Article type: Brief Report

Psychometric Evidence for the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB Screener): An Online Assessment of Mathematical Learning Difficulties

Airi Hakkarainen[1]

Eija Väisänen[2]

Pirjo Aunio[2]

Heidi Hellstrand[3]

Mikko-Jussi Laakso[4]

Anu Laine[2]

Pekka Räsänen[4]

Johan Korhonen[3]

[1]Affiliation: Department of Education, University of Jyväskylä, Jyväskylä, Finland

[2]Affiliation: Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

[3]Affiliation: Faculty of Education and Welfare Studies, Åbo Akademi University, Vaasa, Finland

[4]Affiliation: Turku Research Institute for Learning Analytics, University of Turku, Turku, Finland

Abstract:

Background: Although several paper-and-pencil and digital online measures have been developed to assess basic numeracy skills and identify mathematical learning difficulties in children, psychometric evidence of these measures is seldom thoroughly reported and published. Establishing the validity and reliability of educational measures is a fundamental part of evidence-based practice.

Objective: This study aimed to examine the test-retest reliability, longitudinal measurement invariance, and convergent validity of a new digital online dyscalculia screener, the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB), targeted to 9–16-year-old children.

Method: The participants were 358 children (165 boys and 193 girls) in grades 3, 5 and 7, who participated in the study at two time points. The children's numeracy skills were measured using two time-limited tests: the FUNA-DB online screener and a standardized paper-and-pencil basic arithmetic test, RMAT.

Results: Our results showed that the FUNA-DB has a strong test-retest reliability, displays measurement invariance over time, and is meaningfully related to a traditional paper-and-pencil screening test.

Conclusion: The psychometric evidence supports using the FUNA-DB to measure school-aged children’s number processing and arithmetical fluency across time.

Keywords: arithmetic fluency, mathematics, number processing, psychometrics, screening

Johan Korhonen
Faculty of Education and Welfare Studies
Åbo Akademi University
PB 311
65101, Vaasa
Finland

johan.korhonen@abo.fi

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Publication Ethics

Informed consent was obtained from all participants and their legal guardians included in the study.

APA ethical standards and the ethical guidelines of the Finnish National Board on Research Integrity were carefully followed in the conduct of the whole project.

Authorship

Airi Hakkarainen, writing-original draft, conceptualization, formal analysis

Eija Väisänen, writing-original draft, conceptualization, investigation

Pirjo Aunio, writing – review & editing, conceptualization, methodology

Heidi Hellstrand, writing – review & editing, conceptualization,

Mikko-Jussi Laakso, writing – review & editing,

Anu Laine, writing – review & editing, conceptualization,

Pekka Räsänen, writing – review & editing, conceptualization, methodology

Johan Korhonen, writing – review & editing, conceptualization, formal analysis, methodology

All authors approved the final version of the article.

Psychometric Evidence for the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB Screener): An Online Assessment of Mathematical Learning Difficulties

Together with number processing skills (Hellstrand et al., 2024; Skagerlund & Träff, 2016), arithmetic fluency is a key marker for mathematical learning difficulties (MLD), as it can be seen as a genetically distinct dimension of mathematics skills yet related to other mathematical performance indicators (Petrill et al., 2012). Although several paper-and-pencil and online measures have been developed to assess these basic numeracy skills and identify MLD in children aged 9–12 years, only limited psychometric evidence of these measures has been published in peer-reviewed journals (Hakkarainen et al., 2023). Standards for educational measurement encourage researchers to report various reliability and validity evidence for measures developed to assess children's mathematical performance to secure the trustworthiness of decision-making in educational practice (American Educational Research Association [AERA], American Psychological Assessment [APA], and National Council on Measurement in Education [NCME], 2014). In addition, AERA, APA, and NCME (2014) recommended publishing information about digital measures' specific features or requirements compared to traditional paper-and-pencil measures. For the identification process of children with MLD, digital online measures provide a valuable opportunity to analyze the response time, as a slow response time indicates nonfluency in basic arithmetic and number processing, which is an indicator of MLD (De Smedt et al., 2013).

Previous studies concerning mathematical measures have mostly reported psychometric evidence related to structural validity and internal consistency, content validity , and known group validity (Karagiannakis et al., 2014, 2017; Li & Yang, 2010), but concurrent and longitudinal evidence related to structure has been seldom reported. Longitudinal evidence and convergent validity are highly relevant for identifying and predicting mathematical learning difficulties (Mazzocco & Räsänen, 2013).

To further strengthen the existing psychometric evidence of FUNA-DB (Hellstrand et al., 2024; Räsänen et al., 2021), this study aimed to examine (1) test-retest reliability, (2) longitudinal measurement invariance (LMI), and (3) convergent validity of the FUNA-DB.

Method

Samples

The data were collected with the help of voluntary teachers in the fall semester of 2021 in five comprehensive schools in a middle-sized town in Southern Finland. The teachers received written instructions on administering the FUNA-DB and RMAT tests. A total of 358 children (165 boys and 193 girls) from third (n = 93; girls 50.5%), fifth

(*n* = 131; girls 51.9%), and seventh (*n* = 134; girls 58.2%) grade participated in the study at two measurement points. In the Finnish context, children in grades 3–7 are generally 9–14 years old. The inclusion criteria were that the participants attended general education classrooms and were able to use a computer or tablet to complete tasks independently. There were no explicit exclusion criteria; however, the teacher could opt out of conducting assessments with students with severe difficulties (e.g., perceptual or motor challenges). Informed consent was obtained from participants and their legal guardians.

Measures

FUNA-DB

The FUNA-DB (Räsänen et al., 2021) aims to identify children with MLD from the third to ninth grades (9–16 years). It runs on an online collaborative learning platform called ViLLE (Laakso et al., 2018), works on all widely used internet browsers and operating systems, and can be used on a computer or tablet. The FUNA-DB includes two subscales: *number processing skills* and *arithmetic fluency*. Number processing skills include two tasks: number comparison (NC; 52 items) and digit–dot matching (DM; 42 items). For the analysis, an efficiency index was calculated for each task by dividing the median response time of the correct responses by the accuracy scores. The arithmetic fluency scale includes four time-limited tasks: number series (NS; 3 min time limit), single-digit addition (ADD; 2 min time limit), single-digit subtraction (SUB; 2 min time limit), and calculations (CA; mixed addition/subtraction; 3 min time limit). The sum scores of correctly answered tasks were calculated for each task. The internal consistency was good at both measurement points for NS (Cronbach's alpha = .90 for both), ADD (Cronbach's alpha = .97 for both), SUB (Cronbach's alpha = .97 for both), and CA (Cronbach's alpha = .94 for both).

RMAT

RMAT (Räsänen, 2004) test is a widely used standardized test of basic arithmetic skills in Finland. It is a time-limited (10 min) paper-and-pencil test for grades 3–6 to assess fluency in basic arithmetic skills (max 56 pts). The internal consistency was good, with a Cronbach's alpha of .79. According to the test manual, evidence for RMAT's test-retest reliability with a 6-month interval is good (*r* = .83; Räsänen, 2004). All descriptives and correlations of the measures can be found in Table S1 in the electronic supplementary material (Korhonen et al., 2024a).

Analytic Approach

We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses, including all tested models. For correlations, we report exact p values and effect sizes in terms of correlation coefficients. The analyses were carried out using SPSS (version 28) and Mplus (version 8.9) statistical software. General guidelines for conducting longitudinal structural equation modelling guided our sample size (Wolf et al., 2013). Missing values analysis indicated that data were missing completely at random (MCAR), with Little's MCAR test yielding $\chi 2(193{,}398) = 90{,}753.58$, $p = 1.000$. Full information maximum likelihood (FIML), which uses all available data, was used as the estimator in all the analyses. To date, two large-scale studies (Hellstrand et al., 2024 [N=18,405]; Räsänen et al., 2021 [N=4,265]) have been published that support the 2-factor structure of the measure reported in this manuscript. Hence, proceeding directly with CFA in the main analyses was justified. The LMI of the FUNA-DB over two measurement points was tested using confirmatory factor analysis (CFA). Specifically, the configural, metric, and scalar invariances were tested over the 2–4-week interval. In LMI, the data are fitted with a series of nested models ranging from a model with no invariance constraints to a model in which factor loadings and intercepts are forced to equality across the measurement points. In the longitudinal measurement model, the number comparison and digit-dot matching tasks were specified as factor indicators for the number processing, and number series, single-digit addition, single-digit subtraction, and calculation tasks for the arithmetic fluency factor (see Räsänen et al., 2021) at both measurement points. The corresponding indicator residuals were allowed to correlate over time to account for the method effects in the measures.

We used the chi-square test (χ^2), root mean square error of approximation (RMSEA) with a value below .06, comparative fit index (CFI), and Tucker–Lewis index (TLI) with values above .95, and the standardized root mean square residual (SRMR) with a value below .90 as indicators of a good fitting model. To compare the nested models, the change in the CFI and RMSEA was considered: a more parsimonious model would show a change in the CFI (ΔCFI) of less than 0.01 or a change in the RMSEA (ΔRMSEA) of less than 0.015 (Chen, 2007). To examine the convergent validity of the FUNA-DB, the number processing skills and arithmetic fluency factors at both measurement points were correlated with the sum score of RMAT.

Results

Table 1 summarises the LMI results across the two measurement points for the FUNA-DB. Metric invariance across time for the FUNA-DB was supported since the model with equal factor loadings fitted the data and the confi-

gural model (ΔRMSEA < 0.015, ΔCFI < 0.01). Scalar invariance across time was also supported, as constraining factor loadings and item intercepts to equality across time points did not significantly worsen model fit (ΔRMSEA < 0.015, ΔCFI < 0.01). Therefore, the results indicate that the FUNA-DB's two-factor structure holds across time, and consequently, developmental changes in the two subskills can be reliably detected using this online screener. The FUNA-DB displayed excellent test-retest reliability, as the arithmetic fluency (r = .90) and number processing (r = .84) factors had high correlations between time points. Similarly, the test-retest correlations were also high for the individual tasks (see Table S1 in Supplementary material). Carry-over effects (i.e., improved performance due to repeated testing, reflected in significantly improved mean value at the latent level) were observed in the number processing factor but not in the arithmetic fluency factor. Convergent validity was supported as both number processing (r = -.69), and arithmetic fluency (r = .81) were meaningfully related to RMAT at time point 1. As the time interval between the measurement points varied across classes, we reran the final model with a time difference score as a covariate for the T2 constructs. Adding the covariate did not have a significant effect on the parameter estimates. We report this additional analysis in Table S2 in the supplementary material (Korhonen et al., 2024a).

Discussion

This study aimed to evaluate the psychometric evidence of a digital online math screener by examining its test-retest reliability, longitudinal measurement invariance, and convergent validity. Our results indicate that the FUNA-DB's two-factor structure, number processing and arithmetic fluency, are invariant over time and show strong test-retest reliability. In addition, we demonstrated the convergent validity of this measure, as the performance in both number processing and arithmetic fluency was in line with the performance in the standardized basic arithmetic test RMAT.

These findings show that number processing and arithmetic fluency can be reliably measured in school-aged children. Consequently, developmental changes in these skills can be detected using the FUNA-DB. In addition, two previous studies showed that the FUNA-DB can be used reliably among different groups (e.g., gender and language) (Hellstrand et al., 2024; Räsänen et al., 2021). We stress the importance of examining the stability of the structure of a measure first and then the test-retest reliability, as Brown (2006) recommended. Even though psychometric evidence has been reported related to other digitalized MLD screeners, test-retest and longitudinal invariance have seldom been reported, even though their relevance for the identification of learning difficulties (De Smedt et al., 2013; Mazzocco & Räsänen, 2013). Similar tasks included in FUNA-DB (e.g., number comparison, digit-dot matching, and fluency measures) have been used extensively in research on MLD, and our findings support the continuous use of these tasks.

To conclude, by confirming the psychometric evidence of FUNA-DB, we can assert that the test is reliable and valid, measuring school-aged children's number processing and arithmetical fluency across time. FUNA-DB shows the stability required for measures used to evaluate the effectiveness of educational interventions. Further research is needed regarding sensitivity to detect changes in different types of educational interventions. Our results confirm that FUNA-DB is a potential screening toolfor clinical and educational practice in identifying children with MLD.

References

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

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford.

Chen, F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14,* 464e504. https://doi.org/10.1080/10705510701301834

De Smedt, B., Noël, P-L., Gilmore, G., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Hakkarainen, A., Cordier, R., Parsons, L., Yoon, S., Laine, A., Aunio, P., & Speyer, R. (2023). A systematic review of functional numeracy measures for 9–12-year-olds: Validity and reliability evidence. *International Journal of Educational Research, 119,* 102172. https://doi.org/10.1016/j.ijer.2023.102172

Hellstrand, H., Holopainen, S., Korhonen, J., Räsänen, P., Hakkarainen, A., Laakso, M., Laine, A., & Aunio, P. (2024). Arithmetic Fluency and Number Processing Skills in Identifying Students with Mathematical Learning Disabilities. *Research in Developmental Disabilities, 151,* 104795. https://doi.org/10.1016/j.ridd.2024.104795

Karagiannakis, G., & Papadatos, Y. (2014). Mathematical learning difficulties subtypes classification. Frontiers in Human Neuroscience, 8, 63709. https://doi.org/10.3389/fnhum.2014.00057

Karagiannakis, G., Baccaglini-Frank, A. & Roussos, P. (2017). Detecting strengths and weaknesses in learning mathematics through a model classifying mathematical skills. *Australian Journal of Learning Difficulties, 21*(2), 115–141. https://doi-org.ezproxy.jyu.fi/10.1080/19404158.2017.1289963

Korhonen, J., Hakkarainen, A., Väisänen, E., Aunio, P., Hellstrand, H., Laakso, M.-J., Laine, A., & Räsänen, P. (2024a). *Supplementary materials for: Psychometric Evidence for the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB Screener): An Online Assessment of Mathematical Learning Difficulties*. PsychArchives. https://doi.org/10.23668/psycharchives.15757

Korhonen, J., Hakkarainen, A., Väisänen, E., Aunio, P., Hellstrand, H., Laakso, M.-J., Laine, A., & Räsänen, P. (2024b). *Dataset for: Psychometric Evidence for the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB Screener): An Online Assessment of Mathematical Learning Difficulties* [Data set]. PsychArchives. https://doi.org/10.23668/psycharchives.15755

Korhonen, J., Hakkarainen, A., Väisänen, E., Aunio, P., Hellstrand, H., Laakso, M.-J., Laine, A., & Räsänen, P. (2024c). *Code for: Psychometric Evidence for the Functional Numeracy Assessment Dyscalculia Battery (FUNA-DB Screener): An Online Assessment of Mathematical Learning Difficulties*. PsychArchives. https://doi.org/10.23668/psycharchives.15756

Laakso, M.-J., Kaila, E., & Rajala, T. (2018). ViLLE–collaborative educational tool: Designing and utilizing an exercise-based learning environment. *Education and Information Technologies, 23*, 1655–1676. https://doi.org/10.1007/s10639-017-9659-1

Li, M. N. F., & Yang, D. C. (2010). Development and validation of a computer-administered number sense scale for fifth-grade children in Taiwan. *School Science and Mathematics*, *110*(4), 220–230. https://doi.org/10.1111/j.1949-8594.2010.00024.x

Mazzocco, M., & Räsänen, P. (2013). Contributions of longitudinal studies to evolving definitions and knowledge of developmental dyscalculia. *Trends in Neuroscience and Education*, *2*(2), 65-73. https://doi.org/10.1016/j.tine.2013.05.001

National Council of Teachers of Mathematics. (2023). *Common core state standards for mathematics*. Common Core State Standards. https://www.nctm.org/ccssm/

Petrill, S., Logan, J., Hart, S., Vincent, P., Thompson, L., Kovas, Y., & Plomin, R. (2012). Math fluency is etiologically distinct from untimed math performance, decoding fluency, and untimed reading performance: Evidence from a twin study. *Journal of Learning Disabilities*, *45*(4), 371–381. https://doi.org/10.1177/0022219411407926

Räsänen, P. (2004). *RMAT – laskutaidon testi 9–12 –vuotiaille* [Basic arithmetic test for 9–12-year-olds]. Niilo Mäki Instituutti.

Räsänen, P., Aunio, P., Laine, A., Hakkarainen, A., Väisänen, E., Finell, J., Rajala, T., Laakso, M-J., & Korhonen, J. (2021). Effects of gender on basic numerical and arithmetic skills: Pilot data from third to ninth grade for a large-scale online dyscalculia screener. *Frontiers in Education, 6*, 683672. https://doi.org/10.3389/feduc.2021.683672

Skagerlund, K., & Träff, U. (2016). Number processing and heterogeneity of developmental dyscalculia: Subtypes with different cognitive profiles and deficits. *Journal of Learning Disabilities, 49*(1), 36-50. https://doi.org/10.1177/0022219414522707

Wolf, E., Harrington, K., Clark, S., & Miller, M. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 73*(6) 913–934.

Open Science Section

Open Data: The information needed to reproduce all the reported results are openly accessible (Korhonen et al., 2024b). Open Materials: The information needed to reproduce all the reported methodology is openly accessible. Open Analytic Code: The code is openly accessible (Korhonen et al., 2024c). This study was not preregistered.

Table 1

Summary of Goodness-of-Fit for the Models Used in the Measurement Invariance Analysis Across the Two Measurement points for FUNA-DB's Subscales (Number Processing and Arithmetic Fluency)

Model	χ^2	df	*p*	RMSEA	CFI	TLI	ΔRMSEA	ΔCFI
Configural invariance	87.16	42	< .001	.055	.983	.973		
Metric invariance	94.25	46	< .001	.054	.981	.973	.001	.002
Scalar invariance	105.40	50	< .001	.056	.979	.972	.002	.003

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; TLI = Tucker-Lewis index.