FOR MORE, go to http://www.brookespublishing.com/NSS

Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition

by

Nancy C. Jordan, Ed.D.
University of Delaware

and

Joseph J. Glutting, Ph.D.
University of Delaware

with

Nancy Dyson, Ph.D.
University of Delaware

·P A U L·H·
BROOKES
PUBLISHING Co.®

Baltimore • London • Sydney

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Contents

Appendixes

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

About the Authors

Nancy C. Jordan, Ed.D., Professor, School of Education, University of Delaware, 206E Willard Hall, Newark, DE 19716

Nancy C. Jordan is Principal Investigator of the Number Sense Intervention Project (funded by the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development) as well as the Center for Improving Learning of Fractions (funded by the Institute of Educational Sciences). She is author or coauthor of many articles in mathematics learning difficulties and has recently published articles in *Child Development*, *Journal of Learning Disabilities*, *Developmental Science*, *Developmental Psychology,* and *Journal of Educational Psychology*. Dr. Jordan holds a bachelor's degree from the University of Iowa, where she was awarded Phi Beta Kappa, and a master's degree from Northwestern University. She received her doctoral degree in education from Harvard University and completed a postdoctoral fellowship at the University of Chicago. Before beginning her doctoral studies, she taught elementary school children with special needs. Dr. Jordan served on the Committee on Early Childhood Mathematics of the National Research Council of the National Academies.

Joseph J. Glutting, Ph.D., Professor, School of Education, University of Delaware, 221E Willard Hall, Newark, DE 19716

Joseph J. Glutting is a quantitative psychologist. He is a former project director of clinical and industrial measurement for The Psychological Corporation. He is also a certified school psychologist with 5 years' full-time experience in the public schools. He previously taught classes in child psychopathology, intelligence testing, and child personality assessment. Dr. Glutting specializes in applied multivariate statistics and test construction. He developed several standardized measures of intelligence, occupational interest, and attention-deficit/hyperactivity disorder (ADHD) including the *Wide Range Intelligence Test* (*WRIT;* Wide Range, 2000), *Wide Range Interest and Occupation Test–Second Edition* (*WRIOT2;* Wide Range, 2003), and *College ADHD Response Evaluation* (*CARE;* Wide Range, 2002). He coauthored the *Number Sense Battery* (*NSB;* Merrill Publishing, in press) with Nancy Jordan and published more than 100 peer-reviewed journal articles and book chapters. Dr. Glutting currently teaches graduate classes in applied multivariate and univariate statistics, as well as an undergraduate class in tests and measurement. His research is supported by the Institutes of Education Sciences and the National Institutes of Health.

Nancy Dyson, Ph.D., Postdoctoral Researcher, School of Education, University of Delaware, 130 Willard Hall, Newark, DE 19716

Nancy Dyson has been in education for more than 30 years as both a teacher and the director of a parent cooperative school. She recently completed her doctoral degree in education at the University of Delaware with a research focus on students struggling with mathematics.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

CHAPTER 4

Reliability and Validity of the Number Sense Screener™ NSS™

The establishment of a test's reliability and validity is an ongoing endeavor, and it is one that extends well beyond a test's development and standardization (Gregory, 2007). Consequently, it is impossible to present a complete set of psychometric characteristics at the time of a test's publication. This chapter presents current reliability and validity data for the NSS™. The authors, along with independent researchers, will continue to provide new technical information on the NSS as it becomes available. Independent researchers, and other interested users, are encouraged to contact the authors regarding the development and/or outcomes of such studies.

Reliability

Reliability is defined as the consistency of a measure's scores across items and across time (Anastasi & Urbina, 1997; Salvia, Ysseldyke, & Bolt, 2009). Several statistics are useful to describe a test's reliability. Among them are person- and item-separation indices obtained from modern test-score theory. Likewise, the provision of internal-consistency reliabilities from traditional test-score theory, standard errors of measurement, and test–retest stability coefficients are all recommended by *The Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999).

Item Statistics[1]

This section explains the person- and item-separation reliabilities that are important indices in a Rasch analysis (Bond & Fox, 2007; Rasch, 1960) The Rasch analyses were completed using Winsteps (Linacre, 2007) and SAS 9.1 software. The analyses were directed to children who participated in the item tryout phase of the NSS's development (N = 425). These children completed 26 of the 29 items in the final version of the NSS; the three items were not included because they were administered only to children in the NSS's oldest norm group.

The Rasch model was applied because of its desirable properties of linear, interval measurement (Embretson & Reise, 2000) It is important to note that the model provides indices of fit to test model assumptions of appropriate ordering of items and persons, along with issues

[1] The Rasch item analyses were completed by Jonathan Rubright, doctoral candidate at the University of Delaware. We are deeply indebted to him for his expertise and assistance.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

surrounding dimensionality (Wright & Stone, 1979). Results will be presented using infit and outfit mean squares, *z* standardizations, difficulty (location) parameters, standard errors expressed in logits, and person- and item-separation reliabilities (Wright & Stone, 1979). Items with mean squares (i.e., the average squared residual) above 1.4 were considered to show misfit. Lower values indicated an item was performing in expected ways in relation to the other scale items. Standardized values should generally be below 2.0 (Wright & Linacre, 1994).

Item and person reliabilities refer to the replicability of ordering: how well items would be similarly ordered given to a new sample or how well persons would be ordered given alternative yet similar items (Wright & Masters, 1982). In addition, a principal components analysis (PCA) was performed on the model residuals to search for extraneous factors (Linacre, 1998). Lastly, a differential item functioning (DIF) analysis was performed to search for items behaving differently across gender. The Mantel-Haenszel contingency table approach was used to identify items displaying DIF (Mantel & Haenszel, 1959). Magnitude of DIF was assessed using the Educational Testing Service (ETS) difficulty delta index, $D = -2.35 \ln(\alpha_{MH})$ (Dorans & Holland, 1993). Cut points were as follows: Class A (negligible DIF) = $|D| < 1.00$, Class B (moderate DIF) = $1.00 \leq |D| < 1.50$, and Class C (large DIF) = $|D| \geq 1.50$.

Results revealed that item-reliably index was .99 and the person-reliability index was .84, providing evidence of reliability (person index) and validity (item index) of the scale. The variance explained by the measure of interest was 83.4%, with only 7.3% of the variance explained by the first factor of residuals. Fit statistics from the analysis are displayed in Table 4.1 for every

Table 4.1. Number Sense Screener item fit statistics

Item	Location	Location SE	Infit MNSQ	Infit ZSTD	Outfit MNSQ	Outfit ZSTD
1	−3.97	.3	0.85	−0.5	0.43	−1.4
2	−2.34	.17	0.96	−0.4	0.9	−0.2
3	−2.43	.17	0.79	−1.9	0.42	−2.4
4	0.32	.12	0.86	−2.8	0.81	−2.1
5	−0.83	.12	0.91	−1.5	1.23	1.5
6	1.28	.13	1.31	4.5	1.37	2.8
7	−1.87	.15	1.02	0.3	1.28	1.1
8	−0.6	.12	1.23	4	1.35	2.4
9	−0.81	.12	1.14	2.4	1.29	1.8
10	−0.57	.12	1.05	1	1.33	2.3
11	−0.53	.12	1.18	3.3	1.34	2.4
12	−2.55	.18	0.94	−0.4	1	0.1
13	1.13	.12	1.1	1.6	1.1	0.9
14	0.25	.12	0.89	−2.2	0.84	−1.8
15	−1.62	.14	0.98	−0.2	0.93	−0.2
16	0.69	.12	1.08	1.5	1.07	0.7
17	1.94	.14	0.94	−0.7	0.9	−0.6
18	1.25	.13	0.92	−1.4	1	0
19	1.85	.14	1.3	3.8	2.17	5.7
20	1.02	.12	0.96	−0.6	1	0
21	0.24	.12	0.85	−3	0.82	−1.9
22	1.16	.12	0.8	−3.5	0.87	−1.2
23	1.53	.13	0.77	−3.7	0.72	−2.3
24	1.04	.12	0.85	−2.8	0.8	−2
25	1.92	.14	1.01	0.2	1.11	0.7
26	2.5	.15	0.81	−2.1	0.56	−2.4

Note: N = 425.

Key: location, location (difficulty) parameter; location SE, location standard error; infit MNSQ, infit mean squares; infit ZSTD, standardized infit; outfit MNSQ, outfit mean squares; outfit ZSTD, standardized outfit.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.2. Number Sense Screener Mantel-Haenszel bias analysis across gender

Item	$MH\chi^2$	*p*-value	α_{MH}
1	0.85	.36	0.45
2	0.67	.41	1.35
3	1.75	.19	1.67
4	3.17	.08	1.57
5	0	.99	1.00
6	0.07	.80	1.07
7	0.03	.87	1.05
8	0.04	.84	1.05
9	0	.98	1.01
10	0.04	.84	1.05
11	1.23	.27	1.29
12	2.65	.10	0.53
13	0.24	.63	0.89
14	2.86	.09	1.57
15	0.04	.85	1.06
16	2.92	.09	0.65
17	0.68	.41	0.77
18	0.24	.63	0.87
19	0.50	.48	1.20
20	0.16	.69	1.11
21	2.25	.13	0.67
22	1.27	.26	1.42
23	0.34	.56	0.83
24	6.08	.01	0.50
25	0.51	.48	0.82
26	0.76	.38	1.40

Note: $MH\chi^2$= Mantel-Haenszel summary chi-square test, α_{MH} = Mantel-Haenszel alpha. N = 425.

item. Outcomes from the analysis serve to support the NSS's construct validity by showing that the instrument essentially measures one construct.

Table 4.2 presents results from an item-bias analysis. The analyses were completed using Mantel-Haenszel methodology; it compared genders and used males as the reference group. Results revealed that only one item exhibited significant DIF. Specifically, girls answered item 24 correctly about twice as often as boys (1/.50 = 2). In log odds terms, girls found this question to be $\beta = \ln(.50) = -.69$ times easier than boys with equal ability. Using the ETS index, the magnitude of the difference is $D = -2.35 \ln(.50) = 1.62$, a large difference. Nevertheless, only one instance of bias was found across 26 items. Therefore, it is fair to infer that the NSS is essentially free of gender bias.

Internal Consistency

Cronbach's (1951) coefficient alpha was used to calculate internal-consistency reliability. Coefficient alpha provides a lower bound value of internal consistency and is considered to be a conservative estimate of a test's reliability (Allen & Yen, 1979; Carmines & Zeller, 1979; Reynolds, Livingston, & Willson, 2006). Alpha coefficients are presented in Table 4.3 for each of the NSS's three norm groups. Coefficients are also presented separately for males and for females. Lastly, the bottom row of Table 4.3 presents averaged values.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.3. Internal-consistency reliability for the Number Sense Screener

	Demographic cohort		
Norm group	Total sample[a]	Males	Females
Fall of kindergarten	.82	.83	.82
Spring of kindergarten	.86	.89	.85
Fall of first grade	.87	.87	.87
Average[b]	.85	.87	.85

[a] $N = 425$.

[b] Average coefficients were calculated with Fisher's z' transformation.

Table 4.3 demonstrates that, on average, reliabilities increased with children's age. For instance, internal consistency for the total sample was .82 for youngest children who were assessed during the fall of kindergarten. It continued to increase for children evaluated during the spring of kindergarten (.86) and reached its zenith at .88 for children evaluated during the fall of first grade. The average value for the entire sample was .85. As expected, this internal-consistency reliability coefficient aligned well with the person reliability index (.84) from the Rasch analyses.

Averaged values at the bottom of Table 4.3 are appropriately high for the entire sample, and for males and females (all three equal .85 to .87) and exceed the .80 criterion suggested in certain textbooks for achievement measures (e.g., Reynolds et al., 2006). Consequently, results indicate examiners can use the NSS with confidence, because its scores demonstrate high levels of internal-consistency reliability.

Confidence Intervals

The presence of random error ensures that no test can be perfectly reliable. That is, scores of a child who is retested on different occasions with the same instrument or retested with a different set of equivalent items from the same instrument vary somewhat. The most common method of indicating unreliability is to supply confidence intervals for scores.

The standard error of measurement (SEM) provides the foundation upon which confidence intervals are built. The SEM is reflected whenever the 68% confidence level is reported for a score, because the SEM and the 68% confidence limit are the same (see Glutting, McDermott, & Stanley, 1987, for more information). A 68% confidence level means that over a large number of testings, examiners can be 68% confident that a child's true score resides within a specified score range (Dudek, 1979).

Table 4.4 presents confidence interval magnitudes for standard scores ($M = 100$, $SD = 15$) for the NSS. As the plus (+) symbol indicates, these values should be added and subtracted to children's standard scores in order to establish the upper and lower bounds within which their true scores are likely to fall. In addition to presenting the intervals for the 68% confidence level, Table 4.4 provides the magnitudes of the intervals for other common confidence levels, including 90%, 95%, and 99% levels. All values are reported by NSS norm group.

Table 4.4. Confidence interval magnitudes for Number Sense Screener standard scores

	Confidence level			
Norm group	68%	90%	95%	99%
Fall of kindergarten	±6.4	±10.4	±12.5	±16.4
Spring of kindergarten	±5.6	±9.2	±11.0	±14.5
Fall of first grade	±5.4	±8.9	±10.6	±13.4

Note: Confidence intervals are expressed for standard scores ($M = 100$, $SD = 15$) and based on total sample reliabilities, by age, presented in Table 4.3.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.5. Number Sense Screener test–retest reliability coefficients

	Time of administration					
Time of administration	September K	November K	February K	April K	November Gr. 1	February Gr. 1
September K	—	.81	.80	.78	.69	.61
November K		—	.82	.81	.70	.61
February K			—	.86	.77	.70
April K				—	.81	.75
November Gr. 1					—	.80
February Gr. 1						—

From Jordan, N.C., Glutting, J., Ramineni, C., & Watkins, M.W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review, 39*(2), 187; Copyright 2010 by the National Association of School Psychologists. Bethesda, MD. Reprinted with permission of the publisher www.nasponline.org.

Note: K, kindergarten; Gr. 1, Grade 1. *N* = 378.

Test–Retest Stability

Score stability was examined by using data from 378 children who took part in a 4-year longitudinal investigation of children's mathematics development (Jordan et al., 2006). Participants attended the same public school district in northern Delaware. All kindergartners from six schools were invited to participate in the study. There were 378 children who started the study at the beginning of kindergarten and 204 who remained at the end of third grade. Participant attrition was due to children moving out of the school district (typically right after kindergarten), rather than withdrawal from the study or absence on the day of testing. A logistical regression analysis (Jordan, Kaplan, Ramineni, & Locuniak, 2009) revealed that although gender and age do not predict the odds of being absent from the study in third grade, low-income and minority children, respectively, were about 1.2 times more likely to be absent from the study than were middle-income and nonminority children. In third grade, 52% of the children were boys, 45% had minority ethnic backgrounds (63% African American, 26% Hispanic, and 11% Asian), and 23% came from low-income families. Income status was determined by participation in the free or reduced-price lunch program in school, and most low-income children resided in urban neighborhoods.

Table 4.5 presents test–retest reliability coefficients across the six time periods. As expected, stability coefficients are higher for shorter intervals. Reliabilities ranged from .61 to .86. Twelve of the fifteen reliability coefficients were at or above the .70 criterion recommended in certain assessment textbooks (Gregory, 2007; Reynolds et al., 2006). Three coefficients dipped below the .70 criterion. However, this occurred only when the testing period exceeded 1 year. Findings therefore point to the need for annual retesting with the NSS.

Validity

Validity is multifaceted (American Educational Research Association et al., 1999); yet, at its core, the construct can be viewed simply as the extent to which a test measures what it is designed to measure (Gregory, 2007; Salvia, Ysseldyke, & Bolt, 2009). A variety of statistical data was gathered to document the relevance of information provided by the NSS. The NSS's validation strategy is consistent with the substantive-construct model of test development, wherein a test's validity is examined both internally to itself and externally to criterion variables (cf. Cronbach & Meehl, 1955). Consequently, the presentation below is divided into six sections according to the types of evidence testable at the time of the NSS user's guide's publication: 1) developmental changes, 2) content-related validity, 3) discriminant (contrasted-groups) validity, 4) predictive validity, 5) construct validity, and 6) treatment validity.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

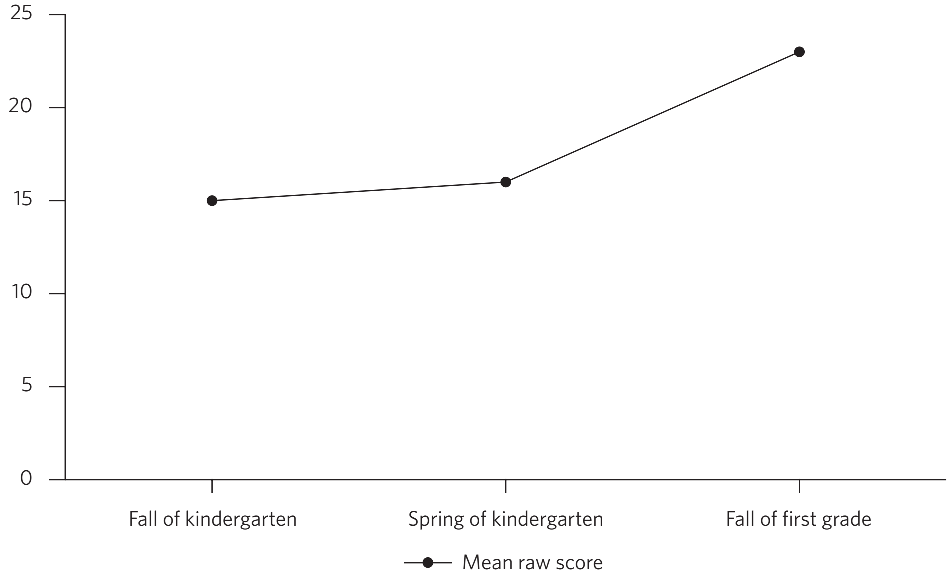

Figure 4.1. Plot of mean Number Sense Screener raw scores across age.

Developmental Changes

Age differentiation is a major criterion employed in the validation of a number of intelligence tests and achievement tests (Anastasi & Urbina, 1997). Because mathematics knowledge is expected to increase with age during childhood, it is argued that valid tests show raw scores that increase with age. Figure 4.1 plots mean raw scores from the NSS's three norm groups. The plot reveals that the NSS possesses considerable developmental validity, because raw scores exhibit clear-cut and consistent age changes.

Content-Related Validity

The content of the NSS is well established by the research (Jordan et al., 2010). The assessment is closely aligned to the Kindergarten Common Core State Standards in Number: Counting and Cardinality, and in Numbers and Operations in Base 10 (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). According to the Common Core, kindergartners are expected to know number names and the count sequence, to count to tell the number of objects in a collection (NSS Counting Skills subarea), and to compare numbers presented as written numerals between 1 and 10 (NSS Number Comparisons subarea). In terms of operations, kindergartners are expected to understand addition as adding to a quantity and understand subtraction as taking from a quantity (NSS Nonverbal Calculation, Story Problems, and Number Combinations). Kindergartners also are expected to work with numbers between 11 and 19 to gain foundations for place value (NSS Number Recognition).

The NSS also lines up with the Kindergarten Focal Points of the National Council of Teachers of Mathematics (2006) in the area of numbers and operations. This includes representing, comparing, and ordering whole numbers (NSS Counting Skills, Number Recognition, and Number Comparisons) and joining and separating sets (NSS Nonverbal Calculation, Story Problems, and Number Combinations).

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Discriminant (Contrasted-Groups) Validity

Campbell and Fiske (1959) introduced the concept of discriminant validity. They stressed that new tests need to be evaluated using both discriminant and convergent validation techniques. One way to establish a discriminant validity is to determine how well scores from a test distinguish (i.e., discriminate) children who meet achievement standards on an external measure from children who fail to meet standards.

Participants

Participants were part of a 4-year longitudinal investigation of children's mathematics development (Jordan et al., 2006). This sample was described previously in the section on test–retest reliability ($N = 378$). Readers are referred to that section for further details about the sample's composition.

Criterion

Mathematics achievement was assessed with the third grade version of the Delaware Student Testing Program (DSTP) in Mathematics (Delaware Department of Education, 2008). The DSTP measures concepts and procedures in accordance with Delaware mathematics standards (i.e., numeric reasoning, algebraic reasoning, geometric reasoning, and quantitative reasoning). It has strong internal reliability (.93) and has established cut scores for meeting state standards and for performing below state standards. The test's cut points and content were fully validated by a panel of experts (Delaware Department of Education, 2008). The DSTP in third grade is highly correlated ($r = .77$, $p < .01$) with scores on the Woodcock-Johnson III—Mathematics (McGrew, Schrank, & Woodcock, 2007), indicating strong criterion validity (Jordan et al., 2009). For the present study, the DSTP mathematics outcome measure was used in categorical form (1 = met standards, 0 = did not meet standards). The original five performance levels were collapsed to two levels to simplify the measurement scale. The performance levels of 3 (*meets the standard*), 4 (*exceeds the standard*), and 5 (*distinguished performance*) were transformed to a 1 on the categorical scale to represent meeting the DSTP standards, whereas the remaining two lower performance, levels 1 (*well below the standard*) and 2 (*below the standard*), were transformed to a 0 on the categorical scale to denote failure to meet the standards on DSTP in mathematics.

Procedure

The NSS was given to children individually in school by one of several trained graduate or undergraduate research assistants. It was administered in September and April of kindergarten and in November of first grade. The state mathematics proficiency achievement test was group-administered by school personnel in April of third grade.

Results

Data were analyzed using a repeated measures analysis of variance (ANOVA). Time was the within-subjects (repeated) measure, and it was evaluated on three occasions: fall of kindergarten, spring of kindergarten, and fall of first grade. Table 4.6 presents means and standard deviations for the groups on the dependent variable, and it does so separately by time period. Mauchly's test indicated that the assumption of sphericity was not violated ($\chi^2 = 2.459$, $df = 2$, $p = .001$). Therefore, there was no need to correct for a sphericity violation.

Figure 4.2 provides a visual representation of the results. It shows children meeting proficiency on the DSTP in third grade had higher NSS scores in the fall of kindergarten than children who did not meet proficiency. This situation occurred again in the spring of kindergarten, and remained true in the fall of first grade. Given this arrangement of scores, it came as no

© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.6. Means and standard deviations for Number Sense Screener scores by group and time

Time	DSTP criterion group	*M*[a]	*SD*
Fall of kindergarten	Failed to meet proficiency	89.1	11.5
	Met proficiency	107.0	14.6
Spring of kindergarten	Failed to meet proficiency	86.4	11.1
	Met proficiency	106.6	15.1
Fall of first grade	Failed to meet proficiency	90.3	10.2
	Met proficiency	105.2	12.4

From Jordan, N.C., Glutting, J., Ramineni, C., & Watkins, M.W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review, 39*(2), 188; Copyright 2010 by the National Association of School Psychologists. Bethesda, MD. Reprinted with permission of the publisher www.nasponline.org.

Note: DSTP, Delaware State Testing Program; *M,* mean; *SD,* standard deviation; mg, milligram.

[a]All numbers rounded at first decimal point for convenient presentation.

surprise that neither was the main effect for time significant (F = .783, *df* [2, 266], p = .001) nor was the group x time interaction significant (F = 2.071, *df* [2, 266], p = .458). Specifically, it was anticipated that a statistically significant main effect would occur only for two groups. In fact, the result revealed that the main effect for group was significant (F = 39.812, *df* [1, 133], p = .001). This *F*-test evaluates whether the dependent variable changes across groups—independent of time (Keppel & Wickens, 2004). So, in the current case, the main effect for time shows that across each of the three time periods (fall of kindergarten, spring of kindergarten, fall of first grade), children meeting proficiency on the DSTP had higher NSS scores.

Partial eta square (η^2) is a common effect-size measure for repeated measures ANOVA. Murphy and Myors (2004) defined the ranges of partial eta square: .01 = a small effect size, .06 = a medium effect size, and .14 = a large effect size. In the current study, the main effect for group represented a very large effect size (i.e., partial eta squared = .23). Consequently, it is reasonable to infer that the NSS possesses substantial discriminant (contrast-group) validity.

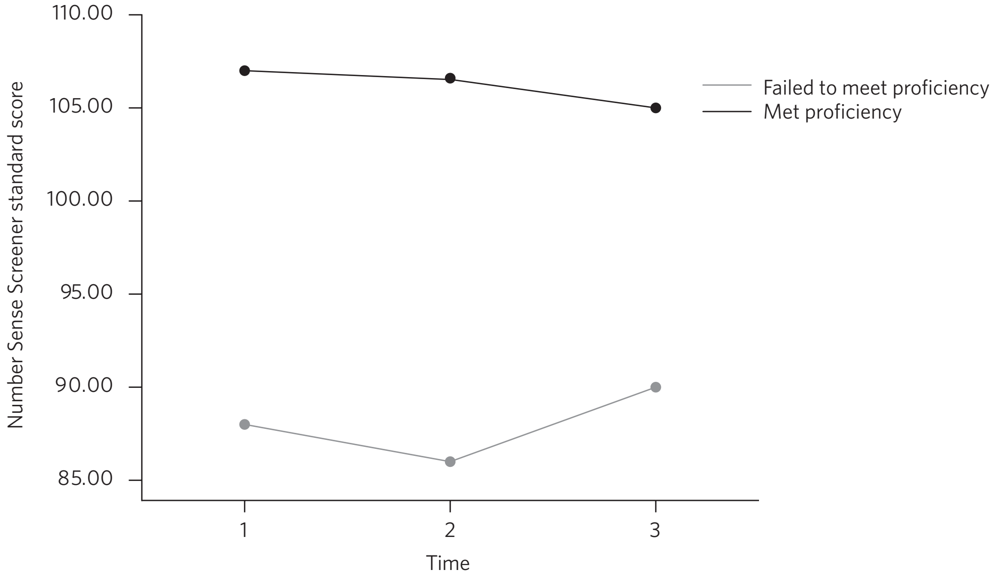

Figure 4.2. Plot of marginal means by group across time.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.7. Demographic information for participants at the end of first grade (n = 279) and the end of third grade (n = 175)

Variable	End of first grade	End of third grade
Gender		
Male	55%	54%
Female	45%	46%
Race		
Minority[a]	52%	42%
Nonminority	48%	58%
Income		
Low income	28%	22%
Middle income	72%	78%
Mean kindergarten start age (*SD*)	5 years 6 months (4 months)	5 years 6 months (4 months)

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

[a] Minority refers to African American (29%, n = 81), Asian (6%, n = 17), and Hispanic (17%, n = 47) at the end of first grade; and African American (25%, n = 44), Asian (6%, n = 11), and Hispanic (11%, n = 19) at the end of third grade.

Predictive Validity

Children were given the NSS at the beginning of first grade, and mathematics outcomes were obtained at the end of both first and third grades (Jordan, Glutting, & Ramineni, 2009). Outcomes included overall mathematics achievement as well as subareas of written computation and applied problem solving. It was hypothesized that number sense proficiency may be more relevant to applied problem solving than written computation, which may be more dependent on learned algorithms. To examine the unique contribution of number sense (as measured by the NSS) to these later mathematics outcomes, we also added the common predictors of age, verbal and spatial abilities, and working memory skills in our analyses.

Participants

Participants were drawn from a multiyear longitudinal investigation of children's mathematics development (Jordan, Kaplan, Ramineni, & Locuniak, 2009). They all attended the same public school district in northern Delaware. Background characteristics of children in first grade (n = 279) and in third grade (n = 175) are presented in Table 4.7. The first graders included children who completed all measures in first grade, and the third graders were children who completed all measures in first and third grade. In the first grade sample, 55% of the children were boys, 52% had minority ethnic backgrounds, and 28% came from low-income families. In the third grade sample, 54% of the children were boys, 42% had minority ethnic backgrounds, and 22% came from low-income families. Income status was determined by participation in the free or reduced-price lunch program in school. Participant attrition was due primarily to children moving out of the school district rather than withdrawal from the study or absence on the day of testing.

Procedure

The measures were given to children individually in school by one of several trained research assistants. The NSS items were given in October of first grade, the cognitive measures (Vocabulary, Matrix Reasoning, and Digit Span tests) in January of first grade, and the math achievement measures in April of first grade and again in April of third grade.

Cognitive Tasks

The Wechsler Abbreviated Scale of Intelligence (Wechsler, 1999) was used to assess oral vocabulary and spatial reasoning. A digit span test (Wechsler, 2009) was used to measure short-term

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.8. Correlations between first grade Number Sense Screener and control variables

Variable	Number Sense Screener correlation
Math composite (end of first grade)	.72
Math applications (end of first grade)	.73
Math calculation (end of first grade)	.58
Math composite (end of third grade)	.70
Math applications (end of third grade)	.74
Math calculation (end of third grade)	.66
Kindergarten start age	.19
Vocabulary	.56
Matrix reasoning	.53
Digit span forward	.34
Digit span backward	.50

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

Note: All correlations are significant, $p < .01$.

and working memory. Digit span forward is a measure of short-term recall and digit span backward a measure of working memory or active recall.

Mathematics Achievement

Math achievement was assessed with the Woodcock-Johnson III (Woodcock, McGrew, Schrank, & Mather, 2007). The composite achievement score (math overall) was the combined raw scores for subtests assessing written calculation (written calculations using a paper and pencil format; math calculation) and applied problem solving (orally presented problems in various contexts; math applications).

Results

Raw scores from the NSS were used for all analyses. Bivariate correlations are presented in Table 4.8 between the NSS raw scores and raw scores on the cognitive measures at first and third grades, as well as between the NSS and age at the beginning of kindergarten. All of the correlations were positive and statistically significant (i.e., all p values $< .05$), with the two lowest correlations being kindergarten start age (.19) and digit span forward (.34) and the highest correlations being math applications in first and third grades (.73 and .74, respectively).

A primary purpose of the study was to determine the unique contribution of the NSS in predicting criterion mathematics performance. Specifically, the study examined the extent to which the NSS predicted mathematics performance above and beyond the contribution of the control (nuisance) variables of age and general cognition related to language (vocabulary), spatial ability (matrix reasoning), and memory (digit span forward and digit span backward). To accomplish these goals, students' scores on the NSS were regressed on a series of established mathematics achievement outcomes (math overall, math calculation, math applications) using the two-stage model. At Step 1 (model 1), the control (nuisance) variables entered simultaneously into an analysis. Step 2 (model 2) comprised entry of the NSS. The analyses were used to predict mathematics achievement in first grade and then in third grade. The independent contributions of predictors were evaluated through the interpretation of squared partial coefficients. Effect sizes were estimated for the predictors using Cohen's (1988) f^2, where values of .02 equal a small effect, values of .15 equal a medium effect, and values of .35 equal a large effect.

Table 4.9 presents the results for predicting criterion performance on the mathematics composite score (math overall). Model 1 (age and general cognitive measures) accounted for

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.9. Results of block entry regression for the end of first-grade math overall and the end of third-grade math overall: regression coefficients and variance explained by each block of variables

	First-grade math overall					Third-grade math overall				
Model	*B*	β	*t* value	*p* value	Effect size[a]	*B*	β	*t* value	*p* value	Effect size[a]
One										
Age	.11	.07	1.50	.14		−.07	−.03	−0.55	.59	
Vocabulary	.27	.30	5.58	0		.30	.24	3.62	0	
Matrix reasoning	.34	.31	6.16	0		.56	.39	6.26	0	
Digit span forward	.06	.02	0.33	.74		.68	.15	2.24	.03	
Digit span backward	.78	.23	4.38	0		.99	.17	2.77	.01	
Two										
Age	.05	.03	0.79	.43	—	−.09	−.04	−0.82	.42	—
Vocabulary	.12	.14	2.71	.01	.03	.14	.11	1.78	.08	—
Matrix reasoning	.18	.17	3.62	0	.05	.36	.25	4.27	0	.08
Digit span forward	.11	.03	0.71	.48	—	.39	.09	1.43	.16	—
Digit span backward	.38	.11	2.31	.02	.02	.41	.07	1.26	.21	—
Number Sense Screener	.53	.48	8.97	0	.29	.78	.46	6.83	0	.21
Model	R square	R square change	*F* change	*df* 1	*df* 2	R square	R square change	*F* change	*df* 1	*df* 2
One	.47		47.99[b]	5	273	.45		28.03[b]	5	169
Two	.59	.12	80.43[b]	1	272	.57	.12	46.70[b]	1	168

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

[a] Cohen's (1988) f^2 statistic where .02 is a small effect size, .15 is a medium effect size, and .35 is a large effect size.

[b] $p < .01$.

47% of the variance in math in first grade ($p < .01$), with vocabulary, matrix reasoning, and digit span backward reaching significance, and 45% of the variance in third grade ($p < .01$), with vocabulary, matrix reasoning, digit span forward, and digit span backward reaching significance. Results showed that the NSS made statistically significant, unique contributions to the prediction at first grade ($p < .01$) and third grade ($p < .01$) outcomes in math overall. In each instance, the NSS accounted for about 12% more criterion variance than the control variables. More important, Cohen's (1988) f^2 represented medium to large effect sizes for both first- and third-grade criterion performance (respectively, .29 and .21).

Table 4.10 presents the results for predicting mathematics calculation. Model 1 (age and general cognitive measures) accounted for 35% of the variance in math calculation in first grade ($p < .01$) with vocabulary, matrix reasoning, and digit span backward reaching significance, and 33% of the variance in third ($p < .01$), with vocabulary and matrix reasoning reaching significance. Model 2 accounted for 41% of the variance in first grade, indicating that the NSS measure accounted for 6% more variance than the control variables. Cohen's (1988) f^2 value for the NSS was .10, which represented a small to medium effect size. Results for third grade were more impressive. The NSS accounted for 14% more variance of math calculation than the control variables and Cohen's (1988) f^2 (.26) represented a medium to large effect size.

Table 4.11 presents the results for mathematics applications where the results were most impressive. Model 1 accounted for 44% of the variance in math applications in first grade ($p < .01$), with vocabulary, matrix reasoning, and digit span backward reaching significance, and 45% of the variance in third grade ($p < .01$), with vocabulary and matrix reasoning reaching significance. Not only did the NSS make significant, unique contributions that accounted for 14% to 17% of the criterion's variance, Cohen's (1988) f^2 represented a large effect size in predicting first-grade NSS performance (.44) and third-grade NSS performance (.45). In sum, results show that the NSS possesses substantial levels of predictive validity.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.10. Results of block entry regression for the end of first-grade math calculation and the end of third-grade math calculation: regression coefficients and variance explained by each block of variables

	First-grade math calculation					Third-grade math calculation				
Model	*B*	β	*t* value	*p* value	Effect size^a	*B*	β	*t* value	*p* value	Effect size^a
One										
Age	.07	.02	0.47	.64		.16	.04	0.69	.49	
Vocabulary	.10	.24	3.96	0		.15	.28	3.72	0	
Matrix reasoning	.14	.29	5.19	0		.21	.34	4.87	0	
Digit span forward	−.07	−.05	−0.77	.44		−.08	−.04	−0.57	.57	
Digit span backward	.40	.26	4.38	0		.25	.13	1.73	.09	
Two										
Age	−.01	0	−0.03	.97	—	.03	.01	0.16	.88	—
Vocabulary	.05	.12	2.02	.05	.02	.06	.12	1.69	.09	—
Matrix reasoning	.10	.19	3.40	0	.05	.11	.18	2.75	.01	.04
Digit span forward	−.05	−.04	−0.61	.54	—	−.04	−.02	−0.30	.77	—
Digit span backward	.27	.18	2.97	0	.03	.07	.03	0.51	.61	—
Number Sense Screener	.17	.33	5.19	0	.10	.33	.48	6.86	0	.26
Model	R square	R square change	*F* change	*df* 1	*df* 2	R square	R square change	*F* change	*df* 1	*df* 2
One	.35		29.46^b	5	273	.33		18.77^b	5	187
Two	.41	.06	26.89^b	1	272	.47	.14	47.12^b	1	186

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

[a]Cohen's (1988) f^2 statistic where .02 is a small effect size, .15 is a medium effect size, and .35 is a large effect size.

[b]$p < .01$.

Table 4.11. Results of block entry regression for the end of first-grade math applications and the end of third-grade math applications: regression coefficients and variance explained by each block of variables

	First-grade math applications					Third-grade math applications				
Model	*B*	β	*t* value	*p* value	Effect size^a	*B*	β	*t* value	*p* value	Effect size^a
One										
Age	.37	.09	1.95	.05		.01	0	0.02	.98	
Vocabulary	.17	.31	5.56	0		.22	.28	4.08	0	
Matrix reasoning	.19	.29	5.49	0		.38	.41	6.51	0	
Digit span forward	.12	.06	1.11	.27		.07	.03	0.38	.71	
Two										
Age	.21	.05	1.27	.21	—	−.21	−.04	−0.78	.44	—
Vocabulary	.07	.13	2.54	.01	.03	.08	.10	1.67	.10	—
Matrix reasoning	.09	.13	2.79	.01	.03	.22	.24	4.20	0	.11
Digit span forward	.16	.08	1.65	.10	—	.14	.05	0.89	.37	—
Digit span backward	.11	.05	1.07	.29	—	.04	.01	0.21	.83	—
Number Sense Screener	.36	.52	9.69	0	.44	.55	.54	9.11	0	.45
Model	R square	R square change	*F* change	*df* 1	*df* 2	R square	R square change	*F* change	*df* 1	*df* 2
One	.44		43.01^b	5	273	.45		30.34^b	5	187
Two	.58	.14	93.89^b	1	272	.62	.17	82.97^b	1	186

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

[a]Cohen's (1988) f^2 statistic where .02 is a small effect size, .15 is a medium effect size, and .35 is a large effect size.

[b]$p < .01$.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Construct Validity

Strong construct validity is suggested whenever there is an appropriate pattern of convergent and divergent associations (American Educational Research Association et al., 1999; Campbell, 1960; Messick, 1989). Convergent validity is demonstrated when a scale correlates highly with other scales with which it shares an overlap of constructs. Discriminant validity is demonstrated when a test does not correlate highly with variables from which it should differ. In the current case, we would expect a pattern of somewhat higher correlations between the NSS and scores from measures of mathematics than between the NSS and scores from measures of reading achievement.

The reason for expecting only "somewhat" higher correlations is because the constructs of reading and mathematics are themselves highly correlated. For instance, the Wechsler Individual Achievement Test–Third Edition (Wechsler, 2009) reports correlations between reading and mathematics composite by grade level from kindergarten through 12th grade. Averaging estimates across ages yields an overall $r = .62$, which is a very large effect size using Cohen's 1988 criteria for correlation coefficients (i.e., $r = .50$). Therefore, given the high redundancy between reading and mathematics, we would expect the NSS to show high, positive correlations with both mathematics and reading scores from other achievement tests—and we further anticipate that the NSS will show somewhat higher correlations with mathematics scores (convergent associations) than with reading scores (divergent associations).

Participants

Participants were drawn from six schools in the same public school district in northern Delaware. The children were part of a longitudinal investigation that evaluated number competencies over six time points, from the beginning of kindergarten to the middle of third grade (Jordan et al., 2009). The schools were selected because they served children from both low-income and middle-income families.

The investigation examined relationships between NSS scores obtained in the fall of kindergarten and reading and mathematics criterion obtained at the end of first grade ($N = 288$) and again at the end of third grade ($N = 211$). In the kindergarten sample, 54% of the children were boys, 56% had minority ethnic backgrounds, and 33% came from low-income families.

Mathematics Criteria

Children were given the Calculation and Applied Problems portions of the Woodcock-Johnson III (McGrew et al., 2007) for a composite mathematics achievement score (WJMath). The Calculation (WJCalc) subtest measures the ability to perform computations in a conventional written format. Applied Problems (WJApp) requires the child to analyze and solve orally presented mathematics problems in various contexts.

Reading Criteria

The Dynamic Indicators of Basic Early Literacy Skills–Sixth Edition (DIBELS; Good & Kaminski, 2002) included measures of letter naming fluency, phoneme segmentation fluency, and nonsense word fluency. The raw score for each measure was the number of letters, phonemes, and nonsense words identified in 1 minute. Scores from the three related measures were totaled for each child and used for the analysis. Average test–retest reliability for the end of kindergarten was .91 (Good, Simmons, Kame'enui, Kaminski, & Wallin, 2002).

The Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999) is a standardized measure. In the sight word subtest (Torgesen et al., 1999), students have 45 seconds to read words. The score is the number of correct words. Test–retest and alternate form reliability is $> .90$.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Table 4.12. Convergent and divergent associations for the Number Sense Screener and first-grade criteria

Criterion	Number Sense Screener
Convergent associations	
WJ grade-based calculation standard score	.60[a]
WJ grade-based applied problem standard score	.69
Average[b]	.65
Divergent associations	
DIBELS phoneme segmentation fluency	.12
DIBELS nonsense word fluency	.38
DIBELS oral reading fluency	.51
DIBELS word use fluency	.36
DIBELS retell	.36
Average[b]	.35

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

Note: $N = 288$; WJ, Woodcock-Johnson; DIBELS, Dynamic Indicators of Basic Early Literacy Skills.

[a] All numbers rounded at second decimal point for convenient presentation.

[b] Average coefficients were calculated with Fisher's z' transformation.

Results

Outcomes from analyses completed at the end of first grade are shown in Table 4.12. The NSS showed high correlations (convergent validity) with mathematics scales from the WJMath designed to measure similar attributes. For example, the two correlations between the NSS and WJMath produced an average $r = .65$. The associations were both appreciable and theoretically congruent. Moreover, no scale from the DIBELS reading measure showed a correlation as high. The average correlation between the NSS and the DIBELS was .35. This average also was theoretically congruent, because it was lower than correlations between the NSS and the WJMath. Consequently, it is reasonable to infer from the pattern of convergent and divergent association in Table 4.13 that the NSS shows substantial construct validity.

Table 4.13 shows outcomes obtained at the end of third grade. The pattern was nearly identical to that found in first grade. Once again, a pattern of higher convergent validity (average $r = .62$) was obtained than divergent validity ($r = .37$), with results serving to further enhance assertions that the NSS possesses high levels of construct validity.

Table 4.13. Convergent and divergent associations for the Number Sense Screener and third-grade criteria

Criterion	Number Sense Screener
Convergent associations	
WJ grade-based calculation standard score	.57[a]
WJ grade-based applied problem standard score	.66
Average[b]	.62
Divergent associations	
TOWRE grade-based standard score	.37
Average[b]	.37

Reprinted from *Learning and Individual Differences, 20,* Jordan, N.C., Glutting, J., & Ramineni, C., The importance of number sense to mathematics achievement in first and third grades, 82–88, (2010) with permission from Elsevier.

Note: $N = 288$; WJ, Woodcock-Johnson; TOWRE, Test of Word Reading Efficiency.

[a] All numbers rounded at second decimal point for convenient presentation.

[b] Average coefficients were calculated with Fisher's z' transformation.

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved

FOR MORE, go to http://www.brookespublishing.com/NSS

Treatment Validity

Dyson, Jordan, and Glutting (in press) examined the effects of an 8-week number sense intervention to develop number competencies among low-income kindergartners (N = 121). The intervention purposefully targeted whole-number concepts related to counting, comparing, and manipulating sets. Children were randomly assigned either to a number sense intervention or a "business as usual" contrast group. The intervention was carried out in small-group, 30-minute sessions, 3 days per week for a total of 24 sessions.

The intervention was based on the premise that weaknesses in key number competencies underlie mathematics difficulties and that these competencies can be developed early through purposeful instruction. It targeted number concepts related to counting, comparing, and manipulating sets. The study used a pretest, posttest, and delayed posttest design. Children were randomly assigned either to the intervention condition or a business as usual control group. Dependent variables included a validated assessment of numeracy indicators as well as a standardized measure of mathematics achievement.

Participants

Children were recruited from kindergarten classes in five schools serving high-risk children from low-income urban families. All schools were in the same district. A total of 121 participants completed the study. Fifty-two of the children were girls (43%) and sixty-nine were boys (57%). Sixty-seven of the students were identified as African American (55%), forty-five as Hispanic (37%), seven as Caucasian (6%), one as Asian, and one as biracial, all by teacher report. Thirty of the students (25%) were identified as English language learners and were enrolled in designated kindergarten classrooms for English language learners.

In each of the five participating schools, about half of the participants within each classroom were randomly selected for the intervention, whereas the other half were assigned to a business as usual control group. Because the interventions were carried out in groups of four, there were a few extra children in the various schools who were assigned to the control group, accounting for the unequal numbers in the intervention and control conditions. Participants also were stratified according to kindergarten class.

Measures

The NSS was used as an outcome measure. Mathematics achievement was assessed using the Woodcock-Johnson III Tests of Achievement Form C Brief Battery: Applied Problems and Calculation subtests (WJ; Woodcock, McGrew, Schrank, & Mather, 2007).

Design and Procedures

A pretest, immediate posttest, and delayed posttest design was used. The intervention started in January of 2010 and was carried out in small groups of four children per instructor. The intervention groups met for three 30-minute sessions per week over an 8-week period for a total of 24 lessons.

During the week following the last lesson, children were individually posttested with the NSS and the WJ measures. Approximately 6 weeks later, children were tested again with the same measures.

Intervention

The intervention was designed to augment the regular kindergarten mathematics program with small-group instruction. Skills were reviewed incrementally over the course of the 24 lessons. A compare-and-contrast approach was used throughout the activities. For example, opposites such as before and after, addition and subtraction, $n + 1$ and $n – 1$ were presented simultaneously. The intervention content included the following areas:

Excerpted from Number Sense Screener™ (NSS™) User's Guide, K–1, Research Edition
by Nancy C. Jordan, Ed.D.; & Joseph J. Glutting, Ph.D.; with invited contributor Nancy Dyson, Ph.D.
Brookes Publishing | www.brookespublishing.com | 1-800-638-3775
© 2012 | All rights reserved