# UDL and Academic Achievement: Critical Analysis of the Meta-analysis by King-Sears and Colleagues (2023)

Steve Bissonnette, Ph.D.,
TÉLUQ University

## Introduction

The meta-analysis by King-Sears and colleagues (2023) is one of the most recent studies evaluating the impact of teaching based on Universal Design for Learning (UDL) on learners' academic achievement compared to traditional teaching. This meta-analysis selected 20 experimental studies covering learners from kindergarten to adulthood. Of the 20 selected studies, I conducted a secondary analysis retaining only those carried out in elementary and secondary education:

- Whose interventions lasted at least 12 weeks, as recommended by Robert Slavin in his *Best Evidence*;
- Involving a regular population of students;
- With teaching delivered in person.

Of the 20 studies in this meta-analysis, only two meet these criteria. Both were conducted in elementary school. Below is a critical analysis of these two studies.

## 1. Study by Proctor et al. (2011).

This study evaluated, over a period of 16 weeks, the impact of the digital ICON program, based on Universal Design for Learning (UDL), with 5th-grade students, including bilingual learners. The intervention, conducted in whole-group settings, targeted reading, vocabulary, and comprehension.

### Critical analysis

Methodological weaknesses and sampling limitations: This study relied on a quasi-experimental design in which assignment to the experimental or control groups was done at the class level rather than at the student level. With only 12 classes, the sample limited the statistical power of the study and prevented causal conclusions, potentially inflating the measured effect of the intervention. Additionally, the control group received "usual" instruction that was highly heterogeneous (literacy instruction time ranged from 55 to 135 minutes per day depending on the class). The researchers acknowledge that a control group benefiting

from a rigorously monitored alternative environment would have been more relevant.

Implementation issues (fidelity): The intervention planned the reading of 8 digital texts, but only 54% of the students in the experimental group completed the full program (with an average of 6.96 texts read). This forced the authors to analyze results based on “time on task” rather than simple group comparisons.

Instability of measures and lack of transfer: Although the tool improved vocabulary, it had no significant effect on overall reading comprehension (measured by the Gates-MacGinitie tests), which is the fundamental objective of any vocabulary intervention. To explain this lack of transfer, the authors suggest that standardized tests require secondary skills (such as time management and elimination of distractors) that their program did not teach. Finally, the researcher-developed tests showed serious weaknesses: the vocabulary breadth test (VBT) proved psychometrically unstable (lacking variance to detect effects), while the depth test (VDT) used a very narrow 4-point scale, which may have artificially amplified statistical sensitivity among different classes.

**2. Study by Yu et al. (2021)**

This study took place over a full year with 4th-grade students and aimed to structure scientific problem-solving processes. It evaluated an inclusive digital science notebook (SNUDLE) during sessions combining whole-class learning and autonomous work with technology.

**Critical analysis**

Major implementation issues (fidelity and dosage): The study suffered from insufficient implementation in the classrooms. Over the entire semester, teachers logged in on average only 9.63 days, and students completed only 6.53 investigations. This low engagement was worsened by a natural disaster during the first year (leading to school closures) and inconsistent access to tablets.

Limitations in collecting individual data: The evaluation of individual student use of the software was biased because students often worked in pairs or groups using a single student’s account. Thus, it is impossible for researchers to link specific feature use to individual academic outcomes with certainty.

Absence of overall effect and lack of transfer to national tests: The intervention showed no statistically significant impact on academic performance or motivation for

the general student population. Although the study found very large positive effect sizes (0.82 to 1.01) specifically for students with disabilities and positive effects for multilingual learners, a transfer problem persists. These academic gains were measured only through curriculum-aligned quizzes and district assessments; they did not transfer to the national standardized science test (MAP Science). This suggests the tool helps complete immediate tasks but does not necessarily improve broader standardized mastery of the subject.

## Conclusion

The analysis of these two studies highlights interesting but too limited results to convincingly support the effectiveness of UDL-based interventions in regular school settings. In the Proctor et al. (2011) study, methodological weaknesses—especially the quasi-experimental design, small sample size, and heterogeneity of the control group—limit the study's ability to establish a clear causal link between the digital intervention and learning outcomes. Implementation issues and measurement instability also affect interpretation, particularly the lack of transfer to standardized comprehension tests.

The study by Yu et al. (2021) presents another set of constraints, mainly related to implementation. Low usage of the digital notebook by teachers and students, combined with major external disruptions and limitations in individual data collection, weakens the robustness of the findings. The positive effects observed for certain subgroups, although promising, are closely linked to proximal assessments and do not transfer to broader measures such as national standardized tests. These two studies show that UDL-based interventions have not yet demonstrated strong, generalizable, and lasting effectiveness under real-life classroom conditions. Methodological limitations, implementation challenges, and the absence of transfer to standardized measures all point to the same conclusion: more rigorous, better-controlled, and consistently implemented studies are needed to reliably evaluate the true impact of UDL on academic achievement.

## References

King-Sears, M. E., Stefanidis, A., Evmenova, A. S., Rao, K., Mergen, R. L., Owen, L. S., & Strimel, M. M. (2023). Achievement of learners receiving UDL instruction: A meta-analysis. *Teaching and Teacher Education, 122*, Article 103956. https://doi.org/10.1016/j.tate.2022.103956.

Proctor, C., Dalton, B., Uccelli, P., Biancarosa, G., Mo, E., Snow, C., & Neugebauer, S. (2011). Improving comprehension online: Effects of deep vocabulary instruction with bilingual and monolingual fifth graders. *Reading and Writing, 24*(5), 517–544.

http://mutex.gmu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=ehh&AN=59928743&site=ehost-live.

Yu, J. W., Fikes, A. E., Ferguson, K., Wei, X., Tiruke, T., Hall, T. E., & Blackorby, J. (2019). Efficacy study of the science notebook in a Universal Design for Learning environment: Preliminary findings. *ERIC*. http://mutex.gmu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED605911&site=ehost-live.