

# The Importance of Replicating Meta-Analyses: Commentary on “Conceptual Replication and Extension of the Relation Between the Number Line Estimation Task and Mathematical Competence Across Seven Studies”

Michael Schneider<sup>1</sup>, Simon Merz<sup>2</sup>, Johannes Stricker<sup>3</sup>, Bert De Smedt<sup>4</sup>, Joke Torbeyns<sup>5</sup>, Lieven Verschaffel<sup>5</sup>,  
Koen Luwel<sup>5,6</sup>

[1] Educational Psychology, University of Trier, Trier, Germany. [2] Cognitive Psychology, University of Trier, Trier, Germany. [3] Clinical Psychology, University of Duesseldorf, Duesseldorf, Germany. [4] Parenting and Special Education, University of Leuven, Leuven, Belgium. [5] Instructional Psychology and Technology, University of Leuven, Leuven, Belgium. [6] Educational Research and Development, University of Leuven, Brussels, Belgium.

---

Journal of Numerical Cognition, 2021, Vol. 7(3), 479–482, <https://doi.org/10.5964/jnc.7617>

Published (VoR): 2021-11-30

**Corresponding Author:** Michael Schneider, Educational Psychology, Division I – Psychology, University of Trier, 54286 Trier, Germany. E-mail: [m.schneider@uni-trier.de](mailto:m.schneider@uni-trier.de)

**Related:** This article is part of the JNC Special Issue “Direct and Conceptual Replication in Numerical Cognition”, Guest Editors: Mojtaba Soltanlou & Krzysztof Cipora, Journal of Numerical Cognition, 7(3), <https://doi.org/10.5964/jnc.v7i3>

---

Ellis et al. (2021, this issue; in the following abbreviated as EEA) conducted a conceptual replication and extension of a recent meta-analysis relating number line estimation to broader mathematical competence (Schneider et al., 2018; in the following: SEA). EEA pooled and analyzed data from seven new studies and compared the results to SEA’s findings. As authors of SEA we briefly comment on three aspects of the conceptual replication study.

## 1. Are Replication Attempts of Meta-Analyses Useful?

Replications of meta-analyses and cross-validations of meta-analyses by comparing them against other (e.g., pre-registered original) studies are rare, and their relevance might not be obvious. However, replications and other cross-validations of meta-analyses are highly useful because meta-analyses generate new findings, so-called *synthesis-generated evidence*, for example, the overall effect size averaged across studies, the distribution of effect sizes, between-study heterogeneity, and the proportions of between-study heterogeneity explained by moderator variables. Conducting a meta-analysis and producing such synthesis-generated evidence is a multi-step process. In this process, researchers have to make many methodological decisions regarding the literature search, inclusion criteria, coding manual, bias correction, statistical models, and result interpretation. Further, when many original studies are biased similarly, the meta-analytic results are also biased. Consequently, it is not clear whether replications and cross-validations of a meta-analysis will replicate the results from the original meta-analysis. An impressive example is a recent investigation comparing 15 meta-analyses against preregistered multiple-laboratory replication studies (Kvarven et al., 2020). On average, the effect sizes found in the replication studies were three times smaller than those previously found in the meta-analyses. Thus, replication attempts of meta-analyses, like the one conducted by EEA, are valuable and essential for examining the validity and robustness of synthesis-generated evidence.

EEA did not use meta-analytic statistical methods to integrate the findings from their seven datasets and do not report typical outcomes of meta-analyses, for example, the between-study heterogeneity and distribution of the study

effect sizes. Instead, they analyzed individual participants' data pooled over studies. This limits the comparability of EEA and SEA. Generally, meta-analyses with individual participants' data, such as the data used by EEA, are possible and offer many advantages over meta-analyses averaging across effect sizes, as done by SEA (Cooper & Patall, 2009).

## 2. Commonalities and Differences of the Original and the Replication Study

EEA successfully replicated SEA's main finding that NLE correlates with broader mathematical competence but found differences regarding some of the moderating effects. Overall, we fully agree with the conclusion of EEA: "the current study successfully replicated the overall findings from the Schneider et al. (2018) meta-analysis." This robustness of the NLE-mathematical competence relation in EEA dovetails with the reliable correlations between the two constructs observed across various moderator levels in SEA. Methodologically, the high replicability of the NLE-mathematical competence association may be due to the fact that the NLE task is easy to explain, to understand, to administer, and to score. On the theoretical level, there is no doubt that number line estimation taps into central components of mathematical competence, notwithstanding the fact that there are many open questions regarding the details. The successful replication of the main finding supports this view.

EEA reported some findings regarding moderating effects that diverge from the findings of SEA. The interpretation of these findings is hampered by the fact that EEA did not conduct a direct replication, but a conceptual replication. That is, they tested the same main hypotheses as the original study but used different methods. When a conceptual replication and an original study differ in their findings, it is unclear whether the evidence from one of the two studies is invalid or whether methodological differences can explain the divergences. This is the case here, too. Table 1 lists the many methodological differences between EEA and SEA, potentially explaining differences in the findings.

**Table 1**

*Differences Between the Original Meta-Analysis (SEA; Schneider et al., 2018) and the Replication (EEA, Ellis et al., 2021, this issue).*

Characteristic	Original meta-analysis (SEA)	Replication and extension (EEA)
Study inclusion	PsycInfo search with standardized search string, exploratory search, explicit inclusion criteria	Datasets chosen by authors
Studies	72	7
Effect sizes	263	1
Total sample size	10576	954
Unit of analysis	Effect sizes coded from studies	Study participants pooled over studies
Method of analysis	Meta-analytic two-level regression, effect sizes weighted with inverse of standard error	Correlations with bootstrapped standard errors, multiple regressions
Age distribution	Children < 6 years, 6-9 years, and > 9 years	Mostly children < 6 years and 6-9 years
Number types	Whole numbers and fractions	Whole numbers
Number line tasks	Standard number lines with the endpoint labeled; non-standard number lines with one unit labeled	Standard number lines with the endpoint labeled; non-standard number lines with three points labeled, flashcards, bunny cover story
Mathematical competence measures	Counting, arithmetic, grades, standardized mathematical achievement tests	ANS acuity (Panamath), standardized mathematical achievement tests (WJAP, PENS-B)
Countries	Belgium, Canada, China, England, Germany, Italy, Luxemburg, Netherlands, Sweden, Scotland, USA	USA, Chile

For example, EEA found other age trends than SEA. However, SEA found that the age trends differed between estimation with whole numbers or fractions and only SEA but not EEA included studies with fractions. Other examples of relevant methodological differences are that EEA included a non-standard version of the NLE task involving nonsymbolic numerosities on flashcards, unusual marks on the number line, and a cover story including a bunny hopping on

the line. Further, SEA had explicitly excluded measures of basic numerical cognition from consideration as mathematical competence measures. By contrast, EEA included one such measure (*Panamath*), which assesses the acuity of the approximate number system by a numerosity comparison task and had been found to have a relatively low reliability (Inglis & Gilmore, 2014).

### 3. Extension of the Original Meta-Analysis

SEA reported that number line estimation predicts mathematical competence over time with  $r = .496$ . EEA extended these findings by analyzing how strongly number line estimation predicts later number line estimation and mathematical competence. Because they analyzed individual participants' data, they could elegantly control the longitudinal relations for potentially confounding variables, such as age and sex. In these analyses, EEA found that number line estimation proficiency was only a very weak predictor of later number line estimation proficiency and mathematical competence. These findings are surprising because many of the studies synthesized by SEA found that the number line estimation task assessed a high proportion of systematic variance, had a good reliability, and a good predictive validity. One explanation of the weak relations found by EEA is that their person-level data allowed them to control for person characteristics, that might have inflated the correlation in other studies. Another possible explanation is that a substantial proportion of the longitudinal data analyzed by EEA had been obtained using a non-standard number line task and using basic ANS acuity rather than a more classical measure of mathematical competence, for example arithmetic tasks or a school achievement test.

## Conclusion

Overall, EEAs' replication of the previous meta-analysis on number line estimation and broader mathematical competence is an innovative and useful contribution to the literature on numerical cognition. Any interpretation of the findings needs to carefully consider the many study characteristics that might moderate the relation between number line estimation and mathematical competence.

---

**Funding:** The authors have no funding to report.

---

**Acknowledgments:** The authors have no additional (i.e., non-financial) support to report.

---

**Competing Interests:** The authors have declared that no competing interests exist.

---

## References

- Cooper, H., & Patall, E. A. (2009). The relative benefits of meta-analysis conducted with individual participant data versus aggregated data. *Psychological Methods, 14*, 165-176. <https://doi.org/10.1037/a0015565>
- Ellis, A., Susperreguy, M. I., Purpura, D. J., & Davis-Kean, P. E. (2021). Conceptual replication and extension of the relation between the number line estimation task and mathematical competence across seven studies. *Journal of Numerical Cognition, 7*(3), 435-452. <https://doi.org/10.5964/jnc.7033>
- Inglis, M., & Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychologica, 145*, 147-155. <https://doi.org/10.1016/j.actpsy.2013.11.009>
- Kvarven, A., Strømmland, E., & Johannesson, M. (2020). Comparing meta-analyses and preregistered multiple-laboratory replication projects. *Nature Human Behaviour, 4*, 423-434. <https://doi.org/10.1038/s41562-019-0787-z>
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of number line estimation with mathematical competence: A meta-analysis. *Child Development, 89*, 1467-1484. <https://doi.org/10.1111/cdev.13068>



*Journal of Numerical Cognition* (JNC) is an official journal of the Mathematical Cognition and Learning Society (MCLS).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.