Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 25-1209

Bridging Literacy Gaps: The Impact of AI-Driven Personalised Learning on Reading Skills and Educational Equity

Pilar Cuevas-Ruiz
Universidad de Sevilla

Luz Rello
Instituto de Empresa

Ismael Sanz
Universidad Rey Juan Carlos

Almudena Sevilla
London School of Economics

Persistent literacy skills deficits hinder educational attainment, limit labour market opportunities, and exacerbate socioeconomic inequalities. This paper evaluates the causal effect of an AI-driven Computer-Assisted Learning (CAL) program implemented by the Government of Madrid, which features personalised, adaptive content and real-time feedback on students' literacy proficiency. We leverage extensive and unique longitudinal information on student learning outcomes from the software across 264 schools over five school years and exploit exogenous variation in the timing of implementation to address possible selection into program participation and engagement. Our findings show that each additional session increases reading progress by 2.4 per cent of a standard deviation, roughly equal to one month of learning. Our findings highlight how AI-driven CAL tools can offer scalable interventions for effectively designing education policies to reduce educational inequities.

VERSION: May 2025

Suggested citation: Cuevas-Ruiz, Pilar, Luz Rello, Ismael Sanz Labrador, and Almudena Sevilla. (2025). Bridging Literacy Gaps: The Impact of AI-Driven Personalised Learning on Reading Skills and Educational Equity. (EdWorkingPaper: 25 -1209). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/68mk-mk89

"Bridging Literacy Gaps: The Impact of AI-Driven Personalised Learning on Reading Skills and Educational Equity"

Pilar Cuevas-Ruiz (Universidad de Sevilla)
Luz Rello (Instituto de Empresa)
Ismael Sanz (Universidad Rey Juan Carlos & London School of Economics)
Almudena Sevilla (London School of Economics)

Abstract

Persistent literacy skills deficits hinder educational attainment, limit labour market opportunities, and exacerbate socioeconomic inequalities. This paper evaluates the causal effect of an AI-driven Computer-Assisted Learning (CAL) program implemented by the Government of Madrid, which features personalised, adaptive content and real-time feedback on students' literacy proficiency. We leverage extensive and unique longitudinal information on student learning outcomes from the software across 264 schools over five school years and exploit exogenous variation in the timing of implementation to address possible selection into program participation and engagement. Our findings show that each additional session increases reading progress by 2.4 per cent of a standard deviation, roughly equal to one month of learning. Our findings highlight how AI-driven CAL tools can offer scalable interventions for effectively designing education policies to reduce educational inequities.

KEYWORDS
Artificial Intelligence, Personalised Learning, Literacy Skills, Computer-Assisted Learning, Adaptive Feedback

JEL CLASSIFICATION CODES
I21, I24, C55, O33

DISCLOSURE STATEMENT
Luz Rello is the CEO of Change Dyslexia, the organisation that promotes DytectiveU, and the creator of this software.

INTRODUCTION

Persistent deficits in reading skills are linked to lower educational attainment, constrained labour market opportunities, and widening socioeconomic disparities, posing a critical challenge for equitable growth and development (OECD, 2024). A growing body of evidence underscores the foundational role of literacy in shaping long-term economic and social outcomes, highlighting its importance as a key driver of individual and societal well-being (Duflo et al., 2024). Data from PISA 2022 reveals that one in four students in OECD countries fails to achieve basic reading proficiency, with average scores steadily declining since 2012, underscoring the need to address these gaps (OECD, 2023). Against this backdrop, emerging technologies, particularly Artificial Intelligence (AI)-driven Computer-Assisted Learning (CAL) programs, have been increasingly proposed as scalable tools to personalise instruction and adapt to diverse learning needs (Muralidharan et al., 2021; Ferman et al., 2021). This paper aims to determine whether AI-driven personalised learning environments can effectively improve literacy proficiency and reduce educational inequalities in diverse and large-scale settings by exploiting exogenous variation in the timing of program implementation across schools as an instrumental variable to identify causal effects.

The effectiveness of AI-driven CAL can be framed within the skill production model (Heckman et al., 2006), where individual inputs, environmental factors, and past investments influence cognitive and non-cognitive skills accumulation. In this framework, these Programmes operate as an innovation that enhances the efficiency of literacy skill formation by providing personalised, adaptive content and real-time feedback, which aligns with the principles of mastery learning. By enabling tailored instruction, AI-driven CAL addresses common barriers in traditional classroom settings, such as the inability to provide individualised attention or accommodate varying learning paces. Two mechanisms drive the observed improvements in literacy. First, sustained engagement with personalized exercises promotes iterative learning and cognitive reinforcement, accelerating skill acquisition. Second, real-time adaptive feedback dynamically adjusts content to the learner's needs, enhancing motivation and targeting specific deficiencies. These features collectively reduce frictions in the education production function, expanding access to quality learning experiences and mitigating inequalities, particularly for socioeconomically disadvantaged students.

We exploit the introduction of an AI-driven CAL program—DytectiveU—launched in 2019 under the "Help Dyslexia" initiative by the Regional Ministry of Education in Madrid. DytectiveU exemplifies how scalable ed-tech solutions can replicate the benefits of high-dosage tutoring while maintaining cost-effectiveness and implementation feasibility across diverse educational contexts. The program was rolled out on a broad scale, tracking 34,607 primary school students across 264 schools in the Madrid region over five school years (January 2019 – May 2023). The goal of the initiative was to tackle literacy challenges in primary education. DytectiveU offers personalised instruction in reading through AI-powered exercises and real-time feedback, enabling dynamically adaptive learning that responds to each student's progress. The system draws on a database of more than 42,000 linguistically curated activities developed by a team of psychologists and linguists to tailor content to the individual needs of each learner. Designed with scalability and flexibility in mind, DytectiveU runs on a lightweight, web-based infrastructure that minimises hardware requirements and supports use across a variety of devices, including computers, tablets, and smartphones. The program requires minimal teacher involvement and can be integrated seamlessly into existing curricula, whether as part of in-class activities, homework, or independent study, without disrupting traditional instructional methods.

To evaluate DytectiveU's impact on literacy outcomes, we use a rich administrative dataset generated by the program's internal logs. Each student is observed throughout the implementation period, from January 2019 to May 2023, with two key measures: (i) the total number of sessions completed and (ii) the Composite Literacy Proficiency Score, which captures improvement on a scale from 0 to 100, with higher values reflecting more significant gains in reading and writing skills. Session completion, recorded as a cumulative count, serves as our primary measure of program engagement. On average, students completed 20.2 sessions, with 6.7% exceeding the 64-session threshold, the total recommended by the program (Rello et al., 2020). In addition to engagement data, the dataset includes student age, gender, grade, and school identifiers.

We employ an Instrumental Variables (IV) strategy to address potential selection bias and unobserved heterogeneity, common issues in the evaluation of educational technology interventions, where more motivated or higher-achieving students may be more likely to participate (Rouse & Krueger, 2004; Büchel et al., 2022). Our instrument leverages exogenous variation in the timing of the first DytectiveU session, which is determined by administrative and logistical factors such as school scheduling and curricular planning. These constraints are unrelated to student-level characteristics such as prior academic performance or parental background, satisfying the exclusion restriction. To verify the validity of this identification strategy, we conduct two key checks: (i) no within-school variation exists in the timing of first use, as all students begin simultaneously within a given school; and (ii) the timing of implementation is uncorrelated with school baseline observables such as socioeconomic status, gender previous performance. The strength of the instrument is confirmed in the first stage: a one-week earlier implementation leads to an average of 1.64 additional sessions. The Cragg-Donald Wald F-statistic is well above the critical value, confirming its strength (Stock and Yogo, 2005). We additionally include school fixed effects to control for time-invariant institutional characteristics, such as leadership, teaching quality, or infrastructure, that may influence Composite Literacy Proficiency Scores independently of program engagement.

Our main results show that completing an additional session is associated with a 0.548 percentage point increase in Composite Literacy Proficiency Scores. This translates into a 1.50% improvement relative to the average baseline progress of 36.45%. To contextualize this magnitude, it corresponds to approximately 2.4% of a standard deviation in the Composite Literacy Proficiency Score, or roughly equivalent to one month of learning, based on benchmarks in the literature (Jakubowski et al., 2025). Our robustness checks incorporate student-level fixed effects for a subsample of students participating over multiple years. This allows us to control for persistent individual-specific factors, such as baseline ability or family background. Results remain consistent with the main specification, reinforcing the interpretation of the program's impact as causal. This strengthens our identification strategy by addressing potential sources of selection bias and isolating the effect of program participation from time-invariant confounders.

We identify two key mechanisms through which Dytectiveu improves literacy outcomes: (i) personalised content engagement, and (ii) adaptive feedback. We test the mechanism of personalised context by comparing children at different stages of development. Without personalisation, all students are exposed to the same content regardless of their developmental stage. For younger students, who benefit from greater brain plasticity and are at a formative stage of skill development, this approach risks missing a critical window of opportunity. Heckman highlights that early childhood is a period of heightened sensitivity to learning inputs, with the highest potential for skill acquisition (Heckman, 2006). Without tailored instruction, these students may receive either too difficult or not well-targeted content, limiting the returns from their inherently greater capacity to learn. For older students, whose literacy skills are typically more developed and whose learning habits are more fixed, uniform instruction may offer limited new challenges. This can reduce motivation and engagement, leading to lower marginal gains. In both cases, the lack of alignment between content and student potential constrains the overall effectiveness of the intervention. Personalisation addresses this by meeting students where they are—cognitively and developmentally. It ensures that students, especially younger ones with the most to gain, receive instruction that matches their current needs and unlocks their learning potential. We observe larger improvements among younger students at a stage of cognitive development and higher neural plasticity. Personalised content engagement is a key mechanism behind Dytectiveu's impact on literacy outcomes. Students who started earlier and engaged more intensively with the program, particularly younger students at critical stages of cognitive development, exhibited significantly greater improvements in reading skills. This highlights the importance of tailoring educational interventions to students' developmental needs. In contrast, older students whose learning habits are more consolidated, while still benefiting from the Dytectiveu tool, do so to a lesser extent. This pattern of results supports the view that personalisation is especially valuable during the early stages of literacy acquisition.

The second mechanism, adaptive feedback, refers to DytectiveU's ability to dynamically tailor the difficulty of tasks based on student performance in real time. This stands in contrast to traditional instruction, which typically delivers a fixed sequence of content regardless of how well students are progressing. In a fixed instructional model, students who quickly master the material receive no additional challenge, leading to stagnation. Meanwhile, students who struggle may continue receiving tasks that are too difficult, resulting in frustration and disengagement. In both cases, the learning curve flattens rapidly producing diminishing returns as instruction fails to remain aligned with students' evolving needs. We estimate a quadratic relationship between the number of sessions and literacy gains, allowing us to capture diminishing returns to engagement. We find that engagement levels remain linear and positive, suggesting that DytectiveU's adaptive algorithm effectively maintains student engagement and learning across a broad range of skill levels and usage intensity. Our findings reveal that initial sessions are critical for early literacy improvements, while subsequent sessions primarily reinforce and consolidate these skills with diminishing marginal returns kicking in beyond approximately 154 sessions, far beyond the observed range of participation in our sample.

Implementing AI-driven literacy interventions at scale presents two key challenges. First, most existing approaches require either additional instructional staff or significant retraining of teachers to adapt their pedagogy to technology-enhanced learning environments (Muralidharan et al., 2019). This process is both resource-intensive and demands substantial behavioural shifts from educators, which previous evidence suggests is difficult to achieve in practice (Banerjee et al. 2016). Second, many AI-driven interventions struggle with heterogeneity in learning levels, particularly in diverse educational systems where students exhibit significant variation in literacy skills. Our results suggest that leveraging CAL programs like DytectiveU, which personalise instruction through AI-based adaptation, offers a promising pathway for scaling literacy interventions while minimising teacher workload. Furthermore, because students receive individualised support within their regular classroom settings, technology-enabled personalisation retains the benefits of differentiated instruction without disrupting age-based cohort structures. This approach aligns with findings from Muralidharan et al. (2019), who emphasise that technology-driven personalisation can facilitate scalable instructional differentiation without requiring structural changes to traditional education models.

This study advances our understanding of the mechanisms that drive literacy improvements in CAL programs by analysing non-linear effects and heterogeneity across participation levels. Prior research, such as that by Büchel et al. (2022), has focused on the general efficacy of adaptive learning environments but has not addressed the specific dynamics of engagement intensity and personalised content. The gap lies in the lack of analysis of detailed participation data and its role in program efficacy. This gap persists because of the logistical and methodological difficulties of collecting and analysing granular engagement data in large-scale implementations. Using DytectiveU's longitudinal dataset, this research reveals that literacy gains are most significant during moderate engagement, while excessive participation leads to diminishing returns (Bettinger et al., 2023). These findings provide actionable recommendations for optimising CAL program designs and enhancing efficiency and equity in their implementation.

Our findings contribute to the growing evidence base on how AI-driven interventions can be sustainably integrated into public education systems. Unlike previous large-scale implementations, which often rely on external funding and intensive teacher support, DytectiveU operates at minimal cost with limited dependency on educators. The key challenge lies in understanding how AI-driven tools can be effectively scaled in diverse educational contexts while maintaining pedagogical efficacy. This challenge is particularly pronounced because most large-scale CAL evaluations have been conducted in controlled settings rather than government-led implementations. Consistent with the insights of Beg et al. (2023), who highlight the importance of embedding interventions into existing educational structures to ensure long-term sustainability, our study provides evidence that AI-driven literacy programs can achieve significant and scalable improvements while reducing the financial and logistical burdens typically associated with large-scale educational reforms.

The remainder of this paper is organised as follows. DytectiveU: an AI-powered literacy intervention describes the DytectiveU CAL program in detail, focusing on the software design, the adaptive personalisation features, and the implementation strategy carried out in public schools in Madrid. Data sources and sample characteristics present the longitudinal dataset collected over five school years, describe the main dependent and independent variables, and provide a detailed overview of student engagement patterns and demographic characteristics. Empirical strategy outlines the methodology used to estimate the impact of DytectiveU on literacy skills, addressing potential biases due to unobserved confounders, and discusses the specification of control variables and fixed effects. Results report the main empirical findings, showing a positive effect of session attendance on literacy outcomes, and identify diminishing marginal returns after a certain number of sessions. Mechanisms examine the pathways through which DytectiveU improves literacy skills, specifically analysing the roles of personalised content engagement and adaptive feedback in enhancing learning outcomes. Identification challenges for addressing causality explain the construction and validation of the instrumental variable based on the timing of first session use, discuss potential threats to identification, and present robustness checks to confirm the validity of the causal estimates. Finally, the Conclusion and future research directions summarise the key contributions of the study, highlight the implications for educational policy, and propose directions for future research to further understand and expand the impact of AI-driven learning technologies.

DYTECTIVEU: AN AI-POWERED LITERACY INTERVENTION

Software Design

We examine DytectiveU CAL, a language software developed by Change Dyslexia, an independent charity dedicated to reducing the number of students dropping out of school due to reading and writing difficulties in Spain. This computer game features a processing module that dishes out customized instructional material based on student feedback, adapting to each student's progress and providing immediate feedback. DytectiveU consists of two main components: (i) a web-based game for students and (ii) a back-end interface for supervisors such as school counselors, or teachers. Students put together avatars to interact with the game, stepping into a "detective academy" where they work through linguistic challenges in sessions lasting 20 minutes.

These sessions are made up of a set of personalised exercises. Supervisors can tap into the back-end interface to keep track of individual performance and weigh it against peers in the same age group. DytectiveU is packed with a vast collection of 42,000 exercises, put together by linguists and psychologists using two principal language resources: a list of 1,171 errors commonly made by individuals with dyslexia and a set of language resources developed by applying natural language processing techniques. These include lists of frequently used words tailored to different contexts, sets of words with similar phonological and orthographic patterns, and groups of commonly confused words. The exercises, which ramp up in complexity as students progress, are crafted to become more challenging by folding in more linguistic elements and distractors. They are structured in various difficulty levels and are personalised for each student based on factors such as age, the number of sessions completed, and performance in previous sessions. The personalisation process ensures that the exercises focus on areas needing reinforcement, addressing at least three of 17 cognitive abilities and seven performance measures related to literacy.

The content provided by DytectiveU is dynamically adaptive, adjusting to the unique learning needs of each student. As students interact with the exercises, the software collects various performance metrics, including the number of clicks, hits, speed, accuracy, and efficiency. These metrics map onto specific cognitive abilities and literacy performance measures, which inform the customisation of subsequent exercises. Based on a student's performance relative to peers of the same age, DytectiveU selects the next set of exercises to strengthen weaker cognitive skills or provide more challenging tasks for more substantial areas.

The software's user interface is designed to be engaging, providing immediate feedback that keeps students motivated and involved in their learning process. Correct answers are rewarded with points and highlighted in green, while incorrect responses are immediately indicated in red, with the correct answer displayed alongside to aid learning and prevent discouragement. As students complete more exercises, they earn more points, which can be used to personalise their avatars, enhancing the gaming aspect of the learning experience.

Integration of DytectiveU in primary schools is complemented by the Dytective Test, a tool designed to detect learning difficulties such as dyslexia. This web-based game involves linguistically motivated activities that help identify differences between individuals with and without dyslexia, using a Machine Learning model trained on data from both computer and tablet users. Evaluated with over 5,000 participants, the model achieves high sensitivity, identifying individuals with dyslexia with 80% accuracy, depending on the age group (Rello et al., 2020). While not all schools apply the Dytective Test universally, it is a valuable adjunct to the DytectiveU CAL program. Schools also received support through training sessions provided by speech therapists, linguists, and school counselors. These sessions covered the research behind the program and implementation strategies and addressed technical questions, ensuring that educators are well-prepared to effectively use these tools in enhancing literacy education.

Licensing Software and Software Deployment

The DytectiveU CAL language software was launched in public schools in January 2019 under the "Help Dyslexia" project, an initiative by the Ministry of Education of Madrid and the social charity Change Dyslexia. The software was designed to allow flexible use at school and home. Teachers and school counselors could integrate DytectiveU into classroom hours, or students could use it after school under the supervision of counselors, many of whom are qualified psychologists specializing in education. Additionally, students could access the software at home via a computer or mobile application. To fully benefit from DytectiveU, students were recommended to complete 64 challenges—sessions of 20 minutes each—spread across four sessions per week for eight weeks.

A session is a "challenge" consisting of linguistic exercises. Each session is represented as a "magnifying glass" on a pathway. These sessions are uniquely structured, featuring a variety of personalised linguistic exercises that students must complete. The total number of sessions attended quantifies this measure of engagement and participation within the educational program. Additional variables include demographic factors such as age and gender.

The DytectiveU program's licensing strategy is a targeted effort to deploy educational technology that enhances literacy skills among primary school students in Madrid, adapting to the evolving needs of schools and their students. The program began in the 2018-2019 school year with 98 public schools, primarily in socioeconomically disadvantaged areas. Participation in the DytectiveU program was voluntary and open to public primary schools. Enrolment was facilitated through a straightforward online process that required schools to provide contact information and a designated coordinator. Schools also needed to confirm staff support and inform the school board about their participation. The annual licensing agreements allowed schools to use the software throughout the school year, aligning with the school calendar from September to June. As shown in Table A.1, the number of participating schools has varied over the years. For example, nine schools exited the program in the 2019-2020 school year while three new schools joined, resulting in 92 schools continuing their participation. The following year saw an increase, with 93 new schools joining, totalling 174 for the 2020-2021 school year. The 2022-2023 school year saw a decline, with 71 schools exiting and only four new schools joining, reducing the total to 155. The annual review of licensing agreements allowed for flexibility, adjusting terms to better suit each school's needs and goals, ensuring effective integration of the DytectiveU program.

The structured financial model and support system of the DytectiveU licensing agreements were relevant for promoting accessibility and sustained engagement among schools, particularly in disadvantaged areas. The Regional Ministry of Education of Madrid covered the costs, making the program inclusive and accessible. For example, the 2021 renewal extended the program by 11 months with a budget allocation of 96,701.99 euros. Change Dyslexia, S.L., the organisation responsible for developing DytectiveU, provided essential support to schools, including training sessions and technical assistance to help educators integrate the software effectively. This support included best practices for using the software, interpreting data from DytectiveU and Dytective Test, and addressing any technical issues. By removing financial barriers and offering comprehensive support, the program created an inclusive environment where all students, regardless of socioeconomic background, could benefit from innovative educational tools.

The licenses were timed to coincide with the school year, allowing schools to fully integrate the software into their educational planning. In the 2018-2019 school year, the licenses became active on January 14, following the Regional Ministry of Education contract approval in

Madrid (Table A2 in the Appendix). In subsequent years, the licensing period extended from September to June, allowing for comprehensive integration into the curriculum.

The timing of the DytectiveU program's licensing and flexibility in start and end dates provides key insights into the varying levels of engagement and motivation across schools, especially those entering, continuing, or exiting the program. A review of Table A2 reveals distinct patterns in participation that can be linked to each school's commitment to the program. In the 2018-2019 school year, all schools began using DytectiveU on January 14, 2019, following the contract approval by the Regional Ministry of Education in Madrid. Schools that eventually exited the program showed less engagement from the start, as evidenced by their delayed entry on January 22, compared to schools that continued in the program, which started on the first day licenses were activated. Additionally, exiting schools stopped using the software two days earlier than the others, finishing on June 28, 2019. This slightly shorter usage period suggests lower motivation or engagement with the program among existing schools, possibly reflecting less integration of DytectiveU into their educational plans.

In contrast, schools that continued into the 2019-2020 school year showed consistent engagement, with no delay in starting or ending their usage. All schools began on September 1, 2019, but exiting schools again demonstrated lower commitment, concluding their use of the software on June 17, nearly two weeks earlier than others. This pattern of early withdrawal before the official end of the school year suggests that less-engaged schools may exit the program after demonstrating reduced motivation through earlier disconnection dates.

During the 2020-2021 school year, a more noticeable gap appeared in starting school enrolment dates. While continuing schools began promptly on September 1, 2020, entering schools delayed their start until October 17. This suggests that new schools require more time to integrate the program into their curriculum, perhaps reflecting the additional logistical or administrative hurdles associated with adopting a new educational technology. Interestingly, despite this later start, these entering schools this year completed the entire license period, suggesting that those who engaged late demonstrated an effort to maximize the use of the software.

In the 2021-2022 school year, all schools, including those exiting the program, began and ended their participation on the same dates (September 1, 2021, to June 30, 2022), indicating that by this point, even schools with lower future commitment had adhered to the standard usage period. The 2022-2023 school year presented a sharp deviation. Schools continuing the program began on September 1, 2022, but ended earlier than in previous years—on May 22, 2023—while entering schools joined later (September 18) and finished even earlier (April 10). It's important to note that this is the latest available data.

In summary, the timing of schools' start and end dates correlates with their overall engagement and motivation. Schools that exited the program consistently showed less commitment by starting later and finishing earlier than those that continued. Conversely, continuing schools demonstrated higher motivation levels, evidenced by their prompt start and full engagement throughout the license period. While showing later start dates, new schools eventually integrated the software fully but may have required additional time to adapt, reflecting the challenges of new technology adoption.

DATA SOURCES AND SAMPLE CHARACTERISTICS

Data Sources

The primary source of our data comprises detailed logs of student interactions with DytectiveU collected from 264 primary schools across the Madrid region. These logs include session attendance, performance metrics, and progression through various educational challenges designed to enhance literacy and cognitive skills.

Sample and Participants

The dataset covers five school years, from 2018-2019 to 2022-2023, and includes data from 264 primary schools in the Madrid region. Over time, participation varied, with some schools continuing across multiple years while others exited the program. Across these school years, the number of students engaged in the program also varied, reflecting both the participation trends of the schools and individual student involvement levels. In total, 34,607 unique students are represented in the dataset, with some appearing in multiple years due to continued use of the DytectiveU program. Each student's data includes the number of sessions attended per year, which provides a comprehensive view of engagement levels across the program's duration.

The dataset includes only those students who have participated in at least one session of DytectiveU. No students with zero sessions were included in the analysis, ensuring that the sample reflects active engagement with the program. Student data analysis shows varied engagement levels, with a fraction of students participating in multiple sessions across different years. This longitudinal data allows us to track improvements in reading skills over time, correlated with the frequency and intensity of program use. Additionally, it is important to note that some students participated in the program for only a single session. In contrast, others completed the full suite of 64 sessions, and some even exceeded this threshold, which was the recommended engagement level. This variability highlights differences in program adherence and provides a basis for examining the effectiveness of different levels of exposure to DytectiveU.

Variable Description

Dependent Variables: The primary outcome measure, "Composite Literacy Proficiency Score", captures students' cumulative learning across multiple skill domains. This measure spans several literacy and cognitive skill development dimensions, each critically relevant to reading proficiency. These dimensions include:

1. Reading Speed: This metric assesses how quickly a student can read text accurately. It helps identify improvements in fluency, an essential skill for adequate reading comprehension.
2. Reading Comprehension: This dimension evaluates a student's ability to understand and interpret the meaning of texts. It includes the ability to infer, summarize, and draw conclusions from written material.
3. Accuracy: Refers to the precision of reading, specifically how correctly students read words without errors. This metric includes various aspects of literacy, such as decoding skills, which involve translating written words into their spoken equivalents.
4. Spelling: The ability to correctly sequence letters to form words is relevant for reading and writing proficiency. This metric helps track students' knowledge of orthographic patterns and their ability to apply them in practical contexts.
5. Visual Perception: This dimension relates to the ability to distinguish and process visual stimuli, such as differentiating between letters and words. It is vital in recognizing letter shapes and understanding the spatial arrangement of text.
6. Auditory Perception: Assesses the ability to process and understand sounds. This skill is essential for phonological awareness, which underpins effective reading development, especially in decoding and recognizing words.
7. Attention and Working Memory: These are higher cognitive functions that involve focusing on tasks and retaining information for short periods. Attention and working memory are vital for following along with reading passages, understanding context, and integrating new information with existing knowledge.

Composite Literacy Proficiency is measured as a percentage, ranging from 0 to 100, and is observed at the end of the last session. The metric reflects the cumulative effect of all exercises completed, providing a comprehensive view of student progress across various literacy competencies. The emphasis on distinct and quantifiable metrics for literacy progress is critical in shaping educator preparation strategies and student performance outcomes (Riehl and Welch, 2023). These authors show that accountability exams with clearly defined precision metrics, especially those targeting specific proficiency margins, drive targeted instruction and lead to measurable improvements.

Independent Variables: The primary independent variable is the intensity of use, quantified by the total number of sessions completed by each student during the school years from 2018-19 to 2022-23. A session is designed to enhance various literacy skills through personalized, adaptive exercises.

Summary descriptive statistics

Figure A.1 presents the distribution of the number of sessions completed by participants in the DytectiveU program, with 44,997 participants recorded, involving 34,607 unique students across primary grades 1 to 6. Of these participations, 15,842 (35.2%) involved students completing 16 or more sessions, which represents one-fourth of the total recommended by the program. Additionally, 3,015 participants (6.7%) completed 64 or more sessions, meeting the recommended threshold for students at high risk of learning difficulties. Notably, some participations even surpassed the 64-session benchmark, indicating an exceptionally high level of engagement. Conversely, 4,830 participations (10.7%) involved students completing only a single session, after which they did not continue. The median number of sessions completed across all participations is 9, with a mean of 20.2 sessions. These findings indicate a high level of initial engagement yet a notable drop-off after the first session for a significant portion of the participants. It is possible that some students, in consultation with their counselor or teacher, decided not to continue with the program after the first session upon realizing that it was primarily targeted at supporting students with learning difficulties.

Tables 1 and A.3 provide a comparative breakdown of descriptive statistics for students who participated in DytectiveU over multiple school years (Table 1) versus those who participated for just one school year (Table A3). Both tables illustrate variations by year in Composite Literacy Proficiency and session engagement, which can be attributed to external factors such as the depth of collaboration between Change Dyslexia, the NGO behind DytectiveU, and the Madrid Regional Ministry of Education. This includes aspects like the scope of their agreement and the number of schools participating. The variability in the start dates of the first session each year, as shown in Table A1, could reflect different school start dates or varying program rollout strategies. Additionally, the earlier or later start dates across different school years may indicate changes in program roll-out due to contractual renewals or expansions in scope. Such external factors, even though unrelated to the students themselves, could impact their initial engagement with the program. The mean first session date, as presented in Table 1 and A3, provides insight into when students typically begin interacting with DytectiveU each school year. Other external factors related to the schools' commitment to the program, such as how well they integrate DytectiveU into their pedagogical project, also play a relevant role.

Table 1 offers insights into shifts in program deployment and administrative dynamics for students who have participated in the program for more than one year. In the 2018-2019 school year, the mean first session date is April 25, 2019. This was the first year of DytectiveU's implementation following the initial agreement between Change Dyslexia and the Regional Ministry of Education. As it was the inaugural year, it took time to establish the collaboration and roll out the program in schools. In the 2019-2020 school year, the mean first session date moved earlier to November 27, 2019, reflecting a timelier start. By 2020-2021, the mean first session date was later, recorded as March 12, 2021, due to the impact of COVID-19 and adjustments in program administration. The mean first session date then shifts again from 2021-2022 to January 4, 2022, indicating improvements in the roll-out process. Finally, the 2022-2023 school year shows an even earlier start date of November 2, 2022, highlighting continued efficiency in implementing the program.

Table 1 also shows that students who engaged with DytectiveU over multiple years tended to make significant strides in their reading skills. For instance, in the 2019-2020 school year, these students achieved an average Composite Literacy Proficiency of 50.19%, with a standard deviation 21.40. They completed an average of 35.61 sessions, underscoring the importance of sustained engagement with the program for better educational outcomes. The first session date for these students in 2019-2020 was November 27, 2019, while the last session occurred on March 16, 2020, indicating a consistent and sustained use of the program throughout the school year.

In contrast, Table A.3 details outcomes for students who participated in DytectiveU for just one school year. These students generally showed less progress than those who participated for multiple years. For instance, in the 2019-2020 school year, students in this category achieved an average Composite Literacy Proficiency of 31.29%, with a standard deviation 19.26. They also completed fewer sessions on average, with 12.21 sessions in 2019-2020. The first and last session dates for these students suggest a less sustained engagement with the program, which likely impacted their ability to benefit fully from DytectiveU's content. When comparing the data from Tables 1 and A.3, it becomes evident that longer participation and a higher number of sessions correlate with greater improvements in reading skills.

Table 1: Summary statistics by School year. Students participating in DytectiveU more than one School year

	Mean (1)	**s.d (2)**	**Min. (3)**	**Max. (4)**	**N (5)**
		2018-19			
Composite Literacy Proficiency	36.22	20.73	3.33	98.52	3,748
Number of Sessions	16.82	18.78	1	199	3,748
Female	0.479	0.50	0	1	3,748
Birth Day	180.37	106.04	1	366	3,748
First Session	April 25, 2019	34.62 days	January 16, 2019	June 30, 2019	3,748
Last Session	June 16, 2019	27.79 days	February 7, 2019	June 30, 2019	3,748
		2019-20			
Composite Literacy Proficiency	50.19	21.40	3.70	100	4,314
Number of Sessions	35.61	37.65	1	460	4,314
Female	0.486	0.50	0	1	4,314
Birth Day	181.43	105.18	1	366	4,314
First Session	November 27, 2019	62.70 days	September 1, 2019	June 28, 2020	4,314
Last Session	March 16, 2020	73.71 days	September 2, 2019	June 30, 2020	4,314
		2020-21			

	Mean (1)	s.d (2)	Min. (3)	Max. (4)	N (5)
Composite Literacy Proficiency	43.94	23.39	3.33	98.89	5,099
Number of Sessions	31.51	49.20	1	1024	5,099
Female	0.479	0.50	0	1	5,099
Birth Day	174.18	108.84	1	366	5,099
First Session	March 12, 2021	73.13 days	September 1, 2020	June 28, 2021	5,099
Last Session	June 8, 2021	55.02 days	September 1, 2020	June 30, 2021	5,099
		2021-22			
Composite Literacy Proficiency	49.57	21.75	4.07	100	5,672
Number of Sessions	35.51	49.54	1	1154	5,672
Female	0.494	0.50	0	1	5,672
Birth Day	173.72	109.52	1	366	5,672
First Session	January 4, 2022	73.01 days	September 1, 2021	June 29, 2022	5,672
Last Session	May 14, 2022	74.78 days	September 2, 2021	June 30, 2022	5,672
		2022-23			
Composite Literacy Proficiency	51.87	20.17	5.93	98.89	2,774
Number of Sessions	42.04	56.55	2	826	2,774
Female	0.504	0.50	0	1	2,774
Birth Day	173.55	109.25	1	366	2,774
First Session	November 2, 2022	35.06 days	September 1, 2022	February 22, 2023	2,774
Last Session	January 20, 2023	39.40 days	September 2, 2022	May 23, 2023	2,774

Notes. Table 1 reports summary statistics by school year for students who participated in DytectiveU during more than one school year. The table includes mean, standard deviation, and range values for key variables such as Composite Literacy Proficiency, number of sessions completed, gender, birth date (expressed as day of the year), and the dates of first and last session. These descriptive statistics reflect shifts in program deployment and student engagement across years. Sample and data. These statistics are based on panel data constructed from DytectiveU usage logs across five school years (2018–2019 to 2022–2023). The sample includes only those students who were matched across at least two different school years using a unique student identifier. The number of sessions is derived from platform-generated log counts, and Composite Literacy Proficiency is measured as a percentage. Birth day is recorded as the year's day (1–366), and session dates are converted to calendar format.

EMPIRICAL STRATEGY

Our benchmark equation is an OLS model of the form:

$$\mathrm{Y}_{\mathrm{i,s,g,t}} = \varphi + \beta_1 \mathrm{N}_{\mathrm{i,s,g,t}} + \alpha\, \mathrm{X}_{\mathrm{is,g,t}} + \gamma_{\mathrm{s}} + \alpha_g + \delta_t + \gamma_{\mathrm{s,t}} + \gamma_{\mathrm{s,g}} + \alpha_{g,t} + \epsilon_{\mathrm{i,s,g,t}} \quad (1)$$

Where $\mathrm{Y}_{\mathrm{I,s,t}}$: is the Composite Literacy Proficiency in DytectiveU of the last challenge in DytectiveU of student *i* of school *s*, in grade g and year t. Grades, g, range from the first go the sixth grade, covering the entirety of the primary education level. The school years, *t*, included in the study are 2018-2019, 2019-2020, 2020-2021, 2021-2022, and 2022-2023. $\mathrm{N}_{\mathrm{I,s,g,t}}$: is the number of sessions (measured in units) taken up by student i in school *s*, in year *t* which captures a student's intensity of use of DytectiveU. To control for school-year fixed effects γ_{s}, we introduce a dummy for each school year in our sample. This is to control for the school's specific variables that may be correlated with both a student's number of sessions and a student's speed of progress. School fixed effects likely affect students' overall progress and session attendance rates. Among these school-specific variables are the principal's leadership, student socioeconomic backgrounds, or the proportion of lagging students each year. These school-specific variables are likely correlated with an individual's engagement and participation in DytectiveU, and the school's motivation and functioning. Omitting them might bias the estimate of β_1. $\mathrm{X}_{\mathrm{i,s,g,,t}}$: a vector of student i in school *s*, grade g, and year *t* characteristics including gender, birth date and grade in which the student is enrolled. We incorporate a school dummy for each year $\gamma_{\mathrm{s,t}}$ to accommodate varying fixed effects annually, acknowledging that student populations and cohorts may differ from one year to the next within the same school. Additionally, the inclusion of school-year dummies also accounts for the fact that teacher rotations between schools might occur, which could influence student outcomes and DytectiveU engagement. δ_t represents year dummies that control for the variation in the month of contract commencement, which dictates when students can begin engaging with the DytectiveU program. The initiation date of each annual contract is entirely exogenous, bearing no relation to specific characteristics of the schools, the motivations of principals and teachers, or the attributes of the students. These year dummies also capture the interruption of the COVID-19 in the school face to face clasess in 2019-2020 and 2020-21. We have also added grade dummies to the analysis α_g, considering the observed trend that younger primary students engage more with DytectiveU, and the potential for greater impact at these foundational levels. School-grade dummies $\gamma_{\mathrm{s,g}}$ a, are included to control for cohort-specific attributes and teacher influences within schools. Furthermore, grade-year interactions $\alpha_{g,t}$are included to capture the synergies between the curriculum of each school year by grade and the DytectiveU program's effect. Finally, we include grade-by-year fixed effects $\gamma_{\mathrm{s,t}}$ to account for the fact that the influence of each year may vary across schools, particularly considering that events like the COVID-19 pandemic affected schools differently, potentially altering their engagement with DytectiveU and their literacy development trajectories. $\in_{i,s,g,t}$: is the error term.

RESULTS

Table 2. OLS estimates

	(1)	(2)	(3)	(4)
Number of Sessions	0.395*** (0.002)	0.327*** (0.002)	0.313*** (0.003)	0.301*** (0.002)
Female				0.961*** (0.133)
Birthday (in days)				-0.011*** (0.001)
Grade				3.731*** (0.488)
Constant	28.72*** (0.098)	32.95*** (0.734)	25.75*** (1.351)	15.24*** (2.407)
School FE	NO	YES	YES	YES
School*Year FE	NO	NO	YES	YES
School*Grade FE	NO	NO	NO	YES
Year*Grade FE	NO	NO	NO	YES
Observations	44,997	44,997	44,997	44,997
Adjusted R-squared	0.367	0.489	0.525	0.645

Notes: The dependent variable is Composite Literacy Proficiency Score measured as the literacy proficiency score in the final DytectiveU challenge completed by student *i* in school *s*, grade *g*, and year *t*. Estimates are based on the following Ordinary Least Squares (OLS) regression model:

$Y_{i,s,g,t} = \beta_0 + \beta_1 N_{i,s,g,t} + \gamma_s + \delta_t + \alpha_g + \gamma_{s,t} + \gamma_{s,g} + \alpha_{g,t} + \in_{i,s,g,t}$ (1). Where $Y_{I,s,t}$: is the Composite Literacy Proficiency in DytectiveU of the last challenge in DytectiveU of student *i* of school *s*, in grade *g* and year t. Grades, *g*, range from the first go the sixth grade, covering the entirety of the primary education level. The school years, *t*, included in the study are 2018-2019, 2019-2020, 2020-2021, 2021-2022, and 2022-2023. $N_{I,s,g,t}$: is the number of sessions (measured in units) taken up by student *i* in school *s*, in year *t* which captures a student's intensity of use of DytectiveU. $(N_{I,s,g,t})^2$: is the squared number of sessions taken up by student *i* in school *s,* grade *g*, and year *t*. The model progressively incorporates fixed effects by column. The sample consists of 44,997 student-year observations from the 2018–2019 to 2022–2023 academic years, covering grades 1 to 6 in Primary education.
Standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

Table 2 presents the results of an Ordinary Least Squares (OLS) regression analysis of the relationship between the Composite Literacy Proficiency Score and the number of sessions attended, including school-fixed effects. In Column (1), the coefficient of interest, denoted as β_1, is 0.395, is positive and significant at the 1% level. This positive coefficient indicates a direct relationship between the number of sessions attended and a student's reading progress. Specifically, students increase their Literacy Proficiency by 0.395 percentage points for each additional session. Given that the average rate of the Composite Literacy Proficiency Score is 36.45% within the sample, this translates to an approximate increase of 1.08% (0.395/36.45) in their reading skills. When comparing two students—Student A and Student B, who are identical except that Student A has completed one more session—Student A demonstrates 0.395 percentage points more progress in the Composite Literacy Proficiency Score by the end of the school year. The R-squared value of 0.363 indicates the extent to which the number of sessions explains the variation in global reading progress.

Column (2) shows that including school fixed effects reduces the coefficient for the number of sessions by 17%, from 0.395 to 0.327. Nevertheless, it remains positive and significant at the 1% level, with school fixed effects also being jointly significant. This suggests that excluding school-fixed effects led to overestimating the impact of session attendance on reading skills. School management and teachers' interest and emphasis on participation in DytectiveU are linked to students' Composite Literacy Proficiency and the number of sessions they complete using the digital tool. By incorporating school fixed effects, we compare students within the same school while controlling for factors like school infrastructure, teaching quality, and student-to-teacher ratios. The R-squared value increases from 0.367 to 0.489, indicating that the number of sessions and the school context account for a significant portion of the variability in global reading progress.

Column (3) further examines the relationship by adding school-year interactions to the model. The coefficient for the number of sessions remains positive and significant at the 1% level, but decreases slightly to 0.313. These interactions during the school year highlight the importance of considering changes within each school over different school years, such as variations in administration, teachers, or student cohorts. The R-squared value rises to 0.525.

In Column (4) we control for specific student characteristics such as gender, birth date, and the student's degree. These variables may capture underlying differences in the Composite Literacy Proficiency Score, reflecting gender-based disparities in reading, variations in maturity related to age, and potential differences linked to grade. Bedard and Dhuey (2006) highlight that older students perform better in the early grades due to age-based maturity differences. This maturity gap can lead to significant performance disparities, with older students scoring higher in standardised tests. Similarly, gender differences in reading achievement are well-documented, with boys typically lagging behind girls (Aucejo et al., 2022). These gender, age-related, and degree-based differences could potentially introduce bias in our estimate of β_1. For instance, if girls, who perform better in reading, are also more engaged in DytectiveU, leading to more sessions, β_1 might be upwardly biased. Likewise, if younger students, lagging in maturity and reading skills, tend to complete more DytectiveU sessions, not controlling for birth date could bias the estimation of β_1 downwards. Additionally, research by Bird et al. (2024) highlights how predictive analytics in education can reinforce disparities when models are insufficiently calibrated for underrepresented groups, underscoring the importance of addressing algorithmic bias to ensure equitable resource allocation. Amuedo-Dorantes and Wang (2024) also emphasise the importance of addressing confounding factors to accurately estimate causal relationships by carefully accounting group construction and controls. To account for these potentially confounding factors, we estimate the effect of these underlying student attributes on the Composite Literacy Proficiency Score. The gender coefficient is positive and highly significant (0.961), confirming that female students generally outperform male students in reading. The coefficient for birth date is negative (-0.011), showing that younger students within the same grade tend to experience greater reading progress challenges than their older peers. This finding aligns with existing literature on age and educational outcomes, which typically observes that younger students within the same grade lag behind their older counterparts in the same cohort due to differences in maturity. Furthermore, the inclusion of "grade" as a variable reveals a strong positive effect (3.731), emphasising the importance of academic level on reading progress. Adding student characteristics, alongside school, year, and grade fixed effects, slightly enhances the model's predictive power, with an R-squared value of 0.645. In summary, these findings reinforce the positive impact of the DytectiveU program on Composite Literacy Proficiency Score and provide insights into how individual student characteristics interact with the program's effectiveness.

MECHANISMS

Personalised content engagement

To examine the importance of personalised content engagement, we analyse how the effects of DytectiveU vary across students at different developmental stages. Table 3 examines the characteristics of students who complete just one session, between 2 and 64 sessions, and those who complete more than 64 sessions—the threshold identified as necessary for significant improvements in reading and writing skills. This breakdown allows us to investigate whether there are systematic differences between students with varying levels of engagement and assess whether selection bias might be present. Amuedo-Dorantes and Arenas-Arroyo (2019) point out the importance of accounting for unobserved heterogeneity when analysing educational interventions, as neglecting these factors can bias causal estimates and misattribute effects. By comparing these groups, we aim to identify the characteristics driving different levels of participation and determine if these characteristics, rather than session attendance itself, could influence the observed learning outcomes.

Table 3 presents the demographic and engagement characteristics of students in the DytectiveU program, categorised by their participation levels: students who completed one session, those who completed between 2 to 64 sessions, and those who completed more than 64 sessions. Significant differences in reading and writing progress emerge across these groups. Students with just one session had limited progress (10.22 on average), while those completing 2-64 sessions made considerably more progress (37.16). The group with more than 64 sessions saw the most substantial gains, with an average improvement of 74.04. This strong positive correlation between the number of sessions and progress underscores the impact of sustained engagement with the program. Among students who completed one session, 48% were female, compared to 49% in the 2-64 session group, and 46% in the group completing more than 64 sessions. The decrease in the proportion of female students in the highest engagement group (more than 64 sessions) is statistically significant at the 5% level. The gender variable highlights an interesting dynamic: female students, who generally tend to have stronger baseline reading skills, are slightly less likely to complete a high number of sessions (46% in the group with more than 64 sessions compared to 49% in the 2-64 sessions group). If we still find a significant effect of the number of sessions on Composite Literacy Proficiency despite this, it will indicate that the impact of the program is robust and less likely to be driven by endogeneity or selection bias. The observed progress could not simply be attributed to the initial advantage of stronger baseline skills among female students.

The birth year variable reinforces that students who completed more sessions were younger. Those who finished just one session had an average birth year of 2010.98, while students with 2-64 sessions had an average birth year of 2011.28, and those with more than 64 sessions were born, on average, in 2011.44. This finding implies that younger students, typically with less developed reading skills, are likelier to attend more sessions. The grade variable represents the average educational level of students within the groups. For example, students who completed just one session had an average grade of 3.98, roughly equivalent to fourth grade. This period is crucial as students transition from "learning to read" to "reading to learn," making the program particularly relevant. On the other hand, students completing 2-64 sessions were in slightly lower grades (mean grade of 3.68, between the third and the fourth grade), while those completing more than 64 sessions were in marginally higher grades (mean grade of 3.86). It suggests that DytectiveU may be especially beneficial in the lower grades, where students are still learning the basics of reading and writing. Younger students, still at earlier stages of cognitive development, are more likely to engage intensively with DytectiveU, consistent with the idea that they respond more positively to personalised instruction. Younger students could show larger improvements when exposed to tailored content, suggesting that early personalised interventions unlock their greater learning potential (Heckman, 2006).

The day of birth variable compares students of the same grade but born at different points in the year. Students completing just one session were born earlier (average day of June 17th), while those with 2-64 sessions were born slightly later (June 23rd), and those with more than 64 sessions were born around June 22nd. This pattern suggests that students born later in the year tend to participate in more sessions, which is typically associated with lower academic performance due to their younger relative age. Despite their relative age disadvantage, these students exhibited stronger literacy gains, indicating that DytectiveU's personalised learning approach successfully adapted content to their developmental stage. This ability to align instruction with students' cognitive readiness helped bridge initial gaps in skills acquisition, further reinforcing the role of personalisation as a key mechanism behind the program's effectiveness.

First session timing also differs significantly. Students who completed only one session had their first session 202.1 days after September 1st (late March), while those with 2-64 sessions started earlier, around 164.8 days (mid-February). Those completing more than 64 sessions began much earlier, about 81.1 days after September 1st (late November). Significantly, the timing of the first session is not determined by the individual student but by the school, with all students from the same class beginning simultaneously. These differences indicate that earlier engagement with the program is associated with more sessions completed. Regarding the last session variable, those who completed only one session had their previous session on the same day (202.1 days from September 1st). In contrast, students with 2-64 sessions had their last session around 222.5 days (mid-April), and those with more than 64 sessions extended participation to 225.4 days (also mid-April). These differences emphasize that students who stay engaged for longer tend to complete more sessions. Earlier exposure to personalised learning content may have provided these students with a longer and better-aligned instructional trajectory.

In summary, while younger students and those in lower grades are more likely to complete more sessions—suggesting that DytectiveU may be particularly beneficial in the early stages of education—the evidence from birth day comparisons points even more clearly to the conclusion that later-born students engage in more sessions and still perform better than expected. Furthermore, the data reveals significant differences between students who complete only one session and those who complete more. However, there are much smaller differences between the group completing 2–64 sessions and those completing more than 64. This pattern helps to reduce concerns about selection bias. It is important to emphasise that younger students, those enrolled in lower grades, and those born later in the academic year tend to complete more sessions, reinforcing that personalised content engagement is a key mechanism behind DytectiveU's impact on literacy outcomes. Students who engaged with the program at earlier stages of cognitive development achieved greater improvements, confirming that individualised instruction effectively harnesses cognitive plasticity and optimises learning during formative years. Despite these baseline differences, the fact that students still show significant improvements in reading skills suggests that it is session attendance—rather than selection bias—that drives better literacy outcomes.

Table 3. Characteristics of students by session intensity

	1 session (1)	**2-64 sessions (2)**	**More than 64 sessions (3)**	**Difference (2-1)**	**Difference (3-2)**
Composite Literacy Proficiency	10.22	37.16	74.04	26.93***	36.88***
	(4.20)	(19.83)	(16.20)	(0.12)	(0.32)
Female	0.48	0.49	0.46	0.01	-0.03**
	(0.50)	(0.50)	(0.50)	(0.01)	(0.01)
Birth Year	2010.98	2011.28	2011.44	0.30***	0.16***
	(2.05)	(1.98)	(1.65)	(0.03)	(0.03)
Grade	3.98	3.68	3.86	-0.30***	0.18***
	(1.56)	(1.54)	(1.38)	(0.02)	(0.03)
Birth day of the year	169.67	175.88	174.95	6.20***	-0.92
	(110.07)	(108.32)	(108.99)	(1.68)	(2.09)

	1 session (1)	**2-64 sessions (2)**	**More than 64 sessions (3)**	**Difference (2-1)**	**Difference (3-2)**
First session	202.1	164.8	81.1	-37.24***	-83.72***
(From 1st September)	(62.64)	(72.38)	(71.41)	(0.98)	(1.37)
Last session	202.1	222.5	225.4	20.43***	2.85*
(From 1st September)	(62.64)	(69.46)	(83.12)	(0.97)	(1.57)

Notes: This table shows the characteristics of students performing 1 session, 2-64 sessions, or more than 64 sessions. The first three columns display the means for these groups, while the last two columns present the differences between the groups. The numbers in parentheses represent the standard deviations for the means in the first three columns, and the standard errors for the differences between groups in the last two columns. Statistical significance is denoted by ***p<0.001.

Adaptive feedback

Table 4. Testing the adaptive feedback mechanism

	(1)	(2)	(3)	(4)
Number of Sessions	0.561*** (0.004)	0.547*** (0.003)	0.548*** (0.003)	0.548*** (0.003)
(Number of sessions)^2	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Female				1.006*** (0.119)
Birthday (in days)				-0.011*** (0.001)
Grade				3.834*** (0.436)
Constant	23.491*** (1.234)	14.76*** (1.775)	14.72*** (1.846)	12.82*** (2.152)
School FE	YES	YES	YES	YES
School*Year FE	YES	YES	YES	YES
School*Grade FE	NO	YES	YES	YES
Year*Grade FE	NO	NO	YES	YES
Observations	44,997	44,997	44,997	44,997
Adjusted R-squared	0.604	0.700	0.701	0.716

Notes: The dependent variable is the Composite Literacy Proficiency Score measured as the literacy proficiency score in the final DytectiveU challenge completed by student *i* in school *s*, grade *g*, and year *t*. Estimates are based on the following Ordinary Least Squares (OLS) regression model:

$$\mathrm{Y}_{\mathrm{i,s,g,t}} = \beta_0 + \beta_1 \mathrm{N}_{\mathrm{i,s,g,t}} + \beta_2 (\mathrm{N}_{\mathrm{i,s,g,t}})^2 + \gamma_{\mathrm{s}} + \delta_t + \alpha_g + \gamma_{\mathrm{s,t}} + \gamma_{\mathrm{s,g}} + \alpha_{g,t} + \epsilon_{\mathrm{i,s,g,t}} \quad (2).$$

Where $\mathrm{Y}_{\mathrm{I,s,t}}$: is the Composite Literacy Proficiency in DytectiveU of the last challenge in DytectiveU of student *i* of school *s*, in grade *g* and year t. Grades, *g*, range from the first go the

sixth grade, covering the entirety of the primary education level. The school years, *t*, included in the study are 2018-2019, 2019-2020, 2020-2021, 2021-2022, and 2022-2023. $N_{I,s,g,t}$: is the number of sessions (measured in units) taken up by student *i* in school *s*, in year *t* which captures a student's intensity of use of DytectiveU. $(N_{I,s,g,t})^2$: is the squared number of sessions taken up by student *i* in school *s,* grade *g*, and year *t*. This squared variable is included because of potential decreasing returns to scale. The model progressively incorporates fixed effects by column. The sample consists of 44,997 student-year observations from the 2018–2019 to 2022–2023 academic years, covering grades 1 to 6 in Primary education.
Standard errors are reported in parentheses. Significance levels: *** p<0.01, ** p<0.05, * p<0.1.

In Column (1), the model includes the squared number of DytectiveU sessions to capture potential diminishing returns in students' global reading progress. The linear term coefficient (β_1) is 0.561, positive and significant at the 1% level, indicating that more sessions continue to affect reading progress positively. Conversely, the coefficient for the squared term (β_2) is -0.001, negative and significant at the 1% level, confirming the hypothesis of diminishing returns. This results in an inverted U-shaped relationship, implying that while additional sessions enhance performance, there is a threshold beyond which the impact begins to decline. Column (2) introduces school-grade interactions alongside school fixed effects and school-year interactions. This allows for an examination of differences in reading progress due to varying characteristics of teachers across different grades within the same school. The coefficient associated with the number of sessions (β_1) remains positive and significant at the 1% level, with a value of 0.547. This result suggests that session attendance positively influences reading progress across grade levels. The significant school-grade interactions imply that reading progress may vary across different schools and within different grades in the same school.

Column (3) adds year-grade interactions alongside school, school-year, and school-grade interactions. The coefficient for the number of sessions β_1 remains positive and significant at the 1% level, with a value of 0.548. This slight increase suggests that the relationship between session attendance and the Composite Literacy Proficiency remains robust, even after accounting for year-grade variations. Year-grade interactions are highly significant, highlighting the potential influence of cohort effects or changes in curriculum across different school years and grades. The R-squared value in Column (6) is 0.701, indicating that the inclusion of all interactions significantly improves the model's explanatory power regarding global reading progress.

Column (4) presents the most comprehensive specification, incorporating school, school-year, school-grade, and year-grade fixed effects and student characteristics such as gender, birth date, and grade. The coefficient for the number of sessions remains positive and statistically significant at the 1% level, with a value of 0.548. This confirms the robustness of the positive relationship between session attendance and improvements in the Composite Literacy Proficiency Score, even after accounting for a wide range of fixed effects and individual characteristics. The coefficient for the squared number of sessions remains negative and significant (-0.001), reinforcing the evidence of diminishing returns as the number of sessions increases. The estimated turning point is 274 sessions. Literacy proficiency gains continue to increase up to around 274 completed sessions, beyond which the marginal benefits of additional sessions begin to diminish. Among the student characteristics, being female is associated with significantly higher literacy proficiency gains (1.006). In contrast, a later birthdate within the year is associated with lower reading progress (-0.011), consistent with findings in educational psychology on the relative age effect. The grade coefficient (3.834) remains significant and positive, indicating greater reading gains at higher grade levels. Including all fixed effects and controls raises the adjusted R-squared to 0.716, suggesting that the model explains a substantial portion of the variance in literacy progress.

IDENTIFICATION CHALLENGES FOR ADDRESSING CAUSALITY

Instrumental Variables Approach for Estimating the Impact of DytectiveU Sessions and Robustness Checks

A challenge in identifying causal effects in this context is the presence of unobserved student characteristics that may influence both the number of sessions attended and the rate of progress in Proficiency Literacy. For our study, these confounders could include inherent student traits like motivation, which could simultaneously affect the likelihood of attending more DytectiveU sessions and the observed improvements in reading skills. If unaccounted for, the estimated effect of session frequency on reading progress might reflect these underlying motivational factors rather than the true impact of the sessions themselves. If this is the case, the estimate β_1 in equation (1) will not necessarily be capturing the effect of doing additional sessions on the rate of progress, but rather the impact of being a motivated student. To address these endogeneity issues, we employ an instrumental variables (IV) strategy, leveraging exogenous variation provided by the timing of the first DytectiveU session, which is assumed to be independent of the students' unobserved characteristics. We provide stronger causal evidence by ensuring these instruments are uncorrelated with pre-existing school characteristics that might affect educational outcomes.

Instrument Specification and Rationale

Our choice of instrument, the timing of the first DytectiveU session, exploits administrative decisions external to and independent of any individual student's influence. The initial use date is predominantly determined by factors such as the school's pre-set schedule for language classes or other activities, availability of computer rooms, and the administrative allocation of digital resources, which are scheduled at the beginning of the school year and are unaffected by student or teacher preferences.

First-Stage Regression Model:

$$\mathrm{N}_{\mathrm{i,s,g,t}} = \alpha_0 + \alpha_1 Z_{is,g,t} + \alpha\, \mathrm{X}_{\mathrm{is,g,t}} + \gamma_{\mathrm{s}} + \alpha_g + \delta_t + \gamma_{\mathrm{s,t}} + \gamma_{\mathrm{s,g}} + \alpha_{g,t} + \epsilon_{i,s,t} \quad (3)$$

Where $Z_{i,s,g,t}$ are the instruments for student *i* in the school *s*, in grade *g* and year *t* and α_0 is a constant term. The rest of the variables are defined as in equation (1). In our analysis, we use one instrument. This model quantifies the instrument's effect on the number of sessions attended, isolating the exogenous variation necessary for a valid IV estimation. Following

Collinson et al. (2024), we conduct this analysis to ensure the instrument's validity by testing its relevance and homogeneity.

Table 5. First-stage results

	Number of Sessions (First Estimation)	**Number of Sessions (Second Estimation)**	**Squared Number of Sessions (Second Estimation)**
	(1)	**(2)**	**(3)**
First session	-0.2337***	-0.5538***	-145.447***
	(0.0038)	(0.0105)	(4.7032)
First session²		0.0012***	0.3728***
		(0.00004)	(0.0163)
Female	-0.6966***	-0.7296***	-173.191
	(0.2636)	(0.2603)	(116.762)
Birth day of year	0.0010	0.0013	-0.0624
	(0.0013)	(0.0013)	(0.5731)
Grade	-1.1873	-0.0178	-95.090
	(0.9689)	(0.9578)	(429.552)
Cragg-Donald Wald F statistic		3745.60	58.36
Stock-Yogo weak ID test critical values		10% maximal IV size: 16.38	10% maximal IV size: 7.03
Underidentification test (Anderson LM)		$\chi^2(1)$= 3590.51, P-val = 0.0000	$\chi^2(1)$= 121.26, P-val = 0.0000

Notes: This table reports the first-stage estimates from three instrumental variable specifications. Column (1) presents the results from estimating equation (3), where the number of sessions is the only endogenous variable. Columns (2) and (3) report results from the second specification, in which both the number of sessions and its squared term are treated as endogenous. The instrument used in all columns is the number of days from the beginning of the school year to the first DytectiveU session, and in columns (2) and (3) we also include its square. The regressions include the full set of student controls and fixed effects used in the main specification: female, birth day of the year, school, grade, and year fixed effects, along with school-year, school-grade, and grade-year interactions. Standard errors are clustered at the school level. The sample includes 44,997 student-year observations from 264 public primary schools in Madrid over the 2018–2023 school years. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Validation of the Instrumental Variable

Relevance: The instrument's relevance is evidenced by its significant influence on the number of sessions a student attends. Given the structured academic calendar with a fixed end date, earlier engagement in DytectiveU allows for potentially more sessions. In the first estimation, the only endogenous variable is the number of sessions. In contrast, in the second estimation, both the number of sessions and the squared number of sessions are included and treated as endogenous variables. The instruments used are as follows: in the first estimation, the instrument is the first day of DytectiveU usage, measured in days since the start of the school year (September 1st). In the second estimation, two instruments are employed: the first day of DytectiveU usage and the square of the first day of DytectiveU usage.

The first-stage results in Table 5 show that the instruments have strong explanatory power in both estimations, as indicated by the Cragg-Donald Wald F statistic. The instrument(s) in the first and second estimation pass the weak identification test based on Stock-Yogo critical values, further confirming the validity of the instruments in explaining the variation in the endogenous variables.

In the first estimation, the first day of DytectiveU usage significantly negatively affects the number of sessions, with a coefficient of -0.2337 ($p < 0.01$), indicating that earlier engagement with the program results in more sessions completed. In the second estimation, the linear and squared terms of the first day of DytectiveU usage are significant. The linear term has a negative coefficient of -0.5538 ($p < 0.01$), while the squared term has a positive coefficient of 0.0012 ($p < 0.01$), suggesting a nonlinear relationship between the timing of the first session and the number of sessions completed. Both estimations include the control variables—female, birthday of year, and grade—with similar effects across models.

Table A4 shows that the first session date of DytectiveU use exhibits no within-school variation, but significant between-school variation, further supporting the instrument's validity. In 2019, the average first session occurred around 207.16 days after September 1st, with no within-school variation, indicating that students within the same school start simultaneously. The first session date varies widely across schools, ranging from 135.54 days to 302.92 days. This pattern is consistent across other years in the dataset, where within-school variation is non-existent, but between-school variation is substantial. This reinforces our assumption that the timing of the first session is determined by administrative and logistical factors at the school level, rather than by individual student characteristics like motivation or prior performance. The exogenous variation we exploit in our instrumental variable strategy is not driven by unobserved student characteristics, but rather by factors external to the students themselves, strengthening the causal interpretation of our results.

Exogeneity: The initial use date is determined by factors external to the student and the teacher, such as the school's predetermined weekly timetable for language classes or sessions with school counselors, established at the beginning of the school year. Additionally, the availability of the computer room, which other teachers could pre-book, also influences the first day of DytectiveU usage. Our study focuses on public schools where computer equipment provisioning comes from funding by the Community of Madrid's administration, following objective criteria[1]. This ensures a uniform level of resources across schools, supporting the

[1] Decree 149/2000 of June 28, regulating the legal framework for the management autonomy of non-university public educational schools of the Region of Madrid.

exogeneity of the first day of DytectiveU use as an exogenous variable. Additionally, initiating DytectiveU sessions might coincide with other academic activities, such as assessments or control tests, which could delay the program's start. Such scheduling decisions are made long before the availability of DytectiveU.

To address these endogeneity concerns, the study also incorporates a secondary dataset of comprehensive administrative records and survey data collected from students, families, and teachers across the Madrid region from 2015/16 to 2018/19. This dataset includes detailed performance metrics from standardised tests and socio-economic demographics, which are used specifically to test the exogeneity of the instrumental variable—the timing of the first DytectiveU session. Both datasets are anonymized, containing no identifiers linking data to specific individuals, and only specify the educational school involved, thereby ensuring privacy and confidentiality. This setup prevents direct linking between the datasets, but allows for robust exogeneity testing of the instrumental variable within the context of the broader educational settings in Madrid.

Table 6 provides insight into the various factors, such as language proficiency, academic subjects, student family characteristics, and how they predict the number of sessions. The critical evaluation lies in the second column, which assesses the influence of these same factors on the timing of the first session. The exogeneity criterion requires that the instrumental variable, in this case, the first session, should not be directly influenced by these other factors.

Table 6: Testing Balance

	Number of Sessions	**First Session**
	(1)	**(2)**
Spanish	-2.419	-1.156
	(2.773)	(9.232)
English	-1.205	6.364
	(2.188)	(7.285)
Mathematics	4.665*	-7.204
	(2.477)	(8.246)
1-10 Books at home	-5.576	4.347
	(9.159)	(30.49)
10-50 Books at home	-12.726	-5.821
	(8.917)	(29.69)
50-100 Books at home	-5.383	-9.993
	(9.470)	(31.53)
100-200 Books at home	-15.238	35.002
	(11.210)	(37.308)
Mother with no Education	-21.489	119.145
	(49.665)	(165.35)
Mother with primary Education	-33.285	-35.743

	Number of Sessions	First Session
	(1)	(2)
	(35.875)	(119.44)
Mother with compulsory Education	-21.301	4.661
	(34.779)	(115.79)
Mother with upper secondary education	-19.203	-41.329
	(34.196)	(113.85)
Mother with higher vocational training	-16.056	-6.243
	(37.296)	(124.17)
Mother with Associate's degree	-39.657	17.04
	(35.222)	(117.3)
Mother with Bachelor's degree	-12.873	-92.719
	(34.551)	(115.03)
Mother with Master's degree	-33.334	63.815
	(38.514)	(128.22)
Mother with PhD	8.978	-26.530
	(41.126)	(136.92)
Constant	44.749	198.062*
	(35.275)	(117.44)
Observations	252	252
Adjusted R-squared	0.02	0.01
F-test (16, 235)	1.34	1.08
Prob > F	0.175	0.374

Notes: This table reports results from two Ordinary Least Squares (OLS) regressions that test the exogeneity of the instrumental variable, first session date (measured in days from September 1st). Column (1) regresses the number of DytectiveU sessions completed by the student on student-level background characteristics, including standardised test scores in core subjects (Spanish, English, and Mathematics), number of books at home, and maternal education. Column (2) uses the same covariates to predict the first session date, the instrument used in our main IV strategy. The exogeneity assumption requires that the timing of first DytectiveU use be uncorrelated with variables that affect learning outcomes. As expected, several covariates are predictive of program engagement in Column (1) (e.g., mathematics scores), but critically, none are jointly predictive of the instrument in Column (2). The F-test for joint significance in Column (2) is not statistically significant (F = 1.08; p = 0.374), indicating that the instrument is uncorrelated with baseline academic and socioeconomic variables. This supports the claim that the first session date is determined by school-level logistical factors (e.g., classroom scheduling, device availability) rather than student-level attributes, and can be treated as an exogenous source of variation. Standard errors are reported in parentheses. Both models include 252 school-grade-year cells. *** p<0.01, ** p<0.05, * p<0.1.

The first column of Table 6, Number of Sessions, reports how various student characteristics—such as academic performance in subjects like Spanish, Mathematics, and Technology, as well as family background (e.g., number of books at home, mother's education level)—are related to the total number of sessions attended. The second column, First Session, shows how these same variables influence the timing of the first session.

In the first column ("Number of Sessions"), the F-test is statistically not significant ($F = 1.34$, $Prob > F = 0.175$). The math score is individually significant (4,665); schools with a one standard deviation higher math score have almost 5 Dytective-U sessions on average. This is expected, as factors like student performance could reasonably influence participation levels in the educational program. The F-test for the second column (First Session) is not statistically significant ($F = 1.34$, $Prob > F = 0.374$). This non-significance means that we cannot reject the null hypothesis that the coefficients of the predictors are jointly equal to zero, validating the assumption that the first session is exogenous to these variables included in the model. This finding is relevant for the analytical framework of the study. The first session can be considered as a valid instrumental variable for estimating the causal effect of session attendance on reading skills improvement. External factors, such as academic proficiency or socio-economic status, do not directly influence when a school begins using DytectiveU, making the First session a strong candidate for a valid instrument.

Empirical Strategy and Detailed Implementation

Building on the first stage results, our second-stage regression uses the predicted number of sessions to ascertain the impact on Profiency Literacy:

$$\mathrm{Y}_{\mathrm{i,s,g,t}} = \varphi_0 + \varphi_1 \left(Predicted\ N_{\mathrm{i,s}}\right) + \varphi_2 Predicted(N_{i,s})^2 + \gamma_{\mathrm{s}} + \alpha_g + \delta_t + \gamma_{\mathrm{s,t}} + \gamma_{\mathrm{s,g}} + \alpha_{g,t} + \epsilon_{\mathrm{i,s,g,t}} \quad (4)$$

Where $\mathrm{Y}_{\mathrm{i,s}}$ is defined as in equation (1), the Composite Literacy Proficiency in DytectiveU of the last challenge in DytectiveU of student *i* of school *s* in grade *g* and year *t*. This specification allows us to isolate the effect of the number of sessions on reading skills development, adjusting for other covariates and ensuring that the estimated effects are purely attributable to the program's exposure, not confounded by pre-existing conditions or selection biases. By rigorously testing these conditions with empirical data, we enhance the robustness of our findings by careful validation of each IV assumption—relevance and exogeneity—and by rigorously testing these conditions.

Following Muralidharan et al., (2019) and Büchel et al., (2022) instrumental variable estimations, φ_1 can be interpreted as the impact of the treatment under the assumption of full participation in the DytectiveU Program. This implies a scenario where treated students engage with all 64 challenges in the program. The estimation can be described as an Intention-to-Treat (ITT) analysis, a "catch-all" measure regardless of the treatment or lack of treatment received, and regardless of their withdrawal from the treatment based on an average number of sessions of 20.2.

In conducting our analysis, we consider additional assumptions beyond the typical monotonicity and independence as highlighted by Angrist and Pischke (2008), Muralidharan et al. (2019) and Büchel et al. (2022). One key assumption is the homogeneity of treatment effects across students, while another pertains to the functional form of the relationship between session attendance and the Composite Proficiency Literacy Score. Similar to findings in Büchel et al. (2022), our data suggests deviations from these assumptions. Particularly, the relationship between session attendance and Proficency Literacy appears to be non-linear, as detailed in Table 4, showing decreasing returns from attending additional sessions. This pattern implies a potential downward bias in our Instrumental Variable (IV) estimates, potentially leading to an understatement of the true impact of full participation.

Main Results

Table 7 presents the results of the second-stage instrumental variable (IV) regression, examining the relationship between the Composite Literacy Proficiency Score and the number of sessions in the DytectiveU program. The analysis employs the timing of the first session as an instrumental variable, controlling for potential confounders and addressing issues related to endogeneity. The results are robust, showing a consistent positive relationship between the number of sessions attended and the Literacy Proficiency, even after accounting for a broad set of student, school, and cohort characteristics. The linear specification (Column 1) shows that an additional session is associated with a 0.518 percentage point increase in Composite Literacy Proficiency ($p < 0.001$). This relationship is both statistically and economically significant, reflecting the substantial benefits of regular engagement with DytectiveU exercises. The positive coefficient implies that, for each additional session completed, students experience meaningful improvements in their reading and writing skills, which include better reading speed, comprehension, and accuracy.

In the non-linear specification (Column 2), the coefficient for the number of sessions increases to 1.413 ($p < 0.001$), and the squared term is negative and significant at -0.0046 ($p < 0.001$), indicating diminishing returns to session attendance. This non-linear relationship suggests that while early increases in session attendance led to significant gains, the marginal benefits taper off as students complete more sessions. The optimal number of sessions, calculated as the turning point where additional sessions no longer yield proportional gains, is estimated at 154 sessions. This threshold indicates the maximum level of beneficial engagement, beyond which further participation may lead to saturation or cognitive overload. Additionally, the coefficients for control variables such as gender and birth date reinforce previous findings in the literature. Female students show a higher average progress rate in Literacy Proficiency compared to their male counterparts (1.105 in the linear model and 0.976 in the non-linear model), while students born later in the school year tend to have lower Compositive Literacy Proficiency scores, likely reflecting the maturity and developmental differences often observed in younger students.

The IV estimates might indicate an upper bound of DytectiveU's impact, capturing the Local Average Treatment Effect (LATE) on student performance. Unlike the OLS estimates, which assume a homogeneous treatment effect across all students, the IV estimates pertain specifically to a subgroup more likely to benefit from the program. This subgroup includes students who began using DytectiveU as soon as it became available at their school, potentially reflecting those with a higher need. Utilising these IV estimates to forecast the effects of varying attendance also demands careful consideration of treatment effect heterogeneity across students, as the Average Causal Response is identified only for a subset of compliers, not the entire sample. These findings underscore DytectiveU's significant role in enhancing Literacy Proficiency.

We analyse the heterogeneity in literacy gains by segmenting students based on their session attendance into three groups: low (25th percentile, 3 sessions), moderate (median, 9 sessions), and high (75th percentile, 22 sessions), as well as the recommended 64 sessions. According to the non-linear model in Column 2 of Table 7, students in the low attendance group (3 sessions) improve by 4.20 points, with a marginal effect of 1.39 points per session. Students in the moderate attendance group (9 sessions) experience a total improvement of 12.34 points, with a marginal impact of 1.33 points per session. The total improvement for the high-attendance group (22 sessions) reaches 28.86 points, with a marginal effect of 1.22 points per session. The total predicted improvement at the recommended 64 sessions is 71.59 points, with a marginal impact of 0.83 points per session. Although diminishing returns are observed as the number of sessions increases, the impact remains significant even at higher session counts, such as the 64-session threshold. These results reveal a pattern of diminishing marginal returns, where early sessions yield the most significant impact on literacy outcomes, while later sessions primarily reinforce previously acquired skills. This finding highlights the critical importance of early participation while illustrating that sustained engagement continues to offer meaningful benefits.

Table 7: 2 SLS Regression

	Composite Literacy Proficiency (Linear)	**Composite Literacy Proficiency (Non-linear)**
	1	**2**
Number of Sessions	0.518***	1.413***
	(0.0088)	(0.0836)
Squared Sessions	-	-0.0046***
	-	(0.0004)
Female	1.105***	0.976***
	(0.1428)	(0.3337)
Birth Day	-0.0114***	-0.0129***
	(0.0007)	(0.0016)
Grade	4.340***	3.272***
	(0.548)	(1.2281)
Constant	12.013***	6.339
	(2.589)	(6.0800)
Observations	44,997	44,997

Notes:
Column (1) reports the second-stage results from the linear specification of equation (1), where the output is the Composite Literacy Proficiency Score, measured at the last challenge completed in DytectiveU. Column (2) presents results from the non-linear specification, which includes a squared term for the number of sessions. Both models control for school fixed effects, grade dummies, year dummies, grade-by-year dummies, school-by-year dummies, and school-by-grade dummies (controls omitted from table). Equation (1) estimates the effect of the number of sessions on literacy skills, isolating the program's impact by adjusting for observable covariates and addressing potential selection biases through instrumental variable (IV). Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Robustness Checks

To corroborate the results obtained in Table 7, we conducted additional robustness checks, outlined in Table A5. The table explores the relationship between session attendance and Proficiency Literacy, considering two distinct samples: excluding students with only one session and introducing student fixed effects for students participating more than one year. These analyses reinforce the robustness and generalizability of our main findings. Excluding students with only one session addresses concerns related to selection bias and ensures that the analysis focuses on students who actively participated in the DytectiveU program. As discussed in Table 3, significant differences exist between students who completed just one session and those who completed multiple sessions regarding reading and writing progress. The coefficient for the number of sessions is 1.2532 ($p < 0.001$), higher than in the main specification, suggesting that regular engagement with the program is critical for achieving significant literacy gains. The coefficient for the quadratic term remains negative at -0.00396, reinforcing the presence of diminishing returns.

The inclusion of a dummy variable for each student in the panel data analysis controls for individual-specific characteristics such as motivation, concentration capacity, allowing for a precise estimate of within-student progress over time. This approach ensures that the estimated effects reflect the evolution of the same student's Literacy Proficiency, rather than differences between students and provides further evidence that student unobserved characteristics did not drive OLS results. Column 2 includes all students participating for more than one year, yielding a coefficient of 1.0000 ($p < 0.001$). The quadratic term remains negative at -0.0035, with a threshold effect of 142.83 sessions. This threshold is lower than the one observed in the main results, suggesting that the benefits of regular attendance may be more affected by diminishing returns to scale among students who have participated in Dytectiveu for several school years. Overall, these robustness checks validate the main results and provide additional insights into DytectiveU's effectiveness across different engagement levels and student populations. The findings suggest that sustained participation is key to maximizing the program's benefits.

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

The results of this study indicate that the DytectiveU program significantly enhances reading skills among primary school students. By leveraging adaptive and personalised features, AI-assisted technologies have proven effective in addressing educational gaps that have traditionally been challenging to bridge through conventional methods. Each additional session contributes positively to Literacy Proficiency, though an optimal threshold was identified, beyond which the benefits begin to diminish. These findings align with previous literature emphasising the importance of tailoring educational interventions to the specific needs of each student (Sevilla et al., 2023). Similar results were observed in studies on math-assisted programs using digital tools, which highlight the need for balanced time allocation to optimise the effectiveness of these technologies. One critical factor identified in both studies is the need for a systemic approach involving both students and educators in integrating technology.

From a policy perspective, these results suggest that education authorities should consider incorporating AI-assisted learning programs into public school systems in a planned and strategic manner. Collaboration between software developers, researchers, and educators is essential to maximize the impact of such programs, while also preventing educational overload and disengagement that can occur when digital tools are not adequately aligned with students' capabilities and motivations. Continuous evaluation and flexible implementation strategies should be integral to any initiative aimed at digital transformation in education.
Ultimately, this study reinforces the potential of AI-assisted programs to complement conventional education, as long as they are implemented with a clear understanding of the constraints and needs of the educational context. The findings suggest that AI-based educational interventions can enhance learning outcomes by providing individualised support and personalised learning paths, helping students achieve greater academic success.

References

Abbey, C., Green, E., Mo, D., Lai, F., Bai, Y., Zhang, L., Bianchi, N., Ma, Y., Feng, Y., Clark, T., Fafchamps, M., & Yang, S. (2024). The effectiveness of EdTech on student learning outcomes in China: A systematic review and meta-analysis. *Computers and Education Open, 6*, 100161.
Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics*. Princeton University Press.
Amuedo-Dorantes, C., & Arenas-Arroyo, E. (2019). Immigration enforcement and children's living arrangements. *Journal of Policy Analysis and Management, 38*(1), 11–40. https://doi.org/10.1002/pam.22106
Amuedo-Dorantes, C., & Wang, C. (2024). Intermarriage amid immigration status uncertainty: Evidence from DACA. *Journal of Policy Analysis and Management, 1*(1), 1–30. https://doi.org/10.1002/pam.22640
Aucejo, E. M., Fruehwirth, J. C., Kelly, S., & Mozenter, Z. (2022). Teachers and the gender gap in reading achievement. *Journal of Human Capital, 16*(3), 372–403. https://doi.org/10.1086/719731
Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics, 122*(3), 1235–1264.

Bedard, K., & Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics, 121*(4), 1437–1472. https://doi.org/10.1093/qje/121.4.1437
Beg, S. A., Fitzpatrick, A. E., & Lucas, A. (2023). Managing to learn (Working Paper No. 31757). *National Bureau of Economic Research*. https://doi.org/10.3386/w31757
Beg, S. A., Lucas, A., Halim, W., & Saif, U. (2019). Engaging teachers with technology increased achievement, bypassing teachers did not (Technical Report No. w25704). *National Bureau of Economic Research*.
Bettinger, E., Fairlie, R., Kapuza, A., Kardanova, E., Loyalka, P., & Zakharov, A. (2023). Diminishing marginal returns to Computer-Assisted Learning. *Journal of Policy Analysis and Management, 42*(3), 552–570. https://doi.org/10.1002/pam.22442
Bird, K. A., Castleman, B. L., & Song, Y. (2024). Are algorithms biased in education? Exploring racial bias in predicting community college student success. *Journal of Policy Analysis and Management*. https://doi.org/10.1002/pam.22569
Büchel, K., Jakob, F., Bühnhan, S., Steffen, T., & Brunetti, A. (2022). The relative effectiveness of teachers and learning software: Evidence from a field experiment in El Salvador. *Journal of Labor Economics, 40*(3), 737–777.
Collinson, R., Humphries, J. E., Mader, N., Reed, D., Tannenbaum, D., & van Dijk, W. (2024). Eviction and poverty in American cities. *The Quarterly Journal of Economics, 139*(1), 57–120. https://doi.org/10.1093/qje/qjad042
Duflo, E., Dupas, P., Spelke, E., & Walsh, M. P. (2024). Intergenerational impacts of secondary education: Experimental evidence from Ghana (Working Paper No. 32742). *National Bureau of Economic Research*. https://doi.org/10.3386/w32742
Ferman, B., Lima, L., & Riva, F. (2021). Artificial intelligence, teacher tasks and individualized pedagogy. *SocArXiv*. https://doi.org/10.31219/osf.io/qw249
Heckman, J.J. (2006). Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 312(5782), 1900–1902. DOI: 10.1126/science.1128898
Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics, 24*(3), 411–482. https://doi.org/10.1086/504455
Hirata, G. (2022). Play to learn: The impact of technology on students' math performance. *Journal of Human Capital, 16*(3), 437–459.
Jakubowski, M., Gajderowicz, T., & Patrinos, H. A. (2025). *COVID-19, school closures, and student learning outcomes. New global evidence from PISA*. *npj Science of Learning*, 10 (1), 5. https://doi.org/10.1038/s41539-024-00185-z
Muralidharan, K., Singh, A., & Ganimian, A. J. (2019). Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review, 109*(4), 1426-1460.
OCDE. (2023). *PISA 2022 Results (Volume I). The state of learning and equity in education, PISA*. *OECD Publishing*. https://doi.org/10.1787/53f23881-en
OECD. (2024). *Do Adults Have the Skills They Need to Thrive in a Changing World?* Paris: OECD Publishing. https://doi.org/10.1787/b263dc5d-en
Rello, L., Baeza-Yates, R., Ali, A., Bigham, J. P., & Serra, M. (2020). Predicting risk of dyslexia with an online gamified test. *PLOS ONE, 15*(12), e0241687. https://doi.org/10.1371/journal.pone.0241687
Riehl, E., & Welch, M. (2023). Accountability, test prep incentives, and the design of math and English exams. *Journal of Policy Analysis and Management, 42*(1), 60–96. https://doi.org/10.1002/pam.22417
Rodriguez-Segura, D. (2022). EdTech in developing countries: A review of the evidence. *World Bank Research Observer, 37*(2), 171–203.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg* (pp. 80–108). Cambridge University Press.

Table A.1: School Participation in the DytectiveU Program by Year

School year (1)	Previous year (2)=(3)+(4)	Exiting (3)	Continuing (4)	Entering (5)	Total in the current year 6=(4) + (5)
2018-2019	-	-	-	98	98
2019-2020	98	9	89	3	92
2020-2021	92	11	81	93	174
2021-2022	174	26	148	74	222
2022-2023	222	71	151	4	155

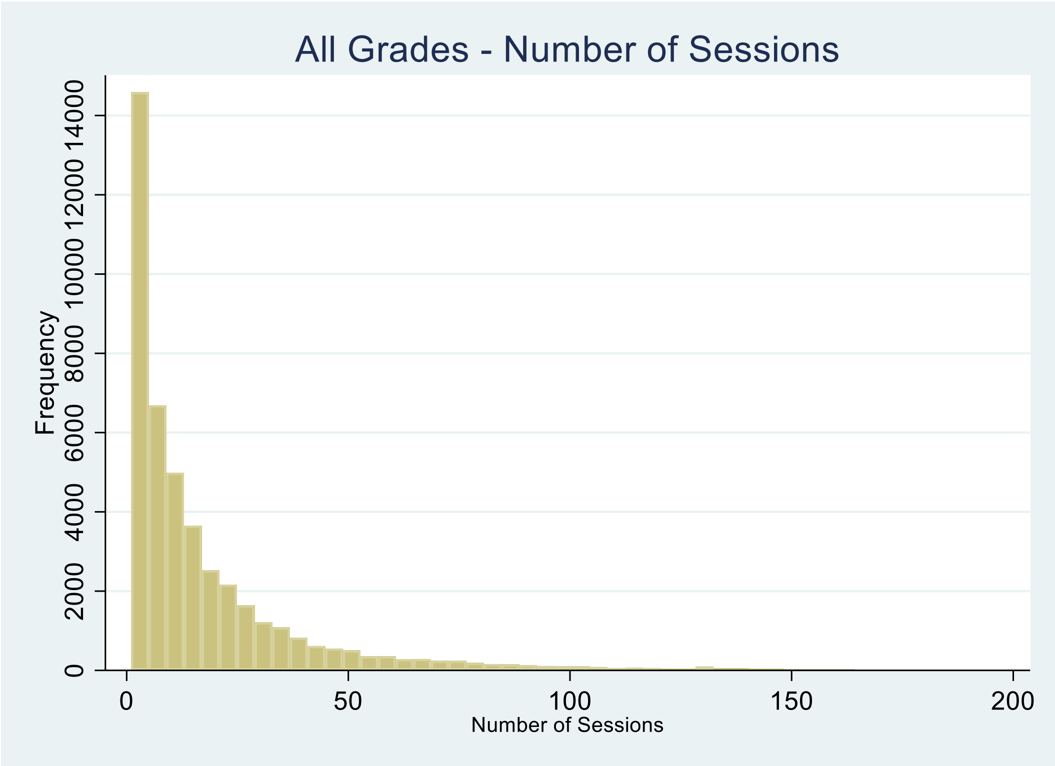
Notes. Table A.1 shows annual participation of public primary schools in the *Help Dyslexia* program using the DytectiveU software from the 2018–2019 to the 2022–2023 school year. Column (1) lists the total number of schools from the previous academic year. Column (2) decomposes this number into those that exited the program (Column 3) and those that continued (Column 4). Column (5) shows the number of new schools joining the program in that academic year, and Column (6) gives the total number of schools participating in the current year.

Table A.2: Licensing Dates for DytectiveU Program by School Type and Year

	Starting date of the license	Latest day of use
	2018-2019 school year	
Entering schools	14th January 2019	30th June 2019
	2019-2020 school year	
All the schools	1st September 2019	30th June 2020
Exiting schools	1st September 2019	17th June 2020
Continuing schools	1st September 2019	30th June 2020
Entering schools	1st September 2019	30th June 2020
	2020-2021 school year	
All the schools	1st September 2020	30th June 2021
Exiting schools	3rd September 2020	30th June 2021
Continuing schools	1st September 2020	30th June 2021
Entering schools	17th October 2020	30th June 2021
	2021-2022 school year	
All the schools	1st September 2021	30th June 2022
Exiting schools	1st September 2021	30th June 2022
Continuing schools	1st September 2021	30th June 2022
Entering schools	2nd September 2021	30th June 2022
	2022-2023 school year	
All the schools	1st September 2022	22nd May 2023
Exiting schools	-	-
Continuing schools	1st September 2022	22nd May 2023
Entering schools	18th September 2022	10th April 2023

Notes. Table A.2 displays the licensing dates for DytectiveU software across five academic years (2018–2019 to 2022–2023), disaggregated by school type: entering, continuing, or exiting. For each category and year, the table shows the starting date of the license, defined as the earliest date of software use by any student in a school, and the latest day of use, corresponding to the last day any student used the software. These dates were extracted from administrative usage logs as processed in the DytectiveU program database. The earliest and latest usage dates were obtained by calculating the minimum and maximum session dates for students within each school-year and participation group.

Figure A.1. The distribution of the number of sessions -challenges



Notes. Figure A.1 displays the distribution of the number of sessions completed by students using the DytectiveU program across all primary school grades (1 to 6). The histogram is truncated at 200 sessions to improve visibility, given the presence of extreme right-tail values. **Sample and data.** The underlying data were constructed by merging administrative records from the five school years between 2018–2019 and 2022–2023. The final dataset includes 44,997 participations (student-year observations), corresponding to 34,607 unique students in grades 1 through 6. The number of sessions per student was computed using the index variable renamed as sessions, based on log data from the DytectiveU platform.

Table A.3: Summary statistics by School year. Students participating only one School year

	Mean (1)	s.d. (2)	Min. (3)	Max. (4)	N (5)
			2018-19		
Composite Literacy Proficiency	29.37	19.80	1.48	96.67	5,662
Number of Sessions	8.49	10.98	1	129	5,662
Female	0.502	0.50	0	1	5,662
Birth Day	167.21	109.72	1	366	5,662
First Session	April 28, 2019	33.06 days	January 16, 2019	June 28, 2019	5,662
Last Session	June 1, 2019	36.71 days	January 23, 2019	June 30, 2019	5,662
			2019-20		
Composite Literacy Proficiency	31.29	19.26	3.70	98.15	1,366
Number of Sessions	12.21	18.10	1	353	1,366
Female	0.485	0.50	0	1	1,366
Birth Day	161.44	111.57	1	366	1,366
First Session	January 12, 2020	62.51 days	September 1, 2019	June 26, 2020	1,366
Last Session	March 14, 2020	65.72 days	September 2, 2019	June 27, 2020	1,366
			2020-21		
Composite Literacy Proficiency	25.38	18.34	3.33	96.67	5,641
Number of Sessions	7.84	11.72	1	146	5,641
Female	0.475	0.50	0	1	5,641
Birth Day	171.54	109.99	1	366	5,641
First Session	May 17, 2021	36.62 days	September 15, 2020	June 30, 2021	5,641

	Mean (1)	s.d. (2)	Min. (3)	Max. (4)	N (5)
			2018-19		
Last Session	June 7, 2021	35.65 days	September 27, 2020	June 30, 2021	5,641
			2021-22		
Composite Literacy Proficiency	29.02	18.79	3.33	100	8,973
Number of Sessions	10.22	15.61	1	214	8,973
Female	0.486	0.50	0	1	8,973
Birth Day	178.79	108.00	1	366	8,973
First Session	February 26, 2022	53.73 days	September 1, 2021	June 29, 2022	8,973
Last Session	May 16, 2022	57.48 days	September 2, 2021	June 29, 2022	8,973
			2022-23		
Composite Literacy Proficiency	20.99	14.53	3.33	99.63	1,748
Number of Sessions	6.48	12.05	1	156	1,748
Female	0.477	0.50	0	1	1,748
Birth Day	187.91	106.11	1	366	1,748
First Session	November 19, 2022	36.68 days	September 1, 2022	April 22, 2023	1,748
Last Session	January 25, 2023	35.83 days	September 2, 2022	May 28, 2023	1,748

Notes. Table A.3 reports summary statistics by school year for students who participated in DytectiveU during only one school year. The table includes mean, standard deviation, and range values for key variables such as Composite Literacy Proficiency, number of sessions completed, gender, birth date (expressed as day of the year), and the dates of first and last session. Sample and data. The statistics in Table A.3 are based on cross-sectional data extracted from DytectiveU platform usage records over five school years (2018–2019 to 2022–2023). The sample includes only those students who appear in the data for a single school yearAs in Table 1, the number of sessions is calculated from platform log entries, Composite Literacy Proficiency is expressed as a percentage, and birth day is encoded as a day-of-year value. Calendar dates for the first and last sessions are derived from timestamp fields, with standard deviations reflecting the variation in engagement timing across the sample.

Table A.4: Descriptive Statistics for the Instrument Variable First Session Date (Days from September 1st)

Mean (1)	s.d. (2)	Min. (3)	Max. (4)	Within Variance	Between Variance	N
				2018-19		
207.16	33.64	135.54	302.92	0.00	33.64	9,970
				2019-20		
93.09	63.81	0.01	299.53	0.00	63.81	6,037
				2020-21		
210.50	59.81	0.40	301.86	0.00	59.81	11,401
				2021-22		
157.14	67.34	0.38	302.71	0.00	67.34	15,378
				2022-23		
68.29	36.82	0.04	221.71	0.00	36.82	4,821

Notes: This table presents summary statistics for the instrument used in the first-stage regression: the number of days between September 1st and the date on which each student completed their first DytectiveU session. Statistics are reported by academic year for the five school years between 2018–2019 and 2022–2023. The sample includes 44,997 student-year observations from 264 public primary schools in the Madrid region. The instrument exhibits zero within-school variation in all years, as all students within a school begin DytectiveU on the same day. By contrast, the between-school variance is substantial and increases in years with broader rollout. Standard deviations are reported in column (2); the between- and within-school variance components are calculated from decomposed variance of the first session variable.

Table A5. Robustness checks.

	Excluding Students with One Session (1)	Student fixed effects (students participating more than one year) (2)
Number of Sessions	1.2532***	1.0000***
	(0.0801)	(0.2640)
Number of Sessions²	-0.00396***	-0.0035***
	(0.00035)	(0.00097)
Female	1.0919***	-
	(0.3167)	
Birth Day of Year	-0.0137***	-
	(0.00156)	
Grade	3.9196***	0.7878
	(1.1516)	(3.2727)
Individual fixed effects	NO	YES
Observations	40,167	21,607
Number of Students (Groups)		9,047
Underidentification Test	101.531 (p-value = 0.0000)	12.220 (p-value = 0.0005)
Weak Identification Test	48.653	5.631

Notes:
Column (1) reports the second-stage IV estimates from the sample excluding students who completed only one session in DytectiveU, in order to focus on students with sustained engagement and minimize selection biases. Column (2) reports results from a student fixed-effects specification, using panel data from students who participated for more than one year, controlling for unobservable student-specific characteristics such as motivation, concentration capacity, and access to digital devices. The dependent variable is the Composite Literacy Proficiency score at the last completed challenge in DytectiveU. Both columns adjust for school fixed effects, grade dummies, year dummies, grade-by-year dummies, school-by-year dummies, and school-by-grade dummies (controls omitted from the table). Standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1.