International Journal of Learning, Teaching and Educational Research
Vol. 15, No. 12, pp. 48-66, November 2016

Psychometric Properties of a Screening Tool for Elementary School Student's Math Learning Disorder Risk

Sinan OLKUN
TED University,
Ziya Gokalp Cad. No: 48 Kolej 06420
Ankara, TURKEY

Arif Altun and Sakine Göçer Şahin
Hacettepe University, Beytepe,
Ankara, TURKEY

Galip Kaya
Havelsan, Mustafa Kemal Mh. 2120 Cad. No: 39 Çankaya-
Ankara, TURKEY

This study has been supported by a grant from the Scientific and Technological Research Council of Turkey (TUBİTAK) under the grant number 111K545.

Abstract. This study reports the psychometric properties of a Basic Number Processing Test (BNPT), which was developed in order to determine elementary school students at risk for Mathematical learning disorder. A total of 478 primary school students were selected from 12 different public schools with an attempt to get a representative sample. The reliability and validity of the Dyscalculia Screening Tool were assessed with approximately 120 students from each of the First to Fourth grade. Results showed that, except for the First grade, the test scores predict the significant portion of the student's curriculum based Math achievement scores for Second, Third and Fourth graders with having the highest variance in the Second grade. These findings indicate that BNPT could be used as a screening tool in order to determine the students at risk for Mathematics learning disorders in those grades. It could also be deduced that at least very important portions of the causes of low achievement in Mathematics might originate from either the core systems of number or the system for accessing numbers from symbols. It is also suggested that symbolic or non-reading measurement paradigms would be more appropriate for screening First graders.

Keywords: Basic number processing skills, Math learning disorder, low Math achievement, reliability, validity

1. Introduction

The prevalence rates for Math learning disorder reported in existing research range from 3-6% (Shalev, Auerbach, Manor, & Gross-Tsur, 2000) to 5-14% (Barbaresi, Katusic, Colligan, Weaver, & Jacobsen, 2005) for children at the ages between 6-11. The structure of Mathematics is highly hierarchical. Therefore, if no adjustments or intervention programs are offered for those who experience Math learning difficulties at early grades, it would be difficult for them to attend and benefit from regular classes with their peers. Middle school or high school levels would be too late for interventions. However, if diagnosed at an earlier age, these students would benefit from guidance or somewhat individualized education depending on their cognitive characteristics. In other words, this would let the development of suitable and supportive learning environments if those students, who are slow in learning Math or find learning Math difficult, were screened at earlier ages.

It is possible that the majority of those students who experience low achievement in Mathematics might have reasons different from their own cognitive characteristics. While a partial of those reasons in the low achieved group could be attributed to neuropsychological sources in nature, the other part could be a result of mere poor instruction (Butterworth, 2010) and poor emotional involvement (Auerbach, Gross-Tsur, Manor, & Shalev, 2008). Therefore, it is both an important and difficult task for educators to pinpoint the exact reasons underlying the low achievement in Mathematics.

Recent research indicates that Math learning difficulties carry some neuro-psychological reasons (Rosenberg-Lee et al., 2015). The findings point out that both in humans and animals there is a number module (NM) which carries out number processing (Butterworth, 2009). Different from animals, humans have a capacity for abstraction and transformation to symbols or to use symbolic representations for numbers and operations. The differences in capacity within this module, which had three subsystems (approximate number system, exact number system, and access to symbols), are thought to be the sources of individual differences in numerical cognition. Research also shows that one's success at those tasks designed according to the principles of these subsystems is strong predictors of Math achievement (Geary, Bailey, & Hoard, 2009). This system with its subsystems as a conceptual framework for this study will be described briefly below.

1.1 Conceptual Framework

It is argued that there is a system, onset at birth, responsible for numerical processes, (Feigenson, Dehaene, & Spelke, 2004), and is comprised mostly of three subsystems. One of them is responsible for approximate quantities (Approximate Number System (ANS)); the other deals with exact quantities (Exact Number System (ENS)). The Third subsystem, on the other hand, is known to be the system which enables an access to matching the symbols to the quantity or vice versa. Literature includes some evidences that these subsystems function independently from each other (Iuculano, Tang, Hall, & Butterworth, 2008). There is a possibility that any deficit or malfunction in either one or more of these subsystems leads to Math learning disorders.

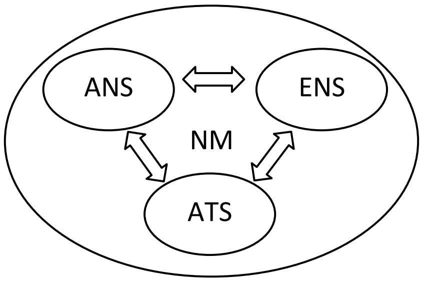


Figure 1: Components of Number Module (NM, Number module; ANS, Approximate Number System; ENS, Exact Number System; ATS, Access to Symbols)

Generally, number line (Dehaene, Izard, Spelke, & Pica, 2008; Van't Noordende, van Hoogmoed, Schot, & Kroesbergen, 2015) or analog quantity comparison tasks (Gimbert, Gentaz, Camos, & Mazens, 2016; Mussolin, Mejias, & Noël, 2010; Park & Starns, 2015) are administered in order to measure about the ANS. Subitizing, dot counting, and symbolic arithmetic calculations (Iuculano vd., 2008) are used for the measurement of the ENS. For accessing the symbols, on the other hand, either Stroop or an equiavilant number-symbol comparison tasks are used (Heine et al., 2010).

Students with Math learning disorder (MLD) have difficulties in subitizing, approximate numbers and estimation, comparing symbolically represented numbers, especially the numbers with close proximities (distance effect), and they use more primitive strategies and spend longer times in calculation than do their classmates. These students also make little transfers among number facts or rarely use derived number facts. If these students are identified earlier in their education, it could be possible to improve their number sense so that they might catch up with their peers. It would also be possible to make some adjustments during their assessment of Mathematical learning. The available screening tools are not comprehensive enough to assess the students' approximate number system, exact number system and access to symbols within a single tool that can be administered in relatively shorter period of time in computer environment. For these types of tasks, latency is more important than accuracy. In this respect, computer based tools are more reliable in collecting accurate item level latencies than paper and pencil based tools. The main purpose of this study was to develop a tablet PC based screening tool for determining Math learning disorder tendencies in children between the ages of 6 to 11.

2. Method

2.1 Participants

Participants were 478 primary school students gathered from 12 different public schools with an attempt to get a representative sample. Table 1 represents demographic characteristics of the sample.

Table 1. Participants' demographic characteristics

Grade	SES of School	Gender		Total
		Boys	Girls	
1st	Low	30	29	59
	Middle low	15	15	30
	Middle	15	15	30
	Total	60	59	119
2nd	Low	30	30	60
	Middle low	15	15	30
	Middle	15	15	30
	Total	60	60	120
3rd	Low	29	31	60
	Middle low	15	15	30
	Middle	14	16	30
	Total	59	61	120
4th	Low	30	29	59
	Middle low	15	15	30
	Middle	15	15	30
	Total	60	59	119
Grand Total		239	239	478

As seen from Table 1, there were approximately 120 students in each grade level from First to Fourth grade. There were nearly equal number of boys and girls in each grade level and in total. The number of students from each socio economic strata was also proportional to population. The information for the socio economic status of the schools was provided by the municipal body of the Ministry of Education. The selection procedure of the sample and the number of students from each grade both contributes to the generality of the results. The legal permission for data collection was also taken from the same authority. The ethical permission was taken from the ethical committee of Ankara University. The teachers in schools were very helpful in providing necessary information for the students.

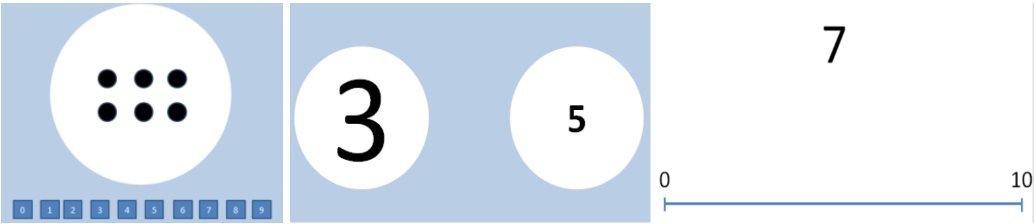
2.2 Data Collection Tools

The Mathematics Achievement Test (MAT): The test has been developed by Fidan (2013) based on the current Turkish State Curriculum (MEB, 2004). Each grade level has different achievement tests. The tests contained 13 items for the First, 15 items for Second, 16 items for Third, and 24 items for the Fourth grades. Open-ended, short answer form was used for all the questions. The MAT is an untimed test. The administration took one class hour (approximately 40 minutes) for the sample. Various methods were used to examine the content, construct and criterion referenced validity of the tests. Since all items were not equal in difficulty the reliability of the tests was estimated by KR-20 method for each grade level (Crocker & Algina, 1986). The reliability coefficients were 0.80, 0.92, 0.93, and 0.96 for the 1st, 2nd, 3rd, and 4th grade respectively.

The Calculation Performance Test (CPT): It was another test used for concurrent validity in this study. It is developed by De Vos (1992) and adapted by Olkun, Can, and Yeşilpınar (2013) into Turkish. It has five columns of basic arithmetic operations written in Arabic numerals and arithmetic operation symbols. Each column has 40 operations. Olkun et al. (2013) found the KR-20 coefficients as .95 and .98 for the timed and untimed administrations of CPT. Similar to the original application; the students were given one minute for each column. The main difference between MAT and CPT is that MAT has open-ended word problems while CPT has only arithmetic operations with Arabic numerals.

The Screening Test: It has three subtests; canonic dot counting (CDC), symbolic number comparison (SNC), and mental number line (MNL) test. It was individually administered and both students' reaction times and correct responses were recorded as data points while administering with tablet PCs.

The data were collected by trained research assistants being unaware of the study aims. Since the data were collected directly in the tablet PCs as the students individually answered questions immersed in custom software there was no need for manual coding or data entry. Therefore, there was no bias in data collection.



a) CDC Task b) SNC Task c) MNL task

Figure 2. Three sample tasks from the three sections in the screening tool

Raven Standard Progressive Matrices (SPM) Test: It was used to measure students' general reasoning abilities. There are 5 sets of a total of 60 diagrammatic puzzles with increasing difficulty in the test. Raven (Raven, 2000) reported that the test has good reliability and validity measures.

2.3 Procedures

Item Analysis: During the piloting phase, first the items, which did not function well, were determined by using ITEMAN software at the grade level. Item discriminations were calculated by bi-serial correlation. The results indicated that the lowest discrimination value for CDC for the First grade was .67, for Second grade it was .50; for Third graders only two items were .09, but the others were higher than .62; for the Fourth graders the lowest discrimination value was .25 and the rest were higher than .46.

The lowest discriminant value for SNC for the First grade was .20, whereas the other values were above .54; for Second grade, only two items (item 14 and 24) yielded negative values (-0.07 and -0.15 respectively), but the others were above .50; for the Third graders, the item 14 had a discriminatory value of -0.07, but the others were higher than .50; for the Fourth graders, the lowest discrimination value was .09 and the rest was higher than .50.

The lowest discrimination value for MNL1 for the First grade was .025, whereas the other values were above .35; for Second grade, only two items (item 1 and 8) yielded negative values (-0.04 and -0.024 respectively), but the others were above .50; for the Third graders, the item 9 had a discrimination value of 0.108, but the others were higher than .30; for the Fourth graders, the discrimination values fall between .36 and .684.

The discrimination values for MNL3 for the First grade was between .078 and 0.653; for Second grade, only one item (item 1) yielded a negative value (-0.187), but the others were above .30; for the Third graders, the values ranged between -0.233 and .646; for the Fourth graders, the discrimination values fall between -0.190 and .655.

MNL4 has been administered only to Third and Fourth graders. For the Third grade, only one item (item 9) yielded a negative value (-0.114), but the others were above 0.37; for the Fourth grade, again, only one item yielded a negative value (-0.117), but the others were above 0.30. The highest discrimination value for this grade was found to be 0.686.

Taking the item statistics and expert views into consideration, necessary changes have been made and the tests were made ready to be administered.

Reliability Results: Table 2 presents the reliability coefficients for the tests. For MNL1-3 and 4 are multi-graded tests Cronbach alpha coefficient was calculated; the other tests were binary, therefore, KR-20 reliability coefficient was calculated.

The results obtained through item-responses, CDC test, which had lower than .70, could be considered medium level reliability. MNL1, MNL3, and SNC had good reliability scores; MNL4, on the other hand, was a highly reliable test. When the time spent on tasks is taken into account, the reliability scores for MNL1 and MNL3 were good; CDC, SNC, and MNL4 had high reliability scores. Reliability for tests in which scores are used to make serious decisions is expected to be between 0.85 and 0.95; however, values equal to or higher than 0.50 for tests prepared by instructors may be deemed as reliable. From this perspective, it can be said that the tests developed for this research has adequate reliability (Frisbie, 1988).

Content Validity: In order to investigate the content validity, an expert panel was formed and item analyses had been performed. In this study, content validity was assessed through expert opinions and item analysis as mentioned above.

Expert Panel: Once the item development process had been completed and the test had been developed to be run on a tablet PC, an expert panel was formed to solicit their evaluations based on a predefined criterion. The panel consisted of an expert from psychology, special education, educational psychology, Math teachers, and a measurement and evaluation expert. The criteria included items related to chosen colors, questions, buttons, and instructions. The feedback had been reviewed and the necessary changes had been applied by the Project team.

Table 2. Reliability results of the tests

	# of items	Internal Consistency Coefficients	Reliability with timed tests
CDC	14	0.69	0.92
SNC	24	0.79	0.93
MNL1	11	0.75	0.75
MNL3	11	0.72	0.72
MNL4	11	0.96	0.96

Criterion-Based Validity: In addition to the content validity of the tests, criterion-based validity has also been investigated. Correlations between curriculum-based Math achievement test (MAT), (Fidan, 2013) and calculation performance test (CPT) (Olkun et al., 2013) as an evidence to ensure its concurrent validity were performed. Furthermore, the predictive validity of the test has also been investigated by analyzing how the tests are predictive on students' Math achievement scores. For this purpose, the battery has been given to 478 elementary school students, its Math achievement predictive coefficients were calculated, and its mean and standard deviations were calculated for the selected sample. Once the data were gathered, students' inverse efficiency scores (IES) related to CDC and SNC tests and absolute error scores (AES) of MNL tests had been calculated.

Townsend and Ashby (1978) proposed the "inverse efficiency score" [IES; see also (Townsend & Ashby, 1983)] as an observable measure that measures the average energy used by the system over time. IE scores were calculated based on students' responses and times elapsed answering each question. It is calculated separately for each test and each subject by dividing the total time for answering the test items by the percentages of correct responses (Bruyer & Brysbaert, 2011). Lower values on this measure indicate a better performance. This measure is used to markdown possible criterion shifts or speed vs. accuracy balances in performance.

IE Scores were then converted into z scores. Outliers were omitted based on each individual's z-score range. The upper and lower range values were determined as ± 5. In other words, standard scores beyond $\pm$ 5 were considered as outliers. If there were only one or two values over the cut-off point these values were removed from the data. A total of 11 data points from CDC and 21 data points from SNC tests were deleted. In addition to outlier analysis, normality assumptions were investigated.

The newly developed screening tool is different from the one already in the literature (Butterworth, 2003) in two respects. First, this one has a subtest for ANS while the other does not. Second, the dot counting subtest used in this screening tool utilizes subitizing ability more than the other tool. As discussed earlier in this paper both ANS and subitizing are found to be related to

Mathematics learning disorders. The results of the correlation and regression analysis were given in findings section for validity.

3. Findings

Descriptive statistics for each grade level were analyzed in order to explore the psychometric characteristics of the test. The findings are presented in Table 3.

Table 3. Students' performances in each subtest: mean scores and standard deviations

	Grade 1			Grade 2			Grade 3			Grade 4		
	N	M	SD	N	M	SD	N	M	SD	N	M	SD
CDCies	119	90.50	26	120	73.50	20.2	120	59.4	13.4	119	52.4	14
SNCies	119	112.40	29.70	120	91.80	24	120	83	17.4	119	74.7	15
MNL1	119	18.62	9.52	120	11.86	7.06	120	8.8	6.89	119	6.57	5.27
MNL3	119	457	209	120	318	170	120	235	138	119	191	103
MNL4							120	3041	1970	119	2315	1480

Both Table 3 and Figure 3 indicate that students' scores have a steady developmental trajectory starting from Grade 1 to Grade 4. These findings provide evidence that the test measures the developmental changes in those skills in students.

Table 4. Multiple regression results for predicting First graders' Math achievement (MAT)

	Unstandardized Coefficients		Standardized Coefficients			
Variable	B	Standard error	Beta	t	p	Partial r
Constant	10.522	.884		11.908	.000	
CDCies	-1.71E-005	.000	-.183	-1.708	.090	-.276
SNCies	5.79E-006	.000	.046	.428	.670	-.160
MNL1	-.033	.027	-.129	-1.204	.231	-.288
MNL3	-.002	.001	-.202	-1.742	.084	-.334
R=0.384	R^2=0.148					
F(4,118)=4.937	p=0.001					

The predictive validity results from regression and correlation analyses are presented in Table 4 through 11. In regression analyses, students' curriculum-based Math achievement scores were taken as dependent (criterion/predicted), and the subtest sores were taken as predictor variables.

In the First grade, only 15% of the variability in Math achievement as measured by the curriculum-based MAT was explained by the subtests (see Table 4). The order of importance according to the standardized regression coefficients (Beta) are as follows: MNL3, CDC, MNL1 and SNC. T-test results as well as the significance test results indicate that those tests were not among the important variables to predict and explain curriculum-based Mathematical achievement in the First grade.

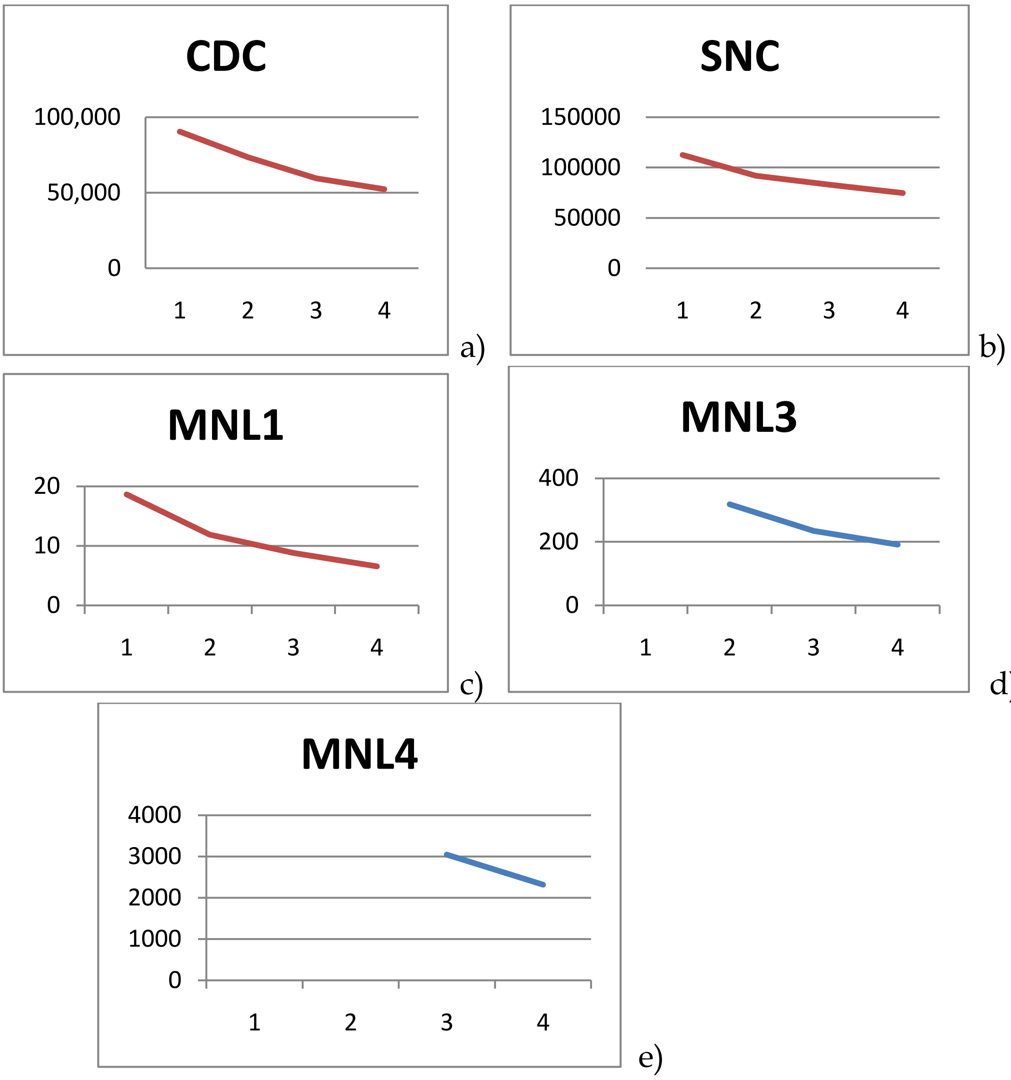


Figure 3. Score changes across grades in subtests [a). Canonic dot counting, CDC, b). Symbolic number comparison, SNC, c) Estimation between 0-10 on number line, MNL1, d). Estimation between 0-100 on number line, MNL3, e). Estimation between 0-1000 on number line, MNL4]

Table 5. Multiple regression results for predicting First graders' calculation performance (TTR)

	Unstandardized Coefficients		Standardized Coefficients			
Variable	B	Standard error	Beta	t	p	Partial r
Constant	18.667	1.972		9.467	.000	
CDCies	-3.69E-005	.000	-.174	-1.649	.102	-.319
SNCies	-1.50E-005	.000	-.052	-.496	.621	-.259
MNL1	-.028	.061	-.048	-.455	.650	-.253
MNL3	-.007	.003	-.254	-2.227	.028	-.374
R=0.420	R^2=0.176					
F(4,118)=6.092	p=0.000					

As seen in Table 5, the CDC, SNC, and MNL test scores explain 18% of the variance in the CPT test, which does not require any reading-writing but includes only symbolic arithmetic calculations (R=0.420, R^2=0.176). Only the MNL3 is significant, and CDC is approaching to significance in explaining the test performance.

Table 6. Multiple regression results for predicting Second grader's Math achievement

Variable	Unstandardized Coefficients B	Unstandardized Coefficients Standard Error	Standardized Coefficients Beta	t	p	Partial r
Constant	19.063	.917		20.777	.000	
CDCies	-6.27E-005	.000	-.331	-4.224	.000	-.634
SNCies	-5.29E-005	.000	-.225	-3.000	.003	-.565
MNL1	-.051	.036	-.095	-1.434	.154	-.399
MNL3	-.008	.001	-.366	-5.499	.000	-.594
R=0.770	R^2=0.593					
F(4,119)=41.805	p=0.000					

In the Second grade, the tests altogether explained 60% of the Math achievement (see Table 6) measured by MAT. The order of importance according to the standardized regression coefficients (Beta) are as follows: MNL3, CDC, SNC, and MNL1. When the significance levels in t-test results were considered, it is only the MNL1 test which did not contribute to the prediction; whereas the others were of importance in explaining the variance in Math achievement.

In the Third grade, the tests altogether explained 45% of the Math achievement (see Table 7) measured by MAT. The order of importance according to the standardized regression coefficients (Beta) are as follows: CDC, MNL1, MNL4, SNC and MNL3. When the significance levels in t-test results were considered, it is only the CDC and MNL1 tests that were found to have contributions to the predictive model.

Table 7. Multiple regression results for predicting Third graders' Math achievement

Variable	Unstandardized Coefficients B	Unstandardized Coefficients Standard Error	Standardized Coefficients Beta	t	p	Partial r
Constant	18.676	1.625		11.493	.000	
CDCies	-8.94E-005	.000	-.289	-3.189	.002	-.564
SNCies	-3.10E-005	.000	-.076	-.953	.343	-.362
MNL1	-.156	.053	-.260	-2.963	.004	-.544
MNL3	-.002	.004	-.059	-.453	.651	-.441
MNL4	.000	.000	-.168	-1.161	.248	-.529
R=0.669	R^2=0.447					
F(5,119)= 18.436	p=0.000					

In the Fourth grade, the tests altogether explained 39% of the Math achievement (see Table 8) measured by MAT. The order of importance according to the standardized regression coefficients (Beta) are as follows: CDC, MNL4, SNC, MNL3, and MNL1. When the significance levels in t-test results were considered, it is only the CDC test that was found to have a significant contribution to the predictive model while MNL4 is approaching the significance.

Table 8. Multiple regression results for predicting Fourth grader's Math achievement

Variable	Unstandardized Coefficients B	Unstandardized Coefficients Standard Error	Standardized Coefficients Beta	t	p	Partial r
Constant	28.352	2.121		13.366	.000	
CDCies	.000	.000	-.341	-3.228	.002	-.560
SNCies	-5.16E-005	.000	-.080	-.856	.394	-.375
MNL1	-.052	.095	-.048	-.551	.582	-.320
MNL3	-.004	.007	-.079	-.621	.536	-.469
MNL4	-.001	.000	-.226	-1.764	.080	-.511
R=0. 624 R^2=0.389						
F(5,118)= 14.386 p=0.000						

The correlations between calculation performances (CPT) and Math achievement scores with the subtests in the battery are presented in Table 8. First of all, the correlations are all negative as expected. There is a significant correlation between CDC and Math achievement scores in First grade (r=-0.276, p<0.01). The correlation between Math achievement and SNC is not significant (r=-0.160, p>0.05). The correlation between MAT and MNL1 is significant (r=-0,288, p<0.01); with MNL3, the correlation is again statistically significant (r=-0.334, p<0.05).

As presented in Table 9, there is a negative and significant correlation found between CPT and CDC scores (r=-0.319, p<0.01). The correlation between CPT and SNC is in the negative direction, weak but significant (r=-0.259, p<0.05). Likewise, there is a negative, weak and significant correlation between CPT and MNL1 (r=-0.253, p<0.01). The correlation between CPT and MNL3 scores are also negative, moderate and significant (r=-0.374, p<0.01).

Table 9. Correlations among scores for the First graders

BNP Tests	MAT	TTR
CDCies	-.276	-.319
(p)	.002	.000
SNCies	-.160	-.259
(p)	.119	.004
MNL1	-.288	-.253
(p)	.001	.006
MNL3	-.334	-.374
(p)	.000	.000

Table 10 shows the correlations between MAT and CPT scores, and BNP Tests for Second graders. There is a negative, moderate and significant correlation between MAT and CDC IES scores (r=-0.634, p<0.01), between MAT and SNC IES (r=-0.565, p<0.01), MAT and MNL1 scores (r=-0.399, p<0.01), and between MAT and MNL3 scores (r=-0.594, p<0.01).

Table 10. Correlation among scores for the Second graders

BNP Tests	MAT	TTR
CDCies	-.634	-.536
(p)	.000	.000
SNCies	-.565	-.495
(p)	.000	.000
MNL1	-.399	-.111
(p)	.000	.227
MNL3	-.594	-.346
(p)	.000	.000

As depicted in Table 10, there is also negative, moderate and significant correlation between CPT and CDC IES scores (r=-0.536, p<0.01), between CPT and SNC IES (r=-0.495, p<0.01), and between CPT and MNL3 scores (r=-0.346, p<0.01). The correlation between CPT and MNL1 scores is negative, weak and significant (r=-0.111, p>0.05).

Table 11. Correlation among scores for the Third graders

BNP Tests	MAT	TTR
CDCies	-.564	-.495
(p)	.000	.000
SNCies	-.362	-.313
(p)	.000	.000
MNL1	-.544	-.476
(p)	.000	.000
MNL3	-.441	-.342
(p)	.000	.000
MNL4	-.529	-.418
(p)	.000	.000

Table 11 shows the correlations between BNP Tests and, MAT and CPT scores for Third graders. There is a negative, moderate and significant correlation between MAT and CDC IES scores (r=-0.564, p<0.05), between MAT and SNC IES (r=-0.362, p<0.05), MAT and MNL1 scores (r=-0.544, p<0.01), MAT and MNL3 scores (r=-0.441, p<0.05), and between MAT and MNL4 (r=-0.529, p<0.05).

As shown in Table 11, there is also negative, moderate and significant correlation between CPT and CDC IES scores (r=-0,495, p<0,01), between CPT and SNC IES (r=-0.313, p<0.01), between CPT and MNL3 scores (r=-0,342, p<0,01), between CPT and MNL1 (r=-0,476, p<0,01), and between CPT and MNL4 (r=-0,418, p<0,01).

As seen in Table 12, there is a negative, moderate and significant correlation between Fourth graders' MAT and CDC IES scores (r=-0.560, p<0.01), between MAT and SNC IES (r=-0.375, p<0.01), MAT and MNL1 scores (r=-0.320, p<0.01), MAT and MNL3 scores (r=-0.469, p<0.01), and between MAT and MNL4 (r=-0.511, p<0.01). Overall, there is a negative and significant relation between MAT scores and CDC, SNC and MNL tests, which are based on either IES or absolute error scores.

Table 12. Correlation among scores for the Fourth graders

BNP Tests	MAT	TTR
CDCies	-.560	-.567
(p)	.000	.000
SNCies	-.375	-.377
(p)	.000	.000
MNL1	-.320	-.196
(p)	.000	.033
MNL3	-.469	-.268
(p)	.000	.003
MNL4	-.511	-.338
(p)	.000	.000

Similarly, as depicted in Table 12, there is negative, moderate and significant correlation between CPT and CDC IES scores (r=-0.567, p<0.01) and between CPT and SNC IES (r=-0.377, p<0.01). The relation between CPT and MNL3 scores (r=-0.268, p<0.01), between CPT and MNL1 (r=-0.196, p<0.05), and between CPT and MNL4 (r=-0.338, p<0.01) is negative, weak and significant.

Considering these results altogether, as expected, CDC, SNC and MNL test scores, which had either IES or absolute error scores, have negative correlations with both curriculum based Mathematics achievement as measured by MAT and calculation performance, as measured with CPT.

Table 13. ANCOVA results for CDC test scores adjusted according to Raven scores

Grade	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
1	raven	1772249843.66	1	1772249843.658	2.870	.093	.027
	Group	7072363604.97	2	3536181802.484	5.727	.004	.098
	Error	64830143430.32	105	617429937.432			
	Total	978029827305.20	109				
2	raven	912190137.94	1	912190137.944	4.624	.034	.046
	Group	5353808450.39	2	2676904225.193	13.571	.000	.222
	Error	18739525260.78	95	197258160.640			
	Total	535539634729.70	99				
3	raven	78008294.40	1	78008294.398	.620	.433	.006
	Group	3366715933.13	2	1683357966.564	13.371	.000	.203
	Error	13219402083.21	105	125899067.459			
	Total	402000955484.06	109				

In order to test whether the source of the variance is coming from the tests themselves or participants' general abilities, and whether the BNP tests would yield similar results when the general ability as measured by Raven SPM test were controlled, an ANCOVA test was executed. In the analyses, it was assumed that MAT and CPT tests were standardized, since their validity and reliability had been tested earlier. Mean scores were computed from First and Second applications for each grade level. Depending on the MAT scores, lower 27% is considered low achievers whereas upper 27% is considered as normal achievers. Researchers usually use the upper 27% and the lower 27% in order to separate the tail from the mean of the normally distributed data (Cureton, 1957). If the CDC, SNC and MNL tests are measuring the same construct then the group differences would be the same among tests. This would be considered as an evidence of validity.

After controlling for student's general abilities with Raven SPM (see Table 13), there was still significant differences among the groups based on CDC test in the First grade [$F_{(2,105)}$=5.727, p<.05, η^2=0.098], Second grade [$F_{(2,95)}$=13.571, p<.05, η^2=0.222], and Third grade [$F_{(2,105)}$=13.371, p<.05, η^2=0.203]. That means, in all the grades tested, lower groups showed less efficiency in CDC test.

Table 14. ANCOVA results for SNC test scores adjusted according to Raven test scores

Grade	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Sq.
1st	raven	229882783.962	1	229882783.962	.617	.434	.006
	Group	2942071757.621	2	1471035878.810	3.950	.022	.070
	Error	39102371517.164	105	372403538.259			
	Total	640519571159.637	109				
2nd	raven	103724769.429	1	103724769.429	.454	.502	.005
	Group	2145180219.755	2	1072590109.877	4.697	.011	.090
	Error	21691655781.145	95	228333218.749			
	Total	351457658059.796	99				
3rd	raven	146015262.380	1	146015262.380	1.218	.272	.011
	Group	710430961.322	2	355215480.661	2.964	.056	.053
	Error	12585666394.209	105	119863489.469			
	Total	315144743780.813	109				

Table 14 shows the ANCOVA results for lower and upper 27% groups for First and Second graders when their Raven SPM test scores were controlled. There is a significant difference between the upper and lower groups in First graders [$F_{(2,105)}$=3.950, p<.05, η^2=0.070] and Second graders [$F_{(2,95)}$=4.697, p<.05, η^2=0.090] in their SNC scores. In both grades, the SNC score averages are higher in lower group than the upper one. When the Raven SPM test scores were controlled, a marginally significant difference between lower and upper groups was observed for the Third graders in their SNC scores [$F_{(2,105)}$=2.964, p<.05, η^2=0.053].

After controlling for student's general abilities (see. Table 15) with Raven SPM, there was still a significant difference among the lower and upper achieving groups based on MNL-3 test in the First [$F_{(2,105)}$=3.349, p<.05, η^2=0.060], Second [$F_{(2,93)}$=11.752, p<.05, η^2=0.202], and Third grade [$F_{(2,105)}$=5.670, p<.05, η^2=0.097]. In all the grades tested, lower groups got higher TAE scores, in other words, they made less accurate estimations on MNL-3 (0-100 number line) test.

Table 15. ANCOVA results for MNL-3 test scores adjusted according to Raven scores

Grade	Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
1	raven	236467.827	1	236467.827	6.449	.013	.058
	Group	245576.951	2	122788.475	3.349	.039	.060
	Error	3850175.029	105	36668.334			
	Total	27133300.210	109				
2	raven	124227.377	1	124227.377	6.640	.012	.067
	Group	439728.032	2	219864.016	11.752	.000	.202
	Error	1739951.487	93	18709.156			
	Total	11634213.560	97				
3	raven	74405.394	1	74405.394	4.696	.032	.043
	Group	179682.037	2	89841.018	5.670	.005	.097
	Error	1663577.571	105	15843.596			
	Total	8393318.150	109				

Table 16. Third Graders' ANCOVA results for MNL-4 test scores adjusted according to Raven scores

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
raven	8189246.549	1	8189246.549	2.759	.100	.026
Group	52890994.510	2	26445497.255	8.908	.000	.145
Error	311700713.973	105	2968578.228			
Total	1465329741.490	109				

Again there was still significant difference (see. Table 16) between the lower and upper achievement groups based on MNL-4 test in the Third grade [$F_{(2,105)}$=8.908, p<.05, η^2=0.145] after controlling for the students' general abilities as measured by Raven SPM. Students in the lower group got higher TAE scores, in other words, they made less accurate estimations on MNL-4 (0-1000 number line) test.
Since we did not administer Raven SPM to Fourth graders we ran t-tests to analyze the differences between the lower and upper 27 percentiles.

Table 17. Fourth graders' t-test results based on MAT scores between upper and lower groups for CDC, SNC and MNL 1-3-4 tests

Subtests	Group	N	Mean	SD	DF	t	p
CDC	Lower Group	31	62139.3540	17163.67989	71	4.617	0,00
	Upper Group	42	45579.1818	11880.65529			
SNC	Lower Group	31	82336.1797	20651.97347	71	3.026	0,00
	Upper Group	42	69876.8086	11578.18597			
MNL1	Lower Group	31	8.1774	6.85642	71	2.710	0,00
	Upper Group	42	4.6405	2.80636			
MNL3	Lower Group	31	259.7968	123.29097	71	4.141	0,00
	Upper Group	42	153.7214	83.45923			
MNL4	Lower Group	31	3420.7935	1933.20610	71	5.034	0,00
	Upper Group	42	1584.0643	726.95047			

Fourth graders were grouped according to their MAT scores (see Table 17), and a t-test was run to see whether their CDC, SNC, MNL1, MNL3, and MNL4 test scores would differ accordingly. The results indicated that there is a significant difference in CDC IES test scores between groups (t=4.617, p<0.05). When their mean scores were examined, the difference is observed in favor of the lower group. Also, the lower group completed the test in a longer period of time. Similarly, there is a significant difference between the groups in terms of SNC IES scores (t=3.026, p<0.05), MNL1 test scores (t=2.710, p<0.05), MNL3 test scores (t=4.141, p<0.05), and MNL4 test scores (t=5.034, p<0.05) in favor of the lower group. In other words, these findings indicate that students in lower groups scored higher in these tests meaning that they had more estimation errors compared to the upper achievement group.

Table 18. Fourth graders t-test results for CDC, SNC and MNL1-3 tests between upper and lower groups based on TTR tests

Subtests	Group	N	Mean	SD	DF	t	p
CDC	Lower Group	32	62186.3715	16212.00981	63	6.133	0,00
	Upper Group	33	42822.5899	7607.24493			
SNC	Lower Group	32	81339.2179	18370.05264	63	4.054	0,00
	Upper Group	33	66645.6972	9242.30986			
MNL1	Lower Group	32	8.3000	7.09861	63	2.218	0,032
	Upper Group	33	5.2273	3.36835			
MNL3	Lower Group	32	229.6813	123.35055	63	2.310	0,024
	Upper Group	33	166.2636	95.78197			
MNL4	Lower Group	32	3005.4000	1824.52973	63	2.833	0,006
	Upper Group	33	1894.8000	1280.55589			

When the Fourth graders were grouped based on CPT test scores (see Table 18), we found a significant difference in CDC IES test scores between groups in favor of the lower group (t=6,133, p<0.05). Similarly, there is a significant difference in SNC IES test scores between lower and upper groups (t=4,054, p<0.05). These findings indicate that lower group students completed the test in a longer period of time. When the MNL test scores are considered, there is a significant difference between groups in MNL 1, (t=2,218, p<0.05), in MNL3 (t=2,310, p<0.05), and in MNL4 (t=2,833, p<0.05) in favor of the lower group. All these findings show that students in the lower group made more erroneous estimations in all MNL tests.

When MAT and CPT tests are considered as standard tests, the results indicate a validity evidence for CDC, SNC, and all MNL tests. Although there was no significant difference in MNL1 for Second graders, there is a significant difference between lower and upper groups when grouped according to CPT test. However, when the grade level increased, the number line with a wider range was more discriminative.

4. Discussion

In this study, a screening tool was developed for determining the tendency of Math learning disorder in children from the age of 6 to 11. We evaluated the reliability and validity of the battery consisted of three subtests as CDC, SNC and MNL1-2-3. The KR-20 method was used for dichotomous data and the Cronbach Alpha for polytomous data. Regression analysis was run for predictive validity measures. Correlations and mean differences were calculated for additional validity. Expert opinions were obtained for content validity.

The results of this study showed that Basic Number Processing Test (BNPT) yielded a high prediction rate especially in the Second grade. Similarly, an important portion of the variance in Math achievement could be explained by the administered tests in Third and Fourth grades. The differences among the groups were still significant even after controlling for the students' general abilities as measured by the Raven SPM test. These findings indicate that problems resulting from the poor processing of number could yield to lower Math achievement especially in Second, Third, and Fourth grades.

When the correlations are examined, despite their significance, the correlations are weak at the First grade but moderate and strong at the upper grades. When considering the regression and correlation scores together, it can be said that the findings provide evidence for the validity of the measurement tool, especially for the Second grade and up. This trend may also show that deficits in basic number processing abilities cause students to lag further behind their peers making it impossible to catch up in regular classes if untreated.

The tests, CDC, SNC, and MNL have all strong contributions to the Math achievement at the primary grades. Therefore, it could be concluded that at least very important portions of the causes of low achievement in Mathematics might originate from the deficits in the core systems of number (Feigenson et al., 2004). In other words, it could result from any one of the exact number system (ENS), the approximate number system (ANS), and the access to symbol system (ATS), or any combinations of these three systems.

Relatively lower relationships among the tests were obtained with the First graders. It should be noted that there are various barriers especially in the First semester of the First year in school. Since children already started to read and write, they often find it challenging to solve verbally instructed Math questions. In addition, they start learning numbers and representing them with symbols (numerals) at their First year. This makes it difficult to predict their real Math achievement by using tasks including verbal and/or symbol-based instructions and questions. For example, we found a higher correlation between Math achievements based on CPT and BNP tests, which required only numbers and other symbols for arithmetic operations, but no word reading-writing. In comparison, the MAT test requires reading-writing skills to solve verbal problems. Therefore, it would be necessary to develop new paradigms and/or longitudinal research with First graders, among which could be using interview method or tasks which do not require any reading-writing tasks. Additionally, it would also be a part of research to observe students who experience difficulty in learning and their performances in basic number processing tests.

References

Auerbach, J. G., Gross-Tsur, V., Manor, O., & Shalev, R. S. (2008). Emotional and behavioral characteristics over a six-year period in youths with persistent and nonpersistent dyscalculia. *J Learn Disabil, 41*(3), 263-273. doi:10.1177/0022219408315637

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., & Jacobsen, S. J. (2005). Math learning disorder: incidence in a population-based birth cohort, 1976-82, Rochester, Minn. *Ambulatory Pediatrics, 5*(5), 281-289. doi:10.1367/A04-209R.1

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (ies) a better dependent variable than the mean reaction time (rt) and the percentage of errors (pe)? *Psychologica Belgica, 51*(1), 5-13.

Butterworth, B. (2003). *Dyscalculia screener manual*. The Chiswick Centre, 414 Chiswick High Road, London W4 5TF, UK: nferNelson.

Butterworth, B. (2009). *Dyscalculia: Causes, identification, intervention and recognition*. Paper presented at the Dyscalculia and Maths Learning Difficulties, Holiday Inn, Bloomsbury (nr. Euston Station) London.

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences, 14*(12), 534-541. doi:10.1016/j.tics.2010.09.007

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Philadelphia: Harcourt Brace Jovanovich College Publishers.

Cureton, E. E. (1957). The Upper and Lower Twenty-seven Per Cent Rule. *Psychometrika*(22), 293- 296.

De Vos, T. (1992). *Tempo-test rekenen (Number fact retrieval test)*. Nijmegen: Berkhout.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. *Science, 320*(5880), 1217-1220. doi:10.1126/science.1156540

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307-314. doi:10.1016/j.tics.2004.05.002

Fidan, E. (2013). *İlkokul Öğrencileri İçin Matematik Dersi Sayılar Öğrenme Alanında Başarı Testi Geliştirilmesi.* (Yayımlanmamış Yüksek Lisans Tezi), Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü.

Frisbie, D. A. (1988). *Reliability of scores from teacher-made tests*. Article of NCME Instructional Module. University of Lowa.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting Mathematical Achievement and Mathematical Learning Disability with a Simple Screening Tool: The Number Sets Test. *Journal of Psychoeducational Assessment, 27*(3), 265-279. doi:10.1177/0734282908330592

Gimbert, F., Gentaz, E., Camos, V., & Mazens, K. (2016). Children's Approximate Number System in Haptic Modality. *Perception, 45*(1-2), 44-55. doi:10.1177/0301006615614448

Heine, A., Tamm, S., De Smedt, B., Schneider, M., Thaler, V., Torbeyns, J., . . . Jacobs, A. (2010). The numerical stroop effect in primary school children: a comparison of low, normal, and high achievers. *Child Neuropsychol, 16*(5), 461-477. doi:10.1080/09297041003689780

Iuculano, T., Tang, J., Hall, C. W., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*(5), 669-680. doi:10.1111/j.1467-7687.2008.00716.x

MEB. (2004). *İlköğretim Matematik Dersi Öğretim Programı*. Ankara: Milli Eğitim Basımevi.

Mussolin, C., Mejias, S., & Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition, 115*(1), 10-25. doi:10.1016/j.cognition.2009.10.006

Olkun, S., Can, D., & Yeşilpınar, M. (2013). *Hesaplama Performansı Testi: Geçerlilik Ve Güvenilirlik Çalışması*. Paper presented at the USOS 2013 Ulusal Sınıf Öğretmenliği Sempozyumu, Aydın, TR.

Park, J., & Starns, J. J. (2015). The Approximate Number System Acuity Redefined: A Diffusion Model Approach. *Front Psychol, 6*, 1955. doi:10.3389/fpsyg.2015.01955

Raven, J. (2000). The Raven's progressive matrices: change and stability over culture and time. *Cogn Psychol, 41*(1), 1-48. doi:10.1006/cogp.1999.0735

Rosenberg-Lee, M., Ashkenazi, S., Chen, T., Young, C. B., Geary, D. C., & Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. *Dev Sci, 18*(3), 351-372. doi:10.1111/desc.12216

Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: prevalence and prognosis. *Eur Child Adolesc Psychiatry, 9 Suppl 2*, II58-64.

Townsend, J. T., & Ashby, F. G. (1978). Methods of modeling capacity in simple processing systems. In J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 200-239). Hillsdale N.J.: Erlbaum.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

Van't Noordende, J. E., van Hoogmoed, A. H., Schot, W. D., & Kroesbergen, E. H. (2015). Number line estimation strategies in children with mathematical learning difficulties measured by eye tracking. *Psychol Res*. doi:10.1007/s00426-015-0736-z