RENAISSANCE

EFFICACY STUDY | APRIL 2021

Efficacy Study of the Lalilo Teaching Assistant on Reading Decoding Skills in CP (French 1st grade) During the Lockdown Period in France

Thomas Sergent | Data Scientist at Lalilo & Educational Data Mining PhD Candidate at Sorbonne University | Paris, France

Abstract

From January to June 2020, we conducted a randomized controlled efficacy study of the Lalilo software on learning to read. The results suggest a positive impact of using Lalilo in the classroom and at home when schools closed. These results are encouraging for Lalilo and suggest that use in the context of pedagogical continuity has a positive impact on literacy skills. Here we present our methodology and discuss the limitations of this study.

Keywords: *Blended Learning - Literacy – Efficacy Study- Lalilo – Remote Learning*

Introduction

Since the creation of the Conseil Scientifique de l'Education Nationale (CSEN) [French Scientific Council of National Education] on January 10, 2018, research has been at the heart of educational decisions taken by the Ministry of Education and Youth in France. During the lockdown, the will to base recommendations for pedagogical tools on methodologically robust experiments has been reaffirmed. "Only controlled experimentation makes it possible to verify that an educational tool works. We would have liked to present here only resources that have demonstrated their effectiveness in randomized controlled trials. This was not possible - there is clearly a great deal of research to be done in this area (Vallejo-Gomez, 2020)."

At the end of November 2019, as a winner of the Partenariat d'innovation et intelligence artificielle [Partnership for Innovation and Artificial Intelligence] (P2IA), we launched a call for participation in an efficacy study in our partner academias [FR districts] - Paris, Créteil, Versailles, Montpellier, Orléans-Tours and Nancy-Metz.

Methodology

The initial objective was to evaluate the efficacy of the Lalilo software for decoding progression in a randomized controlled trial that would take place from January to June 2020. After the initial tests, classes in the test group would have worked an average of 10 to 15 minutes per day on Lalilo - during their class time, while classes in the control group would not have had access to Lalilo.

Table 1. Gaant Chart (including expected and actual timeline of intervention period)

	January	February	March	April	May	June
Initial Tests						
Group Announcements						
Register Teacher on Lalilo						
Intervention Period (estimated)						
Intervention Period- In class (actual)						
Intervention Period – at home (actual)						
Final Tests						

Because of the COVID-19 pandemic, most of the time students spent on Lalilo was spent at home. In this study, we therefore propose to measure the efficiency of Lalilo software in promoting decoding progress at home compared to students who would not have had access to it.

Tests

The tests chosen had to meet several criteria:

- be recognized by the scientific community and, if possible, have already been used in other efficiency studies (e.g., GraphoGame or Kalulu; Ruiz et al., 2017; Watkins et al., 2020);
- be short enough to be passed on to teachers (they were the ones who administered the tests);
- target decoding skills, which are at the heart of Lalilo's scope and sequence.

For the CP (French 1st grade) class we have thus chosen the familiar and invented word tests of Sprenger-Charolles, Theurel and Gentaz (2017): the "BEGINNING-END CP" test in January and the "END CP" test in June. The child must read a maximum of words in one minute and then a maximum of pseudo-words in one minute.

There are 35 words and 30 pseudowords in the initial test and 60 words and 60 pseudowords in the final test. The words and pseudowords are all mono- or bisyllabic. All words are taken from the Manulex database (Lété, Springer-Charolles, & Colé, 2020).

For the CE1 we have chosen the Lefavrais Alouette test (Lefavrais, 1965). It is an asemantic text of 265 words with an increased presence of infrequent words, forcing the child to use the grapho-phonological decoding procedure rather than his lexical knowledge. This test is very sensitive to reading difficulties (Bertrand et al., 2010).

All tests were done on paper.

Groups

We sent an initial questionnaire to teachers to collect:

- the number of students of each level in their class;
- the classification of the school (non REP/REP/REP+) [French equivalent of Title I schools];
- the number of computers and tablets connected to the internet in their class. A minimum ratio of one available computer equipment for every four students was requested so that teachers could use Lalilo for about one hour per week without disrupting their teaching practices.

We matched the classes according to these criteria and then drew one class from each pair to be in the test group and the other to be in the control group. This way of forming the groups, called *Pair Matching in Cluster-Randomized Experiments*, allows the characteristics of each group to be balanced and allows for rigorous comparison of pre- and post-intervention results as shown by Imai et al. (2009).

Given the change in focus during the study, the criteria of the number of students in the class and the number of computers and tablets lost their relevance because the use of Lalilo at home does not depend on it.

Note that we did not randomly assign the groups at the student level - as was done in Kalulu's efficacy study - but at the teacher level (Watkins et al., 2020).

Table 3. Teachers flow chart

	Sent back the initial tests	Final tests
103 teachers	54 in the test group	28 sent back the final tests
		11 reported not being able to do the final tests
		15 did not answers to e-mails anymore
	49 in the control group	9 sent back the final tests
		9 had used Lalilo during lockdown and therefore did not send the final tests
		17 reported not being able to do the final tests for another reason
		14 did not answers to e-mails anymore

We were able to verify the registration of teachers in the test group on the platform as well as the non-registration of teachers in the control group. We had authorized the teachers of the control group to use Lalilo during the lockdown with their students. We removed them from the study. In addition, we had authorized the teachers not to return the final tests without asking for justification.

We can observe significantly higher teacher losses in the control group, even without taking into account the teachers in the control group who used Lalilo during the confinement. Our interpretation is that classes are less involved because they are in the control group and therefore feel less accountable for the experiment, in a context where the pandemic has monopolized a significant part of their time.

CP students (French 1st grade)

Inclusion criteria for students in the study were identified. Each student is eligible if:

- We had both initial and final tests. (word and pseudoword for each test).
- They did at least 300 exercises (about 300 minutes) on Lalilo if they were in the test group.

Our hypothesis is that a student who did less than 300 exercises during the entire lockdown did not use Lalilo significantly enough.

Table 4. Student flow chart

<table>
<tr><td rowspan="3">281 students returned the 2 tests and could be identified on the Lalilo platform if they were in the test group.</td><td>70 students of the testing group had done less than 300 exercises on Lalilo</td><td></td></tr>
<tr><td>122 students of the testing group had done more than 300 exercises on Lalilo</td><td rowspan="2">210 students were eligible</td></tr>
<tr><td>88 students of the control group</td></tr>
</table>

Table 5. Characteristics of the two groups after application of the eligibility criteria

	Intervention group	Control group
N	122	88
Age in months (std)	83.9 (3.5)	83.5 (4.5)
REP [French Title I]	26%	25%
number of exercises done in Lalilo (1 exercise lasts approx. 1 minute). Standard deviation in parenthesis	488 (158)	-

CE1 students (French 2nd grade)

For the CE1 students, we recovered 142 initial and final tests from CE1 students. Of these, 126 belonged to classes in the "testing" group and only 16 to classes in the "control" group.
Given the imbalance between the two groups, we preferred not to carry out a statistical study on the results of CE1 students.

CP results

By controlling on the **initial** number of familiar or invented words read by the students, ANCOVA gives the following results (statistical processing was done with the penguin Python package software; Vallat, 2018):

Table 6. Influence of different factors in the final score for decoded words

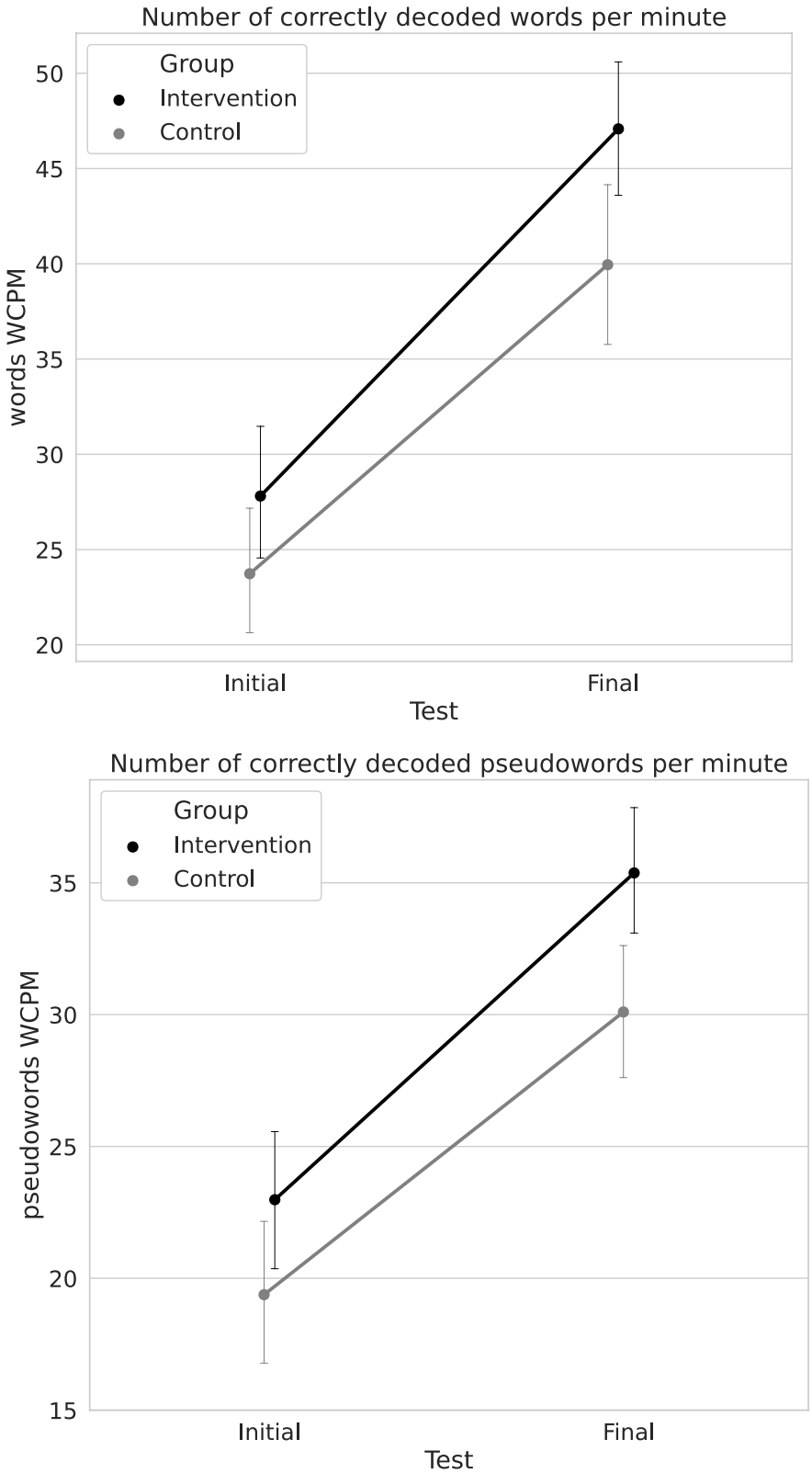
	SS	DF	F	p-unc	np2
Testing vs control group	600.236466	1	4.637025	3.245673e-02	0.022119
Initial number of decoded words	52746.490241	1	407.484021	1.289834e-50	0.665297
Age	147.877590	1	1.142403	2.864021e-01	0.005542
School rating (Title I or not)	31.428162	1	0.242793	6.227235e-01	0.001183
	26536.084725	205	NaN	NaN	NaN

Table 7. Influence of different factors in the final score for reading pseudowords

	SS	DF	F	p-unc	np2
Testing vs control group	372.039187	1	5.462221	2.039675e-02	0.025953
Initial number of decoded pseudowords	20946.346560	1	307.530952	1.157503e-42	0.600024
Age	36.298939	1	0.532935	4.662088e-01	0.002593
School rating (Title I or not)	61.448727	1	0.902181	3.433161e-01	0.004282
	13962.825567	205	NaN	NaN	NaN

For both familiar and invented words, the influence of initial reading level is significant ($p < 0.001$) as expected (blue dotted boxes). Similarly, the group by test interaction results are significant ($p < 0.05$) for both words and pseudowords (red boxes).

Figure 1. Initial and final test results for words (top) and pseudowords (bottom) depending on group (testing group in bold)



It can be observed that the groups are not balanced for the initial tests. ANCOVA allows to control this imbalance and to affirm the significant positive impact of Lalilo on the level in the final reading tests.

Discussion

The results of this randomized controlled study suggest a significantly greater increase in the number of familiar and invented words read in one minute for students in the group using Lalilo during confinement than in the control group not using it.

These results are encouraging for Lalilo and suggest that use within the framework of pedagogical continuity has a positive impact on learning to read. An effect is observed after an average of 488 exercises on Lalilo at home (8 hours at a rate of one minute per exercise) over the period March-June. It is likely that a regular and more intense workout in the classroom will have a greater impact.

We identify the following limitations to our study:

- the original purpose of the study was changed when lockdown was decreed;
- the conditions under which the students used Lalilo during the lockdown were probably different;
- the random assignment to the test and control groups was not made at the student level but at the classroom level, which prevents the teacher effect from being ruled out;
- teachers themselves (and not outsiders) administered the tests to the students;
- we removed students from the test group who had performed less than 300 exercises (i.e., less than 5 hours at a rate of one minute per exercise) on Lalilo during the confinement. It is possible that these were also the students who did not have digital equipment at home, but we did not remove the population of students without digital equipment at home from the control group;
- we allowed teachers to videotape the final tests to their students (i.e., not all students took the final tests under the same conditions);
- the test used was probably not relevant for students who read too poorly or compared to their peers at mid-CP, which removed many students from the analyses.

In order to confirm the trend observed and the positive impact of the use of Lalilo in the classroom as a complement to the teachers' pedagogical practices, we plan to carry out new randomized controlled experiments in the coming years in classes with larger class sizes and closer activity for the control group, so as not to measure an effect linked to the use of a technological tool (which can increase motivation) but to the pedagogical content proposed by Lalilo as such. For example, the control group could be made to work on mathematics software.

Acknowledgments

We would like to thank the teachers, the DANE of the partner academias, the IEN, ERUN and CPC and more generally all those who helped us design and conduct this study.

References

Bertrand, D., Fluss, J., Billard, C., & Ziegler, J. C. (2010). Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture. *L'Année psychologique*, *110*(2), 299–320.

Imai, K., King, G., & Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Statistical Science*, *24*(1), 29-53.

Lefavrais, J. (1965). *Test de l'Alouette* (révision 2005). Paris: ECPA.

Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers, 36*(1), 156–166.

Ruiz, J. P., Lassault, J., Sprenger-Charolles, L., Richardson, U., Lyytinen, H., & Ziegler, J. C. (2017). GraphoGame: un outil numérique pour enfants en difficultés d'apprentissage de la lecture. *ANAE-Approche Neuropsychologique des Apprentissages Chez L'enfant*.

Sprenger-Charolles, L., Theurel, A. & Gentaz, E. (2017). Evaluer les capacités de lecture et les capacités reliées chez des enfants de 6-7 ans: Cadre théorique et bilan des études de Gentaz et al. (2013 et 2015). Document édité par la FAPSE-UNIGE.

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), 1026, https://doi.org/10.21105/joss.01026

Vallejo-Gomez, N. (2020). *Recommandations de ressources pédagogiques*. Conseil scientifique De l'éducation nationale. https://www.reseau-canope.fr/conseil-scientifique-de-leducation-nationale-site-officiel/outils-pedagogiques/recommandations-de-ressources-pedagogiques.html/

Watkins, C. P., Caporal, J., Merville, C., Kouider, S., & Dehaene, S. (2020). Accelerating reading acquisition and boosting comprehension with a cognitive science-based tablet training. *Journal of Computers in Education, 7*(2), 183–212.

©Copyright 2021 Renaissance Learning, Inc. All rights reserved. | (800) 338-4204 | www.renaissance.com

All logos, designs, and brand names for Renaissance's products and services are trademarks of Renaissance Learning, Inc., and its subsidiaries, registered, common law, or pending registration in the United States.