficient, and Raju's beta, were determined to be above the critical point of .80. Accordingly, the items on the scale instrument were acknowledged to meet the criteria of internal consistency reliability. In addition, the reliability coefficients calculated for the whole of the scale instrument

through the same methods were determined to range between .876 and .884. Accordingly, the results obtained from the scale instrument are regarded as reliable. The original validity and reliability analysis of Number Sense Screener was performed on the children in the same sample group throughout different periods of time. The scale instrument was applied to children during three different time periods, namely fall of kindergarten (mean age = 5.8, SD = 3.7), spring of kindergarten (mean age = 6.2, SD = 3.7), and fall of first-grade (mean age = 6.8, SD = 3.7). In consequence of performed analyses, the reliability coefficient for the whole of the scale instrument has been determined to be .82 for fall of kindergarten, .86 for spring of kindergarten, .87 for fall of first-grade, and .85 overall (Jordan et al., 2012). In a longitudinal validity and reliability research carried out by Jordan et al. (2010b) over six time periods spanning kindergarten and primary school mid-first grade, the test-retest reliability coefficient for the whole of Number Sense Brief (33 items) were determined to be between .61 and .86. These results bear similarities to the reliability coefficients obtained in order to adapt the scale instrument to Turkish children.

The infit and outfit values for the majority of the items of the Turkish version of NSS are seen to have acceptable range values for measurement (Linacre, 2002). Although the UMS values of the items no 2 and 3 are above the critical point of 1.50, the fit statistics of the items have been concluded to be extremely good due to 1.50 – 2.00 range being regarded as acceptable (Linacre, 2002). Also, the infit and outfit values have been determined to be between 0.50 and 1.50 based on the results of Rasch Analysis of the original scale instrument performed using Winsteps measurement software, except for the items no 1 and 3 (outfit of the item no 1 = 0.43, outfit of the item no 3 = 0.42) (Jordan et al., 2012).

The results from DIF analysis, performed in order to determine whether the NSS items were gender-biased, indicate that the items no 6, 7 and 24 are moderately biased with B+ levels, while the items no 13, 17, 28 and 29 are moderately biased with B- levels. An item needs to represent at least C-level DIF values in order for it to be regarded as biased towards a specific group (Koyuncu et al., 2018). Therefore, all 29 items on the Turkish version of the scale instrument can be concluded not to be gender-biased in favor of or to the detriment of a specific group. Also in the original form of NSS, the items were analyzed through Mantel-Haenszel method to determine whether they were gender-biased, and accordingly, the scale instrument was acknowledged to be gender-neutral since only one item was determined to have negligible differential item functioning (Jordan et al., 2012).

Also, based on the results from the validity and reliability analysis performed for the Turkish version of NSS adapted for children aged 60-71, it was determined that the reliability coefficients for the whole of the scale instrument ranged between .826 and .837, while the infit and outfit values were between 0.50 – 1.50. In this regard, it can be said that the fit statistics regarding the items on the scale instrument are between ideal ranges (Inal Kızıltepe, 2018).

In light of the results obtained from the study, the Turkish version of NSS has been concluded to be a valid and reliable scale instrument to measure 72-83-month-old children's number sense. Introducing a scale instrument to be used for measuring the number sense skills of children aged 72-83 months to the Turkish literature constitutes one of the strong aspects of this study. In addition, the validity and reliability of data obtained from NSS being proven through more than one analysis methods represent another strong aspect of the study. Besides these aforementioned strong aspects, the study has some limitations as well, which bring along various recommendations for future studies. In the first place, the results from this study have been obtained only from 672 children aged 72-83 months studying at primary schools in Afyonkarahisar province, Turkey. In this respect, it is thought that extending sample groups to span various other provinces in Turkey and retesting the validity and reliability of the scale instrument with new studies will provide important contributions to the scale instrument itself, and in turn the field. Considering the fact that number sense is related to demographic variables such as at-risk children and later mathematical achievement (Jordan & Levine, 2009; Jordan et al., 2010a; Dyson, Jordan & Glutting 2013), studies can be performed to determine what kind of a link exists between number sense and such variables. In addition, with intercultural studies, number sense levels of children can be compared through several versions of NSS in multiple languages. Conducting studies in which NSS is used is extremely important since it will contribute to the measurement capability of the scale instrument.

References

- Aunio, P., Hautamäki, J., & Van Luit, J. E. H. (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>
- 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>
- Bowman, B. T., Donovan, M. S., & Burns, M. S. (Eds.). (2001). *Eager to learn: Educating our preschoolers*. Washington, DC: National Academy Press.

- Buyukozturk, S., Kılıc, C. E., Akgun, Ö. E., Karadeniz, S., & Demirel, F. (2012). *Scientific research method*. Ankara: Pegem Akademi Publishing.
- Charlesworth, R., & Lind, K. K. (2007). *Math & science for young children* (Fifth edition). USA: Thomson Delmar Learning.
- Cıngı, H. (1994). *Sampling theory*. Ankara: Hacettepe University Faculty of Science Publications.
- Colton, D., & Covert, R. W. (2007). *Designing and constructing instruments for social research and evaluation*. San Francisco: John Wiley & Sons Inc.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York, NY: Oxford University Press.
- Denton, K., & West, J. (2002). *Children's reading and mathematics achievement in kindergarten and first grade*. National Center for Education Statistics (ED), Washington, DC. Retrieved from <https://files.eric.ed.gov/fulltext/ED461438.pdf>
- 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>
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314. <https://doi.org/10.1016/j.tics.2004.05.002>
- Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *The 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>
- Griffin, S. (2004). Building number sense with number worlds: a mathematics program for young children. *Early Childhood Research Quarterly, 19*(1), 173-180. <https://doi.org/10.1016/j.ecresq.2004.01.012>
- Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom lessons: integrating cognitive theory and classroom practice* (pp. 24–49). Cambridge, MA: MIT Press.
- Guzeller, C. O., Eser, M. T., & Aksu, G. (2018). *Measurement applications with JMetrik*. Ankara: Maya Akademik.
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature International Journal of Science, Nature 455*(7213), 665–668. <https://doi.org/10.1038/nature07246>
- Harc, S. (2010). Analysis of the current situation of the 6th grade students in terms of number sense concept. Unpublished Master Thesis. Marmara University, Institute of Educational Sciences, İstanbul.
- 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>
- Inal Kızıltepe, G. (2018). Validity and reliability study for the Turkish version of number sense screener for 60-71 months old children. *Journal of Education and Training Studies, 7*(1), 24-35. <https://doi.org/10.11114/jets.v7i2.3888>
- Ivrendi, A. (2011). Influence of self-regulation on the development of children's number sense. *Early Childhood Educational Journal, 39*(4), 239–247. <https://doi.org/10.1007/s10643-011-0462-0>
- Jordan, N. C., & Dyson, N. (2016). Catching math problems early: Findings from the number sense intervention project. In Avishai Henik (ed.) *Continuous Issues in Numerical Cognition How Many or How Much*. (pp. 59-79). <https://doi.org/10.1016/B978-0-12-801637-4.00003-2>

- Jordan, N. C., & Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Developmental Disabilities Research Review*, 15(1), 60–68. <https://doi.org/10.1002/ddrr.46>
- Jordan, N. C., Glutting, J. J., & Dyson, N. (2012). *Number Sense Screener™ (NSS™) User's Guide, k–1, Research Edition*. Baltimore: Brookes Publishing.
- Jordan, N. C., Glutting, J., & Ramineni, C. (2010a). The importance of number sense to mathematics achievement in first and third grades. *Learning and Individual Differences*, 20(2), 82–88. <https://doi.org/10.1016/j.lindif.2009.07.004>
- Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010b). 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. Retrieved from: <https://eric.ed.gov/?id=EJ891847>
- Jordan, N. C., Huttenlocher, J., & Levine, S. C. (1994). Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children. *Learning and Individual Differences*, 6(4), 413–432. [https://doi.org/10.1016/1041-6080\(94\)90003-5](https://doi.org/10.1016/1041-6080(94)90003-5)
- 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., Oláh, L. N., & Locuniak, M. N. (2006). Number sense growth in kindergarten: A longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, 77(1), 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. <https://doi.org/10.1037/a0014939>
- Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for the development of mathematical understanding: Rational numbers and functions. In S. M. Carver & D. Klahr (Eds.), *Cognition and instruction* (pp. 1–38). Mahwah, NJ: Lawrence Erlbaum.
- Kayhan, A. M. (2010). An investigation of middle grade students number sense in terms of grade level gender and components of number sense. Unpublished PhD Thesis. Hacettepe University, Institute of Social Sciences, Ankara.
- 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>
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011) Preschool acuity of the approximate number system correlates with school mathability. *Developmental Science*, 14(6), 1292–1300. <https://doi.org/10.1111/j.1476-7687.2011.01080.x>
- Libertus, M. E., Feigenson, L., & Halberda, J. (2013) Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences*, 25, 126–133. <https://doi.org/10.1016/j.lindif.2013.02.001>
- Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. Retrieved from: <http://www.raschsig.org/rmt162.pdf>
- Malofeeva, E., Day, J., Saco, X., Young, L., & Ciancio, D. (2004). Construction and evaluation of a number sense test with Head Start children. *Journal of Educational Psychology*, 96(4), 648–659. <https://doi.org/10.1037/0022-0663.96.4.648>
- Mazzocco, M. M. M., & Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learning Disabilities Research & Practice*, 20(3), 142–155. <https://doi.org/10.1111/j.1540-5826.2005.00129.x>
- McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed frame work for examining basic number sense. *For The Learning of Mathematics*, 12(3), 2–44. Retrieved from: <https://eric.ed.gov/?id=EJ466000>
- Moomaw, S. C. (2008). Measuring number sense in young children. An Published Doctoral Dissertation. Education Faculty, University of Cincinnati, USA.
- National Council of Teachers of Mathematics (NCTM). (2000). *The principles and standards for school mathematics*. Reston, VA: Author. Retrieved from: <https://www.nctm.org/Standards-and-Positions/Principles-and-Standards/>

- National Mathematics Advisory Panel [NMAP] (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel*. Washington, DC: U.S. Department of Education. Retrieved from: <https://www2.ed.gov/about/bdscomm/list/mathpanel/report/final-report.pdf>
- National Science Board. (2003, August 14). The science and engineering workforce: Realizing America's potential. Retrieved from <http://www.nsf.gov/nsb/documents/2003/nsb0369/.pdf>
- Okamoto, Y., & Case, R. (1996). II. Exploring the microstructure of children's central conceptual structures in the domain of number. *Monographs of the Society for Research in Child Development*, 61(1-2), 27–58. <https://doi.org/10.1111/j.1540-5834.1996.tb00536.x>
- Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37(1), 239–253. <https://doi.org/10.1016/j.ssresearch.2007.06.012>
- Reys, B. J. (1994). Promoting number sense in middle grades. *Mathematics Teaching in the Middle School*, 1(2), 114-120. Retrieved from: <https://eric.ed.gov/?id=EJ500139>
- Reys, R. E., Reys, B. J., 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>
- Sengul, S., & Gulbagcı, H. (2012). An investigation of 5th grade Turkish students' performance in number sense on the topic of decimal numbers. *Procedia Social and Behavioral Sciences*, 46, 2289-2293. <https://doi.org/10.1016/j.sbspro.2012.05.472>
- 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>
- Szkudlarek, E., & Brannon, E. M. (2018). Approximate arithmetic training improves informal math performance in low achieving preschoolers. *Frontiers in Psychology*, 15(9): 606. <https://doi.org/10.3389/fpsyg.2018.00606>
- Van de Rijt, B. A. M., Van Luit, J. E. H., & Pennings, A. H. (1999). The construction of the Utrecht Early Mathematical Competence Scales. *Educational and Psychological Measurement*, 59(2), 289–309. <https://doi.org/10.1177/0013164499592006>
- Zhao, Z. (2006). The study on the development of number sense and estimation in young children aged 3-6. Unpublished master thesis, East China Normal University, Shanghai, China.

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](#) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.