Closing Learning Gaps Early: How Technology Drives Long-Term Literacy Success Across Demographics

Haya Shamir[1], Erik Yoder[1], Mikayla Tom[1], David Pocklington[1], and Alexander Houchins[1]

[1] Waterford Institute, Taylorsville, Utah, USA
{hayashamir, erikyoder, mikaylatom, davidpocklington, alexanderhouchins} @waterford.org

Abstract. This study examined the efficacy of Waterford Early Learning (WEL), a computer-assisted instruction (CAI) program, on K-3 student's literacy skills. Students from a school district in Louisiana ($N = 2{,}419$) were assessed to examine the long-term and early effects of the CAI program used in the classroom. Analysis revealed that students who used WEL for two and three years significantly outperformed students who used WEL for one year. Additionally, students who used WEL for one year and were tested a year after using the program significantly outperformed students who did not use WEL. Results were analyzed across demographics, including gender, race/ethnicity, special education status, and multi-lingual/monolingual learners. For all demographics assessed, students in the experimental group outperformed their control group counterparts. The results of this study contribute to the growing literature on the longitudinal impact of early educational CAI programs.

Keywords: Literacy, Technology, K-3.

1 Introduction

The acquisition of foundational literacy skills, including explicit knowledge of discrete phonological, morphological, and graphic units, starts early [1]. Literacy exposure before the start of formal education can have a lasting impact on student development. Research has indicated that early exposure to print has a moderate to strong impact on the later development of early literacy skills [2]. Mol and Bus (2011) theorized about a virtuous self-reinforcing cycle where early exposure promotes further incremental growth that in turn facilitates further exposure and further growth.

Effective early intervention can have an outsized impact [3] and forestall the need for later remediation [4]. A recent meta-analysis that aggregated research on educational apps that were targeted at children in elementary school grades found positive results ($d = 0.35$), but it also found larger effects for studies involving students in pre-kindergarten compared to later grades [3]. When 300 children who received a standard literacy intervention at six years old were revisited in a quasi-experimental design, the students who received this intervention significantly outperformed comparison students and had a notably lower proportion of students assessed as requiring additional services

by the time they were 16 years old [4]. In theory, this timely 'stitch in time' style of intervention should yield the best possible outcomes, but that is not always born out by the available body of research. When Fikrat-Wevers and her co-authors [5] reviewed the findings of 48 experimental studies, they found meaningful average effects (d = 0.50) when assessed at post-test, but they also found that these results were, on average, negligible when followed up on some time after the intervention.

Literacy development can differ between students for a wide range of factors, and these distinctions can become apparent early on. A quasi-experimental study examining over 5,000 students as they entered kindergarten found significant differences on individual literacy strands between male and female students [6]. In this study, McTigue and colleagues also gathered scores through first and second grade to conduct a longitudinal analysis and found gender-based disparities in literacy skill growth. Likewise, differing circumstances may cause students to have diverse reactions to the same intervention. The review by Fikrat-Wevers and colleagues [5] of literacy interventions found modest, but meaningful, effects on overall literacy domain skills and notably improved outcomes for students from low-SES backgrounds. In similar terms, a synthesis of research considering literacy interventions involving a subsample of dyslexic students found that students with dyslexia benefitted more from those interventions than their peers [7]. Effective early intervention can be a difference maker, enabling equitable outcomes for students from every background.

While there is a broad consensus that early interventions do yield positive effects in general and at least in the short-term it has become apparent in recent years that there is something of a credibility, or at least a comprehensibility, crisis when it comes to measuring learning outcomes. It is a difficult thing to take the complex and dynamic ecosystem of a classroom and to pare out the effect of an intervention apart from what a classroom, or an educator, and efforts to do so have not been without controversy [8], [9]. Studies have found promising initial effects vanishing when revaluated after a delay [5]. Research has indicated a particularly troubling negative correlation between the rigor of a study's design and the strength of its reported outcomes, with the less rigor in design equating to more profound results reported [10]. Other studies have indicated that, once the variability from individual assessments is taken into account, the magnitude of effect reported for literacy interventions drops off a very steep cliff [11]. Additionally, research into the longitudinal effects of literacy interventions have continually struggled to define a consistent relationship between the relative level of learning time and student outcomes, causing some authors to speculate that the existing body of research has explained only the possible floor effect associated with mere exposure [3].

This level of ambiguity in the available literature makes the need for further research explicit. Educators who work with young children have a limited window which must be leveraged to the greatest impact. Accordingly, these educators need to know that the tools they are using have been shown to be effective through rigorous testing. Longitudinal data needs to be evaluated to demonstrate that effects can persist over time. Dosage, and the effects of early intervention, needs to be studied to provide a rationale for extended use of any intervention.

The current study assessed the longitudinal efficacy of Waterford Early Learning (WEL), a computer-assisted instruction (CAI) early literacy intervention. It was

hypothesized that (a) students who used WEL over multiple years would outperform students who did not use the WEL, (b) students who started WEL earlier in their academic careers would benefit in ways that students who started later would not, and (c) all students who used WEL would benefit from its use.

2 Methods

2.1 Materials

Waterford Early Learning (WEL). This computer-assisted instruction (CAI) program delivers an adaptive reading curriculum for early learners. WEL uses a wide range of educational media, including interactive songs, stories, and activities; tailored in sequence to match each student's unique abilities and learning pace. The curriculum focuses on five literacy strands, phonics, phonological awareness, comprehension and vocabulary, language concepts, and fluency. Mastery of foundational literacy skills is ensured by the program before students progress to more advanced skills.

Dynamic Indicators of Early Literacy Skills (DIBELS). DIBELS is an assessment that evaluates the early literacy skills of students using short, fluency-based measures. The assessment was designed to track the development of early literacy skills for kindergarten through eighth-grade students. This assessment is administered by the school district as part of regular instruction. Research has found the DIBELS assessment to be both valid and reliable [12].

2.2 Participants

The sample for this study (N = 2,419) consisted of first-, second-, and third-grade students from a Louisiana school district who were assessed at the end of the 2022-2023 school year. To rule out demos of the program as well as accidental and incidental logins, this study employed a minimum usage criterion of 100 minutes of WEL learning time within a given year. Students in each grade were included in either the experimental or control group based on the number of school years they met the minimum usage criterion. In all, there were three pairs of experimental and control groups, one pair corresponding to each grade level, resulting in six analysis groups overall.

The WEL only in second-grade experimental group (n = 754) consisted of students who had more than 100 minutes of WEL learning time during second grade (2021-2022). The no WEL control group consisted of students who did not have any WEL learning time (n = 60). These students were assessed at the end of their third-grade school year and form the experimental and control groups when assessing long-term effects.

The three years of WEL experimental group (n = 722) consisted of students who had more than 100 minutes of WEL learning time during kindergarten (2020-2021), first grade (2021-2022), and second grade (2022-2023). The WEL only in second-grade control group (n = 83) consisted of students who had more than 100 minutes of WEL learning time during second grade (2022-2023). These students were assessed at the

end of their second-grade school year and form the experimental and control groups when assessing sustained early effects.

The WEL in kindergarten- and first-grade experimental group (n = 669) consisted of students who had more than 100 minutes of WEL learning time during kindergarten (2021-2022) and first grade (2022-2023). The WEL only in first-grade control group (n = 131) consisted of students who had more than 100 minutes of WEL learning in first grade (2022-2023). These students were assessed at the end of their first-grade school year and form the experimental and control groups when assessing early start effects.

2.3 Procedure

Kindergarten students were expected to use WEL in the classroom for 15 minutes per day, five days per week, alongside their school's traditional literacy instruction. First- and second-grade students were expected to use WEL in the classroom for 30 minutes per day, five days per week alongside their school's traditional literacy instruction. Learning time was tracked within the program and monitored weekly as well as at the end of the school year. Students who did not have WEL learning time received traditional literacy instruction.

Students used WEL for over 1,600 minutes per year (M = 1,673.73, SD = 723.92). Usage rose slightly from kindergarten (M = 1,540.94, SD = 645.63) into first- (M = 1,820.75, SD = 766.36) and second-grade (M = 1,711.66, SD = 751.14).

DIBELS was administered to students at the end of the 2022-2023 school year for first-, second-, and third-grade students, up to one year after students used WEL. Literacy test scores were analyzed to determine differences between the experimental and control groups.

3 Results

3.1 Third-Grade Students: No WEL vs WEL Only in Second Grade (Long-term Effects)

To analyze long-term effects, the experimental group included students who had more than 100 minutes of WEL learning time during their second-grade school year (2021-2022). The control group included students who did not have any WEL learning time.

Group Differences Using Independent Samples *t*-tests. Independent samples t-tests were conducted to examine group differences in end-of-third-grade scores between students who used WEL in second grade and students who did not use WEL. Analysis of DIBELS composite scores revealed a significant difference between groups, $t(1, 64) = -4.82$, $p < .001$, due to higher end-of-third-grade scores made by students who had used WEL in second grade (M = 456.07) than by students who did not use WEL (M = 425.48). Effect size (d = 0.65).

Independent samples t-tests were also conducted to examine group differences in DIBELS subskills (see Fig. 1). Analysis of DIBELS subskill scores revealed significant differences between groups in all subskills due to higher end-of-third-grade scores made by students who had used WEL in second grade than by students who did not use

WEL. Effect sizes ranged from 0.50 to 0.66, with the largest effect indicated for Oral Reading Fluency - Words Correct.

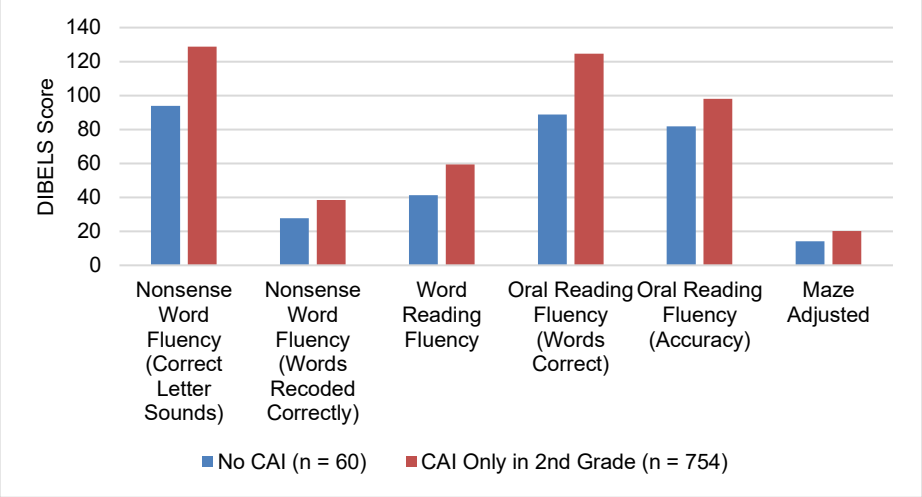

Fig. 1. 2022-2023 End-of-Third-Grade DIBELS Scores by Subskill.

Group Differences Using ANOVAs. Four separate two-way ANOVAs were conducted to examine the long-term effects of WEL and demographics on end-of-third-grade DIBELS composite scores (see Fig. 2). Across all demographics, students in the experimental group outperformed students in the control group. Simple effects analysis showed that for female and male students, African American/Black students, Latino/a students, Caucasian/White students, students with and without special education services, multi-lingual learners (MLL) and non-multi-lingual learners, students who used WEL only in second grade significantly outperformed students who did not use WEL.

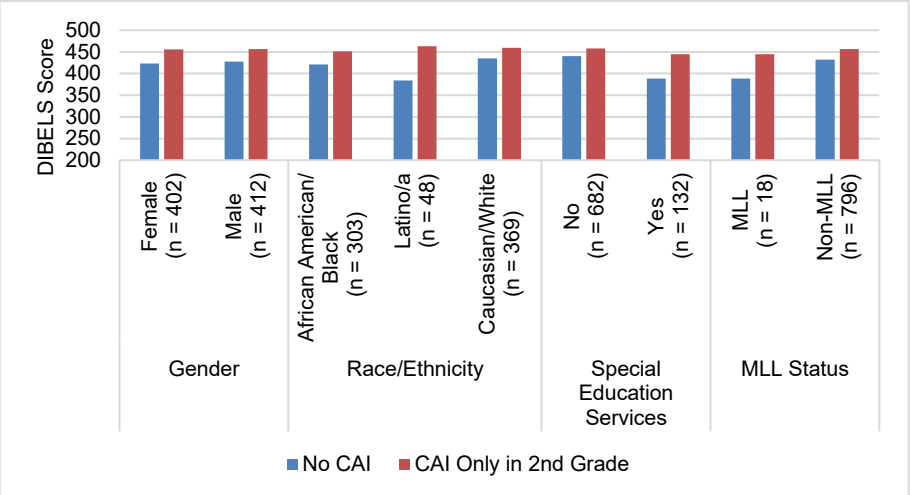

Fig. 2. 2022-2023 End-of-Third-Grade DIBELS Composite Scores by Demographics

3.2 Second-Grade Students: WEL Only in Second Grade vs Three Years of WEL (Sustained Early Effects)

To analyze sustained early effects, the experimental group included students who had more than 100 minutes of WEL learning time during kindergarten (2020-2021), first grade (2021-2022), and second grade (2022-2023). The control group included students who had more than 100 minutes of WEL learning time only during second grade (2022-2023).

Group Differences Using Independent Samples *t*-tests. Independent samples *t*-tests were conducted to examine group differences in end-of-second-grade scores between students who used WEL for three years and students who used WEL only in second grade. Analysis of DIBELS composite scores revealed a significant difference between groups, $t(1, 95) = -4.33$, $p < .001$, due to higher end-of-second-grade scores made by students who had used WEL for three years ($M = 460.13$) than by students who used WEL only in second grade ($M = 439.78$). Effect size ($d = 0.50$).

Independent samples *t*-tests were also conducted to examine group differences in DIBELS subskills (see Fig. 3). Analysis of DIBELS subskill scores revealed significant differences between groups in all subskills due to higher end-of-second-grade scores made by students who had used WEL for three years than by students who used WEL only in second grade. Effect sizes ranged from 0.38 to 0.65, with the largest effect indicated for Maze Adjusted.

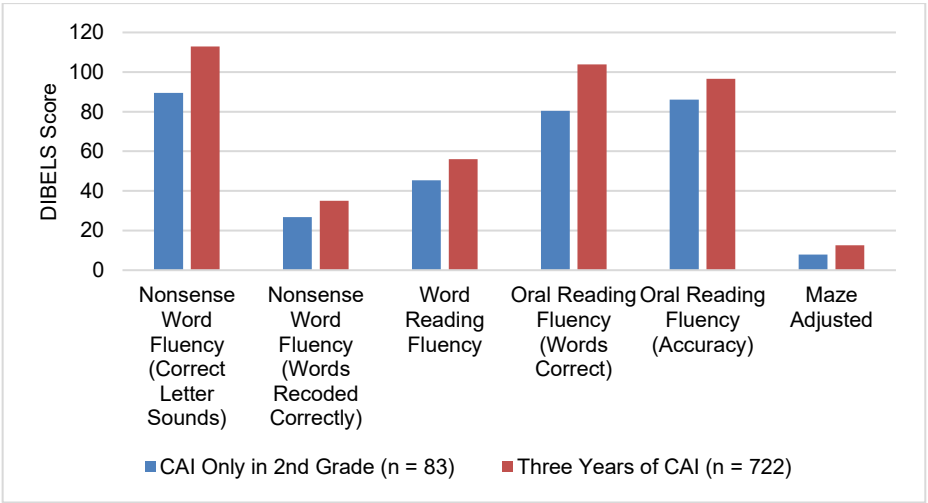

Fig. 3. 2022-2023 End-of-Second-Grade DIBELS Scores by Subskill

Group Differences Using ANOVAs. Four separate two-way ANOVAs were conducted to examine the sustained early effects of WEL and demographics on end-of-second-grade DIBELS composite scores (see Fig. 4). Across all demographics, students in the experimental group outperformed students in the control group. Simple effects analysis showed that for female and male students, African American/Black students, Latino/a students, students with and without special education services, multi-lingual

learners, and non-multi-lingual learners, students who used WEL for three years significantly outperformed students who used WEL only in second grade.

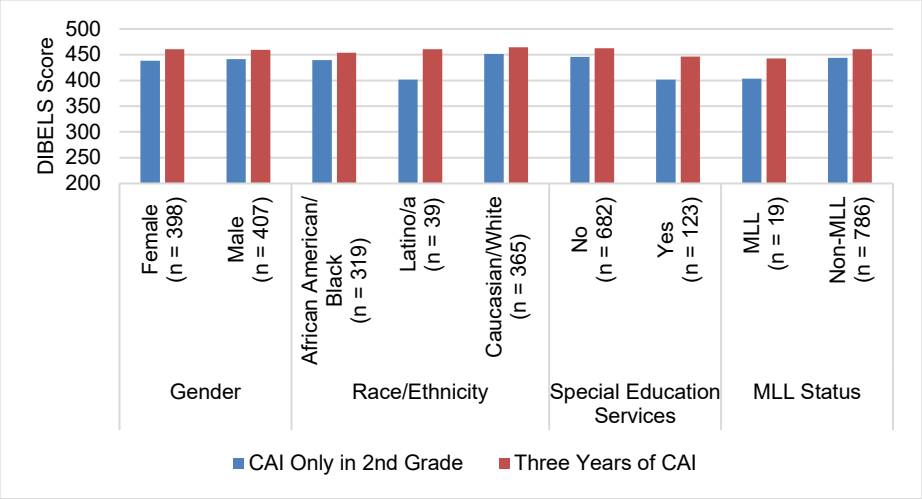

Fig. 4. 2022-2023 End-of-Second-Grade DIBELS Composite Scores by Demographics

3.3 First-Grade Students: WEL Only in First Grade vs WEL in Kindergarten & First Grade (Early Start Effects)

To analyze early start effects, the experimental group included students who had more than 100 minutes of WEL learning time during kindergarten (2021-2022) and first grade (2022-2023). The control group included students who had more than 100 minutes of WEL learning time during only first grade (2022-2023).

Group Differences Using Independent Samples *t*-tests. Independent samples *t*-tests were conducted to examine group differences in end-of-first-grade scores between students who used WEL in kindergarten and first grade and students who used WEL only in first grade. Analysis of DIBELS composite scores revealed a significant difference between groups, $t(1, 798) = -5.66$, $p < .001$, due to higher end-of-first-grade scores made by students who had used WEL in kindergarten and first grade ($M = 487.64$) than by students who used WEL only in first grade ($M = 466.35$). Effect size ($d = 0.54$).

Independent samples *t*-tests were also conducted to examine group differences in DIBELS subskills (see Fig. 5). Analysis of DIBELS subskill scores revealed significant differences between groups in all subskills due to higher end-of-first-grade scores made by students who had used WEL in kindergarten and first grade than by students who used WEL only in first grade. Effect sizes ranged from 0.34 to 0.60, with the largest effect indicated for Oral Reading Fluency - Words Correct.

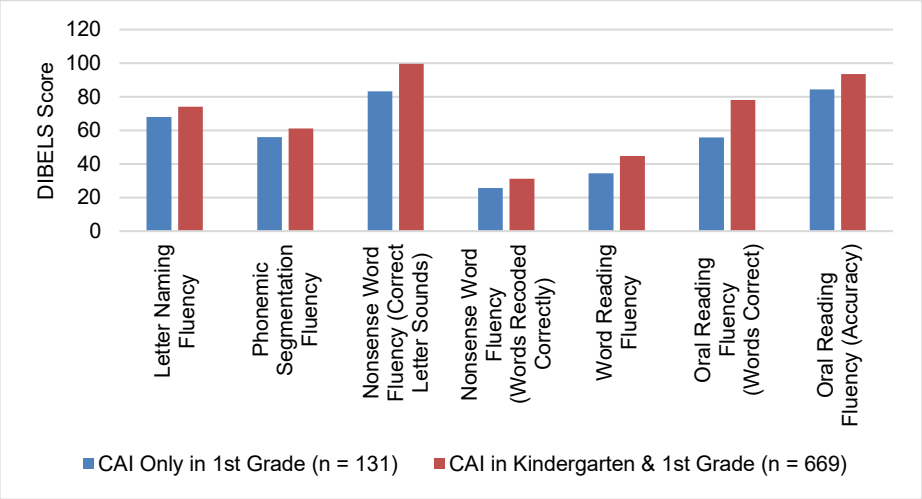

Fig. 5. 2022-2023 End-of-First-Grade DIBELS Scores by Subskill

Group Differences Using ANOVAs. Four separate two-way ANOVAs were conducted to examine the early start effects of WEL and demographics on end-of-first-grade DIBELS composite scores (see Fig. 6). Across all demographics, students in the experimental group outperformed students in the control group. Simple effects analysis showed that for female and male students, African American/Black students, Caucasian/White students, students with and without special education services, and non-multi-lingual learners, students who used WEL in kindergarten and first grade outperformed students who used WEL only in first grade.

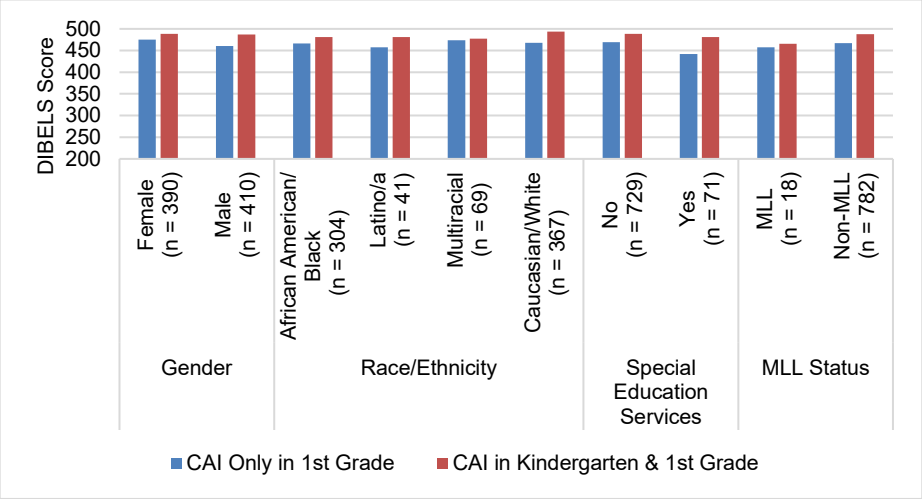

Fig. 6. 2022-2023 End-of-First-Grade DIBELS Composite Scores by Demographics

4 Discussion

This study explored the early and long-term effects of using WEL in the classroom. Results revealed that early and long-term learning time with WEL had beneficial results. Across all comparison groups, students in the experimental group consistently had higher end-of-year literacy scores than control group students.

Students assessed at the end of first grade who had WEL learning time in the classroom for two years significantly outperformed students who only had one year of WEL learning time in composite end-of-year literacy scores (d = 0.54) and across all subskills, with effect sizes ranging from 0.34 to 0.60. Similarly, students assessed at the end of second grade who had WEL learning time in the classroom for three years significantly outperformed students who only had one year of WEL learning time in composite end-of-year literacy scores (d = 0.50) and across all subskills, with effect sizes ranging from 0.38 to 0.65. Oral reading fluency strands had the largest effects in third- (0.50 to 0.66) and first-grade (0.47 to 0.60) students.

These early effects of WEL suggest that starting to use this program at an early age can lead to greater literacy gains. Early exposure to WEL can provide students with more time to develop the foundational literacy skills needed for future academic success. Students who start WEL later may miss these early benefits, which could result in smaller literacy gains.

This result is notable as any significant relationship between dosage and intervention efficacy contrasts with trends in existing literature. While relative starting dates for learning time are more nebulous than concrete total overall learning time [3], the existence of a relationship should not be overlooked. The relationship between learning time and efficacy of a given program remains unreconciled. Future research could benefit from examining the comparative impacts of relative dosage and specific CAI features.

Additionally, WEL's long-term benefits were demonstrated by students assessed at the end of third grade, who had WEL learning time in the classroom during their second-grade school year, significantly outperforming students who did not have WEL learning time in the classroom on composite end-of-year literacy skills, with an effect size of 0.65. These literacy gains were seen across all subskills with high effect sizes ranging from 0.50 to 0.66. These results indicate that WEL has lasting impacts, with even a single year of learning time with WEL yielding benefits evident in students' literacy scores a year later.

This strong positive effect in the second grade is of interest. Prior research had indicated both that similar literacy interventions had more pronounced effects in early grades [3] and that effects tended to drop off notably at follow-up [5]. In this study, a pronounced benefit was still observed in students in later grades. It is possible that this result is capturing a compounding positive effect over multiple years of use. Hypothetically, the greater effects in second grade could be explained by students taking early lessons forward from one year into the next, leaving them in a better position to learn in the next year. Given the limited number of grades involved, it’s impossible to say whether this would have been sustained further into a student’s education. Following students for a more extended period to investigate the benefits of the program multiple years after participation could offer further insights into the lasting effects of WEL.

Ensuring that WEL is effective for students of different genders, races/ethnicities, students receiving and not receiving special education services, and multi- and non-

multilingual learners is crucial to providing students with an equitable education. The findings from this study suggest that WEL is an effective and equitable educational intervention that provides immediate and lasting benefits for diverse learners, with positive outcomes evident for all demographics assessed. It should be noted, however, that the analysis of this study was not intended to draw direct comparisons between students along demographic lines, as was the case with prior research [5], [7]. Additionally, the results highlight WEL's value as a classroom resource with the potential to help bridge the academic achievement gap.

Further, it should be acknowledged that this study assessed natural usage patterns to draw comparisons and did not assign students to usage conditions. As a result, outcomes may have been influenced by factors not accounted for in the analysis, including differences in teacher efficacy or learning environments. Additionally, this study was conducted across a single district. Future research should expand on this study by including a broader sample to provide evidence across multiple school districts.

References

1. Landerl, K., Castles, A., & Parrila, R.: Cognitive precursors of reading: A cross-linguistic perspective. Scientific Studies of Reading 26(2), 111-124 (2021).
2. Mol, S. E., & Bus, A. G.: To read or not to read: a meta-analysis of print exposure from infancy to early adulthood. Psychological Bulletin 137(2), 267-296 (2011).
3. Kim, J., Gilbert, J., Yu, Q., & Gale, C.: Measures matter: A meta-analysis of the effects of educational apps on preschool to grade 3 children's literacy and math skills. Aera Open 7, (2021).
4. Hurry, J., Fridkin, L., & Holliman, A. J.: Reading intervention at age 6: Long-term effects of Reading Recovery in the UK on qualifications and support at age 16. British Educational Research Journal 48(1), 5-21 (2022).
5. Fikrat-Wevers, S., van Steensel, R., & Arends, L.: Effects of family literacy programs on the emergent literacy skills of children from low-SES families: A meta-analysis. Review of Educational Research 91(4), 577-613 (2021).
6. McTigue, E. M., Schwippert, K., Uppstad, P. H., Lundetræ, K., & Solheim, O. J.: Gender differences in early literacy: Boys' response to formal instruction. Journal of Educational Psychology 113(4), 690-705 (2021).
7. Ruan, Y., Maurer, U., & McBride, C.: Effectiveness of reading interventions on literacy skills for Chinese children with and without dyslexia: A meta-analysis of randomized controlled trials. Educational Psychology Review 36, 80 (2024).
8. Education Week, https://www.edweek.org/leadership/opinion-john-hattie-isnt-wrong-you-are-misusing-his-research/2018/06, last accessed 2024/12/11.
9. Robert Slavin's Blog, https://robertslavinsblog.wordpress.com/2018/06/21/john-hattie-is-wrong, last accessed 2024/12/11.
10. Verhoeven, L., Voeten, M., van Setten, E., & Segers, E.: Computer-supported early literacy intervention effects in preschool and kindergarten: A meta-analysis. Educational Research Review 30, 100325 (2020).
11. Silverman, R. D., Keane, K., Darling-Hammond, E., & Khanna, S.: The effects of educational technology interventions on literacy in elementary school: A meta-analysis. Review of Educational Research, (2024).
12. University of Oregon: 8th edition of Dynamic Indicators of Basic Early Literacy Skills (DIBELS ®): Technical Manual, (2018-2020).