PUBLICATION T25-019.9 OCTOBER 2025

Results from a Randomized Trial of the Dysolve Program for Students with Reading Difficulties

Authors:
Henry May, Ph.D., University of Delaware
Sam Van Horne, Ph.D., University of Delaware

Reviewed by:
Gail Headley, Ph.D., University of Delaware
Wei Li, Ph.D., University of Florida

Funding for this Research was Provided by:
EduNational LLC

cresp.udel.edu

Pearson Hall, Ste. 107
125 Academy Street
Newark, DE 19716
Phone: 302-831-2928

CRESP is committed to addressing education and social policy challenges with rigorous, relevant research.

The Center for Research in Education and Social Policy (CRESP) within the College of Education and Human Development at the University of Delaware conducts rigorous research, program evaluation, and policy analysis to help policymakers and practitioners in education, community health, and human services determine which policies and programs are most promising to improve outcomes for children, youth, adults, and families.

Founded in 2013, CRESP recognizes that poverty, educational achievement, and chronic disease prevention are intertwined in a complex social web that challenges communities and policymakers alike. CRESP's mission, values, and scientific priorities seek to inform program and policy development across local, state, and federal levels. We work alongside program professionals, academic leaders, and students to foster engagement in high-quality, practice-driven research, and evaluation. CRESP researchers are trained in evaluation methodology, randomized field experiments, natural experiments, qualitative methods, statistical analysis, mixed-method evaluation, and survey research.

Please feel free to contact us should you have any questions about us or our research.

Center for Research in Education and Social Policy
University of Delaware
Pearson Hall, Suite 107
125 Academy Street
Newark, DE 19716
cresp-info@udel.edu
(302) 831-2928
cresp.udel.edu

CRESP Leadership Team

Henry May, Director (hmay@udel.edu)
Allison Karpyn, Co-Director (karpyn@udel.edu)
Sue Giancola, Senior Associate Director (giancola@udel.edu)
Jeff Klein, Senior Policy Scientist (kleinjef@udel.edu)
Jill Bathon, Assistant Director (jbathon@udel.edu)

Suggested Citation

May, H. & Van Horne, S. (2025) *Results from a Randomized Trial of the Dysolve® Program for Students with Reading Difficulties*. The Center for Research in Education and Social Policy (CRESP), University of Delaware.

ABSTRACT

This technical report presents analyses and results from the first large-scale randomized controlled trial (RCT) of the effect of the game-based, artificial intelligence Dysolve® program on the reading achievement of the lowest-performing students in grades 3-8. Such RCT studies are widely recognized by experts as the "gold standard" for rigorous evaluation of program impacts. Additionally, the artificial intelligence (AI) system behind Dysolve generates single-use games in real time through a patent-protected method during its interaction with each user. Unlike adaptive programs of the past, Dysolve does not select from a premade pool of items or activities. This high responsivity at the individual-specific level is designed to locate and correct language processing deficits underlying each student's reading difficulty.

In Dysolve, games are built in real time based on accumulating data in a user's program. For example, a Dysolve game may assess Phoneme Detection, i.e., auditorily picking out a target sound in single, spoken words. This is important for learning and retaining new words, spelling and reading. In Dysolve, Phoneme Detection is delivered as a Fishing Game. The utterance of a test word is synchronized with a fish swimming across the screen at a set speed. Users 'catch' the fish representing a test word with the target sound. The student is told to listen for a target sound (e.g., /b/). Then they hear audio files of 10 common, single words (e.g., *bed, both*) in succession at a set speed. They tap on the keyboard whenever they detect the target sound in these words. Scoring is done automatically. Depending on the student's game responses, Dysolve AI may decide to deliver another Fishing Game with new test words or a new test sound to verify or explore further or move on to a different activity.

During the 2022-23, through 2024-25 school years, a randomized controlled trial (RCT) was conducted to evaluate the effect of the Dysolve® program on reading achievement scores of students (n=848) enrolled in grades 3-8 in 32 schools from 9 states in the US. Participants included students in grades 3-8 who scored near the 10th percentile on average in reading/ELA for their grade on the previous year state or local reading/ELA test (excluding students with visual impairment, physical hearing problems, or cognitive impairment). Most participants were minority students from low-income districts. Students were randomly assigned to treatment and control conditions within grade and school. Baseline balance was confirmed through a statistical test of pre-intervention reading scores. Recommended minimum dosage was 9 hours of Dysolve in total (e.g., 15 min per day, 4 days per week, for 9 weeks). The trial commenced during the COVID-19 pandemic, and Dysolve was used as a supplemental program to regular reading instruction.

Most students in the treatment group logged a cumulative total of less than 3 hours (less than one-third of the intended minimum dosage). Impacts of the Dysolve intervention were assessed through both intent-to-treat (ITT) and treatment-on-the-treated (TOT) analyses. Results from the ITT analyses revealed a positive and marginally significant (p=.057) intent to treat effect (i.e., +.095 standard deviations). This result suggests that students randomly assigned to Dysolve had posttest reading scores that were slightly higher, on average, than students that did not receive access to Dysolve. Results from the TOT analyses suggested that greater dose is associated with greater gains—the effect of Dysolve increases by .20 standard deviations for each 10-fold increase in dosage. This suggests that the effect associated with a full dose of 27 hours (i.e., 15-20 minutes

per day, 4 days per week, for 6 months) can be projected to produce a +.30 standard deviation (0.095+0.201=0.296) increase in reading scores. However, this conclusion is based on a projection from a dosage-adjusted statistical model. The fact that very few students assigned to the treatment group in this study used Dysolve for more than the minimum dose of 9 hours precludes strong inference about the impacts of full dosage.

Implications from the ITT effects in this RCT suggest that Dysolve, even at low dosage, may have positive impacts on students' performance on standardized reading tests. This is important because Dysolve does not provide direct reading instruction. Dysolve was designed under the assumption that improvement in basic language processing enables reading development. In other words, Dysolve is designed to address language processing deficits shown to be associated with constrained reading development. Thus, to register positive impacts in this RCT, a transfer effect must occur from language processing to the broader reading skills measured by the assessments used in this study.

Unfortunately, the potential impacts of Dysolve at full dose are not well reflected in this study as the results are tempered by the relatively low usage of Dysolve by treatment students. However, as a field trial, this RCT reflects real-world implementation, unlike experimental studies in lab-like settings where researchers have strict control over dosage. This RCT was not conducted in a lab setting and without strict control over treatment dosage. The real-world implementation allowed dosage to vary naturally, albeit with few treated students receiving the recommended dosage. Therefore, this report does not reflect the potential impacts of Dysolve at a full dose. However, as a field trial, the results suggest that future studies of Dysolve are warranted.

This study represents the first external, independent evaluation of the first AI program to deliver individually adaptive intervention to address reading difficulty, without requiring teacher training or instruction as part of the program. Adults without special training (e.g., parents, teachers, tutors, paraeducators) can supervise students using Dysolve. This feature may facilitate adoption and reduce costs in comparison to other programs that target similar outcomes. We plan to conduct additional studies of Dysolve to better understand its impacts under higher dosages, as well as the relative cost-effectiveness of Dysolve versus other interventions for students experiencing reading difficulties.

INTRODUCTION

The purpose of this research study was to examine the efficacy of a game-based learning platform, Dysolve®, which is designed to help reduce language processing deficits in children with reading difficulties, including dyslexia. During the 2022-23, 2023-24, and 2024-25 school years, a randomized controlled trial (RCT) was conducted to evaluate Dysolve's impact through analysis of student's test scores in reading and English language arts (ELA) from state accountability tests and commercially developed standardized assessments.

THE DYSOLVE INTERVENTION

Dysolve is an artificial intelligence program designed to address dyslexia and language-related disorders. Dysolve uses a novel, patent-protected method to generate individually-tailored game-based training tasks intended to target each learner's unique problems. As it probes the root causes of the problems found, Dysolve designs increasingly targeted activities to correct problems experienced by that individual learner. Notably, whereas other adaptive learning programs use a pool of premade items or game activities, Dysolve generates single-use games in real time through a patent-protected method during its interaction with each user. This responsivity at the individual-specific level is designed to locate and correct language processing deficits underlying reading difficulty. The Dysolve program is cloud-based, and students can access it from a computer, tablet, or Chromebook. Dysolve is intended to be used for 15-30 minutes per day. Each student has a unique user ID, and the program tracks individual usage and progress over time.

In Dysolve, games are built in real time based on accumulating data in a user's program. For example, a Dysolve game may assess Phoneme Detection, i.e., auditorily picking out a target sound in single, spoken words. This is important for learning and retaining new words, spelling and reading. In Dysolve, Phoneme Detection is delivered as a Fishing Game. The utterance of a test word is synchronized with a fish swimming across the screen at a set speed. Users 'catch' the fish representing a test word with the target sound. The student is told to listen for a target sound (e.g., /b/). Then they hear audio files of 10 common, single words (e.g., *bed, both*) in succession at a set speed. They tap on the keyboard whenever they detect the target sound in these words. Scoring is done automatically. Depending on the student's game responses, Dysolve AI may decide to deliver another Fishing Game with new test words or a new test sound to verify or explore further or move on to a different activity.

METHODS

STUDY DESIGN

A randomized controlled trial (RCT) experiment was conducted because it provides the strongest support for causal inference about program impacts. Given that many different schools would be recruited to participate in the study, with multiple students participating from each school (see "Participants and Setting" section), the experiment was a multi-site RCT. Within each school, participating students were grouped by grade level, and individual students were randomly assigned to treatment or control conditions within each grade. The study design and analysis plan were preregistered in the Registry of Efficacy and Effectiveness Studies (REES) at https://sreereg.icpsr.umich.edu/sreereg/.

TREATMENT CONDITION

Students randomly assigned to receive the Dysolve intervention were asked to use the Dysolve Program on a computer or Chromebook for 15-30 minutes per day, for 3-5 days per week, over the course of 3-6 months. This translates to a minimum of 9 total hours using Dysolve during the course of the experimental treatment. Engagement with Dysolve was supplemental; it was provided in addition to any other educational activities routinely provided by the schools. Dysolve was offered in the cloud, allowing students to log on at any time, whether at home, at school, or from any other place through a computer or Chromebook with internet access. Given that Dysolve was offered as a supplemental intervention, schools were asked to ensure that children who were selected to use Dysolve did not miss other instructional time. Schools differed in how they promoted/encouraged the use of Dysolve, as there was no specific approach required for schools in this study.

CONTROL CONDITION

Students randomly assigned to the control condition received "business-as-usual" (BAU) instruction and intervention typically provided by their school. After each school completed its participation in this randomized trial (i.e., posttest outcomes were measured), students in the control condition were provided with free access to use the Dysolve program.

PARTICIPANTS AND SETTING

Eligible participants included students in grades 3-8 who scored below the 30th percentile in reading/ELA for their grade on the previous year state or local reading/ELA test and did not have visual or cognitive impairment, nor physical hearing problems. The final recruited sample included 848 students enrolled in grades 3-8 in 32 schools from 9 states (i.e., Mississippi, Illinois, Ohio, Kansas, North Carolina, Wisconsin, Louisiana, New York, New Jersey) and include 13 rural schools, 5 suburban schools, and 14 urban schools. Of these, 19 are public, 1 is private, 11 are charter schools, and 1 is a virtual school. All participants scored below the 30th percentile and, on average, near the 10th percentile in reading/ELA for their grade on the previous year state or local reading/ELA test. Participants included 50% males and 50% females, several races (i.e., 50% black, 20% white, 7% Latinx, 16% Asian, 6% multiracial, and 1% other); 17% were non-native speakers of English, and 96% were economically disadvantaged, 18% were classified with Specific Learning Disability or Speech Language Impairment, and 12% were classified with ADHD, autism, Other Health Impairment, anxiety and emotional disturbance disorders.

POWER ANALYSES AND OUTCOMES DATA

Power analyses were conducted using PowerUp! (Dong & Maynard, 2013) based on a blocked multi-site RCT with random effects for blocks nested within sites, an intraclass correlation (ICC) of 30% for grade-level blocks, and ICC of 10% for school effects, and 50% of the student-level outcome variance explained by a pretest measure. A target sample size of 480 students with 50%

allocation to treatment was determined to produce adequate statistical power (i.e., ≥80%) to detect a treatment effect of .15 standard deviations or larger.

Students' individual scores on state or local reading/ELA tests were provided from each participating school using an anonymized student ID number created for this study. These test scores included the most recent reading/ELA score prior to a student's enrollment in the RCT (i.e., a pretest score), as well as one or more posttest scores after the student had been provided with access to the Dysolve intervention for at least six months. Because the participating schools used several different reading tests, scores on each assessment were rescaled onto a standard normal distribution using grade-level means and standard deviations from published norming data for each assessment (see May et al, 2009 for a discussion of this rescaling method). The most commonly used assessments were MAP Growth (Thum & Kuhfeld, 2020), IXL ELA (Zhao & Mayne, 2024), and AimsWeb (Pearson, 2017).

Usage data in the form of cumulative number of hours using the Dysolve program was extracted from the Dysolve user database. Usage data was then linked to reading/ELA test scores using the anonymized student ID number created for this study.

STATISTICAL ANALYSES

We conducted two types of analyses to estimate the impact of Dysolve on student reading/ELA scores. The first analysis was a test for an intent-to-treat (ITT) effect. It involved comparing the test scores of treatment group students to scores of control group students, regardless of whether or how much of the treatment was experienced by any student. ITT analysis is typically most interesting to policymakers who want to know the impact of offering an intervention (i.e., simply making it available), with the expectation that different individuals will use it more or less. The second analysis was a test of the treatment-on the-treated (TOT) effect. TOT analysis conditions the impact estimate on the actual dosage of the treatment experienced (i.e., the amount of time spent using Dysolve).

For the main analysis of the ITT effect of treatment we used a mixed-effects regression model. We regressed the standardized score at time t for student i from block j and school k on a treatment group indicator (Trt = 1 for Dysolve, Trt = 0 for Control), a time indicator ($Time$ =0 for the pretest, $Time$ =1 for the posttest), and a group-by-time interaction. Random effects were included to represent each random assignment block nested within sites. A primary benefit of this model is that it handles missing pretest or posttest scores through full information maximum likelihood (FIML) estimation (i.e., students with complete data, as well as students missing either a pretest or a posttest score, but not both, are included in the analysis). The mathematical form of the model was:

$$StandardScore_{tijk} = \beta_0 + \beta_1(Trt_{ti}) + \beta_2(Time_{ti}) + \beta_3(Trt \times Time)_{ti} + \alpha_{jk} + \gamma_k + e_{tijk}$$

The parameter β_0 represents the mean pretest reading/ELA score in the control group. The parameter β_1 represents the mean difference in pretest reading/ELA scores between students assigned to the treatment versus control groups (i.e., it acts as a test for baseline equivalence). The parameter β_2 represents the mean change in reading/ELA scores from pretest to posttest for students in the control group (i.e., the gain in reading performance under BAU). The key parameter of interest, β_3 represents the mean difference in reading/ELA score gains for the treatment group versus the control group (i.e., the impact of assignment to the Dysolve treatment). The three residual terms represent the random effect for blocks (α_{jk}), schools(γ_k), and students (e_{tijk}). The student-level residual accounts for repeated measures with separate variance estimates for pretest and posttest scores, and an estimated correlation between pretest and posttest scores.

For the TOT impact analyses, we used an instrumental variables analysis (also incorporating random effects for block by schools) in which treatment assignment is used as an exogenous instrument for the effect of Dysolve dosage on reading/ELA scores. The key assumption of this analysis (i.e., exogeneity) is that random assignment has zero direct impact on reading/ELA scores; instead, random assignment to treatment leads to usage of Dysolve, which then produces impacts on reading/ELA scores. Given that treatment assignment was random, this exogeneity assumption is plausible.

RESULTS

SAMPLE RECRUITMENT

Recruitment of schools and students began with the 2022-23 school year; however, relatively few schools and students were successfully recruited that year. Likely, this was due in part to the lingering effects of the COVID-19 pandemic. Recruitment continued into the 2023-24 school year, and while many more schools were recruited, too few students were recruited to meet the target sample size of 480 students (only about half the target sample was recruited during the first two years of the RCT). The continued difficulty with recruitment was attributed to the parents of eligible students not reading, signing, and returning informed consent forms, despite this opt-in recruitment depending on parents returning consent forms. As such, the RCT was extended into the 2024-25 school year, and an Institutional Review Board (IRB) modification was requested and approved to recruit participants via an opt-out consent process in which information about the study and the Dysolve intervention was sent home to parents as part of an informed consent process, and parents were notified that they could contact the school and choose to opt-out of the study without penalty. Recruitment during the 2024-25 school year was very successful, and the target sample size of 480 students was quickly surpassed.

At the outset, there were 848 students from 32 schools that were recruited and then randomized for this research study (423 to the treatment condition and 425 to the control condition). Student test score data was received from 29 schools. The three schools that did not submit outcomes data dropped out of the study, and the students from those three schools are not included in subsequent analyses. An additional school was excluded from the study because all of its participating students were not enrolled in grades 3-8; thus, they did not meet the eligibility criteria despite having been randomly assigned. The final analytic dataset included 28 schools and 705 students, with 351 in the control group, and 354 in the treatment group.

ATTRITION / MISSING DATA

Table 1 shows the prevalence of missing data and participant attrition in the RCT. Total attrition (i.e., students missing both pretest and posttest data) was 16.9 percent. Another 6.5 percent of participants were missing one test score. There was no significant difference in the pattern of missing data and attrition did not differ across the treatment and control groups, $X^2(3) = 0.38$, $p < .95$. According to What Works Clearinghouse standards (WWC, 2022), these attrition statistics are within the acceptable range (i.e., the attrition is unlikely to bias the statistical analysis).

Table 1. Availability of Data by Treatment Group

	Control		Treatment	
	n	%	n	%
Has Both Pretest & Posttest	324	76.2	326	77.1
Missing Only Baseline	11	2.6	13	3.1
Missing Only Posttest	16	3.8	15	3.5
Missing Both Tests	74	17.4	69	16.3
Total	425		423	

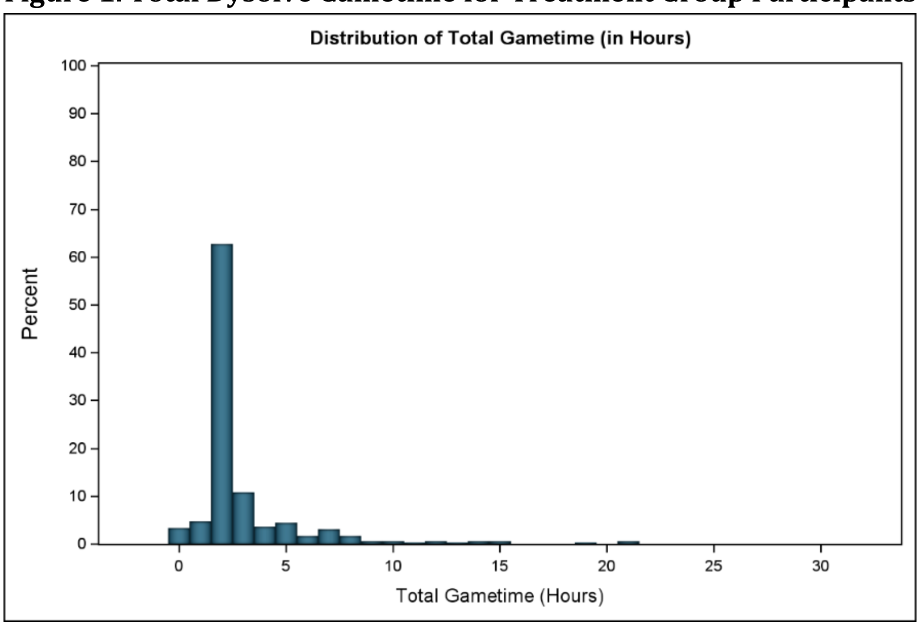
BASELINE EQUIVALENCE

A statistical test of baseline equivalence confirmed that random assignment had achieved equivalence at baseline. There was no statistically significant difference in standardized pretest scores between treatment and control at baseline ($p = .7458$).

TREATMENT FIDELITY / DOSAGE

Engagement with the Dysolve program was far less than the intended dosage for most students. Only three percent of students in the treatment group reached the minimum intended dosage of 9 hours. Approximately 65% of students in the treatment group logged a cumulative total of less than 3 hours using the program. The mean of total gametime for the Dysolve group was 2.66 hours, the standard deviation was 2.86 hours, and the median usage was 1.5 hours. About half the participants joined the program in the final months of this study, limiting the length of time available to engage with Dysolve.

Figure 1. Total Dysolve Gametime for Treatment Group Participants



INTENT TO TREAT ANALYSES

Results from an intent-to-treat (ITT) analysis of the difference in reading score gains for students randomly assigned to use Dysolve versus students assigned to the control condition revealed a positive and marginally significant (p=.057) intent to treat effect. This result suggests that students randomly assigned to Dysolve had posttest reading scores that were higher (i.e., +.095 standard deviations), on average, than students that did not receive access to Dysolve. Parameter estimates from the mixed effects regression model are provided in Appendix A.

The results from the ITT analysis confirmed several things. At pretest, students participating in this study had an average score that was 1.3 standard deviations below average on nationally normed reading tests (β_0=-1.287, p<.001). This corresponds to the 10th percentile in reading, suggesting that students in the study sample were experiencing reading difficulties and were good candidates to use the Dysolve program. Second, the difference in pretest scores between the treatment and control groups was effectively zero (β_1=0.003, p=.945). The lack of statistical significance confirms that random assignment created equivalent treatment and control groups at baseline. Third, the gains made by students in the control condition were positive and significant (β_2=0.286, p<.001), suggesting that other interventions and instruction provided by their schools were helping to improve reading performance overall. Fourth, the gains made by students assigned to use Dysolve were also positive and significant ($\beta_2+\beta_3$ =0.386, p<.001). More importantly, the gains made by the Dysolve group were 33% larger than those of the control group; however, this difference was only marginally significant at the 90% confidence level (β_3=0.095, p<.10). As a sensitivity check, we also estimated this model using fixed block and school effects, and results were nearly identical.

DOSAGE-ADJUSTED ANALYSES

We also conducted an instrumental variables (IV) analysis using two-stage least squares estimation to estimate a potential TOT effect of dose. The Dysolve platform produced a variable that indicated how much total game time each participant had received. Given that the control group was not provided with access to Dysolve, all students in the control condition were assigned a value of 0 for dose. Students in both the treatment and control groups who were missing either a pretest score or a posttest score, and students in the treatment group who were missing usage data, had to be dropped from this analysis. The final analytic sample for the TOT analysis included 662 students (n=331 from the treatment group and n=331 from the control group). The dose variable was highly skewed right due to a small number of participants engaging heavily with the system. To reduce the influence of outliers in the statistical model, the usage variable was transformed via a log base-10 function after adding 1 to all values (i.e., to avoid ln(0), which does not exist). The IV model for TOT effects was estimated using the plm package in R, with random effects for block by site as in the ITT model. Results from the TOT analysis confirmed that group assignment was a significant predictor of dose (F(1,660)= 2569.8, p<.0001), suggesting that assignment to the treatment group led to usage of Dysolve. This satisfies a key assumption of IV analysis (i.e., instrument strength). The second stage model produced a positive coefficient (β = +0.201) suggesting that greater dose may be associated with greater gains; however, this difference was only marginally significant at the 90% confidence level (p = .0973). Given that the dosage variable was transformed using the base-

10 log scale, the coefficient of +.20 suggests that the effect of Dysolve increases by .20 standard deviations for each 10-fold increase in dosage. This suggests that the +.10 standard deviation effect associated with an average usage of 2.7 hours can be projected to increase to +.30 standard deviations (0.095+0.201=0.296) if dosage increases to 27 hours (i.e., 15-20 minutes per day, 4 days per week, for 6 months).

CONCLUSIONS

Results from this RCT suggest that Dysolve, even at low dosage, may have positive impacts on the reading performance of students experiencing reading difficulties. A full dosage of Dysolve is likely to have even larger impacts, but this conclusion is based on a projection from a dosage-adjusted statistical model. The fact that very few students assigned to the treatment group in this study used Dysolve for more than the minimum dose of 9 hours precludes strong inference about the impacts of full dosage.

There are two primary implications of this study. First, there is evidence that the Dysolve program has positive impacts on reading performance among students experiencing reading difficulties. Second, the potential impacts of Dysolve may not be fully reflected in this study because most students in the treatment group did not use Dysolve for the minimum number of hours prescribed.

There are several possible explanations for the low usage; however, this study did not collect any qualitative data that would support empirical statements about the reasons for low usage. As such, the possible explanations discussed here should be considered as hypotheses.

One possible reason for low usage is that young students may be unlikely to engage in Dysolve on their own without prompting by an adult. Unfortunately, school disruptions during the pandemic and staffing shortages post-pandemic made consistent adult supervision challenging during this RCT.

The use of opt-out recruitment in this study may have yielded a sample of participants that were unlikely to engage deeply with Dysolve. While this problem might have been mitigated if schools had set aside in-school, supervised time for treatment students to use the Dysolve program, that was not the case in most schools in this study. Instead, students were typically expected to sign in and use Dysolve unsupervised. Anecdotal reports from the Dysolve implementation team suggested that, although teachers could track each student's daily gametimes on their teacher dashboards, many teachers never logged in to their dashboards. Future studies should communicate expectations more clearly to teachers and parents, and usage should be continuously monitored with feedback to teachers and parents to encourage their students to log in to Dysolve.

Additionally, some sites had high absenteeism among this pool of students who were performing near the 10th percentile. Legal requirements, such as mandated dyslexia screenings, diagnostic assessments, and progress monitoring measures in Ohio, left little time for supplemental interventions like Dysolve during the school day.

With this RCT, Dysolve joins an exclusive group of education interventions that have been subjected to the most rigorous test of their impact on student learning. One might argue that interventions that have no rigorous evidence of impact should not be adopted in schools or classrooms; but unfortunately, too many interventions are adopted in schools and classrooms despite having zero rigorous evidence of impact.

This first RCT of Dysolve provides rigorous evidence that the Dysolve program can improve reading outcomes. Although this study does not provide clear insight into the impacts of a full dose of Dysolve, the finding of a positive impact from a small dose, along with the finding that larger dosage is associated with larger effects, suggests it is important to confirm whether a full dose of Dysolve has a proportionally large effect. Furthermore, this study represents the first external, independent review of the first AI program to address reading difficulty by delivering interventions autonomously (i.e., without relying on teacher instruction as part of the program). Students using Dysolve may be supervised by any adult (e.g., parents, teachers, tutors, paraeducators) without any special training. This has important implications for ease of adoption and cost. We plan to conduct additional studies of Dysolve to better understand its impacts under higher dosages, as well as the relative cost-effectiveness of Dysolve versus other interventions for students experiencing reading difficulties.

REFERENCES

Dong, N., & Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness, 6*(1), 24-67. **https://doi.org/10.1080/19345747.2012.673143**.

May, H., Perez-Johnson, I., Haimson, J., Sattar, S., & Gleason, P. (2009). *Using State Tests in Education Experiments: A Discussion of the Issues*. NCEE 2009-013. National Center for Education Evaluation and Regional Assistance. **https://files.eric.ed.gov/fulltext/ED511776.pdf**

Pearson, N. C. S. (2017). *AIMSwebplus technical manual*. Bloomington, MN: Author.

Thum, Y. M., & Kuhfeld, M. (2020). *NWEA 2020 MAP growth achievement status and growth norms for students and schools*. NWEA Research Report. Portland, OR: NWEA

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0*. U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). This report is available on the What Works Clearinghouse website at **https://ies.ed.gov/ncee/wwc/Handbooks**.

Zhao, Y., & Mayne, Z. (2024). *National norms for IXL's diagnostic in grades K-12*. San Mateo, CA: IXL Learning.

APPENDIX A:
ESTIMATES FROM STATISTICAL MODEL OF INTENT TO TREAT EFFECTS

Table A1 includes the parameter estimates and standard errors from the statistical model of the intent to treat (ITT) treatment effect.

Fixed Effects	Model Parameter	Estimate	Standard Error
Average Pretest (i.e., baseline) Score in Ctrl Group	β_0	-1.287***	0.131
Trt vs. Ctrl Group Difference in Pretest Scores	β_1	0.003	0.049
Pretest->Posttest Gains in Ctrl Group	β_2	0.286***	0.035
Trt vs. Ctrl Difference in Gains (i.e., impact of assignment to the Dysolve treatment)	β_3	0.095~	0.050
Random Effects			
Variance of Block Effects	var(α_{jk})	0.062~	0.045
Variance of School Effects	var(γ_k)	0.381**	0.140
Residual Variance of Pretest Scores	var(e_{0ijk})	0.624***	0.038
Residual Variance of Posttest Scores	var(e_{1ijk})	0.412***	0.024
Correlation Between Pretest and Posttest Scores	corr(e_0, e_1)	0.622***	0.025

Note. $N_{Students} = 705$, ***p < .001, **p<.01, *p<.05, ~p<.10. This model was fit using the PROC HPMIXED procedure in SAS 9.4