School Science and Mathematics Association
Founded in 1901

Photo Credit: Canva

SSMA Annual Convention
Colorado Springs, CO
October 19-21, 2023

PROCEEDINGS OF THE 122ND ANNUAL CONVENTION OF THE SCHOOL SCIENCE AND MATHEMATICS ASSOCIATION

October 19-21, 2023

Colorado Springs, CO

Editors: Rebekah Hammack and Beth Cory

Cite Proceedings Volume as:
Hammack, R. & Cory, B. (Eds.). (2023). *Proceedings of the 122nd annual convention of the School Science and Mathematics Association* (Vol. 10). Colorado Springs, CO: SSMA.

Cite Proceedings Chapter as:
Author. (2023). Title of chapter. In R. Hammack & B. Cory, (Eds.). *Proceedings of the 122nd annual convention of the School Science and Mathematics Association* (Vol. 10). Colorado Springs, CO: SSMA.

Website: www.ssma.org
The proceedings are also available via http://www.ssma.org/convention-proceedings

ISBN: 978-1-7343159-4-3

SSMA 2023 ANNUAL CONVENTION: COLORADO SPRINGS, CO

2023 Convention Coordination

Convention Chairs
Catherine Maiorca, Oklahoma State University
Thomas Roberts, Bowling Green State University

SSMA Leadership

President, 2020-2022
Margaret Mohr-Schroder, University of Kentucky

Past-President, 2018-2020
Christa Jackson, Saint Louis University

Directors-at-Large, 2020-2023
Susan Cooper, Florida Gulf Coast University
Stephen Scogin, Hope College

Directors-at-Large, 2021-2024
Dittika Gupta, Midwestern State University
Rebekah Hammack, Purdue University

Directors-at-Large, 2022-2025
Andrea Foster, Sam Houston State University
Sean Yee, University of South Carolina

Co-Executive Directors, 2019-2023
Stephanie Hathcock, Oklahoma State University
Toni Ivey, Oklahoma State University

Journal Co-Editor, 2021-2026
Christie Martin, University of South Carolina
Bridget Miller, University of South Carolina

Proceedings Editors
Rebekah Hammack, Purdue University
Beth Cory, Sam Houston State University

Special thanks to

The Publications Committee and SSMA members for reviewing and providing valuable feedback on the submitted manuscripts.

Hammack, R. & Cory, B. (Eds.). (2023). *Proceedings of the 122nd annual convention of the School Science and Mathematics Association* (Vol. 10). Colorado Springs, CO: SSMA

School Science and Mathematics Association
Founded in 1901

The School Science and Mathematics Association (SSMA) is an inclusive and thriving professional community comprising educators and researchers dedicated to enhancing the fields of school science and mathematics. Since its inception in 1901, SSMA has remained a steadfast platform for prominent mathematics, science, and STEM educators to showcase their research and publish scholarly works for over 120 years.

SSMA's primary focus lies in fostering research-based innovations within K-16 teacher preparation and continuous professional development in the realms of science and mathematics. This organization caters to a diverse audience, including higher education faculty, K-16 school administrators, and K-16 classroom instructors.

SSMA's mission can be summarized by four key goals:

- Cultivating a close-knit community of educators, researchers, scientists, and mathematicians.
- Advancing knowledge through rigorous research in science and mathematics education and their effective integration.
- Informing and enriching teaching practices by disseminating scholarly works within the fields of science and mathematics.
- Influencing education policies in science and mathematics at local, state, and national levels.

The proceedings of the 122nd Annual Convention represent SSMA's rich traditions and its promising future, serving as a testament to its enduring commitment to the advancement of science and mathematics education.

Margaret Mohr-Schroder
SSMA Presiden

PREFACE

These proceedings are a written record of some of the research and instructional innovations presented at the 122nd Annual Meeting of the School Science and Mathematics Association held October 18-21, 2023, in Colorado Springs, CO. The blinded, peer reviewed proceedings include seven papers regarding instructional innovations and research. The acceptance rate for the proceedings was 64%. We are pleased to present these Proceedings as an important resource for the mathematics, science, and STEM education community.

Rebekah Hammack & Beth Cory
Editors

Hammack, R. & Cory, B. (Eds.). (2023). *Proceedings of the 122nd annual convention of the School Science and Mathematics Association* (Vol. 10). Colorado Springs, CO: SSMA

TABLE OF CONTENTS
VOLUME 10

EVALUATING THE UNIVERSAL SCREENERS FOR NUMBER SENSE: A VALIDATION STUDY

Timothy D. Folger
tdfolge@bgsu.edu
Bowling Green State University

Jonathan D. Bostic
bosticj@bgsu.edu
Bowling Green State University

David Woodward
david@forefront.education
Forefront Education

Abstract

The Universal Screeners for Number Sense (USNS) measure the construct of number sense, which has been characterized in numerous ways over the last 50 years. A purpose of this study is to provide validity evidence regarding the USNS assessments, which may support scholarship on numeracy and practical work in K-12 schools. Findings indicate that the USNS assessments may be used confidently to measure students' growth in number sense.

Keywords: number sense, formative assessment, validity, validation

Introduction

Assessment is central to understanding what and how students learn (Black & Wiliam, 1998). Understanding students as individuals, their reasoning and problem solving skills, and knowledge that students bring to the classroom has potential to enhance students' learning outcomes (National Research Council, 2001). Put simply, teachers are better poised to support student learning when the teachers have greater knowledge of their students. Unfortunately, the information that teachers rely on to understand their students' knowledge and abilities are collected through distal standardized assessments that create vague pictures of student performance (Popham, 2001). Such assessments may not be designed to provide teachers with a detailed understanding of students' strengths and weaknesses. The inferences that teachers draw from assessment results, and how teachers subsequently use that information, reflect issues of validity (AERA et al., 2014; Folger et al., 2023). The Universal Screeners for Number Sense (USNS) are a series of K-6 open source assessments that integrate interview and written tasks. These assessments help teachers to hone in on the key skills, concepts, and developmental milestones of students. Through interview tasks and corresponding item rubrics, teachers are provided with a structure for (1) observing students as they solve problems and (2) listening to student reasoning. This format provides teachers with rich details about their students' number sense.

Purpose

The purpose of this validation study is to evaluate the Universal Screeners for Number Sense (USNS) for use in grades K-5, as a measure of students' number sense. An outcome from this study is to share evidence to support the use and interpretations of the USNS, which may interest

university scholars, teacher educators, and K-12 practitioners and administrators. The following four claims inherent to the USNS are the focus of this study: (1) USNS assess K-5 students' number sense in ways that are aligned to the Common Core State Standards for Mathematics; (2) USNS are effective formative and progress-monitoring assessments; data can be used to interpret student growth between time periods; (3) USNS grade-level assessments demonstrate effective construct development to measure students' number sense; (4) USNS assessment items function reasonably well in collectively measuring a single construct (i.e., number sense).

Related Literature

Validity is an attribute of the interpretation of test scores for proposed uses of tests (AERA et al., 2014). The degree to which an interpretation and use of test scores is valid depends on the quantity and quality of supporting evidence (AERA et al., 2014). *The Standards for Educational and Psychological Testing* describe five sources of validity evidence: test content, response processes, internal structure, relations to other variables, and consequences of testing (AERA et al., 2014). Each source of evidence, briefly described in Table 1, supports differing claims inherent to the interpretation, and subsequent use, of test scores (Folger et al., 2023). For instance, evidence based on relationships to other variables may support claims that performance predicts future academic achievement. An additional example, evidence based on test content may support claims that test items align to a specific set of mathematical standards. Scholars agree that a central goal of validation is to construct and evaluate "arguments for and against the intended interpretation of test scores and their relevance to the proposed use" (AERA et al., 2014, p. 11). One framework to guide a validity argument is an interpretation and use statement, which has a focus on claims about interpretations and uses for an assessment, and supporting evidence (Carney et al., 2022).

Construct Articulation

The USNS assessments measure the construct of number sense, which has been characterized in numerous ways over the last 50 years. Sowder (1989) describes number sense as (a) being able to conceptually relate number and operation properties; (b) the ability to use number magnitude to compare numbers, recognize when calculation results are not reasonable, and to apply non-standard algorithmic strategies when performing mental calculations; and (c) being able to solve problems involving numbers using flexible and creative strategies. Kalchman and colleagues (2001) employ a similar definition: "Number sense is: (a) fluency in estimating and judging magnitude, (b) ability to recognize unreasonable results, (c) flexibility when mentally computing, [and] (d) ability to

move among different representations and to use the most appropriate representation characterized" (p. 2). Berch (2005) posits, "Possessing number sense ostensibly permits one to achieve everything from understanding the meaning of numbers to developing strategies for solving complex math problems" (p. 334). Collectively, these characterizations (Berch, 2005; Kalchman et al., 2001; Sowder, 1989) present number sense as encompassing topics including but not limited to fluency, magnitude, number recognition, representation, and communication. These topics are also taken up in the National Council of Teachers of Mathematics (2000) framing of number sense.

Table 1

Description of the Five Sources of Validity Evidence (AERA et al., 2014)

urce of Validity Evidence	**Description**
Test Content	Test content includes the wording and format of test items or tasks. Validity evidence based on test content would indicate that test items, or test content, align to the construct a test intends to measure.
Response Processes	Response processes describe the alignment between test takers' performance or behavior and the construct a test intends to measure. In cases when a test relies on observers or judges to evaluate test takers, evidence may include "the extent to which the processes of observers or judges are consistent with the intended interpretation of scores" (AERA et al., 2014, p.15).
Internal Structure	Internal structure may indicate the degree to which test items conform to the construct a test intends to measure. Such evidence may be collected through analysis of test dimensionality and item interrelationships.
Relations to Other Variables	Relations to other variables examines the degree to which test scores are, or are not, related to some ancillary variable. The *Standards* describe several examples when relations to other variables may be of interest, such as: (a) hypothesized differences in group performance, (b) the degree to which test scores predict future performance, and (c) whether test scores from different tests measuring a similar construct produce a convergent association.
Consequences of Testing	Consequences of testing presents the intended and unintended consequences following the interpretation and use of test scores. Consequential evidence evaluates "the soundness of [test score] interpretations for their intended uses" (AERA et al., 2014, p. 19).

Operationalization and Administration

The USNS consists of interview-based and paper assessments designed to assess student's number sense. Assessments are administered three times a year: fall, midyear, and spring. The

interview assessments are designed to be completed in approximately five minutes. The written portion is designed to be completed in approximately 10 minutes.

As screening assessments, the USNS are designed to help teachers identify students who might be at risk of accessing and engaging with grade-level mathematics content. However, these assessments also are intended to help teachers perceive trends in students' performance, and identify skills and topics that might need to be addressed in small groups or with their whole class.

Method

This validation study employed a multi-method research design to examine the four validity claims. Two data sources were used for this study. First, a survey was distributed to an expert panel of classroom teachers, teacher leaders, and education professionals to collect quantitative and qualitative data related to USNS. Survey data were used to evaluate claim 1 and claim 2. The expert panel consisted of 19 education professionals who have administered and scored more than 11 different USNS screeners. All expert panel members had a minimum of four years' experience as a classroom teacher, with nearly all (i.e., 98%) panel members having nine or more years of classroom teaching experience. Expert panel members were purposely sampled based on their knowledge and experiences administering USNS. Over 93% of the expert panel had administered USNS more than 20 screeners in their experience, indicating familiarity with the K-5 screeners. Survey questions included both quantitative (i.e., Likert scale) and qualitative items. Quantitative items used a four-point Likert scale to measure panel members' view that USNS aligned to number sense concepts. Additionally, expert panel members identified CCSS standards measured by USNS items. The second data source consisted of de-identified USNS student results. Student data used for this study were collected during the 2021-2022 academic year, and were used to evaluate claim 3 and claim 4.

Data Analysis: Expert Panel Survey Data

Descriptive statistics (i.e., mean and standard deviation) were calculated for all quantitative survey data. Generally, variance, as measured by standard deviation, is considered as the degree to which responses cohere similarly. Standard deviation values for this survey may be considered as low (0-0.49), moderate (0.5-0.99), and high (>1.0). Low variance indicated agreement among experts. Inductive analysis (Hatch, 2002) was conducted by two researchers. Data were initially scanned by them to remove any data that were deemed inappropriate or incomplete. At the second stage, they re-read data and reviewed it for initial codes. Memos were made to record possible ideas that seemed

to highlight major ideas, and to support evidence leading towards possible themes. At the third stage of analysis, data and memos were reviewed and categorized to make initial themes. At the fourth stage, counterevidence was recorded and noted in light of the initial themes. Themes with a preponderance of evidence and minimal counter evidence were retained for further analysis. Quotations were used to contextualize themes and situate findings.

Data Analysis: Student USNS Data

Quantitative data were examined using Rasch (1960/1980) measurement, which constructs a linear statistical model from observed counts and categorical responses (Wright & Stone, 1999). Separation and reliability indices were used to examine the overall functioning of the USNS (i.e., claim 3). Put simply, separation and reliability can be thought of as an indicator of clarity in measuring the construct of number sense. Person separation reflects how well items separate test-takers' performance. Item separation reflects how well a sample of people separate item difficulty. In other words, person-separation values indicate a hierarchy of person ability whereas item-separation indicates a hierarchy of item difficulty. Higher values indicate better separation. Person reliability indicates consistency in how people with similar ability levels perform. Item reliability indicates consistency in item difficulty (or item performance). Reliability and separation indices are respectively classified as excellent at 0.90 and 3.0, good at 0.80 and 2.00, and acceptable at 0.70 and 1.50 (Duncan et al., 2003).

Regarding claim 4: Item fit and point-biserial statistics were used to examine the degree to which USNS assessment items measure a single construct. Item fit statistics and point-biserial correlations are appropriate indicators because "the unidimensionality requirement is satisfied when the data fit the model" (Smith, 1996, p. 26). Effective item fit is observed when the "Infit" and "Outfit" mean square (MNSQ) statistics lie within the range from approximately .5 to 1.5, and the standardized Z-statistics (ZSTD) lie within the range from approximately -2.0 to +2.0 (Linacre, 2002). Point-biserial correlations assess item quality and measure how items function in relation to one another. Correlations range from -1 to 1. Items producing a negative point-biserial are a concern and should be considered for removal from the analysis because such items fail to represent the construct being measured. Varma (2006) suggests good items have point-biserial correlations greater than 0.25. These parameters were examined for each K-5 USNS assessment.

Results and Discussion

Claim 1: USNS assess K-5 students' number sense in ways that are aligned to the Common Core State Standards for Mathematics

Validity evidence based on test content supports the claim that USNS assess K-5 students' number sense in alignment to the Common Core State Standards for Mathematics (CCSSM). The expert panel strongly agreed that the assessments reflect number sense topics, as seen by the mean score of 3.62 units on a four-point scale. Variance was low (SD = 0.49), which indicates substantial agreement across the panel members. Qualitative feedback confirmed the quantitative results. "[USNS] confirms students' understanding of foundational concepts with place value, whole number operations and fractions that will allow the student to progress on to new grade level content in those domains." Another respondent wrote: "Most problems involve assessing if students are flexible with how they mentally compute." Furthermore, expert panel members indicated that USNS items are aligned to the CCSSM. Respondents noted a strength of the kindergarten USNS was a focus on standards primarily located in (1) Counting and Cardinality and (2) Operations and Arithmetic. Alignment between USNS items and CCSSM was consistent across grade levels.

Claim 2: USNS are effective formative and progress-monitoring assessments; data can be used to interpret student growth between time periods

Validity evidence based on consequences of testing support the claim that USNS are effective formative and progress-monitoring assessments; data can be used to interpret student growth between time periods. A consistent theme from qualitative analysis of survey data illuminated a finding that the USNS screeners provided useful information about K-5 students' number sense. As one teacher indicated, "We use them 3 times per year to help us make MTSS [multi-tiered systems of support] decisions." A second panel member added that "teachers plan small group assessments… that align [with] curricular materials as well as different intervention programs." These comments, as well as others, indicated that the K-5 USNS are useful for learning about K-5 students' number sense, which aligns with their intended use.

Similarly, a consistent theme drawn from qualitative analysis indicated that the USNS materials provide information that can inform who may need further instructional intervention or follow-up assessments. One panel member reported that "The screeners are our primary form of data for all our data meetings in K-2." Another panel member reported that "We input data into our district platform and use the data output to group students…as well as looking at trends and needs

by question type". These comments were indicative of the theme that data were used in ways to support instruction throughout an academic year because they provide a measure of growth during the academic year.

Claim 3: USNS grade-level assessments demonstrate effective construct development to measure students' number sense

Validity evidence based on internal structure supports the claim that USNS grade-level assessments demonstrate effective construct development to measure grade-level students' number sense. Rasch separation and reliability indices are reported in Table 2. Person separation and reliability range from good to excellent for each grade-level assessment series (Duncan et al., 2003), indicating that each grade-level assessment series effectively distinguishes between variations in students' number sense. Item separation and reliability statistics are respectively classified as excellent for each USNS grade-level assessment series, suggesting (a) USNS assessments demonstrate effective construct development, and (b) assessment items possess an appropriate hierarchy of item difficulty.

Table 2

Separation and reliability statistics for each grade-level assessment series

	Persons		Items	
	Separation	Reliability	Separation	Reliability
Kindergarten (n = 1,453)	2.36	0.85	14.42	1.00
Grade 1 (n = 1,675)	2.16	0.82	11.45	0.99
Grade 2 (n = 1,524)	2.79	0.89	14.08	0.99
Grade 3 (n = 1,408)	3.33	0.92	14.98	1.00
Grade 4 (n = 1,140)	3.45	0.92	13.63	0.99
Grade 5 (n = 319)	3.80	0.94	8.16	0.99

Claim 4: USNS assessment items function reasonably well in collectively measuring a single construct (i.e., number sense)

Validity evidence based on internal structure supports the claim that USNS assessment items function reasonably well in collectively measuring a single construct (i.e., number sense). Across all 212 USNS items, point biserial correlations ranged from 0.35 to 0.78, suggesting assessment items work together in measuring a single construct. Nine of the 212 USNS items were flagged for

exceeding Mean square (MNSQ) and Z-standardized (ZSTD) fit-statistic parameters. Each of these 9 items possessed MNSQ values between 1.5 and 2.0, which Linacre (2002) suggests is unproductive but not degrading to the measurement model. No other items were flagged for exceeding multiple parameters, supporting the claim of unidimensionality.

Implications

This validation study evaluates four claims inherent to using the USNS as a formative and progress monitoring assessment measuring K-5 students' number sense. Relative to these claims, we present validity evidence based on (1) test content, (2) consequences of testing, and (3) internal structure. Evidence of reliability is also reported alongside validity evidence based on internal structure. Validation is an ongoing process (AERA et al., 2014), and additional validity evidence, particularly evidence of response processes and relations to other variables, would strengthen the argument supporting the use of USNS to measure K-5 students' number sense. However, results presented in this study support the use of K-5 USNS as formative assessments and progress monitoring tools as a part of classroom instruction. Teachers may confidently use the open-source USNS as a part of their instruction and assessment practices to draw detailed inferences of student knowledge and ability, and use this data to support students' individual learning needs.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of learning disabilities*, *38*(4), 333-339.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy & practice*, *5*(1), 7-74.

Carney, M. B., Bostic, J., Krupa, E., & Shih, J. (2022). Interpretation and use statements for instruments in mathematics education. *Journal for Research in Mathematics Education*, *53*(4), 334-340.

Duncan, P. W., Bode, R. K., Lai, S. M., Perera, S., & Glycine Antagonist in Neuroprotection Americas Investigators. (2003). Rasch analysis of a new stroke-specific outcome scale: the Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950-963.

Folger, T., Bostic, J., & Krupa, E. (2023). Defining test-score interpretation, use, and claims: Delphi study for the validity argument. *Educational Measurement: Issues and Practice*, *42*(3), 22-38.

Hatch, J. A. (2002). *Doing qualitative research in education settings*. Suny Press.

Kalchman, M., Moss, J., & Case, R. (2001). Psychological models for the development of mathematical understanding: Rational numbers and functions. *Cognition and instruction: Twenty-five years of progress*, 1-38.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions, 16*(2), 878.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics.*

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. National Academies Press.

Popham, W. J. (2001). Teaching to the Test?. *Educational leadership*, *58*(6), 16-21.

Rasch, G. (1960/1980). Probabilistic models for some intelligence and attainment tests. (Copenhagen, Danish Institute for Educational Research), with foreward and afterword by B.D. Wright. Chicago, IL: The University of Chicago Press.

Smith, R. M. (1996). A comparison of methods for determining dimensionality in Rasch measurement. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(1), 25-40.

Sowder, J. T. (1989). Research into practice: Developing understanding of computational estimation. *The Arithmetic Teacher, 36*(5), 25-27.

Varma, S. (2006). Preliminary item statistics using point-biserial correlation and p-values. Educational Data Systems Inc.: Morgan Hill CA.

Wright, B. D., & Stone, M. H. (1999). *Measurement essentials.* Wilmington, DE: Wide Range.