



HAL
open science

Can technology facilitate reading acquisition? Developing applications for phonics learning and dyslexia screening

Cassandra Potier Par

► **To cite this version:**

Cassandra Potier Par. Can technology facilitate reading acquisition? Developing applications for phonics learning and dyslexia screening. Neuroscience. Université de Paris, 2021. English. NNT : . tel-03702094

HAL Id: tel-03702094

<https://hal.science/tel-03702094v1>

Submitted on 22 Jun 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

École doctorale N°474 FIRE

*Laboratoire INSERM-CEA Cognitive Neuroimaging unit
CEA/SAC/JOLIOT/Neurospin*

Can technology facilitate reading acquisition?

*Developing applications for phonics learning
and dyslexia screening*

Par Cassandra Potier Watkins

Thèse de doctorat de SCIENCES COGNITIVES, PSYCHOLOGIE,
LINGUISTIQUE, PHILOSOPHIE DE LA PENSÉE

Dirigée par Stanislas Dehaene

Présentée et soutenue publiquement le 26 Janvier Devant un jury composé
de :

Johannes Ziegler, DR, Université d'Aix Marseille, rapporteur
Maryanne Wolf, DR, University of California Los Angeles, rapporteur
Liliane Sprenger-Charolles, PR émérite, Université d'Aix Marseille,
examinatrice
Caroline Huron, HDR, Université de Paris, examinatrice



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by-nd/3.0/fr/>

Abstract

This thesis focuses on the application, to French students, of advances in the understanding of how children learn to read, what methods best train literacy and how we can better assess reading deficits— so that these advances can fuel a virtuous circle between cognitive science and educational interventions.

In the case of literacy, there is a global consensus that early explicit phonics instruction is the best means to learning how to read. In this thesis, we propose a tablet-based game to support phonics learning. We also address several unanswered questions of teaching methodology. For example, during what period of time should the teaching of the phonetic method be introduced? Can children learn to read using phonics software before formal literacy? We present the results from our randomized control study with French first graders (N=975). Data collected from pre- and post- tests shows that children only benefited from the phonics game when it was used early in the school year. In a second intervention, we improved the design of the application and research methodology, and then tested the game with French kindergarteners (N=1092). This time students improved in factors that support literacy, irrelevant of the period that the games were used. However, in a follow-up test, benefits disappeared once formal reading education began. In the process of seeking to develop a tablet-based approach to literacy, we also describe a program used by the game engine that automatically creates a phonics progression based on the most frequent and consistent grapheme-phoneme correspondences for any alphabetic language.

Historically, some of the greatest insights into the processes involved in reading have come from the examination of the deficits caused by brain lesions. Based on the hypothesis that reading is a complex cognitive activity, and that a deficit in any part of the process could lead to dyslexia, we developed a screener that looks at the types of errors made by readers to assess for selective deficits. After normalizing the screener with French 6th and 7th graders, we examine the different types of errors made by students previously diagnosed with dyslexia by their school. We present the first French cases of two different types of selective deficits: letter position dyslexia and attention dyslexia.

The body of this work was developed with the goal of practical application. To this end, the tablet-based game and phonics builder code were developed using open-source software and licenses. Our screener for dyslexia is also open for use in collaboration with our lab. In the

general discussion, we discuss ideas for immediate improvements and future studies that could be done using these tools, to the benefit of improved literacy programs for all students.

Résumé

Comprendre comment le cerveau lit est l'une des collaborations réussies entre la recherche scientifique et l'éducation. D'une part, l'imagerie fonctionnelle du cerveau avant et après l'apprentissage de la lecture (Dehaene-Lambertz et al., 2018), et pour les populations analphabètes (Dehaene et al., 2015) a établi les zones corticales dédiées à la lecture et la connexion créée entre les zones visuelles et langagières qui se développe au cours du processus d'apprentissage (Dehaene et al., 2010). D'autre part, la recherche observationnelle et interventionnelle, les procédures de test nationales et les méta-analyses comparatives interlinguistiques ont mis en évidence la manière dont les enfants apprennent à lire et les méthodes qui donnent les meilleurs résultats. Le principal résultat de cette collaboration a conduit à un soutien massif du programme de phonétique comme meilleur moyen initial d'alphabétisation (Groupe de travail du Groupe de travail du Csen, 2019; National Reading Panel, 2000; Rose, 2006), pour stimuler l'automatisation de la sensibilité à l'écrit (Brem et al., 2010) et l'accès au sens dans des aires corticales de langage préalablement établis. Cette thèse s'intéresse à la mise en application, aux élèves français, des avancées dans la compréhension de l'apprentissage de la lecture chez l'enfant. Nous proposons un soutien à l'enseignement de la méthode phonétique, au moyen d'un jeu sur tablette pour l'apprentissage de la lecture, et nous abordons plusieurs questions de méthodologie d'enseignement restées sans réponse. Nous présentons les résultats de deux études contrôlées randomisées auprès d'élèves français, une méthode utilisant l'intelligence artificielle pour développer une progression phonétique basée sur les correspondances graphèmes-phonèmes et un nouveau dispositif de dépistage de la dyslexie.

Dans le **chapitre 2**, nous décrivons ELAN, un jeu adaptatif qui favorise l'acquisition de l'alphabétisation par l'enseignement et la formation phonétique, similaire à celui de GraphoGame (Richardson & Lyytinen, 2014), qui, à l'époque, n'avait pas été développé pour le français. Il fournit un enseignement explicite et systématique par correspondance graphème-phonème et renforce le décodage complet par la pratique de la lecture et de l'orthographe avec un texte 100% décodable, une pratique essentielle pour favoriser l'apprentissage de la lecture (Cheatham & Allor, 2012). Le jeu utilise également des tâches périodiques de décision lexicale pour mesurer la transition entre le décodage lettre par lettre et la reconnaissance fluide des mots (Martens & de Jong, 2006; Zoccolotti et al., 2005). Le

logiciel a été testé dans le cadre d'un essai contrôlé randomisé dans 44 classes de CP (975 enfants français).

Les enfants qui ont utilisé le logiciel ELAN au cours du premier trimestre se sont améliorés par rapport à deux groupes de contrôle : groupe de contrôle actif qui a joué un jeu de maths (www.attrape-nombres.com) ; groupe de contrôle passif des classes sans tablettes "business-as-usual". Les améliorations ont été significatives dans la fluidité de la lecture (lecture de mots et pseudo-mot d'une minute) et la compréhension de la lecture des phrases. Les résultats étaient cohérents avec le « Simple View of Reading », l'idée qu'un décodage amélioré peut aider l'enfant à se concentrer sur la compréhension (Hoover & Gough, 1990). Ces résultats soulignent l'importance d'une formation phonétique précoce, explicite et systématique, et fournissent un nouvel outil logiciel pour la faciliter.

La clé d'un enseignement phonétique efficace est l'enseignement des correspondances graphème-phonème selon une progression explicite et systématique qui commence par les associations les plus fréquentes et les plus cohérentes (de Graaff et al., 2009; Ehri et al., 2001). Cependant, la découverte des règles pertinentes n'est pas une tâche facile et nécessite généralement une analyse subjective par un locuteur natif et/ou un linguiste expert. Dans le **chapitre 3** nous décrivons GPA4.0, un sous-module du modèle de réseau de neurones « Transformer » qui automatise la tâche de transcription et d'alignement des graphèmes en phonèmes (g2p) (Vaswani et al., 2017). Le réseau a été formé avec quatre langues différentes de transparence orthographique décroissante (Espagnol<Portugais<Français<Anglais). Nos résultats montrent que le modèle Transformer améliore l'état actuel de la transcription g2p et que son mécanisme d'attention permet l'alignement des graphèmes à leurs phonèmes correspondants (Toshniwal & Livescu, 2016). À partir des mots alignés en g2p, notre logiciel fournit une progression phonétique ordonnée de façon optimale, basée sur la fréquence et la cohérence des association graphèmes-phonèmes dans la langue cible, ainsi qu'une liste ordonnée de mots que les enseignants peuvent utiliser. Ce travail illustre de manière pratique la façon dont les réseaux de neurones peuvent être utilisés pour développer du matériel pédagogique pour la recherche et les enseignants. Les sous-modules et les résultats de la phonétique sont disponibles à l'adresse suivante : <https://github.com/OlivierDehaene/GPA4.0>.

Dans l'intervention présentée au chapitre 2, notre logiciel pour l'enseignement de la phonétique a été testé avec des élèves de CP. La fluidité et la compréhension de la lecture ont

été améliorées, mais seulement lorsqu'elles étaient utilisées en même temps que l'enseignement de la lecture au début de l'année. L'intervention a également été contrecarrée par plusieurs « bugs » dans le logiciel causant des pertes de données et un contrôle actif qui n'était pas bien assorti. Nous avons demandé si une intervention similaire serait plus efficace si elle commençait un an plus tôt, en Grande Section. Pour cette étude, l'environnement du jeu a été remanié pour inclure un contrôle mathématique apparié. Les suites de jeux de phonétique et de nombres ont été conçues pour développer la compréhension des enfants entre les symboles (graphèmes et nombres) et la signification (phonèmes et quantité).

Quarante classes en France ont été réparties de manière aléatoire pour jouer « Kalulu Phonétique » ou un jeu de contrôle actif « Kalulu Nombres » pendant le premier semestre de l'année. Ces assignations ont été inversées au cours du second semestre. Dix classes non randomisées et sans tablette ont également participé à l'étude (1092 enfants en France). Dans un effet croisé, les enfants qui ont utilisé Kalulu phonétique se sont améliorés dans la dénomination des lettres, la correspondance graphème-phonème et la fluidité de la lecture, tandis que ceux qui ont utilisé Kalulu Nombres ont amélioré leur connaissance des nombres. Ces résultats ont montré que la compréhension des symboles et de leur signification par les enfants s'est améliorée, tant en ce qui concerne la compréhension des graphèmes-phonèmes (lorsque les enfants jouaient avec les jeux de lecture) que les associations entre les nombres et les quantités (lorsqu'ils jouaient avec les jeux de mathématiques). Pour les enfants de grande section, ces résultats ont été obtenus indépendamment du fait que le logiciel ait été utilisé au cours du premier ou du second semestre de l'année. En d'autres termes, nous avons atteint nos deux premiers objectifs, à savoir enseigner aux enfants indépendamment de la session de formation croisée et fournir un contrôle actif comparatif.

Dans le cadre d'un suivi longitudinal, les participants à l'intervention ont été suivis en CP, 4 mois après avoir participé à l'intervention. Il est bien documenté que les connaissances de la plupart des étudiants se dégradent après les vacances d'été, un phénomène connu sous le nom de « summer lag » (décalage estival, en français) (Cooper et al., 1996). Au début de l'année scolaire CP, les élèves qui ont utilisé le logiciel ont continué à surpasser le groupe de contrôle pour deux indicateurs importants de la lecture à l'école : la connaissance des lettres et la conscience des phonèmes (Piquard-Kipffer & Sprenger-Charolles, 2013). Comme ces tests ont été réalisés sur papier, avec des enseignants qui n'étaient pas au courant de l'intervention, nous pouvons affirmer que les jeux ont réussi à aider les élèves à répondre non seulement aux

tests sur tablettes mais aussi aux tests sur papier. Ces premiers résultats positifs en matière de compréhension phonétique n'ont cependant pas aidé les enfants à apprendre spontanément à lire à haute voix plus de mots que le contrôle passif, et le fait de jouer au jeu n'a pas apporté de bénéfices durables une fois l'alphabétisation formelle commencée. Aucun avantage longitudinal n'a été constaté pour les tests de mathématiques. Notre conclusion était que les jeux de tablettes n'étaient pas suffisants pour enseigner la lecture en l'absence d'une véritable pratique de la lecture à voix haute. Ces résultats nous permettent de mieux comprendre quand et pendant combien de temps il faut introduire la méthode phonétique et remettent en question la possibilité qu'une intervention à court terme puisse répondre aux défis complexes des objectifs éducatifs à long terme.

La lecture est un processus complexe qui implique de multiples parcours et étapes de traitement. Ainsi, des types distincts de déficits de lecture peuvent survenir à la suite de déficiences dans l'un ou l'autre de ces processus. Au **chapitre 5**, nous décrivons le test de dépistage Malabi, qui a été développé pour identifier les déficits dans les étapes orthographiques, lexicales et sous-lexicales de la lecture en français, et nous rapportons comment il nous a permis d'identifier une double dissociation entre des types distincts de dyslexie développementale. Inspiré de la littérature neuropsychologique sur les alexies acquises, le Malabi repose sur une analyse minutieuse des types d'erreurs dans la lecture à voix haute. Il comprend plusieurs listes comportant de nombreux mots « pièges », c'est-à-dire des stimuli destinés à susciter un certain type d'erreur et à identifier ainsi une fragilité dans les processus cognitifs correspondants. Les dyslexiques « attentional », par exemple, peuvent mal interpréter la paire de mots "bise vase" comme "vise base", tandis que les dyslexiques « letter position » peuvent mal interpréter "vélo" comme "volé". Nous avons recueilli des normes pour les erreurs de lecture chez les élèves français (élèves de 6^e et 5^e année). Travaillant avec une école spécialisée dans l'aide aux élèves dyslexiques, nous avons identifié trois cas de dyslexie développementale « attentional », compatible avec une déficience dans la liaison des lettres aux mots entre des mots voisins, et un cas de dyslexie « letter position », compatible avec une déficience dans l'encodage de la position des lettres à l'intérieur des mots. Lors de tests post-hoc, nous avons reproduit ces diagnostics et exploré les facteurs qui modulent la migration des lettres entre et dans les mots. Nos résultats concordent avec des preuves antérieures provenant d'autres langues pour de multiples types de dyslexie développementale (Vaswani et al., 2017) tout en fournissant un soutien aux modèles

d'analyse orthographique qui prennent en compte les différentes étapes de l'identification des lettres et de l'encodage de la position des lettres chez les lecteurs normaux (Davis & Bowers, 2004, 2006).

Le **chapitre 6** se termine par une discussion générale des résultats de cette étude, en soulignant plusieurs points clés :

- La combinaison des résultats des deux interventions sur les logiciels de lecture (ELAN et Kalulu) améliore notre connaissance de ce qui fonctionne dans l'éducation phonétique. Il semble que les jeux de phonétique sur tablettes peuvent stimuler l'alphabétisation et aider les enfants à entrer dans la lecture grâce à une meilleure connaissance des graphèmes-phonèmes, mais seulement lorsqu'ils sont utilisés au début de l'éducation à la lecture et, probablement, accompagnés d'un programme plus large qui comprend également des exercices de lecture à voix haute. Cette conclusion provisoire devrait être testée plus avant dans de futures études de contrôle randomisées.
- Une autre contribution de ce projet est venue de nos recherches sur les différents types de dyslexie développementale. Nous avons développé un test qui contient des mots spécifiques pour tester les déficits à différents stades des analyses orthographiques et des stades de lecture (voie lecture sous-lexicale et lexicale) selon le modèle de lecture à double voie (Coltheart, 2005; Friedmann & Coltheart, 2018). Nos résultats ont révélé les premiers cas, en français, de la dyslexie attentionnelle (déficit démontré par la migration des lettres entre les mots, la lecture de "bise vase" comme "vise base") et de dyslexie de position des lettres (déficit démontré par la transposition des lettres lues avec des mots, la lecture de "vélo" comme "volé"). Ce travail s'appuie sur les cas de sous-types de dyslexie déjà rencontrés dans d'autres langues (Friedmann et al., 2010; Friedmann & Gvion, 2001; Güven & Friedmann, 2019; Kohnen et al., 2012). Nos résultats sont encourageants non seulement parce qu'ils nous aident à comprendre les nombreux processus cognitifs impliqués dans la lecture, mais aussi parce qu'ils nous aideront à améliorer la remédiation de la dyslexie. La limite actuelle de ce travail est que nous ne signalons que quatre cas de déficit sélectif. Pour étayer nos conclusions concernant les différents facteurs qui modulent les différents déficits, il sera important de trouver beaucoup plus de cas.

Enfin, l'objectif de cette thèse a été de développer le soutien aux enseignants par la création et la recherche d'outils d'apprentissage applicables. Pour plus d'informations en français concernant le logiciel Kalulu, veuillez visiter le site web du projet pour voir les résultats et télécharger le guide de l'enseignant et le jeu Kalulu : www.ludoeducation.fr.

Références

- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, *107*(17), 7939–7944. <https://doi.org/10.1073/pnas.0904402107>
- Cheatham, J., & Allor, J. (2012). The influence of decodability in early reading text on reading achievement: A review of the evidence. *Reading & Writing*, *25*(9), 2223–2246. ehh.
- Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch1>
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research*, *66*(3), 227–268. <https://doi.org/10.3102/00346543066003227>
- Davis, C. J., & Bowers, J. S. (2004). What Do Letter Migration Errors Reveal About Letter Position Coding in Visual Word Recognition? *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 923–941. <https://doi.org/10.1037/0096-1523.30.5.923>
- Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding: Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 535–557. <https://doi.org/10.1037/0096-1523.32.3.535>
- de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. *Scientific Studies of Reading*, *13*(4), 318–333. <https://doi.org/10.1080/10888430903001308>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, *16*(4), 234–244. <https://doi.org/10.1038/nrn3924>
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, *330*(6009), 1359–1364. <https://doi.org/10.1126/science.1194140>
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, *16*(3), e2004103. <https://doi.org/10.1371/journal.pbio.2004103>
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-

- Analysis. *Review of Educational Research*, 71(3), 393–447. <https://doi.org/10.3102/00346543071003393>
- Friedmann, N., & Coltheart, M. (2018). 35 Types of developmental dyslexia. In E. Dattner & D. Ravid (Eds.), *Handbook of Communication Disorders* (pp. 721–752). De Gruyter. <https://doi.org/10.1515/9781614514909-036>
- Friedmann, N., & Gvion, A. (2001). Letter Position Dyslexia. *Cognitive Neuropsychology*, 18(8), 673–696. <https://doi.org/10.1080/02643290143000051>
- Friedmann, N., Kerbel, N., & Shvimer, L. (2010). Developmental attentional dyslexia. *Cortex*, 46(10), 1216–1237. <https://doi.org/10.1016/j.cortex.2010.06.012>
- Groupe de travail du Csen. (2019). *ÉvalAide, un dispositif scientifique de prévention des difficultés en lecture et en mathématiques en CP et en CE1*. Canopé. https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conference_role_experimentation_domaine_educatif/EvalAide_CSEN_Definitif_Mai2019.pdf
- Güven, S., & Friedmann, N. (2019). Developmental Letter Position Dyslexia in Turkish, a Morphologically Rich and Orthographically Transparent Language. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02401>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127–160. <https://doi.org/10.1007/BF00401799>
- Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When ‘slime’ becomes ‘smile’: Developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681–3692. <https://doi.org/10.1016/j.neuropsychologia.2012.07.016>
- Martens, V. E. G., & de Jong, P. F. (2006). The effect of word length on lexical decision in dyslexic and normal reading children. *Brain and Language*, 98(2), 140–149. <https://doi.org/10.1016/j.bandl.2006.04.003>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769)*. National Institute of Child Health & Development. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- Piquard-Kipffer, A., & Sprenger-Charolles, L. (2013). Early predictors of future reading skills: A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (age 5 to 8). *L'Année Psychologique*, 113(04), 491–521.
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method: The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. *Human Technology*, 10(1). <https://doi.org/10.17011/ht/urn.201405281859>
- Rose, J. (2006). *Independent review of the teaching of early reading: Final report*. DfES. <http://dera.ioe.ac.uk/5551/3/5d970d28fc535dc54eb4dee9995bef36.pdf>
- Toshniwal, S., & Livescu, K. (2016). Jointly Learning to Align and Convert Graphemes to Phonemes with Neural Attention Models. *ArXiv:1610.06540 [Cs]*. <http://arxiv.org/abs/1610.06540>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *ArXiv:1706.03762 [Cs]*. <http://arxiv.org/abs/1706.03762>

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, 93(3), 369–373. <https://doi.org/10.1016/j.bandl.2004.10.010>

Acknowledgements

This body of work has been made possible thanks to the amazing scientists working at the UNICOG-Neurospin lab. I am forever grateful to have had the chance of doing my thesis under the lightly guiding touch of Stanislas Dehaene. He has always made me feel like a peer with whom discussion and criticism are for the sake of science and the benefit of us all. I also thank Ghislaine Dehaene and the baby lab team for adopting me into their fold and growing the group to include children. For their participation in the two large-scale projects ELAN and LUDO, a very special thanks goes to Bernadette Martins for welcoming teachers and Isabelle Denghien for her IT expertise. Finally, a special thanks to Vanna Santoro, our lab manager extraordinaire who is always there to guide us through paperwork and provide us with chocolates and coffee.

This work has been checked, encouraged and critiqued by my supportive thesis advisory committee including Elena Pasquinelli and Elizabeth Spelke. Their insights and encouragement have been helpful in improving the protocol of my research and lifting my confidence to try again when the results have seemed far from encouraging. It is also the fruit of conversations from so many amazing researchers and education specialists that have shared their experiences with me: Lilliane Sprenger-Charolles, Johannes Ziegler, Katerina Lukasova, Bruce McCandliss, Juan Valle Lisboa, Caroline Huron, Veronique Izard, Frank Ramus, Simone Potier, Francois Taddei and Julien Caporal.

I thank Dominique Quéré, Eric Béguin and Cédric Couvrat for overseeing these projects in their school districts. This work is only possible because of the confidence and partnerships they helped me build between school districts, teachers and students. For the work on dyslexia, I thank Hervé Glasel for his tireless efforts to look for new ways to help the students at his school achieve their educational goals, Naama Friedmann for teaching me to look for dyslexia in the details and Noéline Gay for accompanying me to test kids and understand the many difficulties faced by students with dys problems. A very special thanks to the four students spotted for a particular deficit who continued to meet with me on numerous occasions to read more and more tests!

And last, but not least, there is my family. This change in life's work that started seven years ago when I decided to go back to school has been a success thanks to the support of my husband Julien and children Juliette and Maximilien. We will never forget these precious years

of creative chaos in our home! I thank my parents, George Watkins and Jill Edwards who supported my high school experimentations and my godmother, teacher and mentor Marilyn Mosley. You instilled in me the belief that education may be a common good, but it is also a deeply personal struggle for the flourishing of one's mind.

Table of Contents

ABSTRACT	1
RESUME	4
ACKNOWLEDGEMENTS.....	12
TABLE OF CONTENTS	14
RELATED SCIENTIFIC OUTPUT	19
CHAPTER 1 — GENERAL INTRODUCTION	20
PREAMBLE	21
READING IN THE BRAIN	23
TWO ROUTES FOR READING	25
WHERE YOU WERE BORN MATTERS	27
BEST PRACTICES FOR LEARNING HOW TO READ	28
SMALL CHANGES FOR BIG GAINS	29
<i>Avoid crowding</i>	29
<i>Pick-up the pace</i>	30
<i>Get a head start</i>	30
<i>Encode in order to decode</i>	31
GAMING THE SYSTEM.....	31
THE IMPORTANCE OF QUALITY SOFTWARE	35
LET THEM PLAY VIDEO GAMES	37
THE FOUR PILLARS OF LEARNING.....	40
THE CURRENT SITUATION IN FRANCE	42
THE RESEARCH PERFORMED IN THIS PHD THESIS.....	43
REFERENCES.....	46
CHAPTER 2, ACCELERATING READING ACQUISITION AND BOOSTING COMPREHENSION WITH A COGNITIVE SCIENCE-BASED TABLET TRAINING	55
ABSTRACT.....	56
INTRODUCTION	57
THE READING BRAIN	57
THE LANGUAGE CAVEAT.....	58
PRINCIPLES FOR PHONICS INSTRUCTION	58
TABLET-BASED INTERVENTIONS FOR READING INSTRUCTION	59
THE ELAN SOFTWARE.....	59
AUTOMATIZING GRAPHEME—PHONEME ASSOCIATION.....	61

ADAPTING TO THE CHILD’S LEVEL	63
EXPERIMENTAL STUDY	64
MATERIALS AND METHODS	65
<i>Participants</i>	65
<i>Procedure</i>	65
<i>Pre- and post-tests</i>	68
<i>Participant attrition</i>	70
STATISTICAL ANALYSIS	71
RESULTS.....	72
<i>Syllable suppression</i>	73
<i>Phoneme suppression</i>	73
<i>Letter knowledge</i>	73
<i>One-minute word and pseudo-word reading</i>	73
<i>Non-sense text reading</i>	74
<i>Sentence reading comprehension</i>	75
<i>Measuring the evolution of the word length effect</i>	76
<i>Number comparison</i>	79
<i>Subtraction</i>	80
DISCUSSION	80
<i>Summary of results</i>	80
<i>Limitations</i>	82
CONCLUSION.....	83
APPENDIX.....	84
REFERENCES.....	85
CHAPTER 3, AUTOMATIC CONSTRUCTION OF A PHONICS CURRICULUM FOR READING EDUCATION USING THE TRANSFORMER NEURAL NETWORK.....	92
ABSTRACT.....	93
INTRODUCTION	94
EXPERIMENT	95
RESULTS.....	96
<i>G2p Transcription</i>	96
<i>G2p alignment</i>	97
CONCLUSION.....	97
REFERENCES.....	99
CHAPTER 4, A COGNITIVE SCIENCE BASED PHONICS APPLICATION IN KINDERGARTEN: BOOSTING PRE-READING SKILLS, BUT NOT SCHOOL READING GAINS	102
ABSTRACT.....	104

INTRODUCTION	105
<i>Numeric responses to literacy curriculum</i>	106
OUR SOFTWARE: KALULU PHONICS	109
<i>A matched active control: Kalulu numbers</i>	112
<i>An encompassing experiment and control environment</i>	115
METHODS	117
<i>Participants</i>	117
<i>Design</i>	117
<i>Procedure</i>	118
<i>Test measures</i>	119
<i>Statistical analysis for the intervention</i>	122
<i>Intervention participant attrition and base-line characteristics</i>	122
RESULTS	124
<i>Intervention effects</i>	124
<i>Longitudinal follow-up</i>	128
DISCUSSION	131
<i>Future directions</i>	134
CONCLUSION.....	136
APPENDIX.....	137
REFERENCES.....	142
CHAPTER 5, CHARACTERIZING DIFFERENT TYPES OF DEVELOPMENTAL DYSLEXIA: THE MALABI SCREENER..	148
ABSTRACT.....	149
INTRODUCTION	150
DUAL ROUTE MODEL OF READING, AND DIFFERENT TYPES OF DYSLEXIA	151
<i>Letter identification dyslexia</i>	152
<i>Neglect dyslexia</i>	152
<i>Attentional dyslexia</i>	153
<i>Letter position dyslexia</i>	153
<i>Visual dyslexia</i>	154
<i>Surface dyslexia</i>	154
<i>Phonological dyslexia</i>	154
<i>Deep dyslexia</i>	155
SCREENING FOR VARIOUS TYPES OF DYSLEXIA	155
THE MALABI PROJECT	156
EXPERIMENTAL SCREENING	158
METHODS	158
<i>Dyslexic participants</i>	158

<i>Control group for establishing screener norms</i>	159
<i>Procedure</i>	159
<i>Establishing norms</i>	160
SCREENING FOR A SELECTIVE DEFICIT	161
<i>Malabi screener</i>	161
<i>Alouette</i>	161
<i>Oral language and phonological awareness</i>	161
RESULTS.....	162
<i>Did the Malabi identify readers with a selective deficit?</i>	162
<i>Patient data for selective dyslexia candidates</i>	162
WERE THESE STUDENTS CONSIDERED DYSLEXIC BY A TRADITIONAL SCREENER?	164
<i>Could their dyslexia be explained by a phonological deficit?</i>	164
EXPERIMENTAL INVESTIGATION OF ATTENTIONAL AND LETTER POSITION DYSLEXIA PROFILES.....	165
<i>Participants</i>	165
<i>Materials</i>	165
<i>Statistical Analysis</i>	167
RESULTS.....	168
<i>Double dissociation between letter transpositions within words and letter migrations between words</i> .	168
<i>Detailed analysis of migration errors</i>	169
<i>Detailed analysis of transposition errors</i>	174
DISCUSSION	177
<i>Future Directions</i>	178
CONCLUSION.....	180
REFERENCES.....	181
APPENDIX.....	192
CHAPTER 6, GENERAL DISCUSSION	193
OUR CONTRIBUTION	194
FUTURE AVENUES IN PHONICS INSTRUCTION	196
IMPROVING THE GAME ENVIRONMENT	197
CONTRIBUTIONS AND IMPROVEMENTS TO DYSLEXIA	198
HELPING THE CHILDREN THE MOST AT RISK FOR POOR LITERACY SKILLS.	199
AN ADAPTIVE SOLUTION FOR A GLOBAL CRISIS	199
DEVELOPING LOW-COST PAPER-BASED GAMES FOR LEARNING	203
CONCLUSION.....	206
REFERENCES.....	207
APPENDIX – GAMES IN THE PROJECT BIEN JOUÉ !	212

Related Scientific Output

This thesis comprises the following articles that have been accepted or are currently in preparation for publication:

Chapter 2

Potier Watkins, C., 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. <https://doi.org/10.1007/s40692-019-00152-6>

Chapter 3

Potier Watkins, C., Dehaene, O., & Dehaene, S. (2019). Automatic Construction of a Phonics Curriculum for Reading Education Using the Transformer Neural Network. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Éds.), *Artificial Intelligence in Education* (p. 226-231). Springer International Publishing. https://doi.org/10.1007/978-3-030-23207-8_42

Chapter 4

Potier Watkins, C. & Dehaene, S. A cognitive science-based phonics application in kindergarten: Boosting pre-reading skills, but not long-term school reading gains. In preparation

Chapter 5

Potier Watkins, C., Friedmann, N., & Dehaene, S. Characterizing different types of developmental dyslexia: the Malabi screener. In preparation

Chapter 1 — General Introduction

Preamble

Ask any child in their first year of primary school, “What are you learning at school?” and they will probably reply, “I am learning how to read!”. Literacy (along with math skills) is the first and foremost important educational objective in the primary years. It is a cognitive tool, that once acquired, supports all other school learning due to our text-based education system. Beyond school, literacy yields freedom to self-educate and make informed decisions. Children going into school anticipate the importance of learning to read. Since infancy, the little lines, curves and dots of letters decorate picture books, labels, signs and the pages of books, magazines, computers and phones that absorb the regard of their parents. Parents share the joy of reading with their children through storybooks. Classic fairytales and fables are passed from generation to generation to impart historical cultural norms, while modern picture books describe to the child the world around them and expected behavior. Story-time between parent and child is reported as a daily ritual in the majority of households with young kids, generally enjoyed before going to bed (Nichols, 2000). Parents share reading with their children not just as an enjoyable way to pass valuable information about life, emotions and expectations, but also because it is broadly believed that it provides a cognitive advantage important to development. Joint-reading, reading and discussing stories read with the child, have all been attributed with a wide range of learning outcomes, such as language growth, emergent literacy, and reading achievement in school (see, for a meta-analysis, Bus et al., 1995). Joint book reading has also been shown to improve communication skills in children with language delays (Dale et al., 1996; Lavelli et al., 2019).



Figure 1. Children from infancy are introduced to books as a way of learning about past and present. Compare the classic fairytale of Snow White (left-side book) and Mommy Lawyer (right-side book). The juxtaposition between these two books, common to children’s libraries, provides more than just entertainment. From a very young age, reading is an informative tool about our cultural heritage and current ideals.

For its ubiquitous role in childhood development, its importance to school and life success— researchers and educators alike have long been interested in the question of how we learn to read and what type of curriculum best supports the child’s literacy success. Reading has only been a part of human culture for approximately 5,400 years. Throughout much of history, knowing how to read was a job, or an endeavor of the privileged few. The benefits of knowing how to read (access to information, participation in government and commercial exchange), however, quickly made it a necessity to success in life. Today, global literacy is an international goal considered central to poverty relief and accelerating peace and development.¹ It is, in fact, one of the few goals that the international community has been successful in tackling. In the last 200 years, literacy has been lifted from 12% to 86% of the planet’s population (see Fig. 2). This is very encouraging news, but more importantly, a launch pad to encourage further improvements in reading education efforts. Literacy for the remaining 14% of the population remains a goal. Furthermore, access to reading education does not ensure its quality. Reading performance, even in the developed world , is often plagued by disappointing outcomes for low socio-economic populations (Mullis et al., 2017).

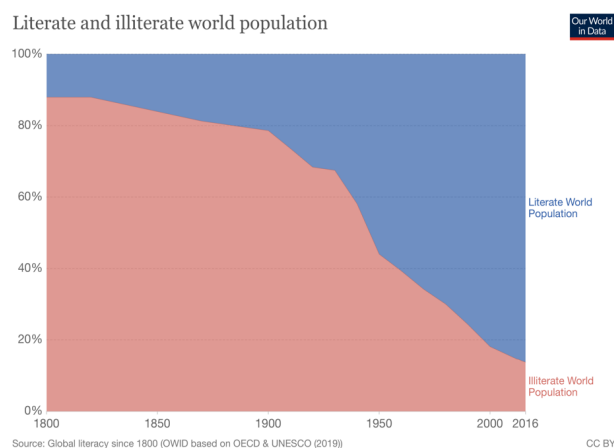


Figure 2. The last 200 years have seen a sharp rise in literacy (www.ourworldindata.org/literacy)

¹ <https://www.un.org/press/en/2002/sgsm8353.doc.htm>

Reading in the brain

The societal benefits of creating a code for communication across time and space are obvious. The universality of reading in such a short period of humanity's existence, however, begs the question: *how did our brains learn to read?* Reading is a complex activity requiring the cooperation between different cortical areas that allow us to hear and access meaning from symbols initially perceived by our visual system. As the Mexican poet Octavio Paz once said, "To read a poem is to hear with your eyes." From a biological perspective, reading is a heritable trait as made evident by genetic studies on reading environments shared by twins (Olson et al., 2014) and research on several genes that are directly responsible for cortical malformations that lead to dyslexia (Galaburda et al., 2006). However, learning how to read is dependent on growing up in an environment where reading is taught. The reading brain is not an evolutionary change as is evident in the fact that reading does not emerge, like language, spontaneously. Reading must be explicitly taught and is acquired with time and effort.

The dominating explanation as to how our brains learn to read has come from Stanislas Dehaene's theory of Neural Recycling (Dehaene, 2010; Dehaene & Cohen, 2007, see Fig. 3). The term neural recycling is used to describe how our brains assimilate cultural tools within a person's lifetime through the use of the brain's plasticity. From infancy, brain organization is highly structured by anatomical and connectional constraints that organize learning through sensory perception and action. Plasticity allows for these anatomical structures, initially evolved for other specific activities, to become reconnected and 'recycled' for novel use (Pascual-Leone et al., 2005). In other words, neural recycling is the ability of very old structures in the brain to learn new tricks.

The theory of neural recycling has been refined by cognitive and brain-imaging research with the goal of uncovering the brain circuits that we use for reading, and how they change as we learn to read. From infancy, long before going to school, the infant brain is already well organized for language learning. Left-lateralized brain regions including the superior temporal and angular gyri involved in adult language processing, are already active in infants (Dehaene-Lambertz et al., 2002). This early interest in language spurs children to become masters at understanding and communicating early in life. For example, young infants are able to discriminate almost all phonetic contrasts, but somewhere between the age of 6-

months (for vowels) and 1-year (for consonants) they become experts in hearing the phonemes of the language they are in born (Dehaene-Lambertz & Spelke, 2015). At the same time that this happens, they lose the ability to discriminate between phonemes that exist in other languages, but not their own mother-tongue (Cheour et al., 1998; Werker & Tees, 1984). Infants as early as 6- months already understand some words, but their comprehension increases substantially in performance in the short period up to 14-months old (Bergelson & Swingle, 2015). Furthermore, language learning is not only parent driven. Infants appear to actively motivate adults to participate in language directed speech from early infancy (Mimeau et al., 2020).

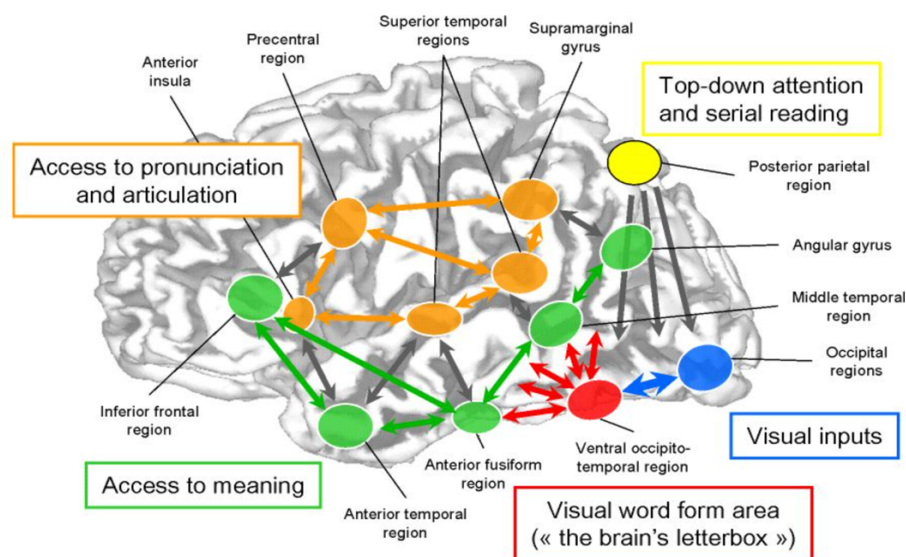


Figure 3. The neurological model of reading described by Stanislas Dehaene in "The Reading Brain". Cortex previously specialized in visual and language faculties interconnect and are 'recycled' for use in reading.

The cortical areas for spoken language processing already exist in infancy, but when we learn to read, we train our visual system to provide a novel input into those language circuits. Advances in functional MRI before and after learning to read have revealed that an area in the left visual cortex, once used to recognize objects and faces, becomes sensitive to letters, displacing the subsequent development of face recognition to the right visual cortex (Dehaene-Lambertz et al., 2018; Monzalvo & Dehaene-Lambertz, 2013). This region is called the Visual Word Form Area (VWFA) or 'the brain's letter box', and it exists in every reader no matter the language. During reading acquisition, this visual area specializing in word perception strengthens its connections to regions specialized in the processing of speech

sounds (Bouhali et al., 2014; Dehaene et al., 2015; López-Barroso et al., 2020; Thiebaut de Schotten et al., 2014). Furthermore, the representation of speech sounds itself changes. Literates, compared to illiterates, have improved phoneme awareness, the ability to dissect the smallest units of the speech stream (Morais et al., 1979, 1986). The plasticity required for this neural recycling is not only limited to children. Functional imaging of the brain also shows that people who learn to read as adults develop sensitivity to print in the visual word form area and improve their activation to spoken language in the left planum temporale (Dehaene et al., 2010), although perhaps not to the same extent as earlier learners. These results are a testimony to the role of plasticity in learning throughout life.

Two routes for reading

Education is critical to learning to read. Most children do not implicitly learn to read from listening to stories and seeing words around them, but must be taught the principles of alphabetic writing and the specific grapheme-phoneme correspondences involved in their language (Byrne, 1992). In the early stages, reading is slow as the learner decodes each grapheme to its phoneme, then listens in on to the combined sounds to access meaning. With time and practice, slow-decoding transforms into fast word recognition. These two types of reading also exist in the expert reader. Reading is fast and fluid until a new word is perceived. Suddenly, reading slows down to decode the sounds of the word. We can measure the transition from decoding to automatic recognition with the lexical decision task. The young learner's reading response time increases linearly with the number of letters in a word. As reading becomes automatized, word length no longer affects reading speed, and the line goes flat. However, pseudoword reading speed is always affected by word length (New et al., 2006; Zoccolotti et al., 2005). When learning to read, it is this slow decoding process that must be explicitly learned and trained. Automatized reading then develops slowly, as the result of practice and having a large oral vocabulary to give meaning to the decoded words. In other words, when children can sound out written words and possess enough spoken vocabulary to understand them, they can also learn a more direct route from the written word to the corresponding meaning – a process that has been termed “self-teaching” (Share, 1995).

The slow and fast routes for reading comprehension are best described by the Dual Route Model of reading (Coltheart, 2005, Fig. 4), which diagrams the various reading steps through

the lexical route (fast word recognition) and sub-lexical (slow decoding) routes, to the mental lexicon and meaning. This model has been used to not only understand reading, but also dyslexia. Dyslexia is defined as:

« Developmental learning disorder with impairment in reading is characterized by significant and persistent difficulties in learning academic skills related to reading, such as word reading accuracy, reading fluency, and reading comprehension. The individual's performance in reading is markedly below what would be expected for chronological age and level of intellectual functioning and results in significant impairment in the individual's academic or occupational functioning. Developmental learning disorder with impairment in reading is not due to a disorder of intellectual development, sensory impairment (vision or hearing), neurological disorder, lack of availability of education, lack of proficiency in the language of academic instruction, or psychosocial adversity (World Health Organization, 2020). »

According to the model, a deficit in one or the other routes to reading causes an over-dependency on the other. This is generally tested by having dyslexic readers read lists of irregular words and pseudo words. Children that cannot recognize the irregular words tend to depend on sub-lexical reading, which causes the over-regularization of letter sounds (e.g. reading 'bread' as 'breed'). This is called surface dyslexia. A deficit in the sub-lexical route curbs the ability to decode. In this case, children need a lengthy period of time to learn how to read, because they must memorize each word in their mental lexicon. This is referred to as phonological dyslexia. Although the dual-route theory was primarily developed on the basis of acquired reading disorders in adults (alexia), developmental cases for both of these deficits have been reported in the literature (Castles, 2006).

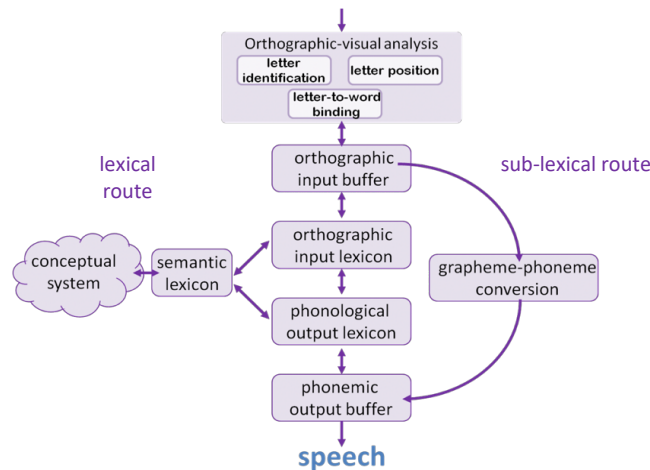


Figure 4, The Dual Route Model of lexical and sub-lexical reading (Friedmann & Coltheart, 2018).

Where you were born matters

The brain's circuits for reading are the same for all individuals, no matter the language. However, linguistic transparency, i.e. the degree to which a language's pool of grapheme-phoneme correspondences suffice to read all words, or suffer a considerable number of additional rules and exceptions is highly predictive of the time it takes for children to learn to read (Ziegler et al., 2010) even, importantly, when factors such as age and school environment are held constant (Spencer & Hanley, 2003). Transparency refers to the degree to which orthography deviates from a one-to-one relationship between letters and their sound. Some languages are highly transparent. For example, in Spanish, Italian and Finnish most letters only make one sound. English, on the other hand, is not considered very transparent. There are many examples of graphemes that make multiple sounds (e.g. the letter 'a' in 'cat', 'Cate', 'catastrophe' and 'caught'). In studies comparing learning to read rates in different European languages, children learning to read in English take almost twice as much time as children from transparent orthographies to obtain the same reading abilities, with French being slightly less extreme (Seymour et al., 2003). The language has many cases of letters that make more than one sound, but unlike in English, the changes are generally consistent and rule-based (e.g. the change in the letter 's', which usually makes the /s/ sound, to the /z/ sound when it is alone between two vowels, like in the words 'bise' or 'poison'). The transparency of a language's code influences the time it takes to learn, there is no evidence of a long-term detriment.

Furthermore, the best practices we will now report have been proven effective for all alphabetic languages.

Best practices for learning how to read

It is clear that reading exists as it does in the brain because it is constrained by the brain's natural organization. The goal of any reading method should therefore be to provide instruction that best supports this architecture. Observing school outcomes in the context of different pedagogies and testing interventions have been central to research on how children should be taught to read, which at its most contentious, erupted in the 1980s as a Reading War between two different opinions for reading instruction (Kim, 2008). On the one side, the 'whole-word' method emphasized learning words through multiple experiences with text (Goodman, 1967). A theory largely argued by the fact that children learn to recognize words they commonly see, such as word signs, words on their morning cereal box and logos. On the other side of the argument, other researchers argued that this was not reading, but a form of object recognition and that the traditional form of reading instructions, the phonics approach, provided the most efficient means to learning to read (Chall, 1967). Phonics instruction means explicitly providing the learner with the alphabetic code by successively teaching all of the grapheme-phoneme correspondences of the language. The gavel eventually did come down. In 1997, at the behest of the United States congress, the National Reading Panel was formed to comb through the literature of educational research and surmise what the data had to say about literacy instruction. In the year 2000, the panel published a work that has since become the foundation for all discussion about reading education (Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, DHHS., 2000). The report tackled: phonemic awareness instruction, phonics instruction, fluency instruction, vocabulary instruction, text comprehension instruction, independent reading, computer assisted instruction, and teacher professional development. The meta-analysis of the available experimental studies showed that phonics was the most successful learning method, particularly for children at risk for reading failure. Phoneme awareness was also shown to be critical to reading success, but only when coupled with learning the letters attached to speech sounds. This conclusion also put to rest a long-held belief that orthographies of varying transparency may benefit from different teaching practices. To the surprise of many English-

speaking educationalists, phonics instruction was also the better reading method for less transparent languages, including English. Furthermore, a sub-committee of the panel demonstrated the importance of explicit systematic phonics instruction (Ehri et al., 2001). Explicit systematic phonics means that the grapheme-phoneme correspondences must be taught taking into account the most frequent and consistent sounds early in teaching.

Small changes for big gains

Since the results from National Reading Panel report, the Reading War argument has been replaced by observational and in-the-field interventions to understand what other factors may influence reading. These factors include organizational and presentation practices that may have a small but significant impact on learning for normal students as well as more targeted interventions for readers at risk. The research in reading education is a crowded field. Without attempting to be exhaustive, we discuss several of the treatments that have a proven track record and are applicable in any educational circumstance.

Avoid crowding

One of the bottlenecks of our visual system is crowding. Crowding is the ambiguity created in recognizing a single item in the presentation of several similar items in a clutter. It affects visual perception and object recognition in a variety of viewing paradigms (i.e. orientation gratings, faces, complex objects, letters, etc., (for a review, Whitney & Levi, 2011). The hypothesis is that crowding occurs due to overlaps in neuronal receptive fields during the early visual perception of a scene (for a review, see "Where is the locus of crowding?" in Levi, 2008). Reading is a minefield for visual crowding, because words typically pack many letters in a close space. Crowding affects the peripheral vision of letter strings for both normal adult (Pelli et al., 2007) and children (Jeon et al., 2010) readers. Reading for dyslexic readers, in particular, is deteriorated by crowding effects (Martelli et al., 2009; Perea et al., 2012; Pernet et al., 2006; Spinelli et al., 2002). One solution has been to increase the space between letters in word presentation. This simple technique increases reading speed in normal adults and children (Perea et al., 2012). It also appears to improve reading in the case of dyslexia (Friedmann & Rahamim, 2014; Perea et al., 2012; Zorzi et al., 2012).

Pick-up the pace

The goal of any good reading program should be to help children read fluently, at an increasingly faster pace. However, educators generally teach grapheme-phoneme correspondences at the slow rate of one per week. Of course, the reasoning is that teachers want to help all students in the class master the new letter-sound relationship before advancing. However, this intuition may be wrong. Phonics instruction has increasingly become the main component of literacy. A legitimate question is how fast and how many should be taught. This was one of the questions reviewed in a large-scale observational study on the factors most likely to affect literacy outcomes (Goigoux, 2016). One of the highlights of the report was to show that French classrooms, on average, teach 11 grapheme-phoneme correspondences in the first 9 weeks of school (close to one per week). Increased ‘tempo’, the rate at which new grapheme-phoneme correspondences were introduced, was found to positively correlate with decoding and spelling performance. The tempo effect was also more recently supported by a study with 952 students in Norway (Sunde et al., 2020).

Get a head start

If providing many grapheme-phoneme correspondences early encourages literacy, the next question we might ask is *when should phonics education start?* Educators and researchers alike have asked whether or not early (before 1st grade) literacy training is beneficial and even feasible (see Ehri, Chapter 16, for a discussion on the arguments for kindergarten reading education, Suggate & Reese, 2012). Do children at that age have the vocabulary and cognitive skills to begin learning to read? In France, formal literacy instruction begins in 1st grade. In the United States, it was previously 1st grade, but then changed to kindergarten. A change prompted as a means to remediate "America's lack of reading proficiency" (Snow et al., 1998). The program has been successful in introducing reading to low-income kindergarten students. In a paper examining the effects of 27 studies on kindergarten literacy programs, children from low-SES backgrounds particularly benefitted from the jump start into reading (Cavanaugh et al., 2004). Importantly, the benefits were specific to programs that included explicit systematic phonics training, small groups and short, but intense, learning sessions. These results supported earlier findings from the National Reading Panel. Furthermore, it has been observed that younger 1st graders benefit more from phonics training than older 1st grade students, who are more likely to improve their reading skills by just reading books (Juel &

Minden-Cupp, 2000). We can conclude that these results merit more research into how kindergarten reading programs should be conducted.

Encode in order to decode

Teachers often hesitate to introduce spelling with reading. The worry is that the fine motor-skills needed for writing may overburden the learner. It turns out, however, that spelling can improve reading ability (Uhry & Shepherd, 1993). One argument could be that writing provides extra motor input to the identification of letters and words in the mental lexicon. For children that find it difficult to write, the solution may be to use pre-written letter cards or magnetic letters. It has been demonstrated that activities requiring children to spell many different words from a set of letter-cards improves full-decoding skills (see Fig. 5 for an example from McCandliss et al., 2003). At the letter level, handwriting has also been shown to improve letter name and sound retrieval (Bara & Gentaz, 2011; Longcamp et al., 2005) as does tracing (Hirsch & Niedermeyer, 1972). Haptic exercises, without writing, have also been shown to help children consolidate letter knowledge and phoneme awareness (Bara et al., 2004).

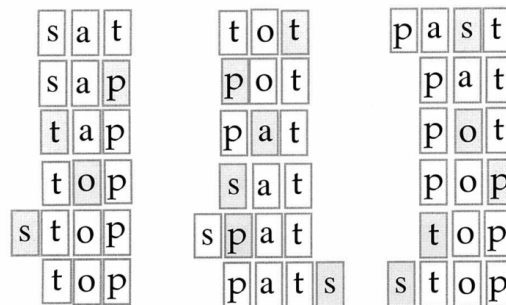


Figure 5, Illustration from the Word Builder intervention (McCandliss et al., 2003). Children use a fixed set of cards to build words focusing on the fact that the letter sounds stay the same no matter the card position.

Gaming the system

One of the topics discussed by the National Reading Panel was the use of technology in literacy education. The report was inconclusive, due the small number of quality papers addressing the subject. The adoption of technology as a tool for classroom education has its roots in cognitive psychology. Sidney Pressey (1888-1979), one of the original founding team members of the American Association of Applied Psychology (AAP), developed the Teaching Machine, an augmented version of multiple-choice questioning (MCQ) tests. The goal of the teaching

machine was to test and teach synchronously. MCQ tests were already popular in schools and with military entry exams when Pressey invented his machine. The machine had a window with a question and four answer buttons. The student pressed the button to the chosen answer, which was then recorded in the machine. Novel to Pressey's MCQ tests was that the machine did not advance to the next question until the correct response was given. By not allowing the student to advance, Pressey believed that this form of answering with immediate feedback on the correct/incorrect quality of the answer would help students to continually re-evaluate and update their knowledge. In other words, the machine would allow for students to learn from their mistakes. Pressey argued that this form of "trial and error" activity would lead to improved school success because the immediate feedback to the current knowledge-state would help the learner make better decisions, leading to a new knowledge state.

Pressey's hope was that the machine would eventually be able to go beyond teaching by deduction to teaching by analytical thought. To do this, he argued, the machine would have to provide the user with feedback to stoke reflection and an improved response. This goal of creating a conversation between machine and learner never came to pass, and eventually Pressey's work would take a back-seat to another famous researcher in behavioral theory. Burrhus Frederic Skinner (1904-1990) was a psychologist interested in the principle of reinforcement learning inspired from the field of behaviorism. Skinner made popular a teaching machine similar to Pressey's, but with a reward feature. Skinner believed that his learning offered several improvements to normal class learning (Skinner, 1968). He argued that it made student learning pleasurable and free of anxiety since immediate feedback was always given, success was eventually evident, and the student could move at their own pace. He also argued for teachers that the learning progression provided by the machines would be developed by skilled instructors and ensure homogenous learning outcomes thanks to the planned instruction. Skinner eventually patented his teaching machine with IBM (Frederic, 1958). Pressey and Skinner's machines never exploded onto the educational scene to replace traditional teacher led book learning, but their arguments continue to hold sway in the hope for a more pleasurable and adaptive learning experience -- an outcome that, despite all intentions and efforts, has remained elusive.

Today, technology has become far more inexpensive, reliable and easy to access. The teaching machine is informatically poor compared to the telephone most people carry in their pocket. In this new paradigm of accessible electronics, several large-scale efforts have been

initiated with the hope that computers and tablets would start an educational revolution of self-determined learners. In 2005, the One Laptop Per Child initiative, founded by MIT professor Nicholas Negroponte, introduced a \$100 portable computer with its own operating system and pre-loaded educational materials (Negroponte et al., 2016). The computer was distributed to children from poor communities around the world. Unfortunately, the gap between a great idea and school results was too far to be crossed. The expectancy that children's natural curiosity would spark self-directed learning turned out not to be true, making the inexpensive computer a very expensive squandering of resources for poor communities (Warschauer et al., 2011).

A similar case of a failed attempt to bring technology to students happened in 2013 in Los Angeles. The Los Angeles unified school district purchased 43,261 Apple iPads loaded with educational materials created by Pearson education. Each pre-loaded tablet cost the school district \$768. The equipment was delivered to classes throughout the city, but unfortunately rarely or never used. Teachers complained that the software loaded into the tablets was inadequate with the classroom curriculum and difficult to understand without training. Furthermore, teachers complained that they did not receive training on how to integrate the tablets into their classroom. Teachers did not provide instruction to students as to how to use the tablets— students, themselves, did not initiate self-directed learning. The tablets quickly became a nuisance as students would use them to access social media and videos, activities that they did not need training for. The Los Angeles school district eventually sued Apple to get a refund on their investment.

These stories mirror a more general acknowledgment by the OECD on technology in the classroom. Following the PISA exams, the OECD assessed scores based on classroom educational technology use (OECD, 2015). The disappointing results did not support past hopes for a technological revolution in the class ICT (information and communication technology) did not improve student outcomes, and above-average use was even detrimental to results (see Fig. 6).

What the data tell us

- Resources invested in ICT for education are not linked to improved student achievement in reading, mathematics or science.
- In countries where it is less common for students to use the Internet at school for schoolwork, students' performance in reading improved more rapidly than in countries where such use is more common, on average.
- Overall, the relationship between computer use at school and performance is graphically illustrated by a hill shape, which suggests that limited use of computers at school may be better than no use at all, but levels of computer use above the current OECD average are associated with significantly poorer results.

Figure 6. OCDE's critical assessment of the use of information and computer technologies in education. http://www.oecd-ilibrary.org/education/students-computers-andlearning_9789264239555-en

The OECD report has since become the sounding alarm of caution to the real benefits of introducing technology in the classroom. Technology may be cheaper than ever, but it remains a heavy investment for any school. Technology today is also quickly obsolete tomorrow. The three examples form a similar conclusion and an important lesson to future directives: the technology itself does not motivate students to become self-learners. Many students interviewed by the OECD felt confused by technology outside of social-media practices. The idea of “digital-natives” engaging freely in self teaching is now considered a “urban legend” and has not come to fruition (Kirschner & van Merriënboer, 2013). Instead, computers and tablets should be seen as tools, like books, that can be brought into the learning environment as supports. Their use should go hand-in-hand with teacher training, evidence-based software and randomized controlled trials to ensure real results (Istance & Kools, 2013). After all, it has also been the case that technology can improve school outcomes. For example, there is evidence of improved school readiness and academic achievement when children are exposed to computer based educational software at home, independent of the child's developmental stage and family socioeconomic status (Li et al., 2004). In these cases, it should be noted that parents conscientiously hand-pick quality software and actively mediate its use with the child. The same quality control provided by parents must be considered when introducing technology to the classroom. We should approach the recommendation of curriculum and materials with the same “do no harm” oath as medicine. This can be done by developing educational software based on cognitive science and evidence-based education practices, then testing the software for a specific outcome in randomized control trials. Once a positive outcome has been found beyond the business as

usual, then we can consider recommending its use, under the critical condition that when and how to use the software is clearly communicated to teachers.

The importance of quality software

Today, many cognitive psychologists are adopting exactly this methodology to develop educational software. In the realm of literacy, the most well-known success story for evidence-based software for reading instruction is the GraphoGame (see Fig. 7, www.graphogame.com). This game was originally developed to support phonological awareness in children at risk for dyslexia by providing intense phoneme awareness training (Lyytinen et al., 2009). Initial success spurred its development into a commercially distributed application for phonics training. The GraphoGame is a suite of mini-games requiring faster and faster responses to hearing and matching grapheme-phoneme correspondences, rhyming words and whole words (for a full explanation of the game theory and environment, see Richardson & Lyytinen, 2014). Since its origins, it has been tested and successful in helping students master phonics in many different languages (Ojanen, Ronimus, et al., 2015), including some of the poorest educational environments (Ojanen, Jere-Folotiya, et al., 2015; Serpell et al., 2017). Along with helping children learn how to read, the GraphoGame has itself become an important tool in research to furthering our understanding of the reading brain. Researchers were able to capture the development of the Visual Word Form Area through functional MRI by scanning non-reading children, then scanning them again after only two weeks of training with the GraphoGame, relative to a control game (Brem et al., 2010).

While the GraphoGame is probably the most well-known success story of software based on principles from cognitive science with a successful learning application, it is only one example. Several applications for math skills have also shown promising results. For example, kindergarten math skills have been shown to get a boost from the application Math Shelf (www.mathshelf.com), a game to help children master number and quantity relationships based on Montessori principles for the presentation of number magnitude (Schacter, 2016). Children from low-SES schools that used MathShelf in kindergarten made learning gains comparable to their kindergarten control counterparts from SES normal schools. Similar outcomes were found for the Number Race game (www.thenumberrace.com) (Räsänen et al., 2009; Wilson et al., 2006, 2009). The Number Race was designed to help children improve

their number sense through numerical comparison, quantity to numeral matching and arithmetic. After a short intervention, children tested in these tasks showed marked improvement over control subjects. In the number comparison task, the child decided which of two symbolic numbers or two quantities was bigger. This task is well documented as an early predictor of math success as number comparison taps into our sense of number magnitude, which is modulated by the size and distance of the numbers being compared (Lyons et al., 2014). Besides showing improved scores in several classical measures of number manipulation, children from low-SES homes also showed marked improvement in number comparison (Wilson et al., 2009). These results were also found with children diagnosed at risk of dyscalculia (Wilson et al., 2006).

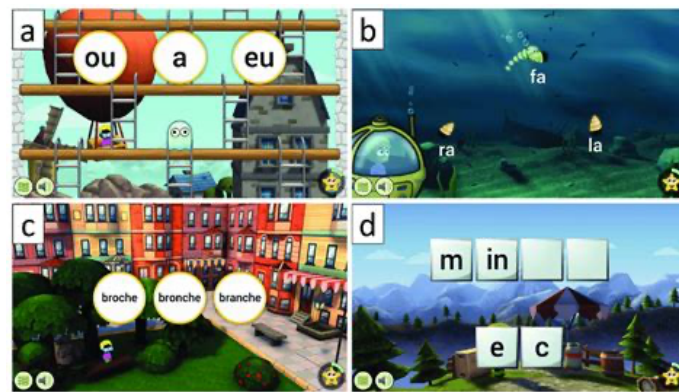


Figure 7, Images from the GraphoGame.

Key to all of these success stories is that, like any educational pedagogy, software should be based on principles from cognitive science on how the brain learns. As engaging as it might seem to include tablets and computers in the classroom, it is not enough to induce learning. For example, researchers compared learning how to read with an application that offered explicit phonics lessons with a systematic progression (a methodology with a proven track record, see Ehri et al., 2001), versus non-systematic phonics. The non-systematic phonics was a popular software distributed in stores. They found that successful reading outcomes were only provided by the systematic phonics software. In other words, only the application based on evidence-based practices was successful in improving reading outcomes (de Graaff et al., 2009).

The examples provided have all been cases where an evidence-based educational principle (i.e. phonics instruction) was adapted for application through the technological

medium. Another way that technology could possibly improve education is to apply evidence-based techniques from the gaming industry to motivate learning. One of the larger commercial successes that has taken this route has been Duolingo (www.duolingo.com). Duolingo is a game for learning to speak a new language that takes advantage of research in motivation to keep users engaged in frequent practice. It was originally created by researchers at Carnegie Mellon not just as a means for learning a second language, but as a covert operation to crowd source document translation. Studies on student use and learning outcomes have shown that the application can improve second-language test scores (Ahmed & Heba, 2016; Loewen et al., 2019). Notably, in all these studies, the application was used by individuals outside of a classroom experiment. In other words, the goal was to see if the game could encourage users in self-directed learning through techniques such as winning prizes, pop-up reminders, quick learning sessions and goal setting. The results show that, without the scaffolding of a classroom, there is still a considerable amount of 'drop-out'. Like we saw before, keeping the user engaged in self-motivated learning is difficult. Duolingo shows that the motivational functions help, but they are not full proof.

Let them play video games

The results from the Duolingo experience are weak, but they point to an interesting avenue of research for educational applications. The medium of video games can also provide a learning experience different from traditional pen and paper. For example, video games can provide motivational tactics, increased exercise item exposure, immediate feedback and individual adaptations such as slowing down or requiring faster responses to develop automatization. Simply translating book learning to an application with 'bells-and-whistles' is a poor use of the potential provided by technology. How our brains are changed by video games is a research field already showing promising outcomes. Deciphering how these outcomes can be adapted and transferred to school skills should be the goal in developing cognitive science-based software curriculum that takes a maximum advantage of computers, tablets and phones.

The video game industry represents 2 billion people around the world and growing. It is expected to produce 196 billion in revenues by 2022.² Today, video games are not only played, but they have also become massive sporting events where people pay to watch others play. For a long time, the video-game industry took a lot of heat for being the bugbear of despondent youth. Video games, especially violent games in the form of first-person-shooter story-lines, were blamed for distracting kids from their studies and promoting violence. Starting in the 1980s, researchers, often avid gamers themselves, decided to explore the possibility that video games were not harming our brains but making them smarter. They got their first win when it was shown that a group of university students that played video games had better eye-hand motor control than a control group of peers (Griffith et al., 1983). Later on in the decade, more and more authors converged in pointing out that medium- to long-term experience playing video games improves motor control (Drew & Waters, 1986; Perzov & Kozminsky, 1989) and visual spatial skills (Gagnon, 1985; Lowery & Knirk, 1982). The theory behind these benefits was that gamers are motivated to play and therefore to learn. Frequent playing has a cumulative effect, and as with everything in life, the cognitive resources trained through regular practice provided by the video games, coupled with intrinsic motivation to win, create an ideal medium for steady improvement.

Today, most of the cognitive benefits found from playing video games have been attributed to action games, mainly in the form of first-person shooter story-lines (for a review, see Eichenbaum et al., 2014). Playing these kinds of action video games requires actively searching for a target while avoiding distractions and dodging attacks. Success takes multiple rounds of trial-and-error. Regular players of action video games show improved visual attentional capacity, enhanced allocation of spatial attention over the visual field and task switching (Green & Bavelier, 2003). Due to the violent content of most action video games, the literature with children is sparse. However, there is evidence from self-reports that students (from elementary to high school) that play videos games, compared to non-gamers also have improved visual spatial attention (Dye & Bavelier, 2010), executive function skills (Homer et al., 2018) and visual rotation facility (Cherney, 2008). These positive results provide support for video games as cognitive enhancers, but perceptual training rarely transfers to

² <https://www.businessinsider.fr/us/video-game-industry-120-billion-future-innovation-2019-9?op=1>

school activities that require high-level thinking and production. Video games, however, may be the exception.

Halo is a first-person shooter game that takes place in a futuristic military world, requiring the player to defend himself from aliens (Microsoft., 2001). It has a huge fan-base and has been used in previous studies examining visual spatial skills. Researchers investigated the possibility that the visual spatial benefits imparted from the game could transfer to thinking about concepts in school that require visuospatial thinking (Sanchez, 2012). Non-gamer students were recruited from a university and randomized into a group to play either 'Halo' or a word game for 25 minutes. After the game period, both groups watched a movie about plate tectonics. Following the movie, the groups were asked to write an essay on the topic, "What caused Mt. St. Helens to erupt?" Students were evaluated on their ability to transfer important factors learned in the film to a new question. The logic of the study was that understanding geological phenomena requires the learner to "integrate and represent these discrete spatial units within their own internal, runnable, mental model that should behave in a consistent fashion with the actual material being learned." The results of the study showed that not only were visual spatial skills improved, but the 25-minute game session also improved scores in the learner's mental representation the geological phenomenon (see Fig. 8).

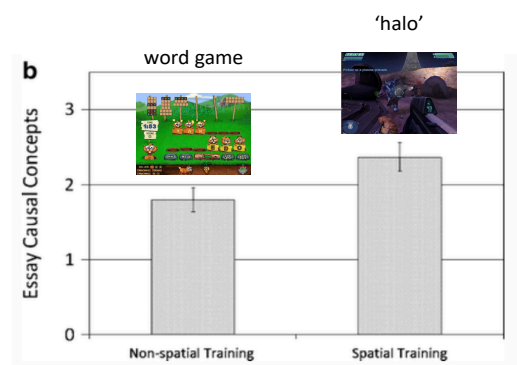


Figure 8, Can acting games help transfer to school skills? Results on the essay writing requiring spatial imagination and manipulation. Playing the action game, Halo, produces significantly better responses.

There is however a major limitation to this study. It is highly likely that the results obtained could have been contaminated by the "Mozart Effect", the consequence of improved spatial abilities due to increased arousal in mood produced by lively music, or in this case, possibly playing an arousing game (Thompson et al., 2001). More evidence is needed to

understand if the cognitive benefits of playing first-person shooter games are not just an artifact spurred by being in an alert state. To do this, more research comparing playing action games followed and educational outcomes is required, including ones that do not require visual spatial abilities is needed.

The four pillars of learning

The evidence, although very limited, from educational games and the research on ways games focus specific cognitive abilities through attention and repetition is a strong argument for investigating their union for enhancing school curriculum. What do the cognitive sciences have to say about learning? Is it possible that the rules of action video games may be more adapted to how our brains learn than traditional pen-and-paper work? If this is the case, then developing ways to combine elements of action games, such visual search, distractor avoidance and task switching into educational paradigms merits our interest. In the book How We Learn (Dehaene, 2020), Stanislas Dehaene argues that there are four pillars critical to learning. These are attention, active engagement, feedback, and consolidation. These four pillars are features common to many video games. Examining how the principles are implemented in video games provides a starting point to understand how we can also apply these techniques to improve educational curriculum.

The first of these pillars is attention. Attention is described as the “filter” that the brain uses to hone concentration and block-out distraction. This system of control develops through childhood and teenage years, through experience. In other words, attention is a cognitive ability that is trained through efforts of sustained focus and the conscious blocking of distraction necessary to accomplish many of the daily objectives in school and in life. Our learning curriculums also have a responsibility in how they orient the attention of the learner. For example, in a study where learners were taught to attend to the global shape or the three letters of a ‘fake’ writing system, scans of their brain showed that they either used the right hemisphere visual cortex, typically used for object identification, or left hemisphere visual cortex, used for letter to sound learning, depending on the instructions received. In other words, the instructions for learning had an immediate effect on what areas of the brain should be used to focus on learning (Yoncheva et al., 2015). Orienting attention through instruction is therefore a technique for teachers to guide students to the right information. As we have

seen, video games are a powerful tool for orienting spatial attention to a goal. It could be that the focused experience could also be helpful in orienting the learner's attention to a goal. However, for this to happen, the information to be learned should be made explicit so that attention is correctly directed to the learning objective.

The second pillar is active engagement. Decades of psychology and cognitive neuroscience experiments have repeatedly demonstrated that learning fails to occur when the brain is just passively exposed to information. In a meta-analysis of 225 studies comparing lecture hall (passive) versus active learning pedagogies for subjects in math and sciences, the latter performed significantly better on class test scores. It is clear that to succeed in a video game, the player must be actively engaged, as each action has a consequence, including doing nothing! Furthermore, active engagement is reinforced through periods of focus and breaks. Staying intensely focused for long periods of time is difficult, but with games, it is easy to provide intense work periods with periodic breaks to assess and regroup. These breaks provide the perfect moment to implement the next pillar of learning, error feedback.

The original Teaching Machines by Pressey and Skinner understood that students learn best when they are not in doubt. The same logic supports the third pillar of error feedback as an essential ingredient of the brain's learning algorithm. Theories of 'predictive coding' indicate that during learning, the brain forms a hypothesis, makes a prediction, and generates an error signal whenever it learns that the prediction is violated (e.g. Friston, 2005). The learning brain is, as stated by Dehaene, "constantly traversed by crisscrossing error signals that adjust its mental models of the world." Educational methods should, like video games, help children make the best use of the information provided by their errors so that they can immediately reevaluate knowledge and make better choices. In games, this can be done through error feedback requiring the learner to reevaluate and make better decision. It can also be achieved by displaying progress. While error feedback provides valuable information for the learner on knowledge to reassess, progress reports provide the positive encouragement needed to stay motivated. As stated, video games, are generally played under a short period of intense time pressure, coupled with pauses and updates on points won (correct behavior) and life-loss (incorrect behavior). The players always know where they stand and where they need to improve. Integrating this idea into educational games can be done by providing the child with many trials, feedback and the ability to try again. This is a training situation that is difficult to produce with traditional pen and paper.

Finally, the fourth pillar presented in Dehaene's book is consolidation. Consolidation is an essential step in transferring information from short-term memory to an acquired knowledge unit. Learning initially draws upon considerable top-down executive attention and conscious effort guided by internal states or intentions moderated by the prefrontal cortex (Miller & Cohen, 2001). Because this is a very slow and costly means of action, information and skills have to be moved to a more permanent memory – a process known as consolidation. Once consolidated, one does not need conscious attention and effort to apply the skill: it becomes automatic. Reading is a great example of this transformation. It is an activity that requires effort by the learner but seems perfectly involuntary by the expert. Sleep plays an essential role in consolidation (Plihal & Born, 1999; Stickgold, 2005). Experiments show that consolidation is aided by alternating periods of explicit teaching with periods of testing (also called 'retrieval practice') that probe the learner's ability to use the learned knowledge in concrete tasks (Cepeda et al., 2006). Thus, to be effective, early intervention programs need not only engage the young viewer, but they must also elicit direct participation. Using games for learning, unlike with listening to a lecture or reading from a book, require the student to grasp an information then quickly apply it. If done well, educational games could be a powerful tool to interleaving periods of learning and testing through required action.

The current situation in France

Every five years the International Association for the Evaluation of Educational Achievement (IEA) conducts the Progress in International Reading Literacy Study (PIRLS), a study of 60 countries on the reading abilities of children in fourth grade. Studies on reading development have shown an important gain in fluent literacy occurs for most children around the fourth-grade year of school, characterized by a developmental change from learning to read... to reading for comprehension and learning. It is at this period that researchers have also described the emergence of the "fourth-grade slump" (Chall et al., 1990). This term describes a divergence in reading comprehension skills between children from low-SES populations compared to their homologues, despite comparable language skills in the previous school years. The repercussions of the fourth-grade slump are considerable, because they have a Mathew effect on school learning in general. In other words, there is a cumulative advantage for children that are strong readers at this age when reading skills become

necessary for the largely text based work and independent study habits required to succeed in most school subjects (Stanovich, 1986). The Mathew effect, in essence describes a point in education when “the rich get richer and the poor get poorer”. A nation’s results on the PIRLS study are not only a measure in reading ability, but a prediction of general educational outcomes. In France, where the research for this thesis has been conducted, the nation’s results on the PIRLS test sounded a devastating alarm and a call to action to improve literacy research and applicable actions. France was also the only country in Europe to see its average score significantly decline from the 2011 edition of the test (see Fig. 9). The country scored in the lower-tail of performance on all measures of reading compared to other European countries (Mullis et al., 2017).

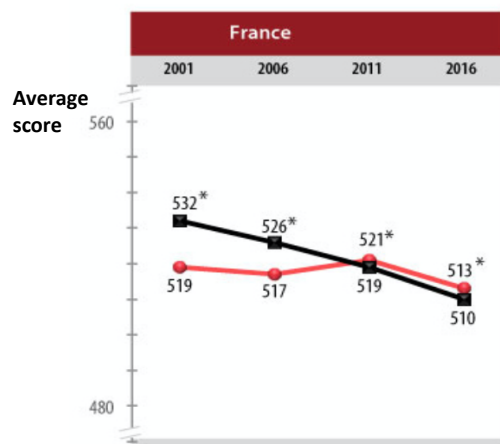


Figure 9, France’s general average score on the PIRLS assessment. The **red line** refers to literary reading and the **black line** refers to informational reading. Scores are for test years 2001, 2006, 2011 and 2016. The * means there was a significant difference between the two reading sub-types. (<http://timssandpirls.bc.edu/pirls2016/international-results/>).

Much of the research that was done for this thesis, that we will detail now, was sponsored by the ministry of education, with the goal of improving education in France, developing applicable evidence-based answers to literacy and testing if technology can be used to provide reading curriculum on a large-scale basis to students.

The research performed in this PhD thesis

The National Reading Panel devoted a chapter to studying the impacts of technology, but lamented that only 5% of the studies respected the selection criteria (i.e. demographic descriptions, sufficient description to allow for replicability, insurance of instruction fidelity

and full description of outcome measures), and concluded that there was not enough data to form a conclusion (National Reading Panel, 2000). Since then, as we have presented in this introduction, there has continued to be a growing interest in developing educational applications to teach children in school and at home. There is also a growing body of data showing that games may provide alternative routes to learning far more efficient than traditional pen and paper techniques. Finally, technology also offers the hope of providing novel means of assessing the difficulties that some children with dyslexia exhibit when they try to learn to read. In the present thesis, we worked on all of these aspects: game development, randomized control trials, and dyslexia assessment.

The research in this thesis partially benefitted from two different initiatives by the French ministry of education to finance projects with the goal of developing and researching numeric solutions for education.³ Through this support, our goal has been to develop a literacy application and test it with children in French schools. We present in the following chapters the development, testing and results of this endeavor. Each of the chapters presented are devoted to a project with concrete application.

In **chapter 2** we start with a description of the first tablet-based game application we designed, called ELAN for reading. ELAN was developed in collaboration with a company specializing in serious games. The application was designed to promote 1st grade literacy. We tested it in a randomized control trial with near 1000 French students, in partnership with the Poitiers school district. The game is based on a phonics approach, similar to GraphoGame, teaching the grapheme-phoneme correspondences of French in an explicit and systematic manner. Our results demonstrate that, relative to a control group using math software, the children who used our application early in the school year show improved decoding and comprehension skills.

Knowing what correspondences of a language should be taught, and in what order is not an easy undertaking, especially for languages with a less transparent orthography, because knowing what graphemes correspond to which phonemes is not a clear-cut relationship. For decades now, a similar conundrum has puzzled neural networks designed for grapheme-to-phoneme conversion. The task of converting graphemes to their phonemes is an important process in text to speech applications. In **chapter 3**, we introduce our solution

³ <https://www.gouvernement.fr/e-fran-3761>

to this roadblock using a neural network for grapheme-to-phoneme transcription of a language's lexicon. Once our algorithm could reliably transcribe each grapheme to its phoneme in a word, we could calculate the frequency and consistency of their occurrences in the language and decide which grapheme-phoneme correspondences should be taught and when. Along with French, we also present the results of our program on Spanish, Portuguese and English orthographies.

While the results from the study led in chapter 2 were positive, there were also several limitations: the control tablet application did not produce results and was generally less appreciated by the classes, we had several bugs in the software that caused a large loss of student data, and children only benefitted from the games early in the year. In **chapter 4**, we present the development of Kalulu, a game environment designed for teaching reading and math using the same game design from ELAN. In this chapter we explain the pedagogical philosophy of the game, and its use in a second randomized control trial, this time in kindergarten. In a longitudinal follow-up, we also look at the results for the participants on national standardized tests one year after the intervention.

Sometimes, despite best practices, children face challenges in learning how to read. In **chapter 5** we describe the theory, testing and results of the Malabi screener. This screener was developed based on the work of Naama Friedmann and colleagues (University of Tel Aviv) on different types of dyslexia. The screener was specifically designed to better understand the types of errors made by dyslexic readers with the goal of providing targeted interventions.

Finally, in **chapter 6** we discuss several of the outcomes of our research and future directions. Educational research is difficult given that it does not exist in the constrained environment of the lab. Our field experiments not only pointed to the importance of phonics instruction, but also to the possibility that this type of instruction is influenced by when children learn to read and how they are supported in this endeavor. All of the work done for this thesis should eventually be made available to the public and other researchers, to the benefit of continuing to improve curriculum and put the child's brain at the center of education.

References

- Ahmed, B. E., & Heba. (2016). Duolingo as a Bilingual Learning App : A Case Study (SSRN Scholarly Paper ID 2814822). Social Science Research Network. <https://doi.org/10.2139/ssrn.2814822>
- Bara, F., & Gentaz, E. (2011). Haptics in teaching handwriting : The role of perceptual and visuo-motor skills. *Human Movement Science*, 30(4), 745-759. <https://doi.org/10.1016/j.humov.2010.05.015>
- Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, 19(3), 433-449. <https://doi.org/10.1016/j.cogdev.2004.05.003>
- Bergelson, E., & Swingley, D. (2015). Early Word Comprehension in Infants : Replication and Extension. *Language learning and development : the official journal of the Society for Language Development*, 11(4), 369-380. <https://doi.org/10.1080/15475441.2014.979387>
- Bouhali, F., Thiebaut de Schotten, M., Pinel, P., Poupon, C., Mangin, J.-F., Dehaene, S., & Cohen, L. (2014). Anatomical connections of the visual word form area. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(46), 15402-15414. <https://doi.org/10.1523/JNEUROSCI.4918-13.2014>
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17), 7939-7944. <https://doi.org/10.1073/pnas.0904402107>
- Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint Book Reading Makes for Success in Learning to Read : A Meta-Analysis on Intergenerational Transmission of Literacy. *Review of Educational Research*, 65(1), 1-21. <https://doi.org/10.3102/00346543065001001>
- Byrne, B. (1992). *Studies in the Acquisition Procedure for Reading : Rationale, Hypotheses, and Data*. Reading Acquisition; Routledge. <https://doi.org/10.4324/9781351236904-1>
- Castles, A. (2006). The dual route model and the developmental dyslexias. *London Review of Education*. <https://doi.org/10.1080/13603110600574454>

- Castro-Caldas, A. (1998). The illiterate brain. Learning to read and write during childhood influences the functional organization of the adult brain. *Brain*, 121(6), 1053-1063. <https://doi.org/10.1093/brain/121.6.1053>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks : A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354-380. <https://doi.org/10.1037/0033-2909.132.3.354>
- Chall, J. S. (1967). *Learning to read: The great debate*. McGraw-Hill. <https://www.weber.edu/wsuiimages/jmitchell/MEDUC%206355/The%20Great%20Debate.pdf>
- Chall, J. S., Jacobs, V. A., Baldwin, L. E., & Chall, J. S. (1990). *The Reading Crisis : Why poor children fall behind*. Harvard University Press.
- Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., & Näätänen, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, 1(5), 351-353. <https://doi.org/10.1038/1561>
- Cherney, I. D. (2008). Mom, Let Me Play More Computer Games : They Improve My Mental Rotation Skills. *Sex Roles*, 59(11-12), 776-786. <https://doi.org/10.1007/s11199-008-9498-z>
- Coltheart, M. (2005). Modeling Reading : The Dual-Route Approach. In M. J. Snowling & C. Hulme (Éds.), *The science of reading : A handbook* (p. 6-23). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch1>
- Dale, P. S., Crain-Thoreson, C., Notari-Syverson, A., & Cole, K. (1996). Parent-Child Book Reading as an Intervention Technique for Young Children with Language Delays. *Topics in Early Childhood Special Education*, 16(2), 213-235. <https://doi.org/10.1177/027112149601600206>
- de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. *Scientific Studies of Reading*, 13(4), 318-333. <https://doi.org/10.1080/10888430903001308>
- Dehaene, S. (2010). *Reading in the Brain : The New Science of How We Read* (Reprint edition). Penguin Books.
- Dehaene, S. (2020). *How We Learn : Why Brains Learn Better Than Any Machine . . . for Now*. Penguin.

- Dehaene, S., & Cohen, L. (2007). Cultural Recycling of Cortical Maps. *Neuron*, 56(2), 384-398.
<https://doi.org/10.1016/j.neuron.2007.10.004>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate : Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, 16(4), 234-244. <https://doi.org/10.1038/nrn3924>
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359-1364.
<https://doi.org/10.1126/science.1194140>
- Dehaene-Lambertz, G., & Spelke, E. S. (2015). The Infancy of the Human Brain. *Neuron*, 88(1), 93-109. <https://doi.org/10.1016/j.neuron.2015.09.026>
- Dehaene-Lambertz, Ghislaine, Dehaene, S., & Hertz-Pannier, L. (2002). Functional Neuroimaging of Speech Perception in Infants. *Science*, 298(5600), 2013-2015.
<https://doi.org/10.1126/science.1077066>
- Dehaene-Lambertz, Ghislaine, Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form : Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, 16(3), e2004103.
<https://doi.org/10.1371/journal.pbio.2004103>
- Drew, B., & Waters, J. (1986). Video games : Utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly. *Cognitive Rehabilitation*, 4(2), 26-31.
- Dye, M. W. G., & Bavelier, D. (2010). Differential development of visual attention skills in school-age children. *Vision Research*, 50(4), 452-459.
<https://doi.org/10.1016/j.visres.2009.10.010>
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic Phonics Instruction Helps Students Learn to Read : Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research*, 71(3), 393-447.
<https://doi.org/10.3102/00346543071003393>
- Eichenbaum, A., Bavelier, D., & Green, C. S. (2014). Video Games : Play That Can Do Serious Good. *American Journal of Play*, 7(1), 50-72.
- Frederic, S. B. (1958). Teaching machine (United States Patent No US2846779A).
<https://patents.google.com/patent/US2846779/en>

- Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci*, 360(1456), 815-836. <https://doi.org/10.1098/rstb.2005.1622>
- Gagnon, D. (1985). Videogames and spatial skills : An exploratory study. 13.
- Galaburda, A. M., LoTurco, J., Ramus, F., Fitch, R. H., & Rosen, G. D. (2006). From genes to behavior in developmental dyslexia. *Nature Neuroscience*, 9(10), 1213-1217. <https://doi.org/10.1038/nn1772>
- Goodman, K. S. (1967). Reading : A psycholinguistic guessing game. *Journal of the Reading Specialist*, 6(4), 126-135. <https://doi.org/10.1080/19388076709556976>
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534-537. <https://doi.org/10.1038/nature01647>
- Griffith, J. L., Voloschin, P., Gibb, G. D., & Bailey, J. R. (1983). Differences in Eye-Hand Motor Coordination of Video-Game Users and Non-Users. *Perceptual and Motor Skills*, 57(1), 155-158. <https://doi.org/10.2466/pms.1983.57.1.155>
- Hirsch, E., & Niedermeyer, F. C. (1972). The Effects of Tracing Prompts and Discrimination Training on Kindergarten Handwriting Performance. <https://eric.ed.gov/?id=ED106864>
- Homer, B. D., Plass, J. L., Raffaele, C., Ober, T. M., & Ali, A. (2018). Improving high school students' executive functions through digital game play. *Computers & Education*, 117, 50-58. <https://doi.org/10.1016/j.compedu.2017.09.011>
- Istance, D., & Kools, M. (2013). OECD Work on Technology and Education : Innovative learning environments as an integrating framework. *European Journal of Education*, 48(1), 43-57. <https://doi.org/10.1111/ejed.12017>
- Kim, J. S. (2008). Research and the Reading Wars. *The Phi Delta Kappan*, 89(5), 372-375.
- Kirschner, P. A., & van Merriënboer, J. J. G. (2013). Do Learners Really Know Best? Urban Legends in Education. *Educational Psychologist*, 48(3), 169-183. <https://doi.org/10.1080/00461520.2013.804395>
- Lavelli, M., Barachetti, C., Majorano, M., Florit, E., Brotto, C., & Miottello, P. (2019). Impacts of a shared book-reading intervention for Italian-speaking children with developmental language disorder. *International Journal of Language & Communication Disorders*, 54(4), 565-579. <https://doi.org/10.1111/1460-6984.12460>
- Li, X., Atkins, M. S., Li, X., & Atkins, M. S. (2004). Early childhood computer experience and cognitive and motor development. *Pediatrics*, 1715-1722.

- Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning : A Duolingo case study. *ReCALL*, 31(3), 293-311. <https://doi.org/10.1017/S0958344019000065>
- Longcamp, M., Zerbato-Poudou, M.-T., & Velay, J.-L. (2005). The influence of writing practice on letter recognition in preschool children : A comparison between handwriting and typing. *Acta Psychologica*, 119(1), 67-79. <https://doi.org/10.1016/j.actpsy.2004.10.019>
- López-Barroso, D., Thiebaut de Schotten, M., Morais, J., Kolinsky, R., Braga, L. W., Guerreiro-Tauil, A., Dehaene, S., & Cohen, L. (2020). Impact of literacy on the functional connectivity of vision and language related networks. *NeuroImage*, 213, 116722. <https://doi.org/10.1016/j.neuroimage.2020.116722>
- Lowery, B. R., & Knirk, F. G. (1982). Micro-Computer Video Games and Spatial Visualization Acquisition. *Journal of Educational Technology Systems*, 11(2), 155-166. <https://doi.org/10.2190/3PAN-CHJM-RTOL-W6AC>
- Lukatela, K., & Liberman, I. Y. (1994). Phonological Awareness in Illiterates : Observations from Serbo-Croatian.
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 17(5), 714-726. <https://doi.org/10.1111/desc.12152>
- Lyytinen, H., Erskine, J., Kujala, J., Ojanen, E., & Richardson, U. (2009). In search of a science-based application : A learning tool for reading acquisition. *Scandinavian Journal of Psychology*, 50(6), 668-675. <https://doi.org/10.1111/j.1467-9450.2009.00791.x>
- McCandliss, B., Beck, I. L., Sandak, R., & Perfetti, C. (2003). Focusing Attention on Decoding for Children With Poor Reading Skills : Design and Preliminary Tests of the Word Building Intervention. *Scientific Studies of Reading*, 7(1), 75-104. https://doi.org/10.1207/S1532799XSSR0701_05
- Microsoft. (2001). Halo : Combat Evolved.
- Miller, E. K., & Cohen, J. D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annual Review of Neuroscience*, 24(1), 167-202. <https://doi.org/10.1146/annurev.neuro.24.1.167>
- Mimeau, C., Cantin, É., Tremblay, R. E., Boivin, M., & Dionne, G. (2020). The bidirectional association between maternal speech and child characteristics. *Journal of Child Language*, 47(2), 435-456. <https://doi.org/10.1017/S0305000919000539>

- Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language*, 127(3), 356-365. <https://doi.org/10.1016/j.bandl.2013.10.009>
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24, 45-64.
- Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, 7, 323-331.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). PIRLS 2016 International Results in Reading. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/pirls2016/international-results/>
- National Reading Panel. (2000). Teaching children to read : An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). National Institute of Child Health & Development. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- Nichols, S. (2000). Unsettling the Bedtime Story : Parents' Reports of Home Literacy Practices. *Contemporary Issues in Early Childhood*, 1(3), 315-328. <https://doi.org/10.2304/ciec.2000.1.3.7>
- Ojanen, E., Jere-Folotiya, J., Yalukanda, C., Sampa, F., Nshimbi, C., Katongo, M., Choopa, M., & Lyytinen, H. (2015). Mobile solution for better reading instruction in rural Africa. 2015 IST-Africa Conference, 1-13. <https://doi.org/10.1109/ISTAFRICA.2015.7190559>
- Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., Kauppinen, K.-P., Ketonen, R., Ngorosho, D., Pitkänen, M., Puhakka, S., Sampa, F., Walubita, G., Yalukanda, C., Pugh, K., Richardson, U., Serpell, R., & Lyytinen, H. (2015). GraphoGame – a catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00671>
- Olson, R. K., Keenan, J. M., Byrne, B., & Samuelsson, S. (2014). Why do Children Differ in Their Development of Reading and Related Skills? *Scientific studies of reading : the official journal of the Society for the Scientific Study of Reading*, 18(1), 38-54. <https://doi.org/10.1080/10888438.2013.800521>

- Pascual-Leone, A., Amedi, A., Fregni, F., & Merabet, L. B. (2005). The plastic human brain cortex. *Annual Review of Neuroscience*, 28(1), 377-401. <https://doi.org/10.1146/annurev.neuro.27.070203.144216>
- Perzov, A., & Kozminsky, E. (1989). The Effect of Computer Games Practice on the Development of Visual Perception Skills in Kindergarten Children. *Computers in the Schools*, 6(3-4), 113-122. https://doi.org/10.1300/J025v06n03_10
- PIRLS 2016—Évaluation internationale des élèves de CM1 en compréhension de l'écrit—Évolution des performances sur quinze ans. (s. d.). Ministère de l'Éducation Nationale de la Jeunesse et des Sports. Consulté 13 octobre 2020, à l'adresse <https://www.education.gouv.fr/pirls-2016-evaluation-internationale-des-eleves-de-cm1-en-comprehension-de-l-ecrit-evolution-des-11429>
- Plihal, W., & Born, J. (1999). Effects of early and late nocturnal sleep on priming and spatial memory. *Psychophysiology*, 36(5), 571-582.
- Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, 24(4), 450-472. <https://doi.org/10.1016/j.cogdev.2009.09.003>
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method : The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. *Human Technology*, 10(1). <https://doi.org/10.17011/ht/urn.201405281859>
- Sanchez, C. A. (2012). Enhancing visuospatial performance through video game training to increase learning in visuospatial science domains. *Psychonomic Bulletin & Review*, 19(1), 58-65. <https://doi.org/10.3758/s13423-011-0177-7>
- Schaadt, G., Pannekamp, A., & van der Meer, E. (2013). Auditory phoneme discrimination in illiterates : Mismatch negativity—A question of literacy? *Developmental Psychology*, 49(11), 2179-2190. <https://doi.org/10.1037/a0031765>
- Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior*, 55, 223-229. <https://doi.org/10.1016/j.chb.2015.09.013>
- Serpell, R., Jere-Folotiya, J., Chansa-Kabali, T., Munachaka, J., Maumbi, M. N., Yalukanda, C., Sampa, F., & Lyytinen, H. (2017). A Culturally Sensitive Approach to Promoting Initial Literacy Development in Africa : Ongoing and Planned Research and Development at the

- University of Zambia's Centre for Promotion of Literacy in Sub-Saharan Africa (CAPOLSA). In A. Abubakar & F. J. R. van de Vijver (Éds.), *Handbook of Applied Developmental Science in Sub-Saharan Africa* (p. 313-334). Springer. https://doi.org/10.1007/978-1-4939-7328-6_17
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143-174. <https://doi.org/10.1348/000712603321661859>
- Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2), 151-218. [https://doi.org/10.1016/0010-0277\(94\)00645-2](https://doi.org/10.1016/0010-0277(94)00645-2)
- Skinner, B. F. (1968). *The Technology of Teaching*. Pearson College Div., 255.
- Spencer, L. H., & Hanley, J. R. (2003). Effects of orthographic transparency on reading and phoneme awareness in children learning to read in Wales. *British Journal of Psychology* (London, England: 1953), 94(Pt 1), 1-28. <https://doi.org/10.1348/000712603762842075>
- Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *READING RESEARCH QUARTERLY*, 48.
- Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063), 1272-1278. <https://doi.org/10.1038/nature04286>
- Thiebaut de Schotten, M., Cohen, L., Amemiya, E., Braga, L. W., & Dehaene, S. (2014). Learning to read improves the structure of the arcuate fasciculus. *Cerebral Cortex* (New York, N.Y.: 1991), 24(4), 989-995. <https://doi.org/10.1093/cercor/bhs383>
- Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/Spelling Instruction as Part of a First-Grade Reading Program: Effects on Several Measures of Reading. *Reading Research Quarterly*, 28(3), 219-233. <https://doi.org/10.2307/747995>
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of Acoustical Society of America*. <https://www.unige.ch/fapse/logopedie/files/1214/1285/1101/article6.pdf>
- Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006). Principles underlying the design of « The Number Race », an adaptive computer game for remediation of dyscalculia. *Behav Brain Funct*, 2(1), 19.
- Wilson, Anna J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an Adaptive Game Intervention on Accessing Number Sense in Low-Socioeconomic-Status Kindergarten

Children. *Mind, Brain & Education*, 3(4), 224-234. <https://doi.org/10.1111/j.1751-228X.2009.01075.x>

Wilson, Anna J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of « The Number Race », an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2(1), 20. <https://doi.org/10.1186/1744-9081-2-20>

World Health Organization. (2020). *International statistical classification of diseases and related health problems (11th ed.)*. <https://icd.who.int/en>

Yoncheva, Y. N., Wise, J., & McCandliss, B. (2015). Hemispheric specialization for visual words is shaped by attention to sublexical units during initial learning. *Brain and Language*, 145-146, 23-33. <https://doi.org/10.1016/j.bandl.2015.04.001>

Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading : A Cross-Language Investigation. *Psychological Science*, 21(4), 551-559. <https://doi.org/10.1177/0956797610363406>

Chapter 2, Accelerating reading acquisition and boosting comprehension with a cognitive science-based tablet training

Potier Watkins, C., 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. <https://doi.org/10.1007/s40692-019-00152-6>

Abstract

Tablets and computers offer opportunities for learning, but their potential is only as great as the quality of the software they propose. Educational games must not only provide an engaging design, but also be based on principles from cognitive neuroscience and education research and be evaluated in large-scale classroom tests. Here, we describe ELAN, an adaptive game that supports literacy acquisition through teaching and training phonics. It provides explicit systematic grapheme–phoneme correspondence instruction and reinforces full decoding through reading and spelling practice with 100% decodable text. The game also uses periodical lexical decision tasks to measure the transition from letter-by-letter decoding to fluent word recognition. The software was tested in a randomized control trial in 44 first-grade classrooms (n = 975 French children). Children who used ELAN software during the first term improved relative to two control groups, respectively, using math software or no-tablet “business-as-usual” classrooms. Improvements were significant in reading fluency (one-minute word and pseudo-word reading) and sentence reading comprehension, consistent with the idea that improved decoding can help the child focus on understanding. These results emphasize the importance of early, explicit and systematic phonics training, and provide a new software tool to facilitate it.

Introduction

International evaluations in schools have reported that major investments in technology for the classroom do not necessarily lead to improvements in math and reading abilities (OECD 2015). This is in part due to the lack of quality research-backed software that not only takes advantage of the experiences that technology can provide, but also incorporates scientific findings from cognitive neuroscience and education about how the brain learns. Previous research has shown that games that adhere to evidence backed pedagogy, gamify drill-and-practice, and use adaptive algorithms to provide individualized learning can lead to improvements in math (Räsänen et al. 2009; Schacter and Jo 2016; Wilson et al. 2006a, b; Wilson et al. 2009) and reading (de Graaff et al. 2009; Kyle et al. 2013; Lyytinen et al. 2007; Saine et al. 2011). The goal of the current project was to test the ELAN software, a tablet-based game designed to support reading acquisition in the classroom. The game's pedagogy is inspired from evidence in cognitive neuroscience and education research as to how children learn to read. The game focuses on phonics teaching and training, an essential steppingstone to reading mastery as made evident by the meta-analyses of the National Reading Panel (Cunningham 2001). ELAN was not designed to supplant the role of the teacher but aims to provide 'drill-and-practice' and tailored game difficulty to accelerate consolidation of the grapheme–phoneme correspondences in initial decoding stage of reading.

The reading brain

The goal of learning to read is to train the visual system to decode letter strings and provide a novel input into the already developed cortical areas for spoken language comprehension, (Dehaene et al. 2010). Advances in functional MRI before and after learning to read have revealed that part of the left occipito-temporal pathway dedicated to visual recognition becomes sensitive to letter strings and develops increasingly efficient connections to regions, such as the planum temporale, specialized in the processing of speech sounds (Dehaene et al. 2010; Dehaene-Lambertz et al. 2018; Monzalvo and Dehaene-Lambertz 2013). In the early stages of reading, the learner must effortfully convert each letter or group of letters (called a 'grapheme') into the corresponding sound unit (called a 'phoneme') while 'listening in' to understand the word. With practice, reading becomes automatic, and words can be treated in their entirety, with all letters analyzed in parallel, permitting rapid access to the lexicon.

During the early decoding stages of reading, reading time increases linearly with word length, with a slope as high as ~200 ms/letter (Zoccolotti et al. 2005), reflecting a slow process of serial letter-by-letter reading. As reading automatizes, that slope decreases until all words between 3 and 8 letters are read equally fast (New et al. 2006), except in young dyslexic readers (Zoccolotti et al. 2005). These results fit with a dual-route model of reading: the phonological path allows for a slow decoding of novel words, while the lexical path lends itself to fluent reading of familiar words (Coltheart 2005). Both neural paths exist in the proficient reader, and the goal of reading instruction should therefore be to quickly establish them. According to meta-analyses by the National Reading Panel (Cunningham 2001), the most efficient way to acquire literacy is phonics instruction, in both normally developing readers and children with increased risk for reading deficits.

The language caveat

The facility with which phonics can be acquired is modulated by the language's orthographic transparency. Transparency describes the degree to which grapheme-phoneme correspondences are consistent enough to support fluent reading of novel words, a key determinant of the speed of reading acquisition (Serrano et al. 2011). For example, English and, to a lesser extent, French have many letters that can make many different sounds. These two languages are considered opaque compared to Spanish or Italian—both of which have a highly consistent orthography, meaning that a given letter (or grapheme) has nearly always the same sound. For transparent languages, knowledge of the basic grapheme-phoneme correspondences suffices to read virtually any word. This difference in the transparency of written language influences the duration and difficulty of reading acquisition (Serrano et al. 2011; Ziegler et al. 2010) and the extent of cortex dedicated to the visual component of reading (Paulesu et al. 2000). Yet in spite of this, across all alphabetic languages, learning the grapheme-phoneme correspondences remain the most economical path to reading all words (Vousden et al. 2011).

Principles for phonics instruction

There are several key principles to a successful phonics instruction. Grapheme-phoneme correspondences should be taught explicitly (Castles et al. 2018; Cunningham 2001). The

explicit teaching of grapheme–phoneme correspondence should follow a systematic progression (Ehri et al. 2001), meaning that they are taught in a carefully planned sequence, within a rational order that takes into account their frequency and their consistency. Systematic phonics software has been shown to provide better results than non-systematic software (de Graaff et al. 2009).

Practicing lesson-related words should also be a part of any systematic phonics instruction (Mason et al. 1974). Children should be presented with decodable text, i.e., using graphemes whose pronunciation has been previously taught. In spite of adequate phonics instruction, many children show difficulties in applying full decoding skills. These children have no trouble identifying the first letter of a word but fail to apply their phonics knowledge to all successive letters in such a manner that they combine them into readable syllables. Reading activities that draw attention to the combinations of letters in a word help children apply the alphabetic principle to all letters (McCandliss et al. 2003). Learning to spell, which requires children to focus on each sound and its corresponding letter, also appears to improve decoding skills (Uhry and Shepherd, 1993).

Tablet-based interventions for reading instruction

Tablet- or computer- based games may provide an excellent medium for helping children to automatize their grapheme–phoneme decoding skills. A prime cross-language example and an important source of inspiration for the present project is the GraphoGame, a Finnish computer/tablet-based game in which the primary goal is to incite the child to automatize the associations between graphemes and phonemes (Richardson and Lyytinen 2014). Functional brain imaging shows that when preschoolers play the GraphoGame for a few hours, the neural circuits for reading begin to emerge (Brem et al. 2010). Playing the GraphoGame helps children considered at risk for reading acquisition in both English (Kyle et al. 2013) and Finnish (Saine et al. 2011) improve their decoding and encoding skills above their respective control groups.

The ELAN software

Since the GraphoGame was at the time was not available in French, our goal for the ELAN software was to provide similar phonics training for children in France. We also included in

ELAN a teaching section that explicitly introduces new grapheme–phoneme combinations in a systematic order, thus allowing children to advance in the game independently of learning in the classroom. ELAN also uses a lexical decision task to measure changes in reading fluency. We now describe these points in turn.

In ELAN, the child must explore 20 different islands where he or she will learn 1 to 5 different grapheme–phonemes relationships. After learning a new grapheme–phoneme lesson, automatization is reinforced through three mini-games that build on skills of syllable and word reading. Passage from one island to the next is allowed once the child has completed all mini-games with a minimum score of 80% correct responses and completed the island “boss”. Children receive awards with each game won.

The systematic ordering of grapheme–phoneme correspondences was determined by examining the frequency and consistency of all grapheme–phoneme correspondences in the Manulex corpus, a corpus of 1.9 million words from French children’s books (Lété et al. 2004). To further facilitate learning, we made several principled adjustments based on the progression usually taught in phonics manuals and teachers reports. In ELAN, the most frequent vowels are taught first. Regarding consonants, frequent fricatives are instructed before stop consonants due to the greater facility in pronouncing such phonemes for a prolonged duration in isolation and therefore in teaching how they blend into syllables (e.g., ‘ffff aaa makes fa’). These adjustments allow ELAN to be used as a practice aid to support general learning in the class, as opposed to replacing methods used by teachers. The appendix presents the full order of grapheme–phonemes taught in ELAN.

Within a given ELAN lesson, a grapheme–phoneme correspondence is introduced in three steps that are commonly used by teachers and supported by research as benefitting phoneme processing and grapheme memorization (Fig. 1). In step 1, the child clicks the grapheme to hear and see a high-quality, noise-free video of an older child pronouncing the corresponding phoneme, with a zoom on the child’s mouth and explicit articulation. Enhancing auditory signal-to-noise, for instance, through auditory aids, is known to facilitate reading acquisition (Tallal 2004). Pictures of distinct articulatory gestures underlying each phoneme facilitate the development of early reading (Boyer and Ehri 2011; Castiglioni-Spalten and Ehri 2003). In step 2, when the child clicks on the grapheme, a picture that starts with the phoneme is displayed and described (e.g., ‘aaa, aaapple’). This type of exercise relies on the acrophonic principle, which is at the origins of alphabetic writing (each letter used to be the

first letter of a common word, e.g., b=“beit”=a house). It is commonly used in the classroom and used here to bind the phoneme to a known word in order to facilitate memorization and to build phonemic awareness of letter sounds in words. The image is there only to ensure understanding of the spoken word. Children can repeat these two steps as much as needed. In step 3, the child traces the grapheme twice in both upper and lower case on the touch screen. Haptic exercises with letters, such as learning to write or to trace letters with the finger are useful activities which have been shown to improve reading skills, probably because they provide an additional motor code to support memory for grapheme (Bara and Gentaz 2011; Bara et al. 2004, 2007, 2016; Longcamp et al. 2005). Because gestures are completely different for visually similar letters such as b and d, tracing may also help surmount the mirror-letter confusions that all children experience, due to the fact that the visual cortex generalizes across viewpoints (Dehaene et al. 2010).

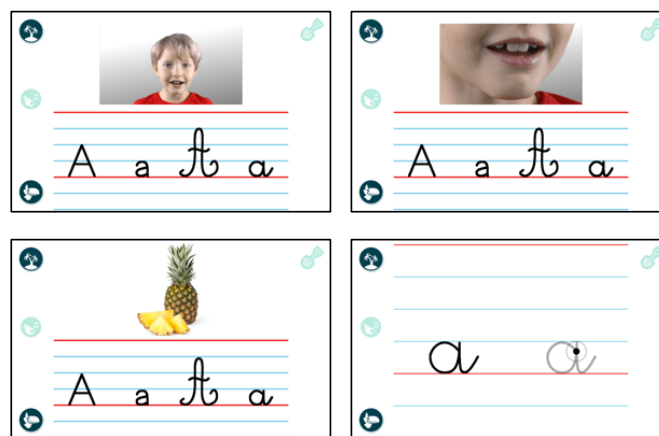


Figure 1, Grapheme–phoneme learning in the ELAN software. The child is introduced to each grapheme–phoneme correspondence by seeing the grapheme and watching a video of another child pronouncing the sound (top images), seeing the grapheme and hearing the sound in a word with accompanying picture (bottom-left image), and writing the letter on the screen (bottom-right image).

Automatizing grapheme–phoneme association

Once the child has completed the lesson, the next goal is to overlearn the correspondence to the point that it becomes automatic. ELAN uses several mini-games that require children to make faster and faster responses to recently acquired grapheme–phoneme relationships, similar to the GraphoGame (see Fig. 2a). In each game, children hear a syllable and must click on the corresponding written stimulus on screen. Syllables can be a single vowel sound, consonant–vowel (CV) or vowel–consonant (VC). Difficulty is created by varying the distractors and the rate of targets and distractors. Distractors can be syllables with the letters reversed,

the vowel replaced, or the consonant replaced (for example, for the target syllable ‘la,’ distractors would include ‘al,’ ‘le,’ and ‘ta’).

Once children have learned a grapheme–phoneme pair, they immediately apply this knowledge to reading and spelling small words (Fig. 2b). Finally, the third type of games in ELAN asks children to read short sentences and show their understanding through an action, for example, choosing a picture that corresponds to the written text, or showing the game character the silent letters in a sentence (Fig. 2c). Phonics methods are often criticized for not also helping children develop text comprehension. The goal of these games is to provide periodic comprehension practice using words that respect the phonics progression of the game. The algorithm for automatic word selection in ELAN ensures that all words that are presented for spelling and reading are 100% decodable.

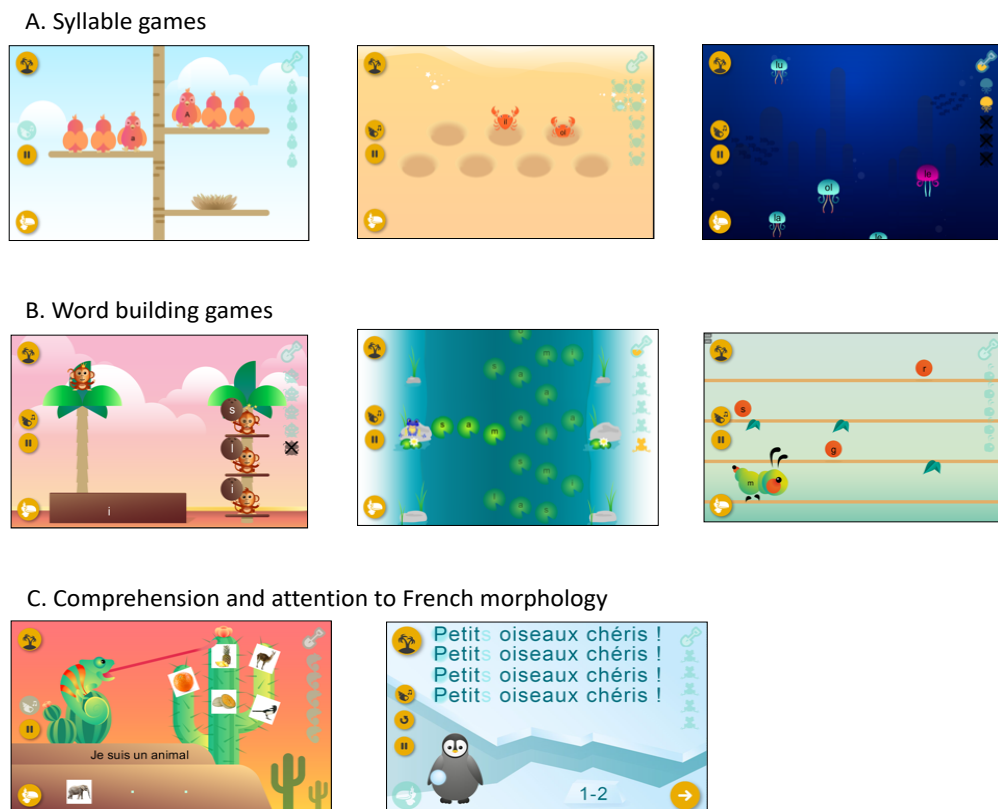


Figure 2 Screenshots of various games. **a** Letter and syllable games. After learning a given grapheme–phoneme correspondence, children practiced automatizing it by playing various games inspired by the GraphoGame, in which they heard the grapheme within syllables and had to find them written in upper and lower-case letters on screen. The number and type of distractors and their rate of their appearance were automatically adapted to each child’s performance. **b** Word building games required children to hear a word and then spell it. Difficulty was adjusted to each child by varying word length and the number and type of distractor letters. **c** Decoding and comprehension games required children to read sentences with 100% decodable text and demonstrate their understanding through an action (e.g., deciding whether the phrase was in the singular or plural).

Adapting to the child's level

ELAN's algorithm provides several types of adaptation to each child's level and errors. The local support, common to adaptive gaming, addresses children's punctual errors: if the child makes mistakes, an error bell and loss of a life-point ensue. If the child makes a second error, the game character appears to remind the child of the game objective and highlights the correct response. If the child responds correctly, the highlight is removed for the next question; otherwise, the game ends and another game proposes the same lesson through a different strategy.

Game difficulty is also adaptive. As the child makes more correct responses, the game requires quicker responses. For each game, five levels are set up. These five levels of difficulty control for, according to the mini-game environment, the number of distractors present, the number of locations the child must monitor in order to find the target, and the time the target is visible. Two successive mini-game wins or losses cause the level of difficulty to change. Requiring faster and faster responses in environments of growing difficulty would be difficult to reproduce with traditional pen and paper tests and is intended to promote automaticity.

Finally, the choice of targets and distractors also adapts to each child. The child is required to constantly maintain a performance level superior to 75% correct for all the grapheme–phoneme pairs learned. If across several games, the child makes several mistakes on the same grapheme–phoneme, its score may decrease below this criterion, causing it to become reintroduced as a target. The knowledge score of each grapheme–phoneme pair is based on the last 5 responses to the pair. At the onset of each mini-game, the game software chooses appropriate targets for the child. This pool of syllable or word items is chosen from the current lesson (80% of possible target items), previously learned items (10%) and items containing a grapheme–phoneme correspondence that tipped below the 75% correct response threshold (10%). If all items are above threshold, then random previously learned items are presented. Using this technique allows for repeated spaced practice, which is known to benefit long-term retention (for reviews, see (Carpenter et al. 2012; Cepeda et al. 2006).

Distractor stimuli are also carefully chosen according to the target and the child's performance history. While only previously learned grapheme–phoneme correspondences can act as distractors (nothing is presented to the child that has not been taught in the game), these are chosen according to their orthographic or phonemic proximity to the target. The

visual similarity matrix was adapted from another study of similarity judgments of letter pairs (Boles and Clifford 1989). The phoneme similarity matrix was adapted from French listeners' confusions in discriminating phonemes in CVC pseudo-words presented in natural background noise (Meyer et al. 2013). Only vowels are presented as vowel distractors and consonants as consonant distractors. Half of distractors were chosen based on the letter shape matrix and half based on the phoneme similarity matrix. Initial items are chosen (when possible) to be >0.75 in similarity to the target (1 being the matrix diagonal, and 0 being maximally different items). If the player chooses an erroneous distractor, then the corresponding threshold on distractor similarity is lowered.

Measuring the word length effect

As mentioned in "The reading brain" section, a marker of efficient reading is the disappearance of the word length effect, all written words being ultimately processed equally fast regardless of their length (between ~3 and ~8 letters). The challenge in ELAN is to measure how reading is affected by word length without requiring the child to actually read aloud (given that high-quality child-oriented speech recognition was unfortunately not available at the time of programming). In order to measure reading time, ELAN uses a lexical decision game, presented as a "boss" allowing passage from island to island (see "The ELAN software" section), and requiring the children to send 'real words' to a green buoy and 'invented words' to a red buoy. In this game, the child is presented with 16 words and 16 pseudo-words (with a single letter change or inversion), 2–5 letters long, in random order. The challenge must be accomplished in under three minutes (indicated to the child by the sun that crosses the sky). This game thus emphasizes fast word recognition and allows us to measure the identification time for words and pseudo-words. Through this game, we aimed to evaluate the reading slope, i.e., the amount of increase in reading time with each additional letter. As described below, this strategy was successful: identification time initially increased linearly with each additional letter, and that slope decreased with automatization.

Experimental study

The Elan software was designed in several successive stages, each involving piloting with first-grade children as well as adult crash tests. The piloting sessions with children were conducted in classrooms with members of the lab observing the children's interactions of the games.

Once the games and their parameters were set up, we tested it in a large test with 1st-grade French schools, contrasting it with both a low-level no-tablet control and with an active control (children playing a previously designed number game, the NumberCatcher). We now describe how this experimental study was organized. The project was approved by the ethical committee CERES (comité d'éthique pour la recherche en santé) on June 28, 2016.

Materials and methods

Participants

Testing of the ELAN software was done in collaboration with the public-school district of Poitiers, a region in central-west France. School district personnel in charge of the project were asked to provide approximately 1000 children from classes representing the region. Teachers were approached by district employees, and 53 teachers agreed to participate, for a total of 53 classrooms from 45 schools and 36 towns (ranging in population from ~ 80,000 to ~ 3000 inhabitants). Representative of these demographics, 44 classrooms agreed to integrate our tablet intervention with their reading instruction curriculum, while 9 control classes (group labeled “control”) followed a “business-as-usual” curricula for the entire school year. Note that the tablet/ no-tablet variable was not randomly assigned, as it was not possible to force teachers to one of these two groups against their will.

Procedure

At the start of the school year, teachers sent home with students a letter explaining the project, accompanied by a ‘letter of opposition’ for parents who did not wish for their child to participate. In classrooms participating in the tablet intervention, the project followed a randomized crossover design (see Fig. 3). Students were randomly assigned to groups of 3 to 8 children playing the same game at the same time in the classroom’s designated ‘tablet zone.’ The only request made by the lab was that teachers check that groups be heterogeneous in ability. These teacher-determined small groups were then randomized by the lab into either a group that played ELAN for the first half of the year, then the control math game for the second half of the year (group 1, labeled “read/math”), or, conversely, the math game first, then ELAN (group 2, labeled “math/read”). At the start of the year, 975 children were announced by teachers as participating in the project: 417 were assigned to read/ math, 357

to math/read, and 201 to control. Whether a group was to play ELAN or the control math game was only announced to teachers on day one of the game intervention.

All children were pretested from September 20 to October 11, 2016, except for 5 children who were tested in the first week of the intervention. This was followed by two weeks of school vacation, and a week to install the application. The two tablet intervention groups then began intervention A as soon as the games were installed. The first period lasted from November 7, 2016, to until the end of the first week of February 2017, for a possible 11 weeks of game play (this excludes the 2 weeks of Christmas break). The first post-test (labeled post-test A) began on January 30, 2017 and ended on February 18, 2017. After post-test A, children had 2 weeks of winter vacation, then switched games. At this point, schools were sent an updated version of ELAN correcting some bugs. Intervention B commenced as soon as the game was updated and lasted from March 13 to June 9, 2017. This second period was one week longer (excluding the 2 weeks of spring vacation), allowing teachers to make up for lost time due to four 3-day weekends during this period. The second post-test (labeled post-test B) began on June 12, 2017 and ended on June 30, 2017. Teachers were asked that, as much as possible, children be allowed to make up for lost days. Our goal was for each child to play for twenty minutes, three times a week during each intervention period. The control game used for the intervention was NumberCatcher (www.thenumbercatcher.com). This game, a sequel to a previous game developed by our lab (Number Race: Wilson et al. 2009; Wilson et al. 2006a, b), was created for children between ages 5 and 10. The goal of NumberCatcher is to train basic concepts of number, arithmetic, and base-10 understanding and to cement the links between symbols and quantities by requiring quicker and quicker responses. No research had been conducted on the benefits of the game for strengthening number understanding at the time of the study.

Study Design

sept.	oct. – jan.	feb.	mar.-may	june
pretest	session A	posttest A	session B	posttest B
	Group 1, Elan Group 2, Math	↔	Group 1, Elan Group 2, Math	
	Group 3, Control		Group 3, Control	



Figure 3, All children took reading and math tests at the beginning, middle, and end of the year (respectively, termed pretest, posttest A, and posttest B). In a classic crossover design, during the first half of the school year, children in the intervention groups (group 1 and group 2) played with either ELAN reading software or the math NumberCatcher software. After posttest A, children switched to the other software for the second half of the year. Control classes (group 3) did not have tablets. Children played in small groups using the same game: ELAN (top picture) or NumberCatcher (bottom picture).

Teachers were instructed not to change their own teaching method in light of the tablet intervention. Adding the tablet atelier was a challenge to incorporate for many teachers but reports from the academy found this change to be overall positive, as it gave more time to take care of other children during this period. The games were created for autonomous play and teachers did not interact with the children while they played, unless asked for help by the student. All teachers reported that, while a subgroup of children was playing with the tablet, they continued their usual learning program with the rest of the class or provided other specialized work groups. When playing ELAN or NumberCatcher, children worked individually, wearing headphones, but could communicate and help each other since all children in a group played the same game.

Most of the participating classes did not have access to wifi, so a child's progression in the game was stored in the application and only occasionally uploaded to our servers. This required that a child always play on the same tablet, logging in by entering a private code. To share the child's progression data with the lab, teachers were asked to connect to the school

wifi and upload the data to the lab server. Periodic reminders were sent to all actors in the project reminding them to upload data.

Pre- and post-tests

Except for a vocabulary test that was only administered at pretest, all tests were given at all three test periods. Forty employees from the Poitiers school district individually tested each child. All tests were done on paper. Twenty-three members of this team had attended a test training day organized by the lab and were in charge of training other testers. These school employees were often in contact with teachers but would only be knowledgeable of a student's group if they purposefully asked, which they were instructed not to do during testing. Given that most children would not yet have had formal reading training at pretest, we emphasized testing known predictive measures as well as measures of reading ability. Although this was not the primary goal of this study, we also included two number knowledge tests to provide a minimal evaluation of the results for our control group.

Passive vocabulary test (30 items) We used a standardized French vocabulary test (TVAP, Deltour and Hupkens 1990). A word was said by the experimenter. The child was asked to choose the best corresponding image from a choice of six pictures. Two points were awarded for the correct response, 1 point for a close response, 0 for all other responses, for a total possible score of 60.

Phoneme (24 items) and syllable suppression (10 items) These two tests were taken from EVALEC, a battery of French predictive reading acquisition measures (Pourcin et al. 2016). The child was asked to suppress the first phoneme/syllable of a pseudo-word pronounced by the experimenter. Ten syllable suppression questions then twelve consonant–vowel–consonant and twelve consonant–consonant–vowel pseudo-words were said to the child. For each type of question, there were two trials with feedback. The child's percentage correct was taken separately for syllable suppression and phoneme suppression.

Letter knowledge (52 items) Children were asked to give the name and sound of all 26 letters. The child's percentage correct out of 52 items was collected. Letters were presented on paper, one-by-one and out of order in lower case. Knowledge of letter sounds is a critical component of early literacy skills and the single best predictor of first-year reading achievement (Neuman and Roskos 1998).

One-minute word and pseudo-word reading (2×1 min reading) The child was asked to read a maximum number of words followed by a maximum number of pseudo-words, each in one minute (Sprenger-Charolles et al. 2017). Children were presented with 35 words and 30 pseudo-words at pretest and post-test A. Sixty words were presented at post-test B. Words and pseudo-words were all either mono- or bi- syllabic. All the presented words were frequent and taken from the Manulex children’s word database (Lété et al. 2004). Pseudo-words were matched to the words in syllable structure and orthographic difficulty. If a child was unable to do these two tests, reading less than four words from the first ten words, then he or she was not asked to continue to the more difficult reading tests that followed. To stabilize the variance, children’s mean combined score for these two tests was analyzed, after checking that they provided consistent results. This combined score from both tests provides us with a measure of reading fluency, in other words, how well the child can read familiar words as well as their ability to decode. Combining these two tests is a main early predictor of later reading comprehension (Gentaz et al. 2015).

Non-sense text reading The child was asked to read a text of 265 words, all of which are frequent. The text, however, is non-sense and thus requires that the child avoid reading words by context. The text, known as the Alouette test, has been standardized with French primary school children and is used to evaluate reading age, diagnose dyslexia (Lefavrais 2005), and predict future reading difficulties (Bertrand et al. 2010; Cavalli et al. 2017).

Sentence comprehension (8 items) Eight items were adapted from the French standardized ECOSSE test and Reading Evaluation Test (Sprenger-Charolles et al. 2017; Lecocq 1996). The child had to read aloud a sentence such as “the man is eating an apple,” then choose the correct corresponding image from a choice of four pictures including semantic and syntactic distractors. Items were changed from post-test A to post-test B. The sentences in this test were in order of increasing syntactic difficulty. The percentage correct score was calculated for each child.

Items from pre and post-test A were not changed as 57% of children did not attempt this test and the average score was below 1 response correct at pretest.

Symbolic number comparison (number of CR in 30 s) In the SYMP test of number comparison, children were asked to cross out the numerically larger of two Arabic one-digit numbers in thirty seconds. This symbolic number comparison test is correlated with math achievement from first to sixth grade in school (Brankaer et al. 2017). 56 pairs of numbers were presented

on a page. Four trials with feedback were provided before beginning. Children did the same test at each of the three tests. The maximum number of correct responses was collected from each child.

Single-digit subtraction (12 items) Single-digit subtraction was used instead of addition, because subtraction provided a better measure of a child's ability to do arithmetic as answers are not typically memorized as with the addition table. Children saw the subtractions, presented one-by-one in $p-q$ format, and heard them orally. There were 18 questions in total: six -1 subtrahend questions, six small subtrahend questions ($-2, -3, -4$), and six large subtrahend questions ($-6, -7, -8$). The test was stopped if the child had less than three correct answers on the first 8 questions. Item order was changed from post-test A to post-test B. The percentage correct score was calculated for each child.

Participant attrition

The goal of this project was to evaluate the possible benefits in reading acquisition for children using the ELAN software. To measure this with confidence, we had set several requirements for inclusion in the research analysis. The first requirement was that all participants complete all three tests so as to ensure an accurate comparison of progression made, given that all children would undoubtedly progress thanks to normal class training. With each test, we lost a number of children due to absence on the testing day, children leaving schools, or testers forgetting to report id information on the tests. From an original count of 975 participants, 728 children completed all three tests (910, pretest; 867, post-test A and 728 post-test B). The large loss between post-test A and post-test B was due to two classes that left the project and three classes that used incorrect post-test B tests, they had re-used post-test A tests.

We also lost participants due to a software bug that damaged datafiles, causing progression data to be lost and sending many players back to square one in the game. If the child reported to the teacher this anomaly, we were able reinitialize the child's game. Unfortunately, not all children reported losing their progression and this problem was not always detected quickly enough. In order to keep a participant in our analysis, we required a normal game progression, meaning the following: (1) the child had spent sufficient time using the game (the threshold criterion was set at a minimum of 330 min of play) and (2) progression had been unhampered by problems with the software. The number of participants remaining from the intervention groups after these criteria was 350 (113 read/math; 103 math/read).

To examine the baseline balance between the two intervention and control groups, we compared the three groups, two-by-two, in terms of baseline demographic variables and a composite z-score of the four predictive reading acquisition tasks at pre-test (vocabulary, syllable awareness, phoneme awareness, and letter knowledge). An ANOVA of this composite

Table 1, Baseline characteristics at pretest

Mean (SD)	Read/math	Math/read	Control	Group difference
<i>N</i>	113	103	127	
<i>N</i> =Male	61	53	63	$X^2(2) = .46, p = 0.80$
Age in months at pretest	75.8 (3.5)	75.7 (3.8)	75.6 (3.5)	$F(2, 340) = .287, p = 0.75$
Time using ELAN (min)	528 (114)	508 (139)	–	$F(1, 214) = 1.36, p = 0.25$
Vocabulary, $x = 60$	41.06 (5.25)	39.60 (5.57)	40.50 (5.80)	$F(2, 340) = 1.89, p = 0.15$
Syllable suppression (%)	52% (41%)	55% (40%)	48% (38%)	$F(2, 340) = 1.14, p = 0.32$
Phoneme suppression (%)	46% (28%)	42% (29%)	39% (30%)	$F(2, 340) = 1.80, p = 0.17$
Letter knowledge (%)	73% (18%)	74% (19%)	71% (20%)	$F(2, 340) = .979, p = 0.38$
Composite predictor	0.08 (0.68)	0.01 (0.73)	– 0.07 (0.77)	$F(2, 340) = 1.20, p = 0.30$

score revealed no significant difference between the two intervention groups [$F(1, 214) = 0.55, p = 0.46$]. A significant difference was however found between the read/math and control groups [$F(1, 245) = 5.70, p = 0.02$], but not between the math/read and control group [$F(1, 235) = 2.47, p = 0.12$]. Seven children with the lowest composite predictor scores in the control group were removed to ensure equal groups, see Table 1 for baseline information of the three groups.

Statistical analysis

Given our 2-period crossover design, for each task, we first analyzed responses from the children randomized into the read/math and math/read groups in a mixed analysis of variance with factors of intervention group (2 levels: read/math versus math/ read) and test period (3 levels: pretest, post-test A, post-test B), corresponding to the randomized part of the experimental design. We then did a larger mixed analysis of variance with 3 levels for the group factors (the 2 intervention groups plus the control group) and the same test period factor. For the tests where reading was required by the child, to factor out pre-schooling differences amongst children, we added a covariable composite z-score computed from the individual's average z-scores on the four predictive reading acquisition tasks at pretest (vocabulary, syllable awareness, phoneme awareness, and letter knowledge). As shown in

Table 2, this covariable was useful since there were significant correlations between the predictive reading acquisition tests and the reading tests.

In a crossover test, we expected to see a 2-step pattern of improvement: children in the read/math group should improve in reading tests at post-test A relative to the math/read intervention and control groups; reading scores should increase for the math/read group relative to the control group at post-test b. An effect of the training is therefore expressed as a significant interaction between the within-participant factor of test time point and the between-participant factor of group. Given a significant interaction effect, post hoc tests were done at each test period to untangle changes in progress made by each group.

Results

Table 2, Correlations (r^2) between reading predictors at pretest and the two post-tests

	Predictors at pretest			
	Vocabulary	Syllable suppression	phoneme suppression	Letter knowledge
Post-test A				
Syllable suppression	0.07	0.20	0.14	0.08
Phoneme suppression	0.06	0.25	0.37	0.23
Letter knowledge	0.04	0.10	0.16	0.22
Word and pseudo-word reading	0.04	0.21	0.32	0.23
Non-sense text reading	0.05	0.17	0.21	0.14
Sentence comprehension	0.04	0.07	0.11	0.05
Post-test B				
Syllable suppression	0.08	0.14	0.09	0.10
Phoneme suppression	0.07	0.19	0.24	0.17
Letter knowledge	0.08	0.10	0.12	0.20
Word and pseudo-word reading	0.02	0.16	0.19	0.14
Non-sense text reading	0.03	0.11	0.12	0.08
Sentence comprehension	0.06	0.04	0.07	0.04

($df = 341$), r^2 scores superior to 0.03 are significant at $p < 0.001$

Figures 4 shows the evolution of performance across the three test periods for each of the read/math, math/read, and control groups, respectively, for the reading and math tests.

Syllable suppression

The analysis was done on the percentage of correct responses from the 10 items. The intervention group ANOVA revealed, as expected, a significant main effect of test session [$F(2, 428) = 94.8, p < 0.001$], consistent with an improvement over time. No group \times test interaction was found [$F(2, 428)=0.35, p=0.705$]. Similar results were found when including the control group in our analysis: there was a significant main effect of test session [$F(2, 680)=176, p<0.001$], but no group \times test interaction [$F(4, 680) = 1.59, p = 0.176$].

Phoneme suppression

The analysis was done on the percentage of correct responses combining both CV and CCV phoneme suppression tests, including a total of 24 questions. Again, the intervention group ANOVA showed a significant main effect of test session [$F(2, 428)=230, p<0.001$]. All children progressed equally in the task as was made evident by the absence of a group \times test interaction [$F(2, 428) = 0.415, p = 0.66$]. Similar results were found when including the control group in our analysis: there was a significant main effect of test session [$F(2, 680)=390, p<0.001$], but no group \times test interaction [$F(4, 680) = 0.333, p = 0.856$].

Letter knowledge

The percentage correct for the child's combined letter name and letter sound score was collected in this test. The intervention group ANOVA showed a significant main effect of test session, [$F(2, 428)=286, p<0.001$], but no interaction between group and test period [$F(2, 428)=1.16, p=0.316$]. Similar results were found when adding the control group: all three groups made progress over time, as shown by a main effect of test session [$F(2, 680)=441, p<0.001$], in the absence of an interaction [$F(4, 680) = 1.44, p = 0.22$].

One-minute word and pseudo-word reading

We analyzed the mean score for word and pseudo-word reading combined. We were notified by testers that on several occasions the one-minute time limit was not observed. These instances were signaled on the tests and the scores of these children removed from analysis, leaving a smaller number of participants per group (read/math, $n=99$; math/read, $n=94$; control, $n=123$). Both groups in the intervention group analysis improved over the three test periods, as revealed by a main effect of test period [$F(2, 382)=1304, p<0.001$]. An effect of the reading predictor covariate was also significant [$F(1, 190)=81.37, p<0.001$]. The group \times test

interaction approached significance [$F(2, 382)=2.61, p=0.075$]. When adding the control group, significant effects were found for the main effect of test period [$F(2, 626)=2164, p<0.001$], the composite reading predictor [$F(1, 312)=178, p<0.001$], and, most importantly, the group \times test interaction [$F(4, 626)=2.96, p=0.019$]. A post hoc ANOVA at pretest showed no difference between the three groups at pretest [$F(2, 313)=0.451, p=0.637$]. An ANCOVA, using the predictive measures covariable, showed that at post-test A the general progress made by all three groups was no longer equal [$F(2, 312)=4.44, p<0.05$] with the read/math group reading significantly more items per minute than the math/read group [$F(1,190)=7.19, p<0.01$] and the control group [$F(1, 219)=6.89, p<0.01$]. In quantitative terms, after 3 1/2 months of game play, the group that had used ELAN improved by 0.26 standard deviations over the control group, and 0.27 standard deviations over the math software group.

The advantage obtained by the read/math group continued into post-test B [$F(2, 312)=4.61, p=0.01$], with the read/math group continuing to read more items on average than the math/read [$F(1, 190)=4.55, p=0.03$] and the control group [$F(1, 219)=9.60, p<0.01$]. In quantitative terms, at the end of the school year, the group that had used ELAN in the first session, after 4 months of no longer playing ELAN, maintained an advantage of 0.34 standard deviations over the control group, and a 0.27 standard deviation advantage over the math software group.

Non-sense text reading

Children that were unable to read more than four words of the first ten words in the one-minute word reading task were not asked to continue with the more difficult text reading tasks. It was also reported that while some children read more than four words on the one-minute reading task, this was done with great difficulty and the child refused to continue to the more difficult reading tasks. As a consequence, the number of children taking the following reading tests were smaller (pretest: read/math=70, math/read=45, control=64; post-test A: read/math=112, math/read=98, control=121; post-test B: read/math=113, math/read=100, control = 126). Children unable to do these tests were scored at '0' in our analysis. Analyses of the intervention group ANCOVA revealed a significant main effect of improvement across the three tests [$F(2, 428)=745, p<0.001$] and an effect of the composite pretest predictor [$F(1, 213)=42.9, p<0.001$]. No group \times test interaction was found [$F(2, 428)=2.39, p=0.09$]. When including the control group, again, there was a main of effect of test period [$F(2, 680)=1182,$

$p < 0.001$], and a significant effect of the composite pretest predictor [$F(1, 339) = 116, p < 0.001$], but no group \times test interaction [$F(4, 680) = 1.69, p = 0.15$].

Sentence reading comprehension

In this test, children read 8 sentences and, for each, selected the picture depicting their meaning among 4 possible choices. The intervention group ANCOVA showed a significant main effect of test period [$F(2, 428) = 596, p < 0.001$] and a significant effect of the composite pretest predictor [$F(1, 213) = 44.38, p < 0.001$]. Most importantly, a significant group \times test period interaction was also found [$F(2, 428) = 3.75, p = 0.02$]. Post hoc tests at each time point revealed no difference between the two intervention groups at pretest [$F(1, 214) = 0.74, p = 0.39$], but a significant difference at post-test A [$F(1, 214) = 11.49, p < 0.001$] and post-test B [$F(1, 214) = 6.41, p = 0.012$].

Similarly, the ANCOVA including the control group also revealed a significant main effect of test [$F(2, 680) = 880, p < 0.001$], an effect of the composite pre-test predictor [$F(1, 339) = 95.95, p < 0.001$], and a group \times test interaction [$F(4, 680) = 2.85, p = 0.023$]. A post hoc ANOVA at pretest showed no difference between the groups at the start of the year [$F(2, 339) = 0.44, p = 0.65$]. By post-test A, however, the general progress made by all three groups was no longer equal, [$F(2, 339) = 7.92, p < 0.001$] with the read/math group answering more questions correct than either the math/read group [$F(1, 213) = 12.34, p < 0.001$] or the control group [$F(1, 237) = 11.75, p < 0.001$]. No difference was found between the math/read and control groups [$F(1, 227) = 0.03, p = 0.87$]. In quantitative terms, after 3 1/2 months of game play, the group that had used ELAN improved by .39 standard deviations over the control group, and .41 standard deviations over the math software group.

The group differences continued into the post-test B [$F(2, 339) = 3.49, p = 0.032$], with the read/math group continuing to answer more items correct than either the math/read [$F(1, 213) = 6.96, p < 0.01$] or the control group [$F(1, 237) = 3.92, p < 0.05$]. No difference was found between the children that used ELAN in the second period and the control group [$F(1, 227) = 0.55, p = 0.46$]. At the end of the school year, the group that had used ELAN in the first session, after 4 months of no longer playing ELAN, maintained an advantage of 0.22 standard deviations over the control group and a 0.31 standard deviations advantage over the math software group.

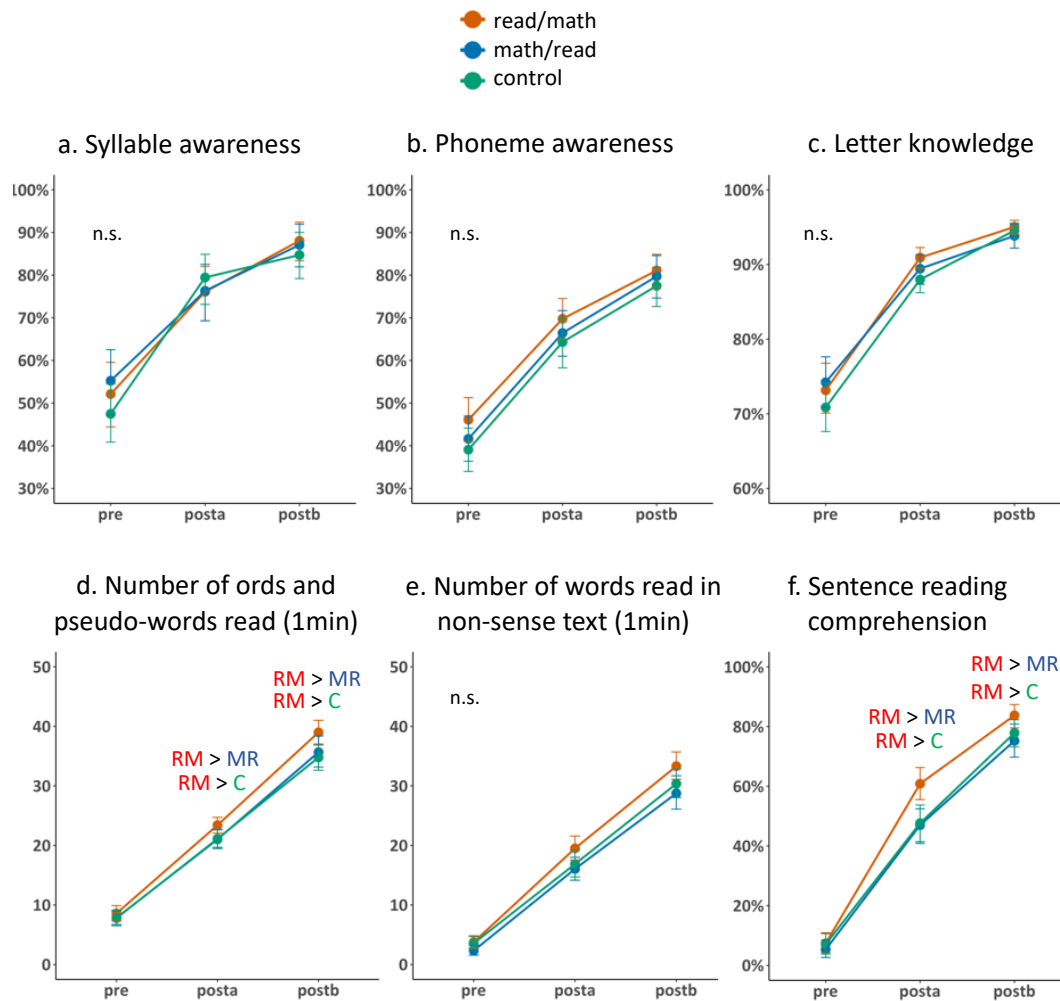


Figure 4 a–f Performance improvements pertaining to reading acquisition. Panels indicate performance in tests of **a** syllable awareness; **b** phoneme awareness (combined score of CV and CVV suppression in pseudo-words); **c** letter knowledge (combined score of letter name and sound knowledge); **d** word and pseudo-word reading (average number of items read in one minute); **e** non-sense text reading (total number of words correctly read in a non-sense text within a 3-minute time limit); **f** sentence comprehension (reading a short sentence and choosing the correct image out of 4 semantically or syntactically related pictures). All tests improved across time. The p-value indicates the significance of the 3 × 3 group × time interaction. Pairwise differences that were significant at $p < 0.05$ are highlighted (RM read/math group; MR math/read group; C control group).

Measuring the evolution of the word length effect

ELAN was developed to help automatize the child’s knowledge of grapheme–phoneme correspondences and word reading. The game pushes the child to respond faster by increasing game difficulty and requiring faster and faster responses. As mentioned in section “Measuring the word length effect”, we used a lexical decision task at the end of each game level to evaluate how reading speed varied with word length.

Our main goal was to measure if the child's word reading speed became progressively less affected by word length over the course of the game, in other words, whether ELAN could be used to measure when students move from slow decoding to parallel word reading. Therefore, we only looked at data from children that had passed at least the first 4 game levels, meaning that these children had learned all the vowels, fricative graphemes, and the most common occlusive consonants. Twenty-one children from the read/math group and 44 children from the math/read group were included in the following analysis. The software measured the time elapsed between the presentation of a target word or pseudo-word and the child's decision. We removed outlier trials with response times below 325 ms (less than 1% of trials) or above a fixed cutoff, defined for a given cell of the design as being 2 standard deviations from the mean of the within-participants medians in that cell (6% of trials). The medians of the remaining correct response times were entered into a mixed ANOVA with participant as the random factor, intervention as the between factor and word length (number of letters, from 2 to 5), test number (test numbers 1 through 4), and lexical status (word or pseudo-word) as within factors. All effects are reported as significant at $p < 0.05$. Degrees of freedom are adjusted to reflect the fact that some cells had missing data for some participants.

The mean of the median response times for each intervention group at each test period and for all lengths is shown in Fig. 5. There was a significant length effect: as expected, response times increased as the number of letters increased from 2 to 5 [$F(3, 175) = 190.21, p < 0.001$]. We also found a main effect of lexical status [$F(1, 51) = 243.81, p < 0.001$] indicating that responses were overall faster for words than for pseudo-words, and a length \times lexical status interaction [$F(3, 180) = 35.39, p < 0.001$], due to the fact that the slope of the length effect was shallower for words than for pseudo-words. Finally, all of these effects were qualified by a length \times test number [$F(9, 557) = 13.06, p < 0.001$] and a triple interaction of length \times test number \times lexical status [$F(9, 543) = 7.12, p < 0.001$]. As seen in Fig. 5, over the four successive tests, the length effect became shallower for words (length \times test number interaction restricted to words only [$F(9, 548) = 21.15, p < 0.001$], whereas this improvement was also present but less pronounced for pseudo-words [$F(9, 562) = 4.79, p < 0.001$]. In agreement with this interpretation, at test number 4, the length effect remained significant for pseudo-words [$F(3, 188) = 36.63, p < 0.001$] but no longer for words [$F(3, 178) = 0.95, n.s.$] (Fig. 5).

All of the above effects were true within both the math/read and the read/math groups. There were no interactions involving this intervention factor. Only a main effect of intervention was observed [$F(1, 51) = 5.52, p = 0.03$], reflecting the fact that the math/read group responded faster, in line with the fact that these children took the tests later in the year than the read/math group.

In summary, the lexical decision task concurred with paper tests in indicating that subjects in group read/math, who started with ELAN software, gained in reading acquisition. Superficially, the absence of any interaction with intervention in the lexical decision task seemed to suggest that the intervention did not have an effect on the evolution of the word length effect, and that the improvements were solely due to time elapsed at school. However, it should be remembered that (for this lexical decision test only) the children in the read/math group were tested within the first half of the year. Thus, their fast progression and the disappearance of the word length effect by test 4 are remarkable and do suggest that the software could have had an effect over and above the improvements due to schooling itself.

To evaluate this, we compared the performance of the read/math group at test 4 with that of the math/read group at test 1. The rationale was that if effects were due solely to time at school, the math/read group should exhibit a better performance on its test 1 (because it was taken later in the year and used grapheme–phoneme correspondences that children would already be very familiar with), whereas if the software intervention was the primary driver of performance, then the converse should hold.

We entered the corresponding median response times into a mixed ANOVA with subject as the random factor, intervention as a 2-level between factor and word length (number of letters, from 2 to 5) and lexical status (word or pseudo-word) as within factors. There were main effects of word length $F(3, 186)=68.72, p<0.001$ and lexical status $F(1, 60)=96.70, p<0.001$. Most importantly, the length \times intervention was significant [$F(3, 186)=3.50, p=0.017$], confirming that, indeed, the read/math group, although tested at an earlier time, had a shallower length effect than the math/ read group. Furthermore, the triple interaction intervention \times length \times lexical status was significant [$F(3, 179) = 6.14, p = 0.001$]. This was due to the fact that, with words, intervention reduced the length effect down to non-significance only for the read/ math group (length \times intervention interaction, restricted to words: [$F(3, 180)=5.46, p=0.001$]), and also reduced the impact of pseudo-word length [$F(3, 188)=2.76, p=0.04$]). In detail, for the read/math group, no effect of word length was found

when reading words on test 4 [$F(3, 53) = .94$, n.s.], while the length remained significant for pseudo-word reading [$F(3, 59)=18.19$, $p<0.001$]. For the math/read group, a significant effect of word length was found at test 1 for words [$F(3, 127)=36.67$, $p < 0.001$] and for pseudo-words $F(3, 129) = 44.36$, $p < 0.001$.

In summary, these results demonstrate that the children who used ELAN in the first session ceased to be affected by word length at an earlier time than children in the math/read group. These results lend support to ELAN as a catalyst to parallel reading.

Number comparison

The intervention group analysis on the number of comparison pairs correctly solved in one

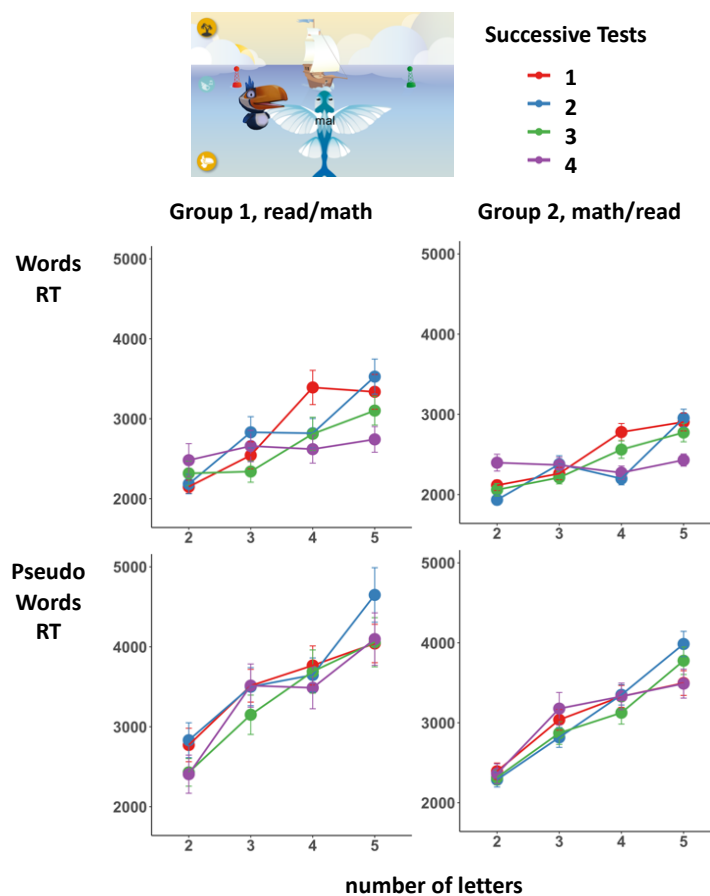


Figure 5, Measuring the speed of reading and its improvement over time using a simple lexical decision game. Passing the lexical decision task was required at the end of each level to proceed to the next. Children had to decide if a word presented was either a real word or an invented word. Thirty-two words were presented: four words of 2, 3, 4, or 5 letters for both words and pseudo-words. All of the words were decodable from the lessons completed by the child. We measured reading response time by word length for both words and pseudo-words of the two different groups. Group 1 played ELAN during intervention A and Group 2 played ELAN during intervention B.

minute revealed a main effect of test period [$F(2, 428)=116$, $p<0.001$], but no significant group \times test interaction [$F(2, 428) = 0.92$, $p = 0.40$]. The 3-group analysis yielded a significant interaction, however: [$F(4, 680) = 3.25$, $p = 0.01$]. Post hoc analysis revealed no difference between the 3 groups at pretest [$F(2, 340)=0.32$, $p=0.73$]. At post-test A, there continued to

be no difference between the two intervention groups [$F(1, 214) = .14, p = 0.71$], but both the read/math and the math/read were superior to the control group [respectively, $F(1, 238)=9.45, p<0.01$; and $F(1, 228) = 12.54, p < 0.001$]. At post-test B, there continued to be no difference between the two intervention groups [$F(1, 214)=0.54, p=0.46$], but the differences relative to the control group remained [respectively, $F(1, 238)=9.28, p<0.001$; and $F(1, 228)=5.79, p=0.02$]. These results point to a possible tablet effect as all children that used the tablets significantly improved over the no-tablet control group (Fig. 6).

Subtraction

The child's percentage of correct responses in subtraction was entered into an intervention group ANOVA, which revealed a significant main effect of the test period [$F(2, 428)=253,$

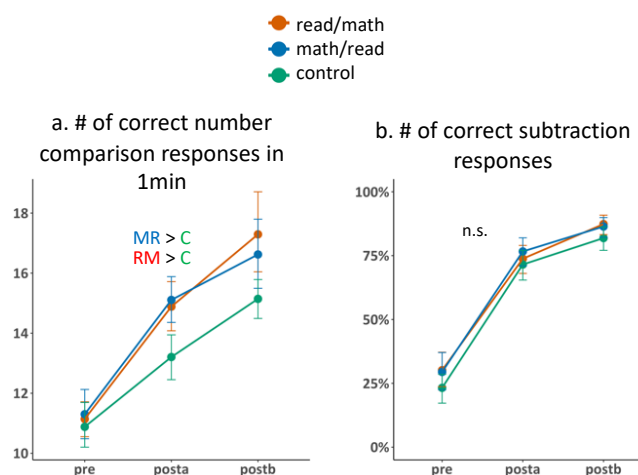


Figure 6, Performance improvements in numerical tests. Two tests were used to measure possible effects of the control game intervention: **a** Symbolic number comparison (% correct in deciding which of two Arabic numerals is larger); **b** Subtraction (% correct). Same format as Fig. 4. All tests improved across time. The p -value indicates the significance of the 3×3 group \times time interaction. Pairwise differences that were significant at $p < 0.05$ are highlighted (RM read/math group, MR math/read group, C control group).

$p<0.001$], but no group \times test interaction [$F(2, 428)=0.314, p=0.73$]. In the 3-group analysis, the main effect of test was significant [$F(2, 680)=417, p < 0.001$], but again, no group test interaction was found [$F(4, 680) = 0.27, p = 0.90$].

Discussion

Summary of results

The progress of the read/math group at post-test A, which continued to post-test B on the one-minute reading test and sentence comprehension, coupled with the evidence from the lexical decision task for a reduced length effect for this group gives strong support for ELAN as

a tool to aiding classroom reading instruction and accelerating first-grade reading acquisition. The observed effect sizes of 0.26–0.41 are considered as small to medium, but more importantly, the benefits seem to extend four months beyond the training period. Interestingly, these results pertain to those children that only used the software in the initial learning stages of reading. Given the short duration of the training period (read/math $x = 8.8$ h, SD 1.9; math/read $x = 8.5$ h, SD 2.3), it is likely that improved results would be obtained with prolonged and/or more intense exposure.

Our results are consistent with previous studies showing that early explicit systematic phonics provides the fastest route to reading acquisition (Cunningham 2001; Ehri et al. 2001). This study also gives strength to a recent observational report of 2500 first-grade students in France, which showed that reading ability improved faster in classes where a larger number of grapheme–phoneme correspondences were taught early in the year (Goigoux, 2016). Phonics methods are often criticized for not supporting comprehension due to the limiting constraints of word choice during the early lessons in methods that strictly adhere to 100% decodable text. In this context, however, it is particularly important to emphasize that sentence reading comprehension improved in the read/math group relative to both active and no-tablet control groups, even though our software was entirely focused on grapheme–phoneme teaching and 100% decoding practice. While we cannot disentangle the benefits of using ELAN from other instructional materials used by the participating teachers, this aspect of our results would seem to support the early use of phonics to help children improve access to comprehension. It is compatible with the *Simple View of Reading* formula that defines reading comprehension as the product of decoding and oral comprehension (Gough and Tunmer 1986; Lervåg et al. 2018). In other words, once a text is decoded, the reader uses the same mechanisms that he or she would apply to hearing language. By facilitating a faster automatization of decoding skills, the reader can focus on what they have read. Early intense practice of phonics may be key to alleviating the difficulty of the phonological path of slow decoding so that the child can quickly access the lexical path of fluent reading for understanding. The goal for developing ELAN was, in fact, that the software help automatize the rote task of grapheme–phoneme correspondence memorization and allow teachers to focus on developing oral comprehension and reading aloud.

While it is disappointing that the NumberCatcher game did not have a more specific influence on number comparison and had no detectable effect on subtraction abilities, the results

suggest instead that tablet software itself may have a generic effect on speeded decision tasks, as both tablet groups improved in the number comparison test. Numerous studies on ‘first-person shooter games,’ video games that require the player to make combative actions from a first-person perspective, have revealed improvements in a number of cognitive functions, including reaction time to ‘go signals’ (Colzato et al. 2013) and use of attentional resources in simple tasks of flanker compatibility and spatial localization (Green and Bavelier 2003, 2006). While neither ELAN nor Number Catcher would be classified as ‘action games,’ they both require players to make faster and faster actions in environments with many distractors. Thus, while our isolated result should be taken with caution, given that the benefit in number comparison was only obtained relative to a non-randomized no-tablet control group, future study of more general cognitive benefits of educational games that require timed responses in noisy environments is warranted.

Limitations

The two main limitations to this project can be attributed to the software design. Nine hundred and seventy-five children were announced as participating in the study. Of the 728 students that completed the three tests, minor software bugs complicated the large-scale deployment of the software to such an extent that only 47% of them experienced a game intervention adapted to their progress and level. While it is virtually impossible to create bug-free software without broad field testing, improved piloting was warranted in this case. Children were sometimes stymied from advancing in the game due to poor ergonomics that were not discovered until a close examination of the children’s game progression.

A second limitation to this study concerns the absence of results with students that used ELAN during intervention B. ELAN adapts to the child’s response rate but requires completion of all lessons to move through the game. It is highly likely, given the improved reading scores for all children at post-test A, that for the math/ read group, the grapheme–phoneme correspondences presented during the second session were mainly a review of already well-learned material in the class. Ideally, ELAN should use an adaptive mechanism that not only provides tailored game difficulty but adjusts more quickly in order to specifically train the grapheme–phoneme correspondences that the child has not learned or needs to practice.

A third limitation to this study is that we measured the benefits of autonomous use of ELAN without taking into account other factors known to influence reading acquisition, such as socio-economic status on academic achievement (for a review of the effects of SES on academic achievement: (Sirin 2005) and the effects of materials and methods already in use by the participating teachers (Braibant and Gerard 1996). The benefits of ELAN might have been partially dependent on the activities already in place by the teacher. For example, it is unknown to us which teachers were already using systematic phonics methods and how many grapheme–phoneme correspondences were taught in the class. If ELAN were to be used in the classroom, it would be important to know if the software itself is sufficient in phonics teaching or if it should be used to provide repetition of the teacher’s lesson.

Conclusion

In a recent review on the science of learning to read, it has been proposed that the “reading wars” debate between whole-word methods and phonics needs to end (Castles et al. 2018). Enough evidence has been accumulated to support explicit systematic phonics and our focus should be on developing guiding principles for “balanced instruction” that acknowledge the complexities of learning that underlie the individual child’s progress from decoder to reading expert. This requires combining an understanding of the biological and cognitive aspects of learning to read, respecting the evidence-based practices already known and furthering research on the bricks of best practice that build a reading method from start to finish. In the case of this project, it appears that intense and focused learning of decoding skills may be more effective when taught early in learning. We also demonstrate that software that focuses on the repetition of decoding skills, and that adapts to the child’s level, can indeed lead to beneficial effects for young readers. More importantly, this work builds onto the evidence that addressing ‘which’ strategies and ‘when’ should now be the focus of reading research.

Appendix

Order of grapheme–phonemes taught in ELAN. Phoneme code used from the MANULEX (Lété et al., 2004).

Lesson number	Grapheme orthography	Phoneme code	Lesson number	Grapheme orthography	Phoneme code (Manulex)
1	A, a	a	31	et, est	e,E
2	E, é	e	32	ent	silent
3	I, i	i	33	an, am	@
4	O, o	o	34	en, em	@
5	U, u	y	35	on, om	§
6	E, e	*	36	in, im, ym	5
7	Y, y	i	37	à, â	a
8	L, l	l	38	au, eau	O
9	M, m	m	39	è, ê, ë	E
10	S, s	s	40	oeu	9
11	R, r	R	41	gn	N
12	Ou	u	42	eConsonantConsonant	E
13	silent letters (e, s, t, x)	silent	43	ai, ei	E
14	J, j	Z	44	ez er	e
15	F, f	f	45	et, fin d'un mot	E
16	V, v	v	46	Hh	silent
17	eu	2,9	47	ph	f
18	Z, z	z	48	ç, ce, ci, cy	s
19	ch	S	49	ge, gi, gy	Z
20	Nn, un	n, l	50	Xx	ks
21	T, t	t	51	silent end letters (all)	silent
22	D, d	d	52	ain, aim	5
23	C, c	k	53	ien	j5
24	P, p	p	54	eConsonant	E
25	Qq, qu	k	55	ti	s
26	B, b	b	56	Yy	j
27	Gg, gu	g	57	il, ill	ij
28	K, k	k	58	il, ill	j
29	s	z	59	ch	k
30	es	E	60	W, w	w

References

- Bara, F., & Gentaz, E. (2011). Haptics in teaching handwriting: The role of perceptual and visuo-motor skills. *Human Movement Science, 30*(4), 745–759. <https://doi.org/10.1016/j.humov.2010.05.015>.
- Bara, F., Gentaz, E., & Colé, P. (2007). Haptics in learning to read with children from low socio-economic status families. *British Journal of Developmental Psychology, 25*(4), 643–663. <https://doi.org/10.1348/026151007X186643>.
- Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development, 19*(3), 433–449. <https://doi.org/10.1016/j.cogdev.2004.05.003>.
- Bara, F., Morin, M.-F., Alamargot, D., & Bosse, M.-L. (2016). Learning different allographs through handwriting: The impact on letter knowledge and reading acquisition. *Learning and Individual Differences, 45*, 88–94. <https://doi.org/10.1016/j.lindif.2015.11.020>.
- Bertrand, D., Fluss, J., Billard, C., & Ziegler, J. C. (2010). Efficacité, sensibilité, spécificité: Comparaison de différents tests de lecture, Abstract. *L'Année Psychologique, 110*(2), 299–320. <https://doi.org/10.4074/S000350331000206X>.
- Boles, D. B., & Clifford, J. E. (1989). An upper and lowercase alphabetic similarity matrix, with derived generation similarity values. *Behavior Research Methods, Instruments, & Computers, 21*(6), 579–586. <https://doi.org/10.3758/BF03210580>.
- Boyer, N., & Ehri, L. C. (2011). Contribution of phonemic segmentation instruction with letters and articulation pictures to word reading and spelling in beginners. *Scientific Studies of Reading, 15*(5), 440–470. <https://doi.org/10.1080/10888438.2010.520778>.
- Braibant, J. M., & Gerard, F.-M. (1996). Savoir lire: Question(s) de méthodes? *Bulletin de psychologie scolaire et d'orientation, 7*-45.
- Brankaer, C., Ghesquière, P., & Smedt, B. D. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods, 49*(4), 1361–1373. <https://doi.org/10.3758/s13428-016-0792-3>.

- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., et al. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences United States of America*, *107*(17), 7939–7944. <https://doi.org/10.1073/pnas.0904402107>.
- Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*(3), 369–378. <https://doi.org/10.1007/s10648-012-9205-z>.
- Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic awareness instruction: Contribution of articulatory segmentation to novice beginners' reading and spelling. *Scientific Studies of Reading*, *7*(1), 25–52. https://doi.org/10.1207/S1532799XSSR0701_03.
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. <https://doi.org/10.1177/1529100618772271>.
- Cavalli, E., Duncan, L. G., Elbro, C., Ahmadi, A. E., & Colé, P. (2017). Phonemic—Morphemic dissociation in university students with dyslexia: An index of reading compensation? *Annals of Dyslexia*, *67*(1), 63–84. <https://doi.org/10.1007/s11881-016-0138-y>.
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>.
- Coltheart, M. (2005). Modeling reading: The dual-route approach. In *The science of reading: A handbook* (pp. 6–23).
- Colzato, L. S., van den Wildenberg, W. P. M., Zmigrod, S., & Hommel, B. (2013). Action video gaming and cognitive control: Playing first person shooter games is associated with improvement in working memory but not action inhibition. *Psychological Research Psychologische Forschung*, *77*(2), 234–239. <https://doi.org/10.1007/s00426-012-0415-2>.
- Cunningham, J. W. (2001). The National Reading Panel Report. *Reading Research Quarterly*, *36*(3), 326–335. <https://doi.org/10.1598/RRQ.36.3.5>.
- de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of systematic phonics instruction. *Scientific Studies of Reading*, *13*(4), 318–333. <https://doi.org/10.1080/10888430903001308>.

- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., et al. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–1364. <https://doi.org/10.1126/science.1194140>.
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, *16*(3), e2004103. <https://doi.org/10.1371/journal.pbio.2004103>.
- Deltour, J.-J., & Hupkens, D. (1990). *Test de vocabulaire actif et passif pour enfants de 3 à 5 ans—TVAP*. Issy-les-Moulineaux: Ed. scientifiques et psychologiques.
- Dye, M. W. G., Green, C. S., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, *18*(6), 321–326. <https://doi.org/10.1111/j.1467-8721.2009.01660.x>.
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the national reading panel's meta-analysis. *Review of Educational Research*, *71*(3), 393–447. <https://doi.org/10.3102/00346543071003393>.
- Gentaz, E., Sprenger-Charolles, L., & Theurel, A. (2015). Differences in the predictors of reading comprehension in first graders from low socio-economic status families with either good or poor decoding skills. *PLoS ONE*, *10*(3), e0119581. <https://doi.org/10.1371/journal.pone.0119581>.
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. <https://doi.org/10.1177/074193258600700104>.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. <https://doi.org/10.1038/nature01647>.
- Green, C. S., & Bavelier, D. (2006). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1465–1478. <https://doi.org/10.1037/0096-1523.32.6.1465>.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. <https://doi.org/10.1007/BF00401799>.
- Kyle, F., Kujala, J., Richardson, U., Lyytinen, H., & Goswami, U. (2013). Assessing the effectiveness of two theoretically motivated computer-assisted reading

- interventions in the United Kingdom: GG rime and GG phoneme. *Reading Research Quarterly*, 48(1), 61–76. <https://doi.org/10.1002/rrq.038>.
- Lecocq, P. (1996). *L'ECosSe: Une épreuve de compréhension syntaxico-sémantique*. Lille: Presses Universitaires du Septentrion.
- Lefavrais, P. (2005). *Alouette-R*. Paris: Test d'analyse de la lecture et de la dyslexie.
- Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 89(5), 1821–1838. <https://doi.org/10.1111/cdev.12861>.
- 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. <https://doi.org/10.3758/BF03195560>.
- Longcamp, M., Zerbato-Poudou, M.-T., & Velay, J.-L. (2005). The influence of writing practice on letter recognition in preschool children: A comparison between handwriting and typing. *Acta Psychologica*, 119(1), 67–79. <https://doi.org/10.1016/j.actpsy.2004.10.019>.
- Lyytinen, H., Ronimus, M., Alanko, A., Poikkeus, A.-M., & Taanila, M. (2007). Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. *Nordic Psychology*, 59(2), 109–126. <https://doi.org/10.1027/1901-2276.59.2.109>.
- Mason, G., McDaniel, H., & Callaway, B. (1974). Relating reading and spelling: A comparison of methods. *The Elementary School Journal*, 74(6), 381–386. <https://doi.org/10.1086/460844>.
- McCandliss, B., Beck, I. L., Sandak, R., & Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the word building intervention. *Scientific Studies of Reading*, 7(1), 75–104. https://doi.org/10.1207/S1532799XSSR0701_05.
- Meyer, J., Dentel, L., & Meunier, F. (2013). Speech recognition in natural background noise. *PLoS ONE*, 8(11), e79279. <https://doi.org/10.1371/journal.pone.0079279>.
- Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language*, 127(3), 356–365. <https://doi.org/10.1016/j.bandl.2013.10.009>.
- Neuman, S. B., & Roskos, K. A. (1998). *Children achieving: Best practices in early literacy*. Newark, DE: Order Department, International Reading Association.

- New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project. *Psychonomic Bulletin & Review*, *13*(1), 45–52. <https://doi.org/10.3758/BF03193811>.
- OECD. (2015). *Students, computers and learning: Making the connection*. PISA. Paris: OECD Publishing. <https://doi.org/10.1787/9789264239555-en>.
- Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., et al. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*(1), 91–96. <https://doi.org/10.1038/71163>.
- Pourcin, L., Sprenger-Charolles, L., El Ahmadi, A., & Colé, P. (2016). Reading and related skills in Grades 6, 7, 8 and 9: French normative data from EVALEC. *Revue Européenne de Psychologie Appliquée/European Review of Applied Psychology*, *66*(1), 23–37. <https://doi.org/10.1016/j.erap.2015.11.002>.
- Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., & Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognitive Development*, *24*(4), 450–472. <https://doi.org/10.1016/j.cogdev.2009.09.003>.
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame method: The theoretical and methodological background of the technology-enhanced learning environment for learning to read. *Human Technology*, *10*(1), 39–60. <https://doi.org/10.17011/ht/urn.201405281859>.
- Goigoux, R. (2016). Apprendre à lire et à écrire au cours préparatoire: enseignements d'une recherche collective. *Revue Française de pédagogie*, (3), 5-6.
- Saine, N. L., Lerkkanen, M.-K., Ahonen, T., Tolvanen, A., & Lyytinen, H. (2011). Computer-assisted remedial reading intervention for school beginners at risk for reading disability: Computer-assisted reading intervention. *Child Development*, *82*(3), 1013–1028. <https://doi.org/10.1111/j.1467-8624.2011.01580.x>.
- Schacter, J., & Jo, B. (2016). Improving low-income preschoolers mathematics achievement with Math Shelf, a preschool tablet computer curriculum. *Computers in Human Behavior*, *55*, 223–229. <https://doi.org/10.1016/j.chb.2015.09.013>.
- Serrano, F., Genard, N., Sucena, A., Defior, S., Alegria, J., Mousty, P., et al. (2011). Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. *Journal of Portuguese Linguistics*, *10*(1), 183–204. <https://doi.org/10.5334/jpl.106>.

- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417–453. <https://doi.org/10.3102/00346543075003417>.
- 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 edited by FAPSE-UNIGE.
- Tallal, P. (2004). Improving language and literacy is a matter of time. *Nature Reviews Neuroscience*, 5(9), 721–728. <https://doi.org/10.1038/nrn1499>.
- Uhry, J. K., & Shepherd, M. J. (1993). Segmentation/spelling instruction as part of a first-grade reading program: Effects on several measures of reading. *Reading Research Quarterly*, 28(3), 219–233. <https://doi.org/10.2307/747995>.
- Vousden, J. I., Ellefson, M. R., Solity, J., & Chater, N. (2011). Simplifying reading: Applying the simplicity principle to reading. *Cognitive Science*, 35(1), 34–78. <https://doi.org/10.1111/j.1551-6709.2010.01134.x>.
- Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., & Cohen, D. (2006a). Principles underlying the design of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2(1), 19. <https://doi.org/10.1186/1744-9081-2-19>.
- Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006b). An open trial assessment of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, 2(1), 20. <https://doi.org/10.1186/1744-9081-2-20>.
- Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain & Education*, 3(4), 224–234. <https://doi.org/10.1111/j.1751-228X.2009.01075.x>.
- Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Fásca, L., et al. (2010). Orthographic depth and its impact on universal predictors of reading: A cross-language investigation. *Psychological Science*, 21(4), 551–559. <https://doi.org/10.1177/0956797610363406>.

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., & Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and Language*, 93(3), 369–373. <https://doi.org/10.1016/j.bandl.2004.10.010>.

Chapter 3, Automatic construction of a phonics curriculum for reading education using the Transformer neural network

Potier Watkins, C., Dehaene, O., & Dehaene, S. (2019). Automatic Construction of a Phonics Curriculum for Reading Education Using the Transformer Neural Network. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Éds.), *Artificial Intelligence in Education* (p. 226-231). Springer International Publishing. https://doi.org/10.1007/978-3-030-23207-8_42

Abstract

Key to effective phonics instruction is the teaching of grapheme-phoneme (GP) correspondences in a systematic progression that starts with the most frequent and consistent pronunciation rules. However, discovering the relevant rules is not an easy task and usually requires subjective analysis by a native speaker and/or expert linguist. We describe GPA4.0, a submodule to the Transformer neural network model that automatizes the task of grapheme-to-phoneme (g2p) transcription and alignment. The network is trained with four different languages of decreasing orthographic transparency (Spanish<Portuguese<French<English). Our results show that the Transformer model improves on the current state-of-the-art in g2p transcription and that the attention mechanism allows for the alignment of graphemes to their corresponding phonemes. From the g2p aligned words, our software provides an optimally ordered phonics progression based on frequency and consistency in the target language, as well as an ordered list of words that teachers can use. This work exemplifies a practical way that neural networks can be used to develop educational materials for research and teachers. Submodules and phonics output are available at, <https://github.com/OlivierDehaene/GPA4.0>.

Introduction

Early phonics introduction is endorsed as the foundation of successful reading instruction in both education research (meta-analysis by the National Reading Panel, 2000, Castles et al., 2018) and cognitive neuroscience (Brem et al., 2010; Dehaene et al., 2010). However, phonics instruction is not universally used. One factor for its relative disaffection could be that knowing what grapheme-phoneme (GP) correspondences to teach, and in what order to teach them, can be a difficult task, given that letter-sound relationships do not all have a one-to-one relationship. Take for example Spanish, a highly *transparent* language, meaning that a given letter is nearly always pronounced the same. In stark contrast is English, which can have many different sounds for a single grapheme (e.g. the ‘a’ in ‘cat’, ‘mate’, ‘what’ or ‘about’). Cross-language research demonstrates that orthographic transparency influences the time and difficulty children have in learning to read (Goswami et al., 1998; Landerl, 2000; Serrano et al., 2011; Seymour et al., 2003; Ziegler et al., 2010).

Orthographic transparency is also a conundrum in neural network text-to-speech applications that rely on grapheme-to-phoneme (g2p) transcription. G2p refers to converting words to their phonemes. The current state-of-the-art applies long short-term memory (LSTM) networks and recurrent neural networks using sequence-to-sequence (seq2seq) modeling combined with an attention-mechanism (Toshniwal & Livescu, 2016). More recently, the Transformer model has brought notable improvements in neural machine language transcription and language parsing (Vaswani et al., 2017). These tasks that are fairly analogous to g2p transcription (both depend on long range dependencies and contextual influences). Improvement made by the Transformer model is in part due to parallel position encoding that curtails the need for recurrence and a self-attention field that enables the concatenation of information between sequences, regardless of their distance. The goal of the current project, GPA4.0 (Figure 1), is to test for g2p transcription improvements, for the first time to our knowledge, using the Transformer model. With this achieved, we take advantage of the Transformer’s attention mechanism to align grapheme input to phoneme output, thus

permitting the construction of a phonics progression based on the frequency and consistency of all found GP correspondences for any alphabetic language word list.

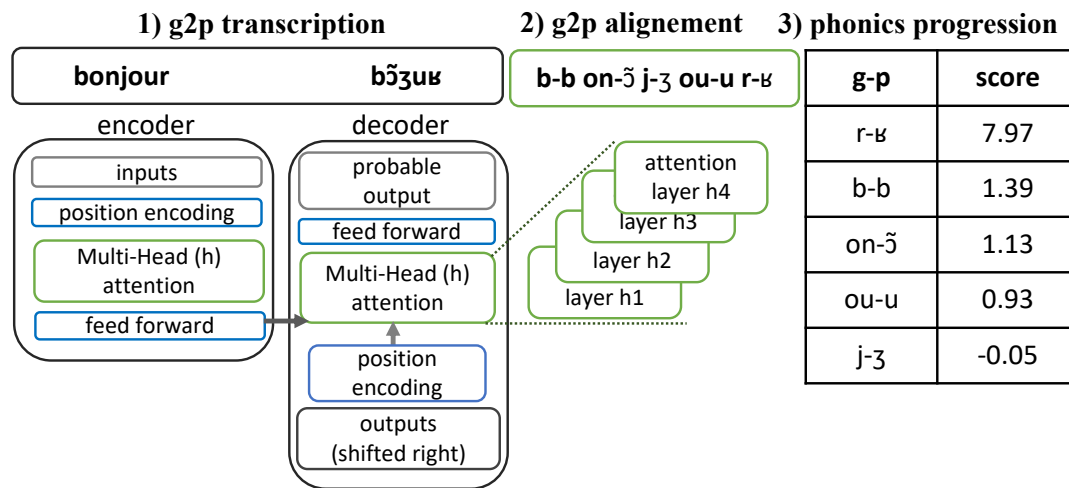


Figure 1, GPA4.0 steps to constructing a phonics progression. 1) g2p transcription is done using the Transformer neural network. 2) g2p alignment uses attention weights to align the 'grapheme inputs' to their 'phoneme outputs'. 3) a phonics progression is built by according each g2p alignment an aggregated z-score based on frequency and consistency in the word corpus.

Experiment

We tested the Transformer model for improved g2p transcription compared to the current-state-of-the-art results (Rao et al., 2015; Toshniwal & Livescu, 2016) on the CMUDict database (Weid, 1998) while also comparing, for the first time to our knowledge, the results of five different languages of varying orthographic transparency: Spanish<Portuguese<French<English. Training was done using one 1080TI NVIDIA GPU on the base models for a total of 10,000 steps. We use Tensor2Tensor (T2T) (Vaswani et al., 2018) an open-source system for training deep learning models in TensorFlow (Abadi et al., 2015). G2p alignment in our model is made possible using the attention weights of the Transformer model. G2p alignment accuracy was analyzed in French, the only language for which we had a reference for comparison. Table 1 describes the word lists used and provides the minor adjustments made to accommodate the small amount of training data. Training was conducted on 80% of the data. The model's performance was tested on the remaining 20% of data.

Table 1, Language wordlists used and adjustments made to the Transformer architecture

language	number of words used for training	number of words used for testing	number of hidden layers	3
Spanish (Corral et al., 2009)	10,400	2,600	hidden size, number of neurons per layer	256
Portuguese (Derived Corpora and Counts)	31,200	7,800	filter size	512
French (Lété et al., 2004)	8,000	2,000	h, number of attention heads	4
English (Masterson et al., 2010)	8,000	2,000	attention dropout rate	0.2
English (Weid, 1998)	95,069	23,767	dropout rate	0.3

To generate a language’s phonics progression, we extract all the GP correspondences in the list of g2p aligned words. For each GP correspondence found, we measure its frequency, g2p consistency and phoneme-to-grapheme consistency. The GP correspondences are then sorted by an aggregate weight of the prementioned measures’ z-scores (we apply weights of 0.7, 0.25 and 0.05 respectively, but these can be adjusted in the code). The weights are designed to 1) give priority to the most frequent GP correspondences when a pair is particularly consistent and less frequent but highly consistent correspondences.

Results

G2p Transcription

The standard measures of word error rate (WER and phoneme error rate (PER) are reported in Table 2. WER is the total number of output errors in which there is at least one phoneme error / total number of words. PER is the Levenshtein distance (Levenshtein, 1966) (the minimum number of single-character edits needed to change one word to the other) of the predicted phoneme sequence to the reference from the original database / the number of phonemes in the reference. Language WER and PER scores reflect, as expected, decreasing

Table 2, Word error rate (WER) and Phoneme error rate (PER) in four languages of decreasing orthographic transparency

	Spanish<	Portuguese<	French<	English	CMUDict
WER	0.38%	2.77%	3.18%	15.04%	20.87%
PER	0.07%	0.55%	0.89%	4.50%	4.59%
<i>Previous best results using the CMUDict database:</i>					WER=21.69% PER=5.04%

orthographic transparency. We report a slight gain over Toshniwal and Livescu’s best prior score on the CMUdict database.

G2p alignment

Figure 2 provides an example of encoder-decoder attention (taken from layer-4 multi-head attention in the decoder, see Figure 1). As the network reads the word “bonjour” or “banane”, it attends to distant information required to know if a vowel followed by the letter ‘n’ will make a single nasal sound (e.g. ‘on’) or two distinct phonemes (e.g. a+ n). GPA4.0 aligns

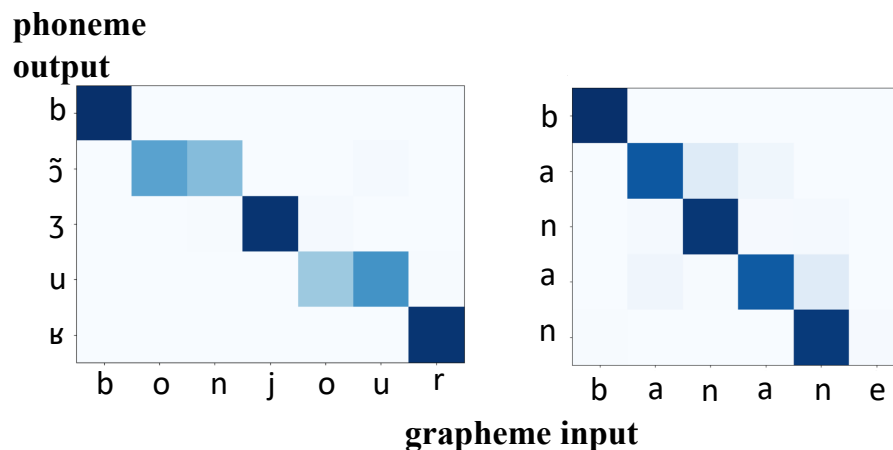


Figure 2, Encoder-decoder attention in g2p transcription

graphemes to phonemes based on the attention carrying the most weight. G2p alignment error rate was assessed for French using the sequence error rate, a correct or incorrect score for each word and the g2p alignment error rate (Levenshtein distance (Levenshtein, 1966)). We report scores of 27.76% and 10.20% respectively. The relatively high sequence error rate compared to the low g2p alignment score is due to the difficulty in parsing silent letters not coded in the phonology of the trained wordlist. 56% of words in the list contain silent letters.

Conclusion

Our results demonstrate improved g2p transcription by the Transformer model. Our submodule, GPA4.0, takes a novel approach to developing applicable phonics tools for the classroom by taking advantage of neural network performance in g2p transcription and, in particular, the attention field for g2p alignment. This work highlights the difficulties for neural networks to learn the GP correspondences in decreasingly transparent languages. The phonics progressions for the four languages analyzed and their ordered wordlists are freely available. These datafiles can be used as a ‘paper’ support to guide reading instruction, or as stimuli for game-based reading applications (e.g. the GraphoGame software Brem et al., 2010;

Richardson & Lyytinen, 2014). We hope that the GPA4.0 submodule will be taken up as a tool for researchers and educators to generate their own phonics lessons with 100% decodable reading materials. GPA4.0 combines cognitive science and neural network technology for evidence-based reading education. Phonics progressions and word lists for the four different languages analyzed in this paper, as well as the submodule code, can be downloaded at <https://github.com/OlivierDehaene/GPA4.0>.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., ... Zheng, X. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Software Available from Tensorflow.Org, 19.
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17), 7939–7944. <https://doi.org/10.1073/pnas.0904402107>
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5–51. <https://doi.org/10.1177/1529100618772271>
- Corral, S., Ferrero, M., & Goikoetxea, E. (2009). LEXIN: A lexical database from Spanish kindergarten and first-grade readers. *Behavior Research Methods*, 41(4), 1009–1017. <https://doi.org/10.3758/BRM.41.4.1009>
- Cunningham, J. W. (2001). The National Reading Panel Report. *Reading Research Quarterly*, 36(3), 326–335. <https://doi.org/10.1598/RRQ.36.3.5>
- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359–1364. <https://doi.org/10.1126/science.1194140>
- Derived Corpora and Counts. (n.d.). Retrieved February 10, 2019, from <https://childes.talkbank.org/derived/>
- Goswami, U., Gombert, J. E., & Barrera, L. F. de. (1998). Children’s orthographic representations and linguistic transparency: Nonsense word reading in English, French, and Spanish. *Applied Psycholinguistics*, 19(1), 19–52. <https://doi.org/10.1017/S0142716400010560>
- Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education*, 15(3), 239. <https://doi.org/10.1007/BF03173177>

- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX: A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 36(1), 156–166.
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- Masterson, J., Stuart, M., Dixon, M., & Lovejoy, S. (2010). Children’s printed word database: Continuities and changes over time in children’s early reading vocabulary. *British Journal of Psychology*, 101(2), 221–242. <https://doi.org/10.1348/000712608X371744>
- National Reading Panel. (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). National Institute of Child Health & Development. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- Potier Watkins, C., 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. <https://doi.org/10.1007/s40692-019-00152-6>
- Rao, K., Peng, F., Sak, H., & Beaufays, F. (2015). Grapheme-to-phoneme conversion using Long Short-Term Memory recurrent neural networks. 4225–4229. <https://doi.org/10.1109/ICASSP.2015.7178767>
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method: The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. *Human Technology*, 10(1). <https://doi.org/10.17011/ht/urn.201405281859>
- Serrano, F., Genard, N., Sucena, A., Defior, S., Alegria, J., Mousty, P., Leybaert, J., Castro, S. L., & Seymour, P. H. K. (2011). Variations in reading and spelling acquisition in Portuguese, French and Spanish: A cross-linguistic comparison. *Journal of Portuguese Linguistics*, 10(1), 183–204. <https://doi.org/10.5334/jpl.106>

- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94(2), 143–174. <https://doi.org/10.1348/000712603321661859>
- Toshniwal, S., & Livescu, K. (2016). Jointly Learning to Align and Convert Graphemes to Phonemes with Neural Attention Models. ArXiv:1610.06540 [Cs]. <http://arxiv.org/abs/1610.06540>
- Vaswani, A., Bengio, S., Brevdo, E., Chollet, F., Gomez, A. N., Gouws, S., Jones, L., Kaiser, Ł., Kalchbrenner, N., Parmar, N., Sepassi, R., Shazeer, N., & Uszkoreit, J. (2018). Tensor2Tensor for Neural Machine Translation. ArXiv:1803.07416 [Cs, Stat]. <http://arxiv.org/abs/1803.07416>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. ArXiv:1706.03762 [Cs]. <http://arxiv.org/abs/1706.03762>
- Weid, R. L. (1998). The CMU Pronouncing Dictionary. <http://www.speech.cs.cmu.edu/cgi-bin/cmudict>
- Ziegler, J. C., Bertrand, D., Tóth, D., Csépe, V., Reis, A., Faísca, L., Saine, N., Lyytinen, H., Vaessen, A., & Blomert, L. (2010). Orthographic Depth and Its Impact on Universal Predictors of Reading: A Cross-Language Investigation. *Psychological Science*, 21(4), 551–559. <https://doi.org/10.1177/0956797610363406>

Chapter 4, A cognitive science based phonics application in kindergarten: Boosting pre-reading skills, but not school reading gains

Potier Watkins, C. & Dehaene, S. A cognitive science-based phonics application in kindergarten: Boosting pre-reading skills, but not school reading gains. In preparation

Abstract

The Kalulu software is a tablet-based suite of phonics and reading-related lessons and minigames. In a previous intervention with a previous version of the software in 1st grade students, fluency and comprehension were boosted, but only when used in concert with reading instruction at the start of the year. Here, we asked whether a similar intervention would be more efficient if it started a year earlier, in kindergarten. Forty classes were randomized into playing Kalulu phonics or an active matched control game (Kalulu numbers) for the first half of the year. Those assignments were reversed in the second half of the year. Ten non-randomized business-as-usual classes also participated (N=1092). In a cross-over effect, children who used the phonics version improved in letter naming, grapheme-phoneme matching and reading fluency, while those with the number version improved in number knowledge. In a longitudinal follow-up, intervention participants maintained an advantage in phoneme awareness at the start of 1st grade, but this advantage failed to translate into school literacy gains in the middle of 1st grade, and no longitudinal benefits were found for numbers. Those results improve our understanding of when and for how long to introduce phonics and question the possibility that a short-term intervention may address the complex challenges of long-term educational goals.

Introduction

As one of the few examples of successful applications of cognitive science to classrooms, phonics instruction has emerged from the “reading wars” as the clear winner in producing the best literacy gains for all children (Anne Castles et al., 2018). This finding is supported by randomized control trials and longitudinal follow-up assessments (for national reports including meta-analyses, see National Reading Panel, 2000 from the United States of America and Rose, 2006 from Great Britain). It converges with functional brain imaging studies which show how a standard reading circuit that involves both grapheme and phoneme processing nodes gets established early on during reading acquisition, and can be particularly boosted by phonics instruction (e.g. Brem et al., 2010; Turkeltaub et al., 2003). For example, studies comparing children before and after they learn to read highlight activation in a part of the left occipito-temporal pathway dedicated to visual recognition that becomes sensitive to letter strings and develops increasingly efficient connections to regions specialized in processing speech sounds such as the planum temporale (Dehaene et al., 2015; Dehaene-Lambertz et al., 2018a; Monzalvo & Dehaene-Lambertz, 2013).

Today, the results of this collaboration between science and education have come full circle, as phonics becomes more and more prevalent in literacy curriculum. Phonics mastery has become central to standardized testing in several different countries (e.g., the Common Core evaluations in the United States; the Phonics Check in the United Kingdom; the EvalAide program in France). The question now is not so much if or why this method works, but what are the best practices for success. Several tenets have emerged. First, cracking the alphabetic code is a cultural learning criterion that must be made explicit, meaning that the grapheme-phoneme correspondences of a language should be taught. Preliterate children exposed to print, and merely instructed to ‘recognize’ words such as *fat* and *bat* fail to transfer their learning when shown the new word, such as *fun*, and asked if the word is “fun” or “bun” (Byrne, 1992). In other words, they do not implicitly learn the letter-sound combinations for transfer to new stimuli. Second, explicit phonics instruction should also follow a systematic progression (Graaff et al., 2009), in which grapheme-phoneme correspondences are introduced in a rational order taking into account their frequency and consistency in a given orthography (Ehri et al., 2001). These two findings have been central to developing phonics

curriculum, but many other procedural questions remain, such as when to start, with which types of activities, for how long, etc.

Numeric responses to literacy curriculum

With the advent of technology to provide on-line classes and individualized learning through adaptive software, teachers and parents are also searching for numeric options to support traditional classroom literacy curriculum. Educational applications have become the most popular downloaded app category after games (AppBrain, 2020). Technology has been important in improving our lives and how we work, but there is no evidence that its increased use in the classroom translates into greater educational gains (OECD, 2015). In fact, there is considerable evidence that the human brain is highly attuned to linguistic and social interactions, and often learns most optimally from another human being (Csibra & Gergely, 2009; Dehaene, 2020). Nevertheless, evidence suggests that technology may usefully complement classroom learning practices, when the learning software is developed using principles from the cognitive sciences (Hirsh-Pasek et al., 2015) and when it is tested through randomized control trials. In fact, phonics instruction appears to be one area where children are aided by learning software, since games can introduce many grapheme-phoneme correspondences requiring active ‘drill-and-practice’ to automatize their assimilation. A successful example is the computer-based phonics method Graphogame (www.graphogame.com) (Ojanen, Ronimus, et al., 2015; Richardson & Lyytinen, 2014). Research has shown that playing Graphogame leads to an increased activation of the left occipito-temporal visual word-form system, a marker of sensitivity to print in the brain, after only a few weeks of use (Brem et al., 2010).

In our own research in France, a randomized control study comparing the phonics software ELAN to a math control game was successful in supporting literacy gains (Potier Watkins et al., 2020). The tablet-based game was specifically built for explicit systematic phonics instruction in 1st grade. Students that used ELAN showed significant improvements in decoding and reading comprehension skills. The comprehension task required correctly choosing a picture that best described a short sentence, amongst distractor pictures that were either syntactic or semantically altered. Notably, significant learning results were only achieved by children that participated in the reading intervention during the first half of the year, coinciding with the time point when children were first receiving phonics instruction in

the class. In France, 1st grade is the first year of formal literacy training. Using the application during the second half of the year, as a review or as an aid to struggling readers, failed to produce significant improvements. Additionally, children that used the reading intervention at the beginning of the year continued to make significantly more correct responses at the end of the year.

Our finding of improved reading comprehension supported the hypothesis that the automatization of decoding skills can free up cognitive resources for comprehension. It fits with the Simple View of Reading, which states that written comprehension is the product of decoding and spoken language comprehension (Hoover & Gough, 1990). But perhaps even more importantly for the goal of better understanding ‘what works’, the aforementioned results highlight a very specific and early window for phonics use. While our initial hypothesis was that phonics practice should help kids improve automatization and improve reading no matter if used during the initial learning phase of school year, or as review in the second half of the year, this was not the case. This finding is also supported by other recent reports that phonics instruction is critical in the early stages of learning. In one intervention study, classrooms that taught many grapheme-phoneme correspondences early in learning (i.e. up to two a week, compared to the frequent practice of once a week) produced greater literacy outcomes (Sunde et al., 2020). In France, an observational study of 1st grade teaching practices and student outcomes also demonstrated improved reading ability in classes where a larger number of grapheme-phoneme correspondences are taught early in the year (Goigoux, 2016). Critically, the National Reading Panel meta-analysis (National Reading Panel, 2000) reported a larger effect size when phonics instruction began early ($d = 0.55$), compared to after 1st grade ($d = 0.27$).

The ELAN intervention (Potier Watkins et al., 2020) provided children with the ability to learn new grapheme-phoneme correspondences at their own pace. During a pilot study, it was observed that children completed a single lesson in 25 minutes on average. Using the software in the classroom for 20-minutes three times a week meant that children could easily practice two new correspondences a week. However, there were two major limitations to clearly understanding if the early practice of many grapheme-phoneme correspondences was critical to our results. Our software taught children letter-sound relationships in a programmed order, without taking shortcuts to adapt to the student’s level. Undoubtedly, students from the 2nd-half of the year did move through the levels quicker, and the game did

adapt to their skills by requiring faster answers, with the goal of building automatization, but the experience was still one of review. If the game could have assessed and provided tailored practice of grapheme-phoneme correspondences that had not been mastered, then our results may have been different. The second limitation to the study was one common to educational research— our intervention provided an extra experience to literacy education in the class, which remained uncontrolled. The ELAN suite was designed to complement many of the most common manuals used in reading education, but it could have been more complementary to certain methods than others.

In the present project, we aimed to address these two limitations and to obtain new data capable of improving our knowledge of the optimal time window for phonics intervention. We therefore decided to retest the pedagogical principles of explicit systematic phonics instruction, but to do so in kindergarten classrooms, one year before formal reading instruction begins in the French school system. Our logic was that an intervention at that age would provide a better controlled, standardized amount of phonics teaching, well adapted to the children’s current instruction level, and would determine if such an approach can yield long term benefits.

In French kindergarten, the national program specifies a compulsory teaching of language development, oral comprehension, phoneme awareness (through oral manipulation) and writing. Children are expected to apprehend the “alphabetic-principle” — understanding that there are predictable relationships between sounds and letters. However, the focus of letter-sound relationships, according to the official program, is geared towards the mere recognition of the printed word as representing sounds, as well as early attempts at writing. In fact, it states that “there is no pre-reading in Kindergarten” (*Programme d’enseignement de l’école maternelle*, 2015). Our first concern was that the absence of reading instruction may be because children are considered developmentally too young. This concern was assuaged by previous research showing benefits of pre-1st grade phonics programs. It has been demonstrated that children can learn letter to sound relationships as early as preschool (Wolf, 2016), mentoring kindergarten teachers in phonics teaching techniques improves reading skills at the end of kindergarten and in 1st grade (Ehri & Flugman, 2018) and early phonics and small group reading interventions provide the biggest effect sizes for kindergarteners at-risk for reading difficulties (for a meta-analysis see (Cavanaugh et al., 2004). Computer based programs for reading acquisition have also shown benefits in

kindergarten literacy programs (Comaskey et al., 2009; de Graaff et al., 2009). Taken together, testing our phonics software seemed relevant and not completely incongruent with the current kindergarten learning goals. Furthermore, any positive impact of the software on reading would occur in the absence of formal reading instruction in the class.

Our software: Kalulu phonics

Based on several teacher recommended adjustments from the ELAN project as well as pilot sessions with a kindergarten class, we made several principled adjustments and rebranded the game under the name “Kalulu” (a clever rabbit character in Swahili culture). Adjustments included reducing the level of difficulty in several word encoding games and removing the need to pass a lexical decision task required in ELAN for passage from one level to the next. Pilot studies showed that while children were easily able to understand and play games of reading syllables and word-building, the double-task of decoding a word and deciding if it was a ‘real’ word or a ‘made-up’ word proved confusing and frustrating. The lexical decision minigame was kept as a measure of reading progress, but not as a test to pass a level.

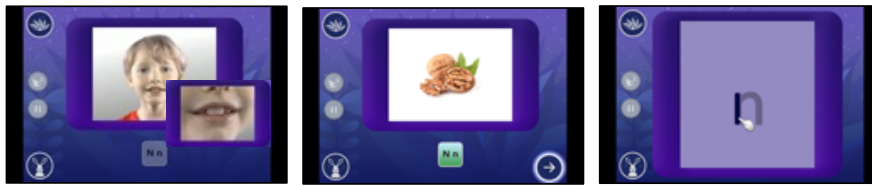
Like its predecessor, Kalulu phonics teaches children the grapheme-phoneme correspondences in an order that takes into account their frequency and consistency in written language. This progression was determined using an automatized neural-network based approach that we developed and can automatically detect the grapheme-phoneme correspondences (GPC) and their frequency and consistency in any language (Potier Watkins et al., 2019). We used the Manulex children’s print database to train our model (Lété et al., 2004). The lesson progression is provided in the supporting materials downloadable here (www.ludoeducation.fr, “TÉLÉCHARGER LE LIVRE”). Several adjustments were made to the ordering of the lessons in concordance with other existing French phonics manuals familiar to teachers (e.g., teaching all the primary vowel sounds in the beginning, introducing fricative consonants, before occlusive ones). Each grapheme-phoneme lesson was introduced in three steps, commonly used by teachers and supported by research as benefitting phoneme processing and grapheme memorization (Fig. 1a). In step 1, clicking on the grapheme starts a video of an older child pronouncing the corresponding phoneme, with a zoom on the child’s mouth and explicit articulation to enhance auditory signal-to-noise and focus attention on the articulatory gestures underlying each phoneme, an early aid in reading acquisition (Boyer &

Ehri, 2011; Castiglioni-Spalten & Ehri, 2003). In step 2, clicking on the grapheme produces a voice emphasizing the phoneme sound and its presence at the onset of the name of a picture (e.g. 'aaa, aaapple') to facilitate memorization of the sound by linking it to a known word (De Graaff et al., 2007; DiLorenzo et al., 2011; Ehri et al., 1984; Shmidman & Ehri, 2010). This type of exercise is commonly used in the classroom. In step 3, the child traces the grapheme in both upper and lower case on the touch screen. Previous research has made the hypothesis that haptic exercises with letters, such as learning to write or to trace letters with the finger improve reading skills because they provide an additional motor code to support memory for grapheme shape (Bara et al., 2004; Bara & Gentaz, 2011). Children can repeat these steps as much as needed.

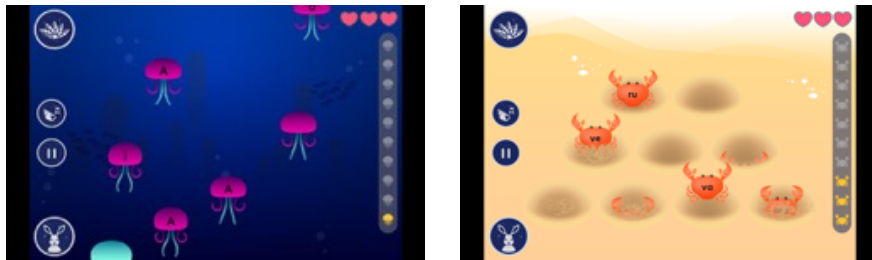
After completing the lesson, three games are proposed to automatize recognition of the grapheme-phoneme correspondence (figure 1, b-d). The mini-games in Kalulu all have the goal of spurring the child to make faster and faster responses and automatize the recently acquired grapheme-phoneme relationships with gradual difficulty. The simplest games develop phonics awareness through syllables. Children hear the grapheme-phoneme correspondence in a syllable, either as a single vowel, consonant-vowel or vowel-consonant, and must click on the corresponding written stimulus on screen. Once children have learned enough grapheme-phoneme pairs to build short words, they practice their knowledge in reading and spelling games. Another game was specifically designed to develop knowledge of French morphology by highlighting the silent letters that mark gender and plurals in words. Difficulty is created by varying the distractors and the rate of targets and distractors. All syllables and words presented in Kalulu are 100% decodable from the lessons learned.

Kalulu phonics pedagogy

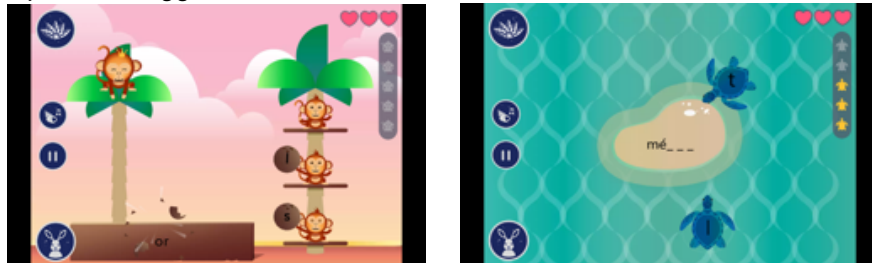
a) 3-step grapheme-phoneme lesson



b) Syllable games



c) Word building games



d) Comprehension and attention to French morphology

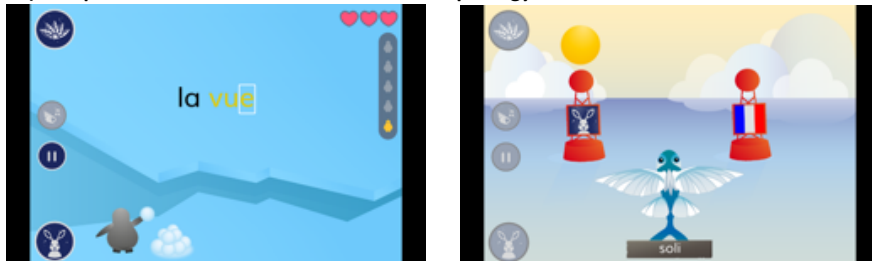


Figure 1 Screenshots from the game. **a)** Introduction to a grapheme-phoneme correspondence (GPC) by seeing a video pronunciation of the phoneme, hearing the sound in a word with accompanying picture, tracing the grapheme in upper and lower case forms. After learning the GPC, different mini-games are proposed to automatize learning. **b)** Syllable games require hearing the new sound in a syllable and finding it amongst distractors on a screen. **c)** Word building games require hearing a word then choosing the correct graphemes to spell the word. **d)** Comprehension and morphology games require reading the word silently, then making a decision of which letters in the word are a silent feature of French or deciding if a presented word exists or was made-up by the game mascot.

The game algorithm aims to provide appropriate distractors to each target. Only previously learned grapheme–phoneme correspondences can act as distractors (only learned material is presented in the mini-games). Furthermore, only vowels are presented as vowel distractors and consonants as consonant distractors. Distractors in a mini-game are chosen from a visual similarity matrix adapted from another study of similarity judgments of letter pairs (Boles and Clifford 1989). Initial items are chosen (when possible) to be >0.75 in similarity

to the target (1 being the matrix diagonal, and 0 being maximally different items). If the player makes a mistake in the mini-game, then the corresponding threshold on distractor similarity is lowered. The game also builds an internal model of the child’s confusion between targets and distractors, by keeping constant track of performance for each GPC and amount of confusion with other graphemes phonemes (figure 2). Across games, the player is required to maintain a >75% score for all GPC. If the child makes too many mistakes across several games, and its score dips below the 75% threshold, then the corresponding grapheme is reintroduced as a target.

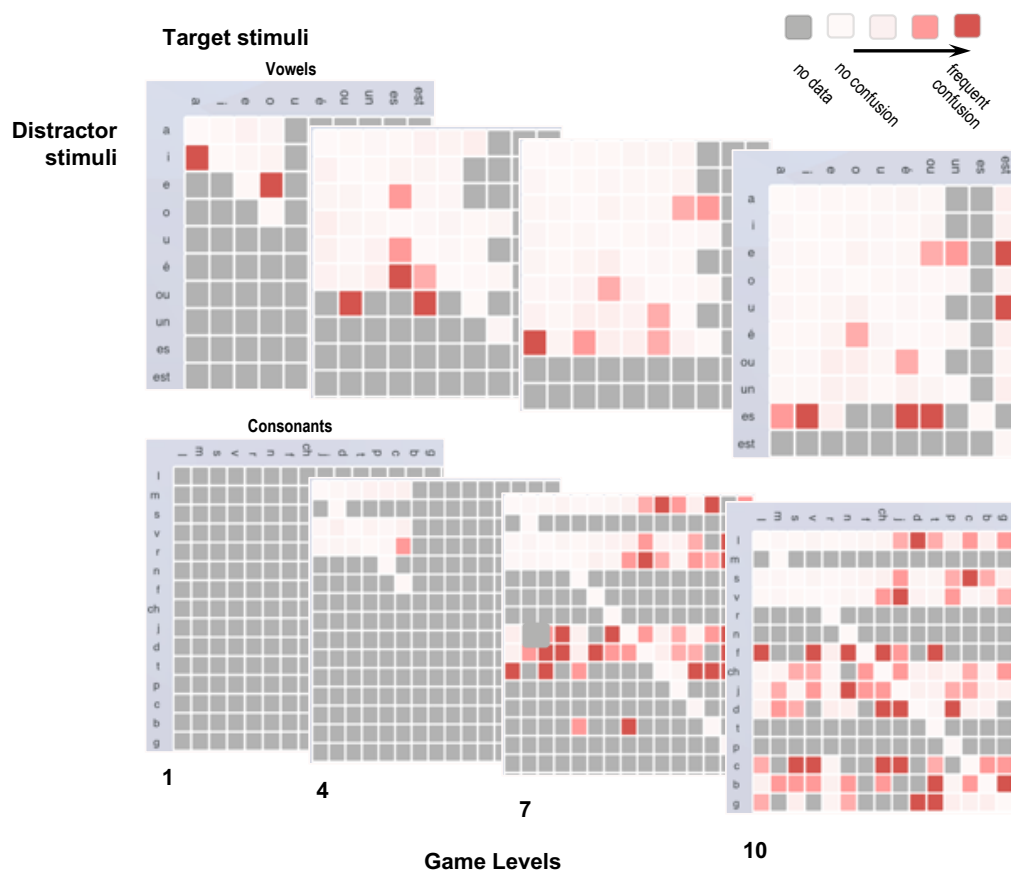


Figure 2, the grapheme-phoneme confusion matrix, The game uses detection theory to record ‘hits’, ‘misses’, ‘correct rejections’ and ‘false alarms’ in the players game responses to a target amongst distractors. On the diagonol is the average combined hit and miss responses to a present target when the target is visible. Outside of the diagonol is the measure of confusion between target and distractor, as the combined average of correct rejection and false alarm to a present distractor for a given target.

A matched active control: Kalulu numbers

A criticism of our previous study was that our control software for enhancing number knowledge, Number Catcher, was not directly matched to the phonics games and did not produce any benefits on our control measures for early math ability. In fact, both the

experimental and control software produced improvements over a passive control on a number comparison task. The hypothesis was that using that using the games had a generic effect on rapid decision-making tasks. Participating teachers also remarked that interest in the control game dissipated quickly as it was always the same (whereas the minigames for reading constantly changed). For the current project, our goal was to use a better matched control. To do this we adapted the mini-games from our Kalulu phonics program into number games that followed similar visuals, rules and procedures to make them comparably challenging and engaging (Fig. 3). We emphasize that the need for a close match between the phonics and number game implied that we could not include all of the principles that are known to facilitate the acquisition of numeracy, as previously implemented for instance in the NumberRace software (Räsänen et al., 2009; Wilson et al., 2006). Rather, for close similarity to the phonics program, we developed our games to build associations between symbol and meaning. In the case of numbers, this meant learning to associate symbolic numbers (either number names, e.g. 'two', or indo-arabic numerals, e.g. '2') to their quantity ('**').

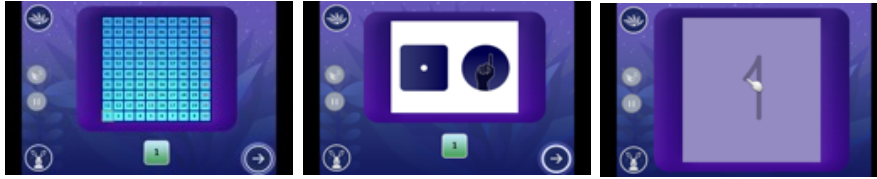
Kalulu numbers teaches numbers 0-100 in forty lessons. The progression is provided in the appendix. Each lesson starts with a 3-part sequence to associate number with meaning using techniques supported by the literature (Fig. 3a). First the number is presented in its counting sequence on a 10 x10 game board. Playing on game boards that demonstrate the linear structure of numbers has been linked with improved numeracy (Laski & Siegler, 2014; Siegler & Ramani, 2009). Next, the child is introduced to different quantity representations of the number. This is a common practice shown to improve math readiness in school (Skwarchuk et al., 2014). Finally, children are asked to trace the number symbol twice to help memorize its form. The hypothesis with letters is that haptic exercises like tracing or writing may provide an extra motor code and reduce instances of mirror-writing (Gottfried et al., 2003). Mirror writing is also phenomena that exists in numeral writing (Arnas et al., 2004; Fischer, 2011), and we reasoned that this practice could help cement the memory for numerals.

Once the learning session is complete, as with the reading path, several mini-games are proposed to cement comprehension (Fig. 3). The first type of mini-game focuses on automatizing fluency of association between number name, number symbol and quantity. Once children learn enough numbers, a counting game is introduced. The counting game requires helping a game character find the correct number in a counting series starting from

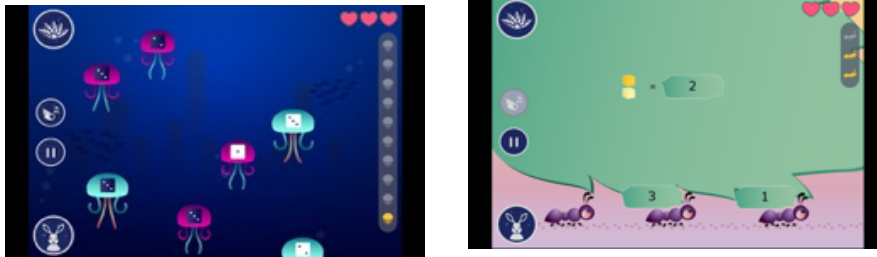
either '0' or another random number and ending at specific number (e.g. "starting from the number 5, count all the numbers up to the number 8"). Once children learn numbers above 10, a game is introduced to associate numerals to quantities, taking into account place value (e.g. if asked to find the number matching the quantity of '28', first the child would match the symbolic number '20', then the symbolic number '8'). Finally, we introduce arithmetic games for small additions (<10). Passage from one garden to the next is tested by a 'choose the biggest number task', where children have 3 minutes to choose the biggest between two numbers, with 20 pairs presented in total.

Kalulu numbers pedagogy

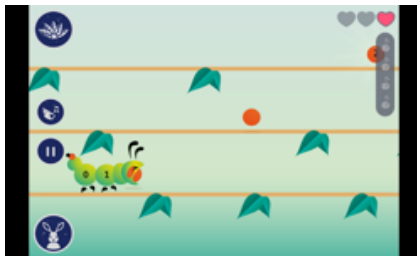
a) 3-step number-quantity lesson



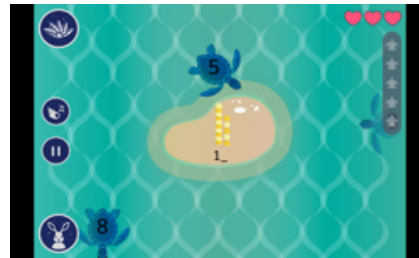
b) Associating number name, quantity and symbol



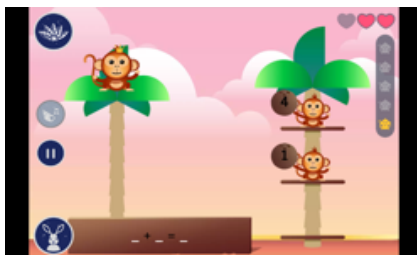
c) Counting



d) Base-10 understanding



e) Arithmetic



f) Number comparison



Figure 3, Screenshots from the game. **a)** Introduction to a number and its quantity by counting to the number on a game-board that highlights the linear order of the numbers, showing different quantity representations of the number and tracing the number symbol. Once the lesson is learned, children are introduced to different games to automatize number and meaning. **b)** Associating number name, quantity and symbol requires hearing or seeing a quantity then matching it with the correct symbol or quantity. **c)** Counting games require finding the correct number in a counting suite amongst distractors. **d)** Base-10 understanding games require associating quantity and symbol, while taking into account place value. **e)** Arithmetic games provide single digit addition problems. **f)** A number comparison task requires choosing the biggest number in a limited amount of time.

An encompassing experiment and control environment

Both the experimental phonics pedagogy and the active control game were presented in the same game environment (Fig. 4). A game panel allowed teachers to select Kalulu phonics or numbers. In both pedagogical goals the game is introduced by the mascot Kalulu, a rabbit that explains that learning is like cultivating the garden of the mind: the more you learn, the more vibrant your garden will grow. The main screen of the game represents a brain in a schematic

format of 20 gardens. The player touches the highlighted garden to zoom-in. Across all gardens runs a letter path and a number path. Nodes along the path each represent a lesson for learning a new grapheme-phoneme relationship (letter path) or a new number (number path). Pressing the node opens a lesson followed by three mini-games to practice the lesson learned. Once the mini-games have been won, at a minimum score of 80% correct, the next node on the path opens. With each lesson learned and game won, the once barren garden flowers with plants. Game play and rules are identical on the two paths. Kalulu was developed using the open-source platform Godot to ultimately allow non-licensed distribution of the game, and is currently available in compiled form at www.ludoeducation.fr.

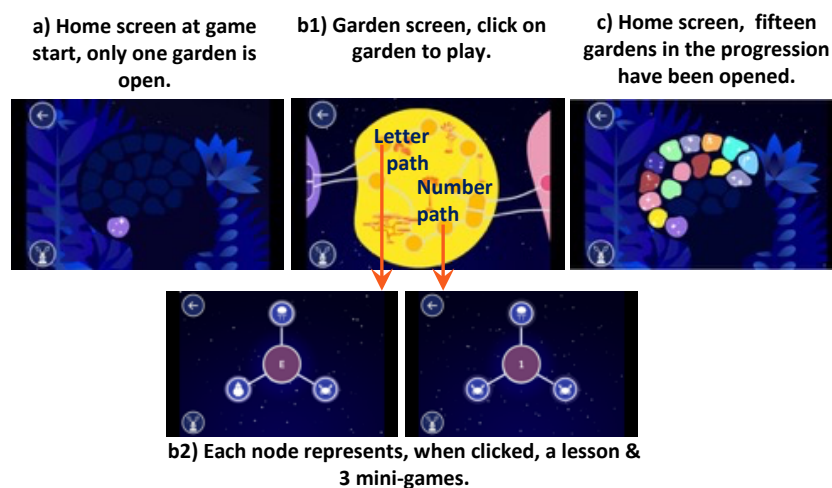


Figure 4, Screen shots of the game environment. Both the e-perimental software (phonics) and active control (numbers) are encased in the same environnemnt, emphasizing learning as tending to the garden of one's mind and building brain power. **a)** The brain screen starts with a single garden open. **b)** Zooming-in, the child can e-lore the 'letter path' or the 'number path'. Touching a node opens a lesson and its accompanying 3 mini-games. Winning games results in plants growing in a once barren garden. **c)** Zooming back out, as the child progresses, the gardens lights up the brain screen to show children their progress along the learning path.

Experimental Study

In summary, the main goal of the project was to test the Kalulu phonics software with Kindergarten children to see if teaching many grapheme-phoneme correspondences at this early age could boost phonics knowledge in kindergarten and literacy in 1st grade. The Kalulu software was adapted from our ELAN software and piloted for conformity to Kindergarten usability as a stand-alone game needing minimal explanation from the teacher. Pilots were conducted for several months in a classroom with a member of the lab observing interactions

and reporting to the developer. Using a cross-over design, we tested the game with kindergarten classes, comparing the phonics software to both an active control (classrooms with the Kalulu math software) and a passive control (classes without tablets). To assess possible long-term benefits, we also followed-up on participants during the subsequent year (1st grade), using the results on national standardized tests, up to a year after exposure to the intervention. The project was approved by the ethical committee CERES (Conseil d'évaluation éthique pour les recherches en santé) on 25/04/2017.

Methods

Participants

Two different school districts in France participated in the project (Nice and Poitiers). Each district recruited twenty classes to participate in a tablet-based intervention, as well as five classes for a tablet-free “business-as-usual” control group. Participation in the study was proposed to teachers by the school district and on a voluntary basis. At the beginning of the year, teachers sent home with students a letter from the lab explaining the project, accompanied with an ‘opposition form’. Parents could return the form if they wished to opt-out. At the start of the school year, 1092 students were initially counted in the intervention.

Design

The school districts required that all children eventually have access to both of the intervention games. Thus, we employed a 2-period crossover design (similar to Potier Watkins et al., 2020). All children in the study were tested at the beginning of the year (labeled pretest), after which each of the 40 intervention classes was randomly assigned to either playing with the Kalulu phonics game for the first half of the year, then the Kalulu number game for the second half (group labeled “read/math”), or vice versa (group labeled “math/read”). Randomization required an equal number of classes in each group and was stratified by the school district (Poitiers or Nice) and the size of the class, which varied from 5 to 28 children. Generally, a smaller class size indicates that it is a class split into two grade levels. For example, a class may have students from middle-preschool and Kindergarten. We also tested 10 ‘business-as-usual’ control classrooms (group labeled “control”). These classes could not be randomized, as the teachers were interested in participating, but did not want to receive

tablets. The inclusion of both an active and passive control allowed us to measure progress due to the pedagogy separate from the effect of merely introducing a tablet intervention to the class. To measure the evolution of learning, children were tested during a 3-week period before switching games (labeled midtest) and during another 3-week period at the end of the year (labeled endtest). To assess long-term benefits to using the software, children’s results on national evaluations given at the beginning and middle of the 1st grade were also examined in a longitudinal follow-up. Figure 5 provides a calendar of intervention and test sessions.

Study Design					1 st grade follow-up 2019-2020	
kindergarten intervention 2018-2019						
sept.	oct. – jan.	feb.	mar. - may	june	sept.	jan.
pretest	session A	posttest A	session B	posttest B	Beginning of the year, National Evaluation	Middle of the year, National Evaluation
	Group 1, Read Group 2, Math Group 3, Control		Group 1, Math Group 2, Read Group 3, Control			

Figure 5, All children took reading and math tests at the beginning, middle, and end of the year (respectively, termed pretest, mid-test, and end-test). In a classic crossover design, during the first half of the school year, children in the intervention groups (group 1 and group 2) played with either Kalulu reading or math software. After the mid-test, children switched to the other software for the second half of the year. Control classes (group 3) did not have tablets. In the following school year, when students no longer had access to Kalulu, we looked at national evaluation scores in a follow-up analysis.

Procedure

Five to eight tablets (depending on class size) were sent to classes at the start of the school year. A lab member contacted each teacher individually to explain the design of the project and introduce the teacher to the project website with supporting information. Teachers were encouraged throughout the year to call or email the designated lab member with their implementation questions. This person was not aware of students’ results on the tests. Within each class, a workspace was set up in a quiet space for tablet sessions. Teachers were asked to organize small groups of students, according to the number of tablets received, that would take turns participating in the intervention session. We also asked that groups be heterogenous in terms of their expected ability by teachers. The goal was for children to help each other and minimize the need for teacher assistance. Teachers in our previous research had reported appreciating the small group play sessions, while teachers in the current project informed us that rotating small groups between activity stations is a common practice in

kindergarten and easy to implement. Research has shown that small group learning is especially beneficial for young children (for a meta-analysis see, Hattie, 2010). We would also like to mention here that small groups rotating at a tablet workstation is far more economical for large-scale implementation than having one tablet per child. Visits to the class confirmed our hope that children would help each other with the games.

Each intervention session lasted ten weeks. Teachers were asked to ensure that students play for 1 hour a week for at least eight weeks of the intervention period. Teachers could organize tablet time into three 20-minute sessions or two 30-minute sessions per week. This flexibility was to take into account absenteeism and time dedicated to class trips or other projects. Each of the test periods lasted three weeks.

Once the pretests were completed, teachers were informed of their intervention group (read/math or math/read). They also received an instruction book (on paper and pdf) detailing the cognitive science behind the pedagogy for their immediate group (phonics or numbers). Before starting the intervention, classes had a 2-week fall break, during which time teachers could familiarize themselves with the booklet and game. Everyone was encouraged to understand the materials in order to support children's learning in the class. However, teachers were under no obligation to modify their own pedagogy in consequence. Communications between the lab and the classes were maintained throughout the year and teachers, as much as possible, uploaded data from the tablets to the lab. In many cases, the Kindergarten classes did not have wifi, and two classes did not enter the correct class id in their tablet. In these cases, data was sent when possible or the lab was kept abreast of progress through phone calls. Teachers were asked to report bugs and criticism of the games throughout the year. Everyone reported positive remarks and high engagement from the class. Only one child in the intervention refused to play half-way through the year. All of the classes remained in the project throughout the intervention year.

Test measures

Our goal was to see if the intervention could boost kindergarten phonics and number knowledge and improve 1st grade formal learning. To assess this question, we decided to use the national standardized evaluation given to all children at the beginning of their 1st grade year, and which assesses several aspects of the expected knowledge and predictors for reading and math achievement. By using this test during the intervention, we could measure

progress through the kindergarten year and follow-up on progress on comparative measures when our participants took the national evaluations at the beginning and middle of 1st grade.

This national evaluation, called EvalAide (Groupe de travail du Csen, 2019), is generally administered by the teacher and done using pencil and paper, but there also exists a tablet version. Testing at all three periods during the kindergarten intervention was done on tablet by employees of each school district specialized in pedagogical teacher support. Testers attended an on-line training day to learn how to conduct the tests. Children did the test individually, wearing headphones and providing answers on their own on the touch screen. The testing administrator was present to repeat instructions if needed and ensure that all children did the test on their own. Data from the tests was subsequently sent to the lab.

The tablet version of the national evaluation for 1st grade is separated into three subtests, two testing language and reading skills (FR1 and FR2) and one testing number and math skills (MA1). Due to time constraints, only a selection of questions was used. At the pretest, to assess differences in baseline knowledge, children's responses were collected on the visual attention, vocabulary and syllable awareness tasks in the FR1 subtest. At pretest, midtest and endtest children answered all the questions in the FR2 (excluding an orally presented text and comprehension task, due to errors in the recordings) and the MA1 subtest (excluding a number comparison task that did not accurately record response time). Each subtest was done in its entirety with a break between the two.

The FR2 subtest measures the abilities that have been shown to be the most predictive of reading outcomes such as oral sentence comprehension, phoneme awareness and letter name knowledge (Sprenger-Charolles et al., 2005). It also looks at phonics knowledge in a grapheme-phoneme matching task. This task is directly taught in Kalulu and is considered the backbone to reading. Grapheme-phoneme correspondences are not required teaching in kindergarten, and children are not expected to have acquired this knowledge before 1st grade. All questions included a trial with feedback. The percentage correct for each question was collected. As we were also interested in learning if the phonics games in Kalulu could spur literacy before formal instruction, we included in our battery of language and reading questions a 1-minute word and pseudoword reading task. This test was done on an individual basis by the tester with the child. Children had 1-minute to read as many words and then matched pseudowords as possible. We looked at the average number of words and pseudowords in a minute as a measure of reading fluency.

The MA2 subtest assesses number understanding, predictive measures of math ability and early problem-solving skills. Only numbers up to ten were shown or required as answers, ten being the highest number of focus in French kindergarten. A trial item with feed-back was provided for each question. The number understanding portion of the test required associating symbolic-numbers to number-name (e.g., “Find the number ‘five’ six symbolic-numbers presented) and symbolic-number to quantity (e.g. “Touch the number that corresponds to the quantity of eggs in a basket”). Math problem solving is assessed using simple everyday word problems (i.e., “Five kids need a pen. There are three pens in the box. How many more pens do they need?”). For each question an image demonstrating the two given quantities of the problem (e.g., In the given example, an image of five kids and three pens) was provided to help the child find the solution. The predictive measures included a number-line and number comparison task. In the number-line task, children were shown a number-line from 0 to 10 with a tick mark crossing the line. Children had to choose from six options the number indicated by the tick mark. An additional number comparison task was attempted, but the results was not included in our analyses due to a design issue that prohibited us from calculating the number of correct responses made in a minute time constraint, the typical measure of analysis for this task.

For the longitudinal follow-up in 1st grade, we are able to access the anonymized data for each of our groups from the national standardized tests. Most questions on the tests were administered by teachers simultaneously to the entire class, with children writing their responses with pen-on-paper. Several questions in the test required reading aloud and these were done individually with the teacher. Teachers uploaded students’ responses on a dedicated platform managed by the *Department of evaluation, prospective and performance (DEPP)* of the ministry of education. The test is given in the beginning of the 1st grade (September) and in the middle of the school year (January). Questions between these two periods test similar categories of reading and math skills, but include more difficult material (i.e. bigger numbers, reading and spelling) to take into account expected learning progress between the two time periods. The mission of these evaluations is to signal teachers to predictive measures of the student’s future performance and measure school progress. The beginning of the year test provides teachers with an objective measure of the student’s achievement after the ‘summer-lag’, defined by the loss skills acquired in the preceding school year after several months without (Alexander et al., 2001; McCoach et al., 2006). The middle

of the year evaluation gives teachers the time to ensure all students have made the expected progress and provide targeted interventions to children in need. In the appendix we have included a detailed list of the questions assessed from kindergarten to 1st grade with the different times periods and administration details.

Statistical analysis for the intervention

Given our 2-period crossover design, for each task, we first analyzed responses for the randomized part of the experimental design using a mixed analysis of variance with a 2-level between-group factor (read/math versus math/ read) and a 3-level within-group test period factor (pretest, midtest, endtest). We then re-ran the same analysis including the non-randomized no-tablet control (3 x 3 Anova). To factor out pre-schooling differences amongst children, we included the covariable composite z-score computed from the individual's average z-scores on the three predictive tasks at pretest (visual attention, vocabulary and syllable awareness).

In a crossover design, we expect to see a 2-step pattern of improvement: children in the read/math group should improve in the language and reading tests at midtest relative to the math/read intervention, whereas the latter should improve on the number and math tests. During the second session of the intervention, we should see the opposite effect of learning for each group so that no difference should eventually be found on all measures at the endtest (unless one period is more conducive to learning, perhaps due to greater readiness for reading). An effect of the training is therefore expressed as a significant interaction between the group and test period. Because all children are expected to learn during the year, the passive control provides us with a measure of normal progression. To correct for multiple comparisons, we used a Bonferroni correction for each comparison by dividing the Type I error rate by the number of summary measures that we assessed, $a/k = .05/5 = .01$ for the language and reading subtest, and $a/k = .05/3 = .017$ for the number and math subtest. The results we report all passed this threshold.

Intervention participant attrition and base-line characteristics

Only children that completed the three tests given during the kindergarten intervention were examined. This included 772 children (read/math= 300, math/read = 320, control = 163). To examine the baseline balance between the two intervention and control groups, we performed an overall F test or Chi-square test (see Table 1), as well as pairwise comparisons

between the three groups in terms of baseline demographic variables, the three predictive variables at pretest (visual attention, vocabulary and syllable awareness) and the pretest results in each domain (see Table 1). No differences were found between groups, except for the fluency task (the average number of word and pseudowords read in a minute). Critically, two by two comparisons revealed that a difference between the two intervention groups, $F(1, 617)=03.86$, $p=.050$ was not significant at the Bonferroni corrected threshold $p_{corrected}=0.01$. The control group did however read more items at this threshold than the math/read group $F(1, 480)=10.71$, $p=.001$, but not the read/math group $F(1, 460)=3.20$, $p=.074$. These differences were mainly due to the fact that few children attempted this test. Children were required to read the first two trial syllables in order to continue, and the proportions of children who did differed slightly across groups (read/math=25%, math/read=20%, control=29% of the participants succeeded in reading the trials). We decided to not adjust groups based on this variable as, on average, all groups read less than one item per minute.

Table 1, Baseline characteristics at pretest for children that completed the intervention.

	read/ math	math/ read	control	
N=	300	320	163	
N=female	152	144	82	$\chi^2(2)=2.33, p=0.317$
Age in months	62 (3.98)	62 (3.54)	63 (3.59)	$F(2, 780)=1.24, p=0.299$
Pretest predictive measures				
visual attention	60%(.27)	63%(.29)	63% (.28)	$F(2, 780)=0.62, p=0.540$
syllable awareness	31% (.20)	31% (.20)	33% (.20)	$F(2, 780)=0.87, p=0.419$
vocabulary	42% (.28)	41% (.29)	45% (.28)	$F(2, 780)=0.63, p=0.531$
Language and literacy				
oral comprehension	49% (.30)	47% (.31)	51% (.30)	$F(2, 780)=0.93, p=0.395$
phoneme awareness	21% (.13)	22% (.16)	22% (.15)	$F(2, 780)=0.35, p=0.703$
letter-name knowledge	26% (.25)	23% (.24)	25% (.23)	$F(2, 780)=0.87, p=0.421$
GP matching	33% (.23)	32% (.24)	31% (.22)	$F(2, 780)=0.54, p=0.585$
One-minute word and pseudoword reading*	.29 (.71) n=25%	.21 (.44) n=20%	.48 (1.3) n=29%	$F(2, 780)=6.41, p=0.002$ control > r/m > m/r
Number and math knowledge				
number knowledge	66% (.23)	67% (.23)	67% (.20)	$F(2, 780)=0.33, p=0.717$
word-problem	30% (.27)	32% (.28)	30% (.27)	$F(2, 780)=0.36, p=0.695$
number-line	23% (.23)	20% (.21)	21% (.20)	$F(2, 780)=1.28, p=0.278$

* Children were not expected to know how to read at the beginning of kindergarten and only attempted the question if they could read the first two trial words, otherwise they were scored 0. The significant group effect can be attributed to the greater percentage of children that were able to attempt this question (r/m = 25%, m/r = 20%, control = 29%). However, as the average number of words read for each group was < 1, we did remove participants to create equal groups.

Results

Intervention effects

Children's progress across the three test periods during the intervention year for read/math, math/read and control group on the language and reading knowledge are provided in Figure 6. Performance in early number understanding and manipulation is presented in Figure 7. There was a significant effect of session on all test tasks, $p < 0.001$, demonstrating normal learning progression through the school year. There was also a significant effect of the composite predictor score for all measures, $p < 0.001$. Over and above those expected effects, we now report for each measure the group by session interactions.

We begin by looking at the language and reading results. On the two orally presented early predictive measures for reading, i.e. oral comprehension and phoneme awareness,

children in both intervention groups performed equally across all three periods, $F(2, 1234)=0.24$, $p=0.788$ and $F(2, 1234)=0.79$, $p=0.455$, respectively. Including the passive control group into the analysis revealed similar results. We did not find a group by test interaction on either of the orally presented measures (oral comprehension: $F(4, 1558)=0.45$, $p=0.769$; and phoneme awareness : $F(4, 1558)=1.40$, $p=0.232$).

A different outcome appeared when we looked at tasks that required children to manipulate letter symbols and their sounds. In letter name knowledge, there was an intervention group by test interaction, $F(2, 1234)=9.22$, $p<0.001$. Post hoc analysis at each test period showed that the intervention groups started the year with similar knowledge, $F(1, 617)=2.04$, $p=0.154$. After the first training period, however, children that used the phonics suite could recognize more letters by their letter names than children that used the number suite, $F(1, 617)=10.64$, $p<0.001$. This initial advantage disappeared at the end of the year, once both groups had had training with the Kalulu phonics suite, $F(1, 617)=2.27$, $p=0.132$. Including the control group to our analyses also produced a group by session interaction, $F(4, 1558)=4.88$, $p<0.001$. Post hoc analysis at each test period confirmed that all three groups began the year on par, $F(2, 779)=1.01$, $p=0.363$. An effect of group on the midtest, $F(2, 779)=5.36$, $p=0.005$ was found, due to an advantage for children in the phonics intervention group compared to the math group. No difference was found between either of the intervention groups and the control (read/math vs. control, $F(1, 460)=1.77$, $p=0.184$; math/read vs. control, $F(1, 480)=1.87$, $p=0.172$). No difference in letter name knowledge was found at the end of the year, $F(2, 779)=1.95$, $p=0.143$. All effects were significant at the Bonferroni corrected threshold $p_{corrected}=0.01$.

We also found a significant intervention group by session interaction for grapheme-phoneme matching, $F(2, 1234), 25.71$, $p<0.001$. Post hoc analysis at each test period showed no effect at pretest, $F(1, 617)=1.10$, $p<0.294$. Children that received the phonics intervention during the first session performed better in grapheme-phoneme matching by the mid-year test, $F(1, 617)=35.72$, $p<0.001$, an advantage that disappeared at the endtest, $F(1, 617)=2.85$, $p=0.090$. Including the passive control group again produced a group by test session interaction, $F(4, 1558)=14.03$, $p<0.001$. All three groups started the year with similar phonics knowledge, $F(2, 779)=1.26$, $p=0.283$. An effect of group emerged at midtest, $F(2, 779)=22.12$, $p<0.001$, showing that children who had received the phonics intervention performed better than children who received the math intervention, but also than the passive control, $F(1,$

460)=25.40, $p < 0.001$, while no difference was found between the active and passive control, $F(1, 480) = 0.14$, $p = 0.704$. An effect of group extended to the endtest, $F(2, 779) = 7.54$, $p < 0.001$. By the end of the year, the read/math showed a marked trend to better grapheme-phoneme understanding over the control $F(1, 460) = 5.88$, $p = 0.016$, while the math/read group now also made significantly more correct matches than the control, $F(1, 480) = 14.68$, $p < 0.001$.

On the measure of reading fluency, the average number of words and pseudowords read in a minute, we found a significant intervention group by test interaction, $F(2, 1234) = 5.47$, $p = 0.004$. At the start of the year, an advantage for children that would go into the read/math group was marginally significant, $F(1, 617) = 3.86$, $p = 0.050$. As mentioned in the section 'Base-line characteristics at pretest', this was mainly due to the fact that slightly different numbers of children attempted this test. By the middle of the year, the group that used the phonics suite read more words than the active control, $F(1, 617) = 17.32$, $p < 0.001$. This advantage disappeared after both groups had used the phonics suite, $F(1, 617) = 0.02$, $p = 0.900$. Those results were confirmed by a significant three group by test session interaction, $F(4, 1558) = 4.68$, $p < 0.001$. Once again, due to the percentage of students able to attempt the test, there was a significant group difference at pretest, $F(2, 779) = 5.97$, $p = 0.003$. Our post hoc analyses showed that the control group started with marginally but non-significantly more words read than the read/math group, $F(1, 460) = 3.20$, $p = 0.074$, and significantly greater early reading ability compared to the math/read group, $F(1, 480) = 10.71$, $p = 0.001$. By the middle of the year, a significant group effect, $F(2, 779) = 11.17$, $p < 0.001$, revealed that children that had used the phonics software also read more words than either the active or the passive control ($F(1, 460) = 11.65$, $p < 0.001$ for the latter test). Both the math/read and control group read an equal number of words on average by the middle of the year, $F(1, 480) = 0.01$, $p = 0.933$. On the endtest, we did not find an effect of group, $F(2, 779) = 1.43$, $p = 0.240$.

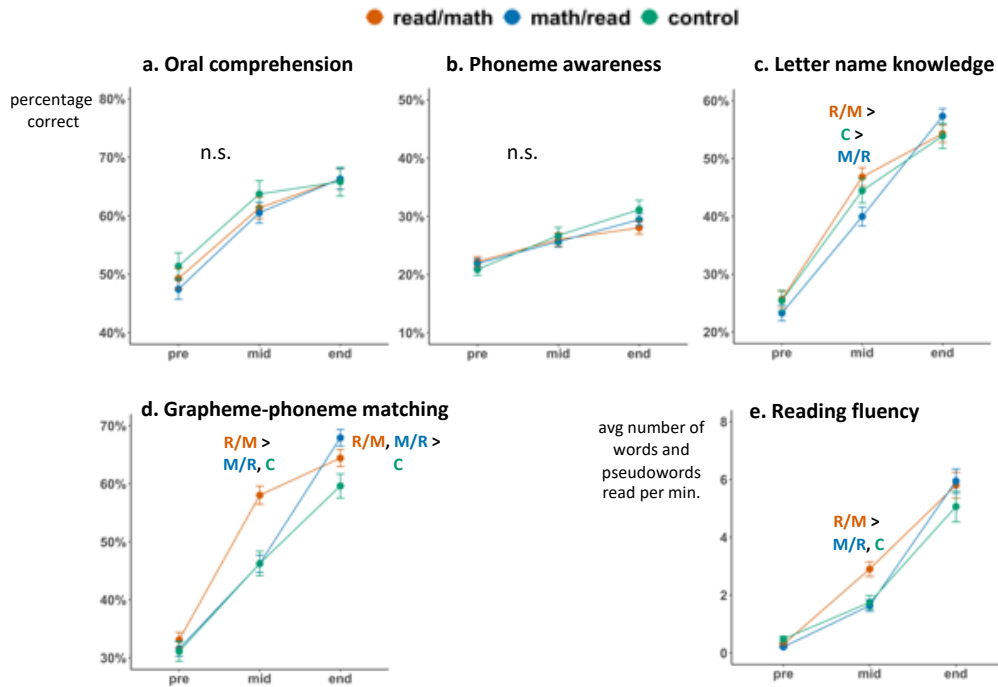


Figure 6 a-e, Group averages on language & reading tasks at the three intervention time-points. All tests improved across time. **a)** oral comprehension (listening to short sentence and choosing the corresponding picture), **b)** phoneme awareness (choosing a picture description of a word that either starts or ends with a target name), **c)** letter name knowledge (hearing a letter name and choosing three different orthographic presentations (upper or lower case; print or cursive) of the correct letter amongst distractors), **d)** grapheme-phoneme matching (hearing a phoneme and choosing the corresponding grapheme), **e)** word and pseudo-word reading (average number of items read in one minute). The p-value indicates the significance of the 3 group \times 3 time interaction. Pairwise differences that were significant at Bonferroni $p_{corrected} < 0.01$ are highlighted (RM read/math group; MR math/read group; C control group).

We now present the results for the number and math knowledge tests. For the combined number knowledge portion of the test, we found a significant group by test interaction, $F(2, 1234)=8.37, p<0.001$. Post hoc analysis showed that both intervention groups started on equal footing at pretest, $F(1, 617)=0.28, p=0.596$. By the middle of the year, children that played with the number games made more correct responses when matching symbolic numbers with their names and quantities, $F(1, 617)=18.28, p<0.001$. This difference disappeared by the end of the year, once both groups had had experience with the number games, $F(1, 617)=0.02, p=0.902$. Including the control group revealed a similar group by test interaction, $F(4, 1558)=5.35, p<0.001$. Post hoc tests revealed no difference between groups at pretest $F(2, 779)=0.16, p=0.855$. At the midtest, a significant group effect, $F(2, 779)=9.94, p<0.001$ highlighted a greater number understanding for children that played with the number games than the control, $F(1, 480)=10.14, p=0.002$. No difference was found between the read/math and control groups $F(1, 460)=0.25, p=0.621$. At the endtest, there was a marginal

group effect, $F(2, 779)=3.90$, $p=0.023$, with both intervention groups making more correct responses than the control (read/math vs. control, $F(1, 460)=6.32$, $p=0.012$; math/read vs. control, $F(1, 480)=6.36$, $p=0.012$). In quantitative terms, at the end of the school year, the group that had used Kalulu number games in the first session, after 4 months of no longer playing, maintained an advantage of 0.18 standard deviations over the control group, while the group that had used the math suite in the second half of the year had a 0.19 standard deviation advantage over the control group.

The word problem solving and number-line tasks did not produce a group by test interaction when taking into account only the two randomized groups (word problem solving, $F(2, 1234)=2.17$, $p=0.115$; number-line, $F(2, 1234)=0.99$, $p=0.372$) or when including the passive control (word problem solving, $F(4, 1558)=1.33$, $p=0.258$; number-line, $F(4, 1558)=0.68$, $p=0.608$).

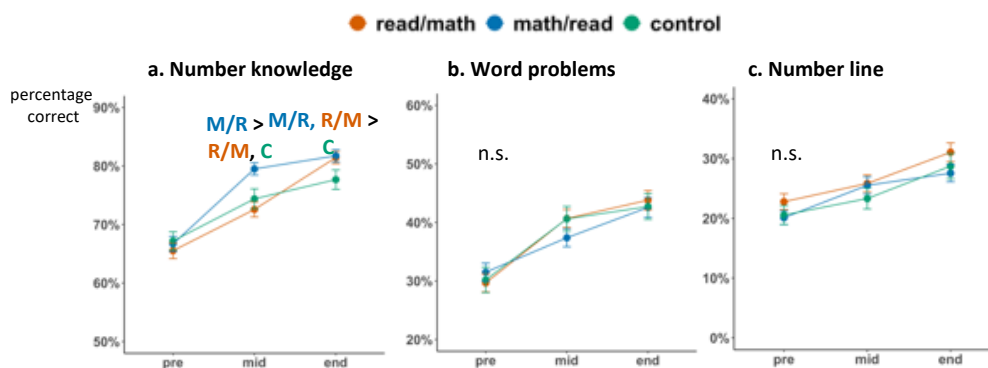


Figure 7 a-e, Group averages on number & math tasks at the three intervention time-points. All tests improved across time. **a)** number knowledge (a combined score of matching number name ('two') and number symbol ('2') and number symbol with quantity ('**')), **b)** word problem solving, **c)** number-line. The p-value indicates the significance of the 3 group \times 3 time period interaction. Pairwise differences that were significant at Bonferroni $p_{corrected} < 0.01$ are highlighted (RM read/math group; MR math/read group; C control group).

Longitudinal follow-up

Did using the Kalulu software boost 1st grade school readiness, leading to better early literacy and math success? To assess this, we looked at the data from beginning and mid- school year national evaluations of the participants from the kindergarten intervention. Data for 94% of the participants was found (read/math=270, math/read=312, control=156). Consistent with the intervention analysis, we first looked at the randomized part of the study in a 2 (group: read/math, math/read) one-way ANOVA on each of the questions, the goal being to assess if children in the 1st grade retained more or less knowledge from the reading or math software

depending on the period of exposure during the intervention year. If no difference was found between the two intervention groups, we collapsed both groups and compared them to the non-randomized condition in a 2 (group: intervention vs. passive control) and repeated the analysis. We used the same predictor composite as a covariate. We report whether the results achieve significance after a Bonferroni correction, obtained by dividing the Type I error rate by the number of measures assessed. For the beginning of the year evaluation, this is $a/k = .05/4 = .013$ for the reading tests and $a/k = .05/4 = .013$ for the math tests; for the mid-year evaluation, this is this is $a/k = .05/6 = .008$ for the reading tests and $a/k = .05/5 = .01$ for the math tests. The composite predictor continued to be significant on all tasks at $p < 0.001$. Degrees of freedom were adjusted to reflect the fact some cells had missing data for certain questions within a test. This could have been due to a child that was absent for a portion of the test or accidental non-reporting by the teacher. Figure 8 shows the performance at the beginning and middle of the 1st grade year for each of the read/math, math/read, and control groups, respectively, for the language & reading and number & math tests.

At the beginning of 1st grade, the two intervention groups showed similar abilities in all of the pre-reading tasks: oral comprehension, $F(1, 569)=0.30$, $p=0.586$; letter-name knowledge, $F(1, 567)=0.09$, $p=0.760$; phoneme-awareness, $F(1, 572)=0.37$, $p=0.545$; and grapheme-phoneme matching, $F(1, 573)=0.14$, $p=0.710$. We therefore collapsed the randomized groups and compared them to the passive control group using the same analysis. No difference was found between the intervention and passive control on the tasks of oral comprehension, $F(1, 725)=1.05$, $p=0.305$ and letter-name knowledge, $F(1, 721)=0.00$, $p=0.965$. In the two tasks requiring phoneme manipulation, a skill directly taught by the Kalulu reading software, the intervention groups did significantly better than the control group, in spite of the summer break: phoneme awareness, $F(1, 727)=8.25$, $p=0.004$; grapheme-phoneme matching, $F(1, 651)=7.27$, $p=0.007$. After Bonferroni correction, both effects remained significant.

Looking at the number and math questions. No difference was found between the randomized intervention groups on any of the tasks: number knowledge, $F(1, 564)=0.14$, $p=0.712$, word-problems, $F(1, 576)=0.02$, $p=0.896$, number-line, $F(1, 573)=1.25$, $p=0.264$, number comparison, $F(1, 569)=0.06$, $p=0.807$. We collapsed these two groups and compared them to the passive control group. At this time, the control group performed comparably to the intervention on tasks of number knowledge, $F(1, 718)=0.01$, $p=0.916$; word-problems, $F(1,$

732)=0.55, $p=0.460$ and number-line, $F(1, 727)=1.02$, $p=0.313$. However, on the number comparison task, children that received the intervention performed significantly better than the control, $F(1, 725)=4.53$, $p=0.034$, but this difference did not survive Bonferroni correction.

By the middle of 1st grade, the two intervention groups still performed equally on the reading measure of oral comprehension, $F(1, 543)=0.72$, $p=0.395$; grapheme-phoneme matching $F(1, 545)=0.01$, $p=0.934$; reading comprehension, $F(1, 556)=0.45$, $p=0.502$; spelling, $F(1, 531)=0.00$, $p=0.949$. On phoneme awareness, this time, students that had used the reading intervention early in their Kindergarten year showed a slight advantage over students that used the method during the second half, $F(1, 556)=5.60$, $p=0.018$ while an advantage in the opposite direction was found for reading speed, $F(1, 548)=5.86$, $p=0.016$. As neither of these group effects were significant after Bonferroni correction, the two intervention groups were collapsed for comparison with the passive control. After half-a-year of formal reading education, all groups performed equally well in oral comprehension, $F(1, 692)=0.11$, $p=0.735$, phoneme awareness, $F(1, 708)=0.75$, $p=0.388$, grapheme-phoneme matching, $F(1, 693)=1.63$, $p=0.202$, reading speed, $F(1, 698)=0.97$, $p=0.326$. In tasks of reading comprehension and spelling, the control group performed better than the intervention group, $F(1, 708)=5.77$, $p=0.017$ and $F(1, 676)=12.14$, $p<0.001$, respectively. After Bonferroni correction for the 6 measures, only the effect of spelling remained significant.

As for math, in the middle of 1st grade, no difference was found between the two intervention groups on number writing, $F(1, 545)=1.36$, $p=0.248$; word-problem solving, $F(1, 556)=0.42$, $p=0.517$; arithmetic, $F(1, 533)=0.23$, $p=0.635$; number-line, $F(1, 555)=0.30$, $p=0.582$; number-comparison, $F(1, 547)=0.50$, $p=0.481$. Collapsing the two intervention groups and comparing them to the control did not produce a group effect on number-writing, $F(1, 693)=0.490$, $p=0.49$; arithmetic, $F(1, 678)=0.68$, $p=0.414$; number-line, $F(1, 706)=0.43$, $p=0.516$; number-comparison, $F(1, 694)=0.10$, $p=0.748$. Children in the control group made more correct responses than the intervention group in solving math word problems, $F(1, 708)=5.23$, $p=0.023$, but this effect was not significant after Bonferroni correction for the 5 questions.

Group average z-scores on tasks from the 1st grade National Evaluations

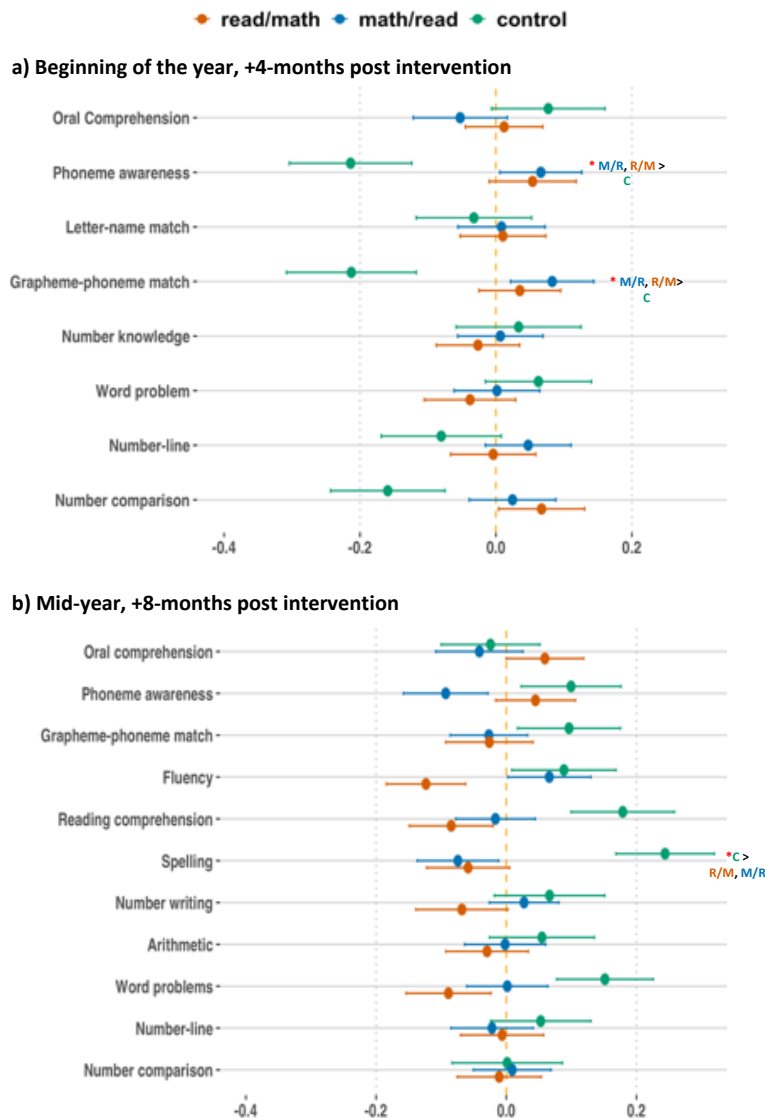


Figure 8 a-b, Group averages on number & math tasks at the two 1st grade evaluation time-points. Z-scores, corrected for the composite predictor from the kindergarten analysis, for each question are presented to allow for comparison across all tasks. * Highlights pairwise differences that were significant at Bonferroni $p_{corrected} < 0.013$ for reading and math tests at the beginning of the year, then $p_{corrected} < 0.008$ for reading and $p_{corrected} < 0.01$ for math at the mid-year test. (RM read/math group; MR math/read group; C control group).

Discussion

This study sought to remediate several limitations brought up in our first intervention with the ELAN software. On the one hand, during the ELAN intervention, we were concerned that our active control was an insufficient comparison in the pronunciation of the results made by the reading software since it did not produce any benefit in number understanding. We therefore

redesigned our software to include the active math control within the same engaging environment. In this matter, the Kalulu intervention was successful as both groups showed an effect of improved symbol (letter, number) to meaning (sound, quantity) dependent on the software used. Compared to the ELAN intervention, this current study was adapted to Kindergarteners. The reason being that in Kindergarten, children would not be receiving formal reading instruction. This would allow us to measure benefits from using the software, minus input from the intense literacy training children receive in 1st grade. Our goal was that children would thus benefit from using the software at both time periods of the year, since the instruction from Kalulu would always be novel. In this matter, the intervention was also successful. Progress was made on the questions tested, relevant to the treatment intervention and independent of session. Furthermore, both intervention groups, irrelevant of session, appear to have equally benefitted by the longitudinal follow-up.

Compared to the passive control group, using the Kalulu software provided a significant gain in improved symbol to meaning knowledge for both grapheme-phoneme and number-quantity knowledge. This benefit imparted from the software was again irrelevant to session used. However, during the kindergarten intervention, an improved understanding of the symbols and their meaning did not transfer to gains in reading or math ability beyond normal learning progression as exemplified by the control. In other words, Kalulu was only successful in imparting knowledge on the learning tasks that were made explicit by the game.

Encouragingly, children that had taken part in the intervention arrived in 1st grade with an advantage in grapheme-phoneme knowledge and phoneme awareness, +4 months after the intervention. Furthermore, in spite of the fact that the intervention and testing during the intervention period had been done on tablets, students were able to transfer their knowledge to pen-and-paper tasks. This outcome was encouraging as it shows that learning by using the tablet was not just context dependent. Both phoneme-awareness and letter to sound knowledge are important predictors in reading success (Bradley & Bryant, 1983; Cunningham, 2001; Melby-Lervåg et al., 2012; Piquard-Kipffer & Sprenger-Charolles, 2013). At the beginning of 1st grade, there was also a trend to improved number comparison ability compared to the passive control. Symbolic number comparison has been previously shown to be a predictor of math success in the early years of elementary school (Lyons et al., 2014).

In spite of these encouraging outcomes at the start of 1st grade, by the middle of the year, there was no longer an advantage to having been exposed to Kalulu in kindergarten.

Furthermore, children from the passive control did better in spelling, and showed a trend in greater reading comprehension and word problem solving. What could have happened from the start to the middle of 1st grade is unbeknownst to us as we do not have access to school information from the data provided by the National Evaluations. Our first consideration was that the Kalulu software produced long-term negative effects. In the past, methods promoting pure phonics instruction have been criticized for possibly developing an over-dependence on decoding skill while ignoring the end goal of fluent reading which is word recognition. We would consider this hypothesis unfounded for several reasons. First of all, the results from our first study showed that comprehension was improved by children that received the phonics intervention, with benefits lasting +4 months after the intervention.

Another explanation, supported by the results from our previous intervention (Potier Watkins et al., 2020), is that this project minimized the importance of introducing grapheme-phoneme correspondences simultaneously with the focused experience of learning to read in the classroom. As stated, the goal of our phonics software was to provide children with many grapheme-phoneme correspondences early in learning. A practice supported by previous observational and intervention studies (Goigoux, 2016; Sunde et al., 2020). And in fact, children that participated in the intervention were exposed to more grapheme-phoneme training than children in the control, as exemplified by the 'endtest' and early 1st grade results. While hypothetically we would assume that this advantage would boost literacy learning, it may have had the adverse effect of drawing-out and decorrelating phonics education with classroom practices, such as reading aloud, necessary for children to learn to read. In other words, our students learned about phonics, but given the limitation that our software does not include voice recognition, children could not apply this new knowledge to actual reading practice. Furthermore, 1st grade teachers would not have taken into account that students from the intervention had already started to build their phonics knowledge and hence provide adapted literacy practice. Students from the intervention probably began their 1st grade year reviewing learned grapheme-phoneme correspondences, while children from the 'business-as-usual' control would have learned this novel information simultaneously with common classroom practices, such as reading-aloud. In this scenario, children from the control condition possibly benefitted from an advantageous situation of introducing phonics in combination with structured reading practice, while the intervention students received review for their first months of literacy education. Our design possibly recreated a similar situation to

the one that we saw in our first experiment, whereby children that used the games as review did not benefit, only in this case, the 1st grade classroom experience was the review.

Another possible explanation is to the credit of the National Assessments. As described earlier, these tests are specifically designed to evaluate and provide individualized feedback in a timely manner. Children from the control group began the year with significantly lower skills in phoneme awareness and grapheme-phoneme matching. It is not impossible that these students received targeted individualized attention to subvert an early deficit. The National Education system of France has provided teachers with both detailed explanations of the tests with corresponding targeted remediation exercises for these cases.

Future directions

Teachers' reports were generally positive, and children learned, played, and enjoyed the games. Our participants made initial gains in reading readiness, and these improvements persisted after the summer-lag. On average, a slight decline across all academic areas is seen in all students over the summer (Borman, Benson, & Overman, 2005; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Heyns, 1987; Murnane, 1975). We cannot say, however, that Kalulu reproduced the results in ELAN. Children improved in tasks of symbol to meaning understanding compared to the control group, but this knowledge did not transfer to reading speed or comprehension improvement in either the short- or long-term results.

Any future direction of this project should seek to reduce the greatest limitation to this study, the disappointing follow-up result that initial boosts in phonics were not a catalyst to literacy, in spite of previous research showing that phoneme awareness and letter-sound knowledge are predictors (Cunningham, 2001). First, there continue to be many aspects of the Kalulu game that could be improved. Primarily, there is no voice recognition due to, to our knowledge, the unreliability in the current technology to recognize nuances in children's reading (i.e., several programs tested could not detect differences in /b/ and /d/ sounds which is important in ensuring that children do not make mirror-errors). Our game requires children to 'read' in their heads and then associate an orally presented syllable or word with its written form. Some of the harder games require encoding. For example, children hear a word then must break it down into its phonemes and assemble one by one the corresponding graphemes. A solution would be to integrate reliable voice-recognition or reproduce the study and have teachers involved in reading aloud practice. If the greater knowledge in grapheme-

phoneme correspondences would have helped the intervention population in reading once it was coupled with reading aloud in the class, then we may have argued that phonics software in kindergarten was enough to spur early literacy. Instead, our results highlight the importance of phonics included as an early piece of a larger curriculum that also includes reading aloud.

One criticism to this may be that providing this type of intervention was just too early, and that learning to read would be better off introduced in 1st grade. In response, we would argue that there is enough research showing that reading instruction can begin as early as Kindergarten. The current national program requires that children apprehend the alphabetic principle and begin spelling. There is no doubt that spelling should be a part of any reading curriculum given the benefits it provides to reading (Cunningham, 2001), but it seems peculiar that teachers would be told to begin spelling without reading. Developing a Kindergarten-based pedagogy that emphasizes learning the grapheme-phoneme correspondences and practice reading decodable text may free-up time for 1st grade teachers to emphasize comprehension and learning of non-transparent site words. Given that all but one child did not continue the Kalulu software, we see no reason why children would not benefit from a more integrated early reading program. Future research would aim to develop a curriculum that could be adopted by the teacher and take into account reading as well as spelling readiness in kindergarten. Future research could go in several different directions, any of which should couple reading aloud practice with the phonics training. For example, if we choose to retest the software in kindergarten, it should include voice recognition with games that provide reading aloud practice. If reliable voice recognition software proves to be unreliable for this age of children or disruptive (groups of children speaking at different time periods may prove too loud while other children are doing other activities) then the intervention could be re-developed as a more complete teacher-led kindergarten curriculum. In this case, the Kalulu software should be played in unison with teaching practices that include reading aloud in the class. Another option is to retest Kalulu in 1st grade, likewise as a complete reading curriculum that includes class activities and time spent in drill-and-practice with the game. The guidebook that we wrote to accompany the game has since been improved and made into a student reader with class lessons, reading aloud exercises and activities, all of which follow the Kalulu software pedagogical progression.

Conclusion

Today it is undisputed that phonics instruction is an essential component in learning to read. Yet, teachers, and scientists, know that learning to read requires more than just cracking the code. For this reason, it is important that field interventions such as this one examine the many facets of when and how to use a curriculum in order to go beyond tenets and provide teachers with usable school materials for concrete pedagogical guidelines. Furthermore, interventions that posit classroom gains should be evaluated using standardized tests and tested for long-term benefits. When the gains are limited to short-term benefits and simply mirror the functions that were taught, they may not be followed by any subsequent classroom-relevant benefits. By evaluating our intervention using expected school-based measures and looking at long term outcomes, we obtained a better picture of the true benefits and shortcomings to improve future implementations of the software. Working in tandem, teachers, researchers and school policy makers should all endeavor to build a system of checks-and-balances with the goal of building a curriculum that provides the best step-by-step support for teachers and the greatest impact on the child's learning.

Appendix

Order of grapheme-phoneme correspondences taught in Kalulu phonics*

a, i, e, o	1	in, en, im, eConsonantConsonant, ç, c/s/	11
u, l, m, s	2	g/dʒ/, s/z/, er	12
é, v, r, ou	3	oi, è, et	13
silent letters, n, un	4	ent, au, ant	14
f, ch, j	5	h, ti/s/, x	15
d, t, p	6	gn, ai, ez	16
c, b, g	7	silent letters, ill/ij/, eu	17
gu, k, q	8	â, oy	18
z, es, est	9	il/j/, ill/j/, ph, ei	19
nasals: en, em, on, om, an, am	10	y, ê, w	20

*In Kalulu phonics, there are 20 levels to complete. In each level, two to six grapheme-phoneme correspondences are taught. Levels presented here represent the gardens found on the game's home screen.

Descriptions of test questions asked at the three time points during the kindergarten intervention and the 1st grade longitudinal follow-up

	Kindergarten (during the intervention)		1 st Grade longitudinal follow-up	
	pretest	midtest & endtest	September	January
Reading:				
symbolic awareness	tablet	-	teacher led in class	-
visual attention	tablet	-	teacher led in class	-
vocabulary	tablet	-	teacher led in class	-
oral comprehension	-	-	teacher led in class	teacher led in class
grapheme-phoneme matching	tablet	tablet	teacher led in class	teacher led in class
letter to letter name matching	tablet	tablet	teacher led in class	-
phoneme awareness	tablet	tablet	teacher led in class	teacher led in class
oral comprehension, short sentence	tablet	tablet	teacher led in class	teacher led in class
one minute word and pseudo word reading	individual, w/ tester	individual, w/ tester	-	individual w/ teacher
syllable & word writing	-	-	-	teacher led in class
text fluency	-	-	-	teacher led in class
Math:				
number to number name matching	tablet	tablet	teacher led in class	-
number writing	-	-	teacher led in class	teacher led in class
number to quantity matching	tablet	tablet	teacher led in class	teacher led in class
word problem	tablet	tablet	teacher led in class	teacher led in class
number line	tablet	tablet	teacher led in class	teacher led in class
number comparison	-	-	teacher led in class	teacher led in class
arithmetic	-	-	-	teacher led in class

Pictures in color are taken from the tablet version of the test and black and white pictures from the paper version. Unless noted, paper and tablet versions were identical in every way.

Symbolic Awareness



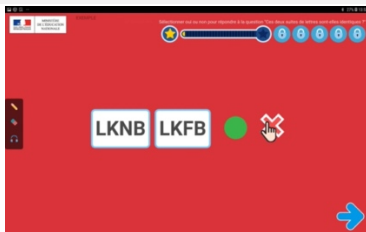
Four pictures are presented by their names. The child then hears a new CVCV target word and must choose the picture that starts with the same first syllable.

Rhyming



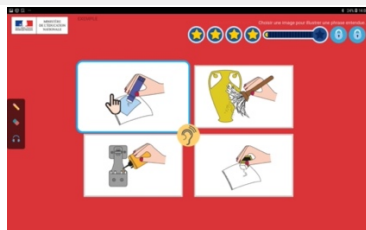
Four pictures are presented by their names. The child is asked to choose which word does not rhyme with the others.

Visual Attention



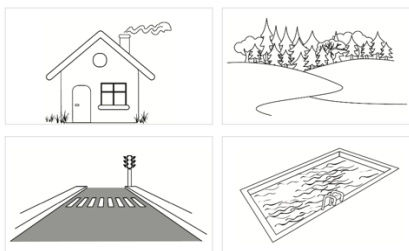
Two letter suites with 3, 4 or 5 letters are presented simultaneously. The child decides if the two letter suites are the same or different.

Vocabulary



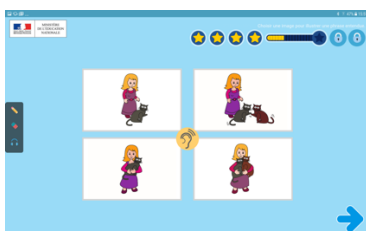
Adapted from the TVAP (Deltour & D, 1990), the child is orally presented a word then asked to find its corresponding picture. Words can describe objects or actions. Distractors share pronunciation, functional semantics or categorical semantics.

Oral comprehension (short story)



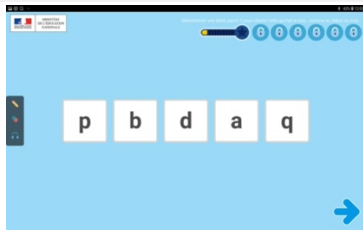
Children hear a short story, then are asked to circle the picture that best answers an orally presented question

Oral comprehension (short sentence)



Adapted from the ECOSSE (Lecocq, 1996). The child hears a short sentence then chooses its corresponding picture. The distractors are of images that are either semantically or syntactically different.

Grapheme-phoneme matching



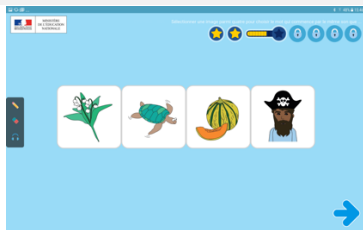
The child hears a phoneme then chooses its corresponding grapheme. The four distractors are either visually or phonetically similar.

Letter to letter name matching

E b a c C d
 b O A e D c
 o e B c a d

The child hears a letter names, then must find its three corresponding identities. Letters are presented in print and cursive, lower and upper case.

Phoneme awareness



Each picture is named, then the child hears a target word and must choose the picture that either starts (or ends) with the same sound.

One-minute word and pseudo word reading

à oû la au tu
 un il été on mur
 ni sur qui vélo par
 feu ce peur ami moto
 peau lune gare lire bon
 mardi col avril roi faire

Children are presented with a list of words and have 1 minute to read as many words as possible, followed by a list of matched pseudowords, for which they have again 1 minute to read.

Syllable & word writing

Syllables or words are said aloud to the child who must spell them.

Text reading

Le renard court dans la forêt. Il arrive à la ferme.
 Va-t-il voler une poule ? Lola l'a vu. Elle crie et
 le chasse.
 « Bravo Lola ! », dit la poule.

The child reads aloud a short text. The teacher scores the reading for number of correctly read words in 1 minute.

Number writing

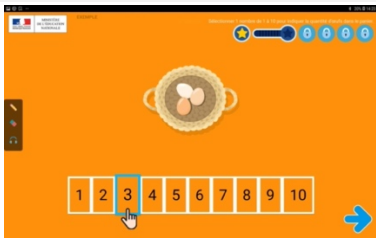
The child hears a number name and must write the number.

Number name to symbolic number matching (pre, mid and end; baseline)

0 6 3 2 5 1

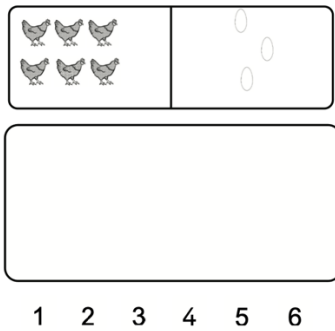
The child hears a number name and choose its corresponding symbol.

Number to quantity matching



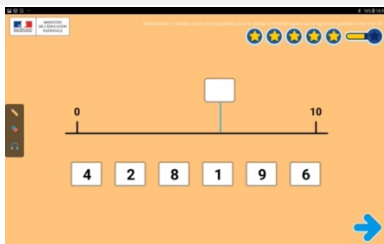
The child is presented with varying quantities of eggs in a basket and must choose the corresponding number symbol.

Math word problem



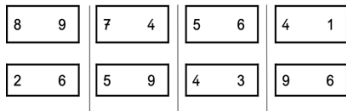
Children are presented with a word problem and must choose the corresponding number to answer the problem (e.g. “There are six chickens that each want an egg. There are three eggs. How many more eggs do you need for each chicken to have an egg?”). Children are provided with a picture example to help them find the solution.

Number line



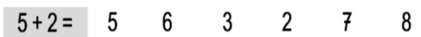
Children must choose the number that best fits the position on a given number-line.

Number comparison



Adapted from the SYMP test (Brankaer et al., 2017). The goal is to circle the biggest number in a pair of numbers. The child has 1 minute to circle a maximum number of items.

Arithmetic



The child circles the response to the corresponding addition or subtraction problem.

References

- Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, Achievement, and Inequality : A Seasonal Perspective. *Educational Evaluation and Policy Analysis*, 23(2), 171-191.
- AppBrain. (2020, novembre 2). Top categories on Google Play. AppBrain. <https://www.appbrain.com/stats/android-market-app-categories>
- Bara, F., & Gentaz, E. (2011). Haptics in teaching handwriting : The role of perceptual and visuo-motor skills. *Human Movement Science*, 30(4), 745-759. <https://doi.org/10.1016/j.humov.2010.05.015>
- Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, 19(3), 433-449. <https://doi.org/10.1016/j.cogdev.2004.05.003>
- Boyer, N., & Ehri, L. C. (2011). Contribution of Phonemic Segmentation Instruction With Letters and Articulation Pictures to Word Reading and Spelling in Beginners. *Scientific Studies of Reading*, 15(5), 440-470. <https://doi.org/10.1080/10888438.2010.520778>
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, 301(5899), 419-421. <https://doi.org/10.1038/301419a0>
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17), 7939-7944. <https://doi.org/10.1073/pnas.0904402107>
- Byrne, B. (1992). *Studies in the Acquisition Procedure for Reading : Rationale, Hypotheses, and Data*. Reading Acquisition; Routledge. <https://doi.org/10.4324/9781351236904-1>
- Castiglioni-Spalten, M. L., & Ehri, L. C. (2003). Phonemic Awareness Instruction : Contribution of Articulatory Segmentation to Novice Beginners' Reading and Spelling. *Scientific Studies of Reading*, 7(1), 25-52. https://doi.org/10.1207/S1532799XSSR0701_03

- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars : Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5-51. <https://doi.org/10.1177/1529100618772271>
- Cavanaugh, C. L., Kim, A.-H., Wanzek, J., & Vaughn, S. (2004). Kindergarten Reading Interventions for At-Risk Students : Twenty Years of Research. *Learning Disabilities*, 13.
- Comaskey, E. M., Savage, R. S., & Abrami, P. (2009). A randomised efficacy study of Web-based synthetic and analytic programmes among disadvantaged urban Kindergarten children. *Journal of Research in Reading*, 32(1), 92-108. <https://doi.org/10.1111/j.1467-9817.2008.01383.x>
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends Cogn Sci*, 13(4), 148-153. <https://doi.org/10.1016/j.tics.2009.01.005>
- Cunningham, J. W. (2001). The National Reading Panel Report. *Reading Research Quarterly*, 36(3), 326-335. <https://doi.org/10.1598/RRQ.36.3.5>
- De Graaff, S., Verhoeven, L., Bosman, A. M., & Hasselman, F. (2007). Integrated pictorial mnemonics and stimulus fading : Teaching kindergartners letter sounds. *British Journal of Educational Psychology*, 77(3), 519-539.
- de Graaff, S., Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. *Scientific Studies of Reading*, 13(4), 318-333. <https://doi.org/10.1080/10888430903001308>
- Dehaene, S. (2020). *How We Learn : The New Science of Education and the Brain*. Penguin Books Ltd.
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate : Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, 16(4), 234-244. <https://doi.org/10.1038/nrn3924>
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form : Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biology*, 16(3), e2004103. <https://doi.org/10.1371/journal.pbio.2004103>
- DiLorenzo, K. E., Rody, C. A., Bucholz, J. L., & Brady, M. P. (2011). Teaching letter–sound connections with picture mnemonics : Itchy’s alphabet and early decoding. *Preventing School Failure: Alternative Education for Children and Youth*, 55(1), 28-34.

- Ehri, L. C., Deffner, N. D., & Wilce, L. S. (1984). Pictorial mnemonics for phonics. *Journal of educational psychology*, 76(5), 880.
- Ehri, L. C., & Flugman, B. (2018). Mentoring teachers in systematic phonics instruction : Effectiveness of an intensive year-long program for kindergarten through 3rd grade teachers and their students. *Reading and Writing*, 31(2), 425-456. <https://doi.org/10.1007/s11145-017-9792-7>
- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic Phonics Instruction Helps Students Learn to Read : Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research*, 71(3), 393-447. <https://doi.org/10.3102/00346543071003393>
- Ehri, L. C., & Wilce, L. S. (1987). Does Learning to Spell Help Beginners Learn to Read Words? *Reading Research Quarterly*, 22(1), 47-65. JSTOR. <https://doi.org/10.2307/747720>
- Goigoux, R. (2016). Apprendre à lire et à écrire au cours préparatoire: enseignements d'une recherche collective. *Revue Française de pédagogie*, (3), 5-6.
- Graaff, S. de, Bosman, A. M. T., Hasselman, F., & Verhoeven, L. (2009). Benefits of Systematic Phonics Instruction. *Scientific Studies of Reading*, 13(4), 318-333. <https://doi.org/10.1080/10888430903001308>
- Groupe de travail du Csen. (2019). ÉvalAide, un dispositif scientifique de prévention des difficultés en lecture et en mathématiques en CP et en CE1. Canopé. https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conference_role_experimentation_domaine_educatif/EvalAide_CSEN_Definitif_Mai2019.pdf
- Hattie, J. (2010). *Visible learning : A synthesis of over 800 meta-analyses relating to achievement* (Reprinted). Routledge.
- Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting Education in "Educational" Apps : Lessons From the Science of Learning. *Psychological Science in the Public Interest*, 16(1), 3-34. <https://doi.org/10.1177/1529100615569721>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, 2(2), 127-160.

- Lété, B., Sprenger-Charolles, L., & Colé, P. (2004). MANULEX : A grade-level lexical database from French elementary school readers. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, 36(1), 156-166.
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 17(5), 714-726. <https://doi.org/10.1111/desc.12152>
- Maurer, U., Blau, V. C., Yoncheva, Y. N., & McCandliss, B. D. (2010). Development of visual expertise for reading : Rapid emergence of visual familiarity for an artificial script. *Dev Neuropsychol*, 35(4), 404-422. <https://doi.org/10.1080/87565641.2010.480916>
- McCoach, D. B., O'Connell, A. A., Reis, S. M., & Levitt, H. A. (2006). Growing readers : A hierarchical linear model of children's reading growth during the first 2 years of school. *Journal of Educational Psychology*, 98(1), 14-28. <https://doi.org/10.1037/0022-0663.98.1.14>
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read : A meta-analytic review. *Psychological Bulletin*, 138(2), 322-352. <https://doi.org/10.1037/a0026744>
- Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language*, 127(3), 356-365. <https://doi.org/10.1016/j.bandl.2013.10.009>
- National Reading Panel. (2000). Teaching children to read : An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). National Institute of Child Health & Development. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- OECD. (2015). *Students, Computers and Learning*. OECD Publishing. <https://doi.org/10.1787/9789264239555-en>
- Ojanen, E., Ronimus, M., Ahonen, T., Chansa-Kabali, T., February, P., Jere-Folotiya, J., Kauppinen, K.-P., Ketonen, R., Ngorosho, D., Pitkänen, M., Puhakka, S., Sampa, F., Walubita, G., Yalukanda, C., Pugh, K., Richardson, U., Serpell, R., & Lyytinen, H. (2015). GraphoGame – a catalyst for multi-level promotion of literacy in diverse contexts. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00671>

- Piquard-Kipffer, A., & Sprenger-Charolles, L. (2013). Early predictors of future reading skills : A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (age 5 to 8). *L'Année psychologique*, 113(04), 491-521.
- Potier Watkins, C., 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. <https://doi.org/10.1007/s40692-019-00152-6>
- Potier Watkins, C., Dehaene, O., & Dehaene, S. (2019). Automatic Construction of a Phonics Curriculum for Reading Education Using the Transformer Neural Network. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Éds.), *Artificial Intelligence in Education* (p. 226-231). Springer International Publishing. https://doi.org/10.1007/978-3-030-23207-8_42
- Programme d'enseignement de l'école maternelle. (2015, mars 25). Ministère de l'Éducation Nationale de la Jeunesse et des Sports. <https://www.education.gouv.fr/bo/15/Special2/MENE1504759A.htm>
- Richardson, U., & Lyytinen, H. (2014). The GraphoGame Method : The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. *Human Technology*, 10(1). <https://doi.org/10.17011/ht/urn.201405281859>
- Rose, J. (2006). Independent review of the teaching of early reading : Final report. DfES. <http://dera.ioe.ac.uk/5551/3/5d970d28fc535dc54eb4dee9995bef36.pdf>
- Shmidman, A., & Ehri, L. (2010). Embedded picture mnemonics to learn letters. *Scientific studies of reading*, 14(2), 159-182.
- Sunde, K., Furnes, B., & Lundetræ, K. (2020). Does Introducing the Letters Faster Boost the Development of Children's Letter Knowledge, Word Reading and Spelling in the First Year of School? *Scientific Studies of Reading*, 24(2), 141-158. <https://doi.org/10.1080/10888438.2019.1615491>
- Turkeltaub, P. E., Gareau, L., Flowers, D. L., Zeffiro, T. A., & Eden, G. F. (2003). Development of neural mechanisms for reading. *Nat Neurosci*, 6(7), 767-773.
- Wolf, G. M. (2016). Letter-Sound Reading : Teaching Preschool Children Print-to-Sound Processing. *Early Childhood Education Journal*, 44(1), 11-19. <https://doi.org/10.1007/s10643-014-0685-y>

Chapter 5, Characterizing different types of developmental dyslexia: the Malabi screener

Potier Watkins, C., Friedmann, N., & Dehaene, S. Characterizing different types of developmental dyslexia: the Malabi screener. In preparation

Abstract

Reading is a complex process that involves multiple pathways and processing stages. Thus, distinct types of reading deficits may arise as a consequence of impairments in any of those processes. Here, we describe the Malabi screener, which was developed to identify deficits in orthographic, lexical and sub-lexical stages of reading in French, and we report how it allowed us to identify a double dissociation between distinct types of developmental dyslexia. Inspired by the neuropsychological literature on acquired alexias, the Malabi rests on a careful analysis of error types in reading aloud. It comprises several lists comprising many word traps, i.e., stimuli designed to elicit a certain type of error and thus identify a frailty in the corresponding cognitive processes. Attentional dyslexics, for instance, may misread the word-pair “bar cat” as “bat car”, while letter-position dyslexics may misread ‘diary’ as ‘dairy’. We collected norms for reading errors in French junior-high students, and identified three cases of developmental attention dyslexia, compatible with an impairment in binding letters to words between neighboring words, and one case of letter-position dyslexia, compatible with an impairment in letter position encoding within words. In post-hoc tests, we replicated those diagnoses and explored the factors that modulate letter migration between and within words. Our results fit with prior evidence from other languages for multiple types of developmental dyslexia, including deficits rooted in impaired orthographic rather than phonological processing, as also supported by the study of acquired alexias and reading process models. We argue that dyslexia screeners can be improved through the careful elicitation and analysis of reading errors and may result in better reading instruction methods and targeted interventions for struggling readers.

Introduction

Despite educational opportunity, 5 to 7% of the population are diagnosed every year with developmental dyslexia, a general term describing difficulty in learning to read despite age and educational experience and in the absence of visible sensory impairment, neurological disorders or language proficiency (Peterson & Pennington, 2015; World Health Organization, 2020). Given the importance of reading in school and in life, a large body of research has gone into describing and attempting to understand dyslexia. This goal, however, has proven elusive. Reading is a complex activity involving many processes and pathways linking the visual and language systems of the brain (Coltheart, 2005; Dehaene et al., 2010), and a fragility in any part of this multiple-route system could result in a different reading deficit, an evidence also supported by the variety of different types of selective reading deficits reported in case studies of acquired alexia following brain damage in adults (Coslett & Turkeltaub, 2016).

It is with this logic that dyslexia is often studied as a blanket term, under which different subtypes of the deficit may exist. Importantly, identifying these subtypes is likely to be a crucial step towards providing efficient targeted interventions to struggling readers. For example, from the perspective of the visual system, multiple theories of dyslexia have surfaced including: the magnocellular visual pathway deficit leading to poor binocular control (Stein & Walsh, 1997), the visual stress condition resulting in a general visual perception discomfort (for a review and rebuttal of its link to dyslexia, Evans & Allen, 2016) and the visual attention span deficiency which attributes dyslexia to a reduction in the number of visual elements that can be processed in parallel (Valdois et al., 2004). In terms of frailty in language processing, dyslexia has also been associated to an underlying phonological deficit (Vellutino et al., 2004). The key characteristic of the phonological deficit is a difficulty in hearing and correctly manipulating phonemes in language tasks (Araújo et al., 2014). Dyslexic readers often have difficulties in phonological perception and processing in language (Peterson & Pennington, 2015; Wagner & Torgesen, 1987). Furthermore, a phonological deficit has emerged as the primary deficit in several large-scale studies. The majority of people that suffer from dyslexia also do poorly in tasks that require phoneme manipulation compared to normal controls (Elbro & Jensen, 2005; Ramus et al., 2003). For example, in a population of 164 dyslexic children, 92% of the population had a phonological deficit, while none of the participants' dyslexia could only be linked to a visual deficit (Saksida et al., 2016). In another

study including 1,114 dyslexic readers from various European orthographies, the majority of participants were also found to have poor phoneme awareness, regardless of the transparency of orthography (Landerl et al., 2013). Intensive phoneme training practices also appear to attenuate certain cases of dyslexia (Alexander et al., 1991; Elbro & Petersen, 2004). Furthermore, phonological skills are predictive in reading outcomes for normal readers (Melby-Lervåg et al., 2012). However, the consensus that phonological deficits underlie the majority of dyslexic cases has also come under criticism in light of studies of normal reading. For example, the direction of causality has been questioned since phonological awareness is poor in illiterates and develops as a consequence of literacy (Castles & Coltheart, 2004; Dehaene et al., 2010, 2015; Froyen et al., 2008; Monzalvo & Dehaene-Lambertz, 2013; Morais et al., 1986; Schaadt et al., 2013) and phoneme awareness is poor in illiterates (Schaadt et al., 2013). Children with dyslexia tend to avoid reading, raising the question of correlation of phonological and reading skills as a “chicken or the egg” problem (Castles & Coltheart, 2004).

Dual Route Model of reading, and different types of dyslexia

To describe expert reading, and the different types of dyslexia observed in the literature, the prevalent model is the Dual-Route model (Coltheart, 2005; Friedmann & Coltheart, 2018) and its computational variants (Coltheart et al., 2001; Plaut et al., 1996; Ziegler et al., 2008; Zorzi et al., 1998); see Figure 1. According to this model, reading begins with visual orthographic analysis. This peripheral reading stage is responsible for abstract letter identification, letter-to-word binding and letter-position encoding within words (Ellis, 2016). Words are then processed by distinct parallel lexical and sub-lexical routes. Reading via the lexical route describes the fast and fluent immediate access of known words to the mental lexicon, from where information about meaning and sound are retrieved. Reading via the sub-lexical route refers to the grapheme-phoneme decoding procedure, which is applied to unknown words and pseudowords, as well as during learning to read in beginner readers. A computational implementation of the Dual Route model, the Dual Route Cascaded Model (Coltheart et al., 2001) successfully explains various aspects of normal reading behavior, including faster reading times for regular words (lexical reading) than for pseudowords (sub-lexical reading) (Rastle & Coltheart, 1999). The Dual Route model and the study of dyslexia have had a mutually beneficial relationship. On the one hand, the model provides a map of the different

types of dyslexia that may exist. This has benefitted educational research and therapists alike in the common goal of improving screeners and remedial training. On the other hand, case studies of both acquired alexia and developmental dyslexia have been critical in refining the model.

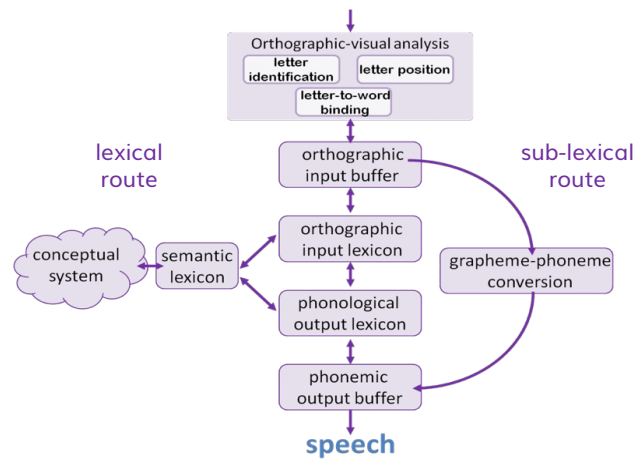


Figure 1, The Dual Route Model of lexical and sub-lexical reading (Friedmann & Coltheart, 2018).

Letter identification dyslexia

Abstract letter identification refers to the ability to recognize letters regardless of their spatial position, size, font or case. Rapid, parallel identification of the letters forming a written word is thought to rely on a series of visual processing stages that culminate in the left lateralized fusiform visual word form area (VWFA) (Cohen et al., 2003; Dehaene et al., 2004). Those processes can be impaired through lesions, as in the case of pure alexia, forcing the reader to recognize letters like other abstract shapes, through stronger reliance on generic object-recognition mechanisms, particularly in the right-hemispheric visual system (Cohen et al., 2004; Finkbeiner et al., 2006).

Neglect dyslexia

In neglect dyslexia, also called neglexia the reader may omit and substitute letters in reading, but unlike in letter identification dyslexia, single letter identity is recognized. Errors are predominantly observed on one side of the word or the other, and particularly for written words compared to other recognition tasks. The reader is aware of the presence of letters but is unable to identify them (i.e. reading 'yellow' as 'pillow', an example from Ellis et al., 1987). Neglexia has been relatively well documented in adult cognitive neuropsychology (for a

review, see Vallar et al., 2010), but few developmental cases with children have been reported (for examples in Hebrew, see Friedmann & Nachman-Katz, 2004).

Attentional dyslexia

Eye-tracking experiments reveal that in expert readers, each eye fixation collects information of letter and word shape from approximately 10-character positions to the right of fixation (McConkie & Rayner, 1975). Word information is first gathered at the peripheral level, before a subsequent fixation brings the word into the fovea. Thus, as we read, multiple words are simultaneously processed (Snell & Grainger, 2019), but their processing is modulated by an attentional window that controls for letter-to-word binding. This theory of the attentional window in letter-to-word binding was the inspiration for naming a deficit characterized by letter migrations between words as attentional dyslexia (for example, reading 'win fed' as 'fin fed', Shallice & Warrington, 1977). Since then, a report of developmental attentional dyslexia has also been described in English (Rayner et al., 1989) and several in Hebrew (Friedmann et al., 2010). The majority of migrations between words maintain their serial position within the new word. This is important because it underscores a difference between coding for position with-in words and

Letter position dyslexia

Evidence of the importance of letter position encoding comes from our ability to distinguish anagrams such as 'diary' and 'dairy'. Since both words contain the same letters, it is their position that permits correct reading. Letter position dyslexia refers to patients who make errors that respect the identity of the letters, but switch their locations, often resulting in the transposition of nearby letters within the same word (e.g., 'form' instead of 'from'). Importantly, this type of error generally occurs when the transposition creates an existing word, as in the aforementioned example. Letter position dyslexia appears to be a problem of under-specification of letter positions in a word and not mis-identification of absolute position (Davis, 2010). A conclusion supported by masked priming experiments in lexical-decision tasks (Davis & Bowers, 2006). In masked priming, preceding a target word with an orthographically similar letter string facilitates response time to target words, compared to unrelated letter string primes (Ferrand & Grainger, 1992, 1993). Crucially, masked priming trials in which the prime is a letter-transposition version of the target (e.g. 'bagde'-->'badge') produces faster response times to targets than an orthographic substitution control (e.g. 'barle'-->'badge'; see

(Perea & Lupker, 2004). Cases of selective developmental letter-position dyslexia have been described in readers of Hebrew (Friedmann & Rahamim, 2007), English (Kohnen et al., 2012), Arabic (Friedmann & Haddad-Hanna, 2012) and Turkish (Güven & Friedmann, 2019).

Visual dyslexia

Visual dyslexia refers to a deficit that causes letter substitution errors in word reading (e.g. reading 'hand' as 'band') that cannot be accounted for by letter position, attentional, or neglect errors. It is also different from letter identification dyslexia, because single letter identification is intact. Case studies of acquired visual dyslexia have been reported (Zoccolotti et al., 1999), but the evidence is rare in developmental cases and confounded with other visual deficits. For example, visual dyslexia has been reported coinciding with a more general developmental deficit in perceiving the location and orientation of visual stimuli (McCloskey & Rapp, 2000).

Surface dyslexia

Surface dyslexia refers to an impairment in the lexical route that pushes the reader to overly depend on slow sub-lexical reading. The resulting outcome is a difficulty in reading irregular words and a frequent over-regularizing of infrequent grapheme-phoneme combinations (e.g. reading 'know' as 'now'). Pseudoword and regular word reading is unaffected in cases of pure surface dyslexia. In highly transparent languages, such as Italian, surface dyslexia is identified as slow reading (Zoccolotti et al., 1999). Cases of pure developmental surface dyslexia have been reported in English (Castles, 1996; Castles & Coltheart, 1993), French (Valdois et al., 2003), Greek (Sotiropoulos & Hanley, 2017) and Hebrew (Friedmann & Lukov, 2008).

Phonological dyslexia

The mirror image of surface dyslexia is phonological dyslexia. This is reported as a deficit in reading tasks that depend upon grapheme-phoneme decoding such as reading novel and pseudowords. Because phonics instruction is generally the first step to literacy, phonological dyslexia generally makes learning to read slow, as all words must be memorized by their orthographic form (Castles & Coltheart, 1993). Phonological dyslexia was first discussed in the case of a patient with alexia that was able to read fluently, but unable to read pseudowords despite of an intact ability to orally repeat and write said items (Beauvois & Derouesne, 1979). Phonological dyslexia is often argued as stemming from an underlying phonological deficit in

language processing as supported by reports of poor phoneme processing and manipulation deficit that underlies the majority of dyslexic cases (Bradley & Bryant, 1983; Bruck et al., 1997; Landerl et al., 2013; Ramus, 2003; Saksida et al., 2016; Wagner & Torgesen, 1987). A finding also supported by the unique importance of oral phoneme manipulation in explaining early literacy skills (Piquard-Kipffer & Sprenger-Charolles, 2013; Torgesen et al., 1997). The argument that the phonological deficit stems from poor phoneme processing has found criticism in the evidence that learning to read appears to be a driver in improved phoneme processing, in other words, dyslexic readers may avoid reading, which therefore leads to poor phoneme processing skills compared to normal readers.

According to the Dual Route Model, phonological dyslexia can surface as either a deficit in input or in output stages of reading. In the case of a deficit in input to the sub-lexical path, the reader can fluently read words stored in the mental lexicon but has difficulty reading novel or pseudowords due to an impairment in grapheme-phoneme conversion (see the developmental case study of Campbell & Butterworth, 1985). In the case of a deficit in the output buffer, the deficit lies in the short-term memory buffer that holds phonological units until their production and assembles phonemes into a word. Individuals with aphasia following brain damage who have impairment in the phonological output buffer demonstrate difficulty in reading aloud long pseudowords or words with complex morphology, as well as difficulties in their oral repetition, as this stage is involved in language as well as reading (Dotan & Friedmann, 2015).

Deep dyslexia

Finally, whether reading by the lexical or sub-lexical route, the ultimate goal of reading is to access meaning. In the rare cases of deep dyslexia, see (Temple, 2006) the deficit is characterized by semantic errors (e.g. reading 'sand' as 'beach'). Developmental deep dyslexia has been reported as occurring in English in children with Williams Syndrome (Temple, 2003, 2006), in Japanese (Yamada, 1995) and Arabic (Friedmann & Haddad-Hanna, 2014a).

Screening for various types of dyslexia

The predominance of the Dual Route model to describe expert reading has led it to also influence how developmental dyslexia is screened. Typical screeners used by neuropsychologists and speech therapists require the reader to read separate word lists of

regular words, irregular word and pseudo words (for examples of screeners in English, see Castles et al., 2009; Parkin, 2018, and in French, Jacquier-Roux et al., 2002). These screeners diagnose dyslexia based on the number of errors made (but not the type of errors) as well as reading speed. The logic is that if the locus of impairment is in the lexical path, demonstrating surface dyslexia, the reader will struggle to read the irregular words. If the dyslexia is phonological, resulting from a deficit in sub-lexical reading, then pseudowords will be impacted. Our criticism with this type of screener is that, as we have seen in the description of the Dual Route Model, it does not take into account the previously reported subtypes of dyslexia, notably, in the orthographic visual analysis stage. Furthermore, as seen in the examples from the previous descriptions of dyslexia, errors can only be classified when the right type of stimuli are provided. This is not the case in screeners that only focus on overall accuracy. Screeners that do not take into the account the rich evidence from case studies in dyslexia research risk biasing teachers and therapists to provide remediation tasks, such as phoneme awareness training, that do not in fact treat the locus of the deficit.

The Malabi project

The present approach, motivated by the Dual Route model and by previous evidence for selective dyslexia detection when appropriate stimuli are provided, was based on the Tiltan reading battery in Hebrew (Friedmann & Gvion, 2003). Instead of measuring overall accuracy and speed, our French screener (christened Malabi, after the eponymous delicious dessert that the last two authors were eating during the design of this project) measures the type of errors made by the dyslexic reader. When the reader shows an abnormal number of errors in one of the predefined categories, and in the absence of errors outside of the normal range in other categories, we may speak of a selective deficit. To this end, the screener contains stimuli carefully tailored to elicit different types of errors. Each word in the screener is strategically selected in order to increase the likelihood of eliciting errors characteristic of different subtypes of dyslexia. In particular, we ensure that, if the reader has a specific type of dyslexia, the error may result in the production of a plausible alternative word (e.g., reading 'signe' (sign) as 'singe' (monkey) in letter position dyslexia or 'orgue' (organ) as 'orge' (barley) in surface dyslexia). This strategy is like laying 'traps' for the deficit. Since falling into the trap produces a word, the reader is less likely to self-correct. The Malabi screener is an oral reading

test that includes three subtests: 161 single words (2–8 letters long, $M = 5.12$, $SD = 1.29$), 40 pseudowords (4–5 letters long, $M = 4.88$, $SD = 0.92$), and 40 word-pairs (2–9 letters long, $M = 5.16$, $SD = 1.62$). Table 1 describes the type of word traps in the Malabi screener.

Table 1, Descriptions of the different types of dyslexia screened by the Malabi. For each type of dyslexia, examples of word 'traps' in French are presented.

Dyslexia type and corresponding stimuli in the Malabi test	Description	'Trap' examples word -> possible error
Orthographic visual analysis stage		
Attentional 44 word-pairs allowing for letter migrations	Errors in reading consist in the migration of letters from neighboring words. Migrations can come from encroachments in the visual field either vertically or horizontally. The position of the letter remains intact.	Word pairs: 'balle selle' -> 'salle belle'
Letter Position 44 single words and 22 pseudo words allowing for letter transpositions	Errors in letter transpositions within words so that a new word is read. Transpositions are also possible in pseudo words.	Single words: volé -> vélo Pseudo-words: fotre -> forte
Neglect 80 single words allowing for left and 60 allowing for right neglect 13 pseudowords with possible neglect	Errors consist in deleting or substituting the left side of a word (or the right side in case of right neglect)	Single words: ruse -> use or muse
Visual All 161 single words	Includes both letter-position and attentional errors, plus omissions, substitutions and additions of letters in reading.	Single words: mille -> bille
Sub-lexical stage		
Phonological deficit 40 single words 40 pseudowords	Errors made in reading new or pseudo words. Reading from the mental lexicon is intact. If the deficit is buffer specific, word length or complex morphology modulate difficulty.	Single words: marcherions -> marchons All pseudowords
Vowel 73 single word 20 pseudo words	Visual errors specific to pseudowords read via the sub-lexical route. If combined with surface dyslexia, then vowel-errors occur also in reading words via the sub-lexical route.	Pseudowords : troche -> triche
Lexical stage		
Surface 97 single words	Errors in reading irregular words. Dependency on grapheme-phoneme conversion leads to a tendency to regularize less frequent sounds.	Single words : vise /viz/ -> visse /vis/
Conceptual stage		
Deep 40 unambiguously imaginable single words	Characterized by reading aloud another word of the same category, morphological errors and difficulty with reading functional or ambiguous words.	Single Words : Boulangerie->pain

The goal of the current project was to assess if the Malabi screener could detect selective subtypes of reading deficits. We report the first cases, to our knowledge, of developmental attentional and letter position dyslexia in French. We then provide an in-depth analysis of the factors that appear to modulate the deficit for these readers, demonstrating a double dissociation between them. This research was carried out in two phases:

Experimental screening, Our first goal was to assess if we could spot selective deficits exposed by the categories of traps in the Malabi screener, in a dyslexic population. To do this, we established norms for the different error categories made by normal readers in junior-high school. We chose the junior-high population in order to ensure a population that should have had enough years in school to have mastered the alphabetic code of French and have a large orthographic lexicon. Our dyslexic population was chosen from a school that had already vetted the students for an absence of other IQ and cognitive deficits. We present the results of this dyslexic population and focus on the students that appear to have a selective impairment. Traditional tests for dyslexia were also used to assess a phonological deficit.

Experimental investigation, Having identified putative cases of selective dyslexia according to the Malabi, our next goal was to revisit these students with a second battery of tests designed to replicate the original diagnosis, clarify the locus of the deficit, and understand the factors that modulate reading for these individuals.

Experimental screening

Methods

Dyslexic participants

Our dyslexic participants all came from CERENE (www.cerene-education.fr) a specialized school for children with developmental disorders (dyslexia, dyspraxia, dyscalculia, etc.) located in five different locations in Paris, and providing 2nd to 8th grade education. Our reason for working with this particular school was to exclude dyslexic readers with other cognitive or environmental confounds that could influence our measures. Admission to CERENE requires a stringent neuropsychological assessment to ensure normal IQ and specific understanding of the student's profile. Dyslexia can often exist in comorbidity with other deficits, such as attention disorders and dyspraxia (Pauc, 2005). The school provides a high-quality education through trained teachers for learning disorders, small classes and providing students with tools that help them compensate for their deficit (i.e., electronic readers that use larger font with spacing and highlighting, audio instructions, adapted instructions, etc.). Furthermore, the school has speech therapists on-site that work individually with every student to ensure targeted training. Testing students at CERENE thereby provided us with a

sample of readers whose difficulty could not be due to low intelligence or a lack of targeted practice. Sixteen students were sent to us by the school (6th grade = 9, 7th grade = 7). All of these students were French natives and in a normal age range for their grade.

Control group for establishing screener norms

The control sample consisted of 138 children from six classes in two junior high schools (in France, junior high begins in 6th grade). Both schools were in towns about an hour outside of Paris (average population for the towns is ~60,000). The schools were situated in lower to middle socio-economic areas. Children that had not been in the French school system for a minimum of four years, or that were receiving special help in language or reading skills, as reported by the school director, were not included. Sample demographics are provided in Table 2.

School grade	Number of participants	Sex (Male)	Age Range years:months
6th	77	43	11:3 – 13:0
7th	61	41	12:3 – 13:5

Procedure

This project received ethics approval from the Comité d'Évaluation Éthique d'Établissement (C3E) de l'Université Paris-Saclay, (February 2018, project number 28). All parents of the participating students were sent a letter explaining the study and given the possibility of 'opting-out' if they did not want their child to participate. Each participant was tested individually in a quiet room in the school. Each of the three sub-tests was printed on paper in 14pt Calibri font, with vertical double spacing between words. No time limit was imposed during testing, but children were timed and told to try and read quickly but accurately. A short break was taken between each of the three tests was presented in a fixed order (single word, pseudoword, word-pair reading). No feedback was given during reading. The testing was carried out over a 3-month period at the end of the French school year. Students reading errors were phonetically noted. Two testers coded all of the error types. For each type of reading error, an official 'error code' was attributed by agreement between three researchers in the project. This method allowed us to eventually automate the error-coding process.

Establishing norms

To obtain norms on the number of errors for each category assessed in the Malabi, we calculated the mean number and standard deviation for each of the error types in our coding scheme. Outliers, defined as students that made a number of errors that was more than 3 standard deviations away from the mean, were removed. After removal, a new mean and standard deviation were calculated, providing our normed data. For each dyslexia type, we report the number of errors a patient would have to make in order to be significantly below the control group ($p < 0.05$) using (Crawford & Howell, 1998).

Combining the number of errors from all three subtests in the screener (single words, pseudowords and word pairs), we did not find a difference in the number of errors between the two different schools, $F(1, 136)=1.44$, $p=0.23$, nor between genders, $F(1, 136)= 0.01$, $p=0.94$. Summary statistics for accuracy in reading aloud from the three tests, for each grade, are presented in Table X. The number of errors made improved by grade, but this was only significant in reading single words, $F(1, 136)= 7.6$, $p=0.01$, and word pairs, $F(1, 136)= 4.66$, $p=0.03$. No difference was found between groups in reading pseudo words, $F(1, 136)=0.8$, $p=0.37$. This is probably explained by the sample of junior high school kids that have had many years of school and reading experience. Their decoding skills had probably plateaued, but their reading of irregular or infrequent words was still affected by age.

Table 3, Mean (SD) number of errors for each error type in each test.						
	6th grade, N=77			7th grade, N=61		
	Single Words N= 161	Pseudo Words = 40	Word Pairs = 88	Single Words = 161	Pseudo Words = 40	Word Pairs = 88
Total Avg Errors	13.18 (10.94)	5.19 (5.19)	4.40 (4.04)	8.72 (7.09)	4.44 (4.50)	3.08 (2.87)
Attention	0.77 (1.28), 3	0.24 (0.49), 2	2.41 (1.95), 6	0.67 (0.88), 3	0.27 (0.55), 2	1.78 (1.62), 5
Letter Position	1.78 (2.66), 7	2.15 (2.60), 7	0.12 (0.33), 1	1.27 (2.11), 5	1.98 (2.57), 7	0.00 (0.00), 1
Neglect	0.79 (1.15), 3	0.16 (0.37), 1	0.40 (0.67), 2	0.60 (0.94), 3	0.16 (0.37), 1	0.31 (0.62), 2
Visual	4.07 (4.27), 12	3.43 (3.53), 10	1.22 (1.40), 4	2.42 (2.71), 7	2.83 (3.10), 9	0.78 (1.09), 3
Phonological	1.06 (1.33), 5	1.03 (1.20), 4	-	0.75 (1.15), 4	0.78 (0.83), 3	-
Vowel	4.12 (4.43), 12	1.52 (1.89), 5	1.94 (1.76), 5	3.07 (3.01), 9	1.21 (1.48), 4	1.23 (1.17), 4
Surface	7.24 (4.82), 16	-	0.76 (1.21), 3	5.05 (3.70), 12	-	0.53 (0.92), 3
Deep	0.09 (0.28), 1	-	0.53 (0.66), 2	0.08 (0.28), 1	-	0.35 (0.55), 2

Numbers in bold/italic are the number of errors needed for a patient to be considered different from the controls according to Crawford and Howell's one-way t-test ($p < 0.05$).

Screening for a selective deficit

Malabi screener

A selective dyslexia was demonstrated by a student making significantly more errors in one of the eight error types established by the normed data, and in the absence of a significantly greater number of errors of another error-type. Only one exception was made to this rule which was the case of a large amount of surface dyslexia errors in concert with another type of error. Surface errors are due to an over-reliance on decoding, which can often lead to regularizing words. In most cases of dyslexia, this is attributed to a deficit in the direct lexical route, which forces students to read via the decoding route and therefore lead to regularization. In other words, readers might produce the most frequent phoneme for a particular grapheme instead of a less frequent rule-based or irregular phoneme. For instance, a common mistake in French would be reading the letter 's' in the word 'vise' as its more common sound /s/ sound (as in the word "vis") instead of following the rule that a single 's' between two vowels is pronounced /z/. Children learn to integrate these rules with reading experience. Given that most children with dyslexia develop strategies to avoid reading, regularization could be due to a lack of practice as opposed to a genuine case of surface dyslexia. In agreement with this hypothesis, note that normal readers also made more surface errors on average than any other type of error. This was most likely the result of not having fully integrated all rule-based grapheme-phoneme correspondences.

Alouette

The "Alouette" reading test (Lefavrais, 2005) is a nonsense text containing many existing but rare words. It forces the reader to read each word carefully since they cannot rely on meaning to extract word identity. Readers are asked to read the text as quickly and accurately as possible in no more than 3 minutes. A combined score of number of words read and time is collected, which can be compared to standardized scores for grade level.

Oral language and phonological awareness

To assess oral language and phonological awareness skills, we presented students with a pseudoword repetition and two phoneme manipulation tasks taken from the Odédys dyslexia test battery (Jacquier-Roux et al., 2002). In the pseudoword repetition task, 20 items were orally presented. The students had to repeat each one. Pseudowords items were between 2

to 5 syllables. In the first phoneme manipulation task, the listener was orally presented a word and had to remove the first phoneme and say aloud the new word (for the word 'brame', respond 'rame') In the second task, two words were orally presented, and the listener had to remove the first phoneme from each word and say aloud the syllable of these two items combined (for the words 'photo – artistique', responds 'fa'). Ten items for each of the phoneme manipulation tasks were presented. For all three tasks, the number of correct responses was recorded. Two trials with feedback were given before starting.

Results

Did the Malabi identify readers with a selective deficit?

The number of errors for each participant is reported in a table in the annex (Table A.1). Four participants were in the normal range for all errors. Another five participants made many errors across many of the different categories in the three tests. Testers noted that, in general, reading was extremely difficult for these students, making their errors often difficult to encode and categorize. These manifestations of multiple dyslexia were outside of the scope of the current project, where we focused on selective deficits.

Seven children were retained because their scores deviated significantly from the controls for a particular error type, suggesting a possible selective deficit: TR, LP and LD for attentional dyslexia and JB, PO and HW for letter position dyslexia. We now describe them in turn.

Patient data for selective dyslexia candidates

Patient TR. Reading was in the normal error range for all error types in single word reading.

On the pseudoword reading test, TR's number of letter transposition errors was at 2.26 SD, which neared the Crawford's significance. However, several of these letter transpositions errors could also be explained as migration of letters from vertically presented words on the list. On the word pair list, TR made a large number of migration errors, equivalent to 4.56 SD from the mean. TR also had a tendency to make more surface errors on this sub-test, equivalent to 1.88 SD from the mean.

Patient LP. Reading errors for all error types in the single word and pseudoword reading tests were in the normal range. In the word pair test, specifically designed with word traps for

horizontal migration errors, LP made 6 errors, equivalent to 1.93 SD from mean and just reaching Crawford's significance.

Patient LD. Reading was very fluid, but migration errors were made on all three tests. The number of errors was not significant in the single word and pseudo word reading tests. Errors in these tests stemmed from vertical migrations. For example, the word 'varie' was read 'marie'. We assumed the 'm' migration came from the word 'morte' located two words previously read on the list. The number of horizontal migration errors made on the word pair tests (e.g., 'fous tour' -> 'tour four' and 'masse cesse' -> 'masse casse') was equivalent to 4.46sd from the mean.

Patient JB. On the single word test, significantly more errors of letter transpositions were made (e.g., magner -> manger; patrie -> parti, cirer -> crier), equivalent to 2.5 SD from the mean. A similar outcome was seen on the pseudoword list, with the number of errors at 3.9 SD from the mean. There was also a significant number of errors in vowel letters on the pseudoword test, but over half were made from letter transpositions. The others were made from dropping double occurrences of a vowel in a word (e.g., anical -> /anicl/). Omitting double letters in a word has been previously reported for Hebrew cases of letter position deficit (Friedmann & Rahamim, 2007). These authors hypothesized that when position encoding is ambiguous, and the only distinguishing feature of a letter is its identity, which can lead to one of the letters to be dropped. Two neglect errors were also produced by JB in the word pair list, but it was unclear if these were truly representative of this category, or just regularizations. For example, the word 'cesse' was read 'ce' and 'puis' was read 'pu'. In French, the letters 'e' and 's' are often silent on the end of words.

Patient PO. Similarly, PO also made significantly more letter transposition errors on the single word test (e.g., poil -> poli; calque -> claque), an equivalent of 5.1 SD from the mean. Nineteen surface errors were also made on this test, which was at Crawford's significance for this error type. A large number of transpositions were also made on the pseudoword list, equal to 3.5 SD from the mean. On the word pair list, PO did not produce a significant number of errors in any of the categories. Importantly this list did not contain words with possible transpositions.

Patient HW. HW did not make a significant number of errors in any of the error types in either the single word or word pair reading subtests. HW's reading of the pseudowords was markedly strained, with the majority of errors scored as letter transposition errors, for a score landing 2.5 SD from the mean. These results possibly support a difficulty in letter position encoding at

the level of the orthographic input buffer, as information from the lexicon appears to have sufficed for a correct reading of words but not pseudowords.

Patient JC. The only reader that made a significant number of errors attributed to central reading skills was JC. On the single word reading test, JC made surface errors on 20% of words with a possible word forming regularized letter (e.g., reading base /baz/ as basse /bas/). JC’s surface deficit was in the absence of a phonological impairment as pseudoword reading was normal. JC did not make a significant number of reading errors in the word pair list, which contained mostly short and frequent words.

Were these students considered dyslexic by a traditional screener?

None of the seven patients described above were considered at risk for dyslexia by the Alouette reading test, (see Table 4, a deficit is considered at <-2SD from the mean). This may seem surprising, but it should be taken into account that all of these students were relatively fluid in reading, even in the Malabi screener. Their errors were few, but selective. However, all of their scores, except for JB, were between the 25th and 50th percentile on the Alouette. This is actually low for students coming from such a highly supportive school system with small classes and personalized training. We would argue that these results are a credit to the school. JB’s percentile was above the mean.

Could their dyslexia be explained by a phonological deficit?

All students performed within the normal range on the tasks testing oral language and phonological awareness (see Table 4, a deficit is considered at <-2SD from the mean). This was an important finding, as a low score in any one of these tests would have ruled out a selective deficit stemming from the visual orthographic stage.

Table 4, Participant scores on standardized French tests for dyslexia detection.					
	grade	Alouette percentile	pseudoword repetition +* N=20 M= 19., SD=1	phoneme fusion* N=10 M=7.5, SD=2.4	phoneme suppression* N=10 M=8.3, SD=2.0
JB	6	(50, 75)	19	9	8
LP	6	[25, 50)	20	9	9
JC	6	[25, 50)	20	8	8
HW	7	[25, 50)	20	8	8
LD	7	[25, 50)	20	10	10

TR	7	[25, 50)	20	7	8
PO	7	[25, 50)	20	7	8

+ Standardized scores are not provided after 5th grade, when children are at ceiling. This was clearly the case for all of our participants.

* Standardized scores are provided for the 7th grade (the screener has not been normed for the 6th grade).

Experimental investigation of attentional and letter position dyslexia profiles

Participants

Having found seven tentative cases of specific dyslexia types using the Malabi screener, we recontacted these students to replicate the findings and shed light on the factors that modulate reading errors. Three of the students who made attentional (migration) errors (LD, TR, LP) and one of the students with letter-position (transposition) errors (JB) on the initial Malabi screener were available. Six control students from Paris schools, matched to the earliest grade-year (8th grade) of our dyslexic participants, were also tested for comparison.

Materials

Attentional dyslexia was addressed by a new list of 528 word pairs. 468 pairs had at least one possible letter migration that would make a new word, and 160 pairs did not have a migratable letter (see Table 5 for details). Pairs of words always had the same number of letters, between 5 and 7 letters. All pairs and their inverse were presented in the list (e.g., lent vois, lent vois) as well as their migratable pair (e.g., vent lois, lent vois), but never on the same page. As all word pairs permitted a double letter swap and each pair and its opposite were presented, we could measure the direction of migration (e.g., reading sale fois as soie fois or sale fais). The words ranged in frequency between 0 and 8,296 per million, $M=158$ (frequencies taken from the Lexique French database; New et al., 2004). There was no difference in frequency between migratable and non-migratable pairs $p=0.38$. There was no difference in frequency between migratable and non-migratable words ($p=0.38$) and no difference in frequency between the position of the letter migration nor in the number of different letters between words.

Table 5, Types of word pairs in the 528 word pair list targeting attentional errors								
	number of letters	number of different letters between words	position of migration	number of items	Target pair	Possible migration		
Migratable N=468	4	all	First	20	vent lois	lent vois		
			Middle	20	sale fois	sole fais		
			Last	20	aime vole	aimé vole		
		one	First	20	loin soin	soin loin		
			Middle	20	fuit fait	fait fuit		
			Last	20	aimé aime	aime aimé		
		two	first_last	12	fini mine	mini, fine fine, mini		
			first_middle	12	poux deux	doux,oux peux, doux		
			middle_last	12	dura dire	dira, dure dure, dira		
		5	all	First	20	tasse pendu	passe tendu	
				Middle	20	repué laver	revue laper	
				Last	20	monté gagna	monta gagné	
	one		First	20	banal canal	canal banal		
			Middle	20	douce douze	douze douce		
			Last	20	marié marie	marie marié		
	two		first_last	12	serai feras	ferai, seras seras, ferai		
			first_middle	12	belle salle	selle, balle balle, selle		
			middle_last	12	bougé boude	boudé, bouge bouge, boudé		
	6		all	First	20	réparé soutes	séparé routes	
				Middle	20	bottés lampes	bottes lampés	
				Last	20	acheté pleura	acheta pleuré	
		one	First	20	otages étages	éttages otages		
			Middle	20	places plages	plages places		
			Last	20	devine deviné	deviné devine		
		two	first_last	12	drainé graine	grainé, draine draine, grainé		
			first_middle	12	centre vendre	ventre, cendre cendre, ventre		
			middle_last	12	marges manger	manges, marger manger, manges		
		Non-Migratable N=60	4			20	joli gros	
			5			20	balai taper	
			6			20	reflet occupé	

Letter-position dyslexia was addressed in a single-word reading-aloud test comprising 406 new words. 304 of the words had a possible letter transposition that would make a new word and 102 words did not have such an anagram (see Table 6 for details). Each word and its

anagram were present in the list. The words ranged in frequency between 0 and 14,662 per million, $M=151$ (frequencies taken from Lexique). For each word and its anagram, of the two, the high frequency word occurrence per million was $M=68$ ($SD=157$) and low frequency word occurrence per million was $M=10$ ($SD=26$). There was no difference in frequency between words with a possible transposition and words that did not allow for a transposition, $p=0.59$.

Type of possible transposition	Number of items	Examples from the test
Middle migrations		
Adjacent consonant-consonant (CC) migration	26	congé -> cogné
Adjacent including a vowel (VV or CV) migration	110	loin -> lion
Non-adjacent consonant-consonant migration (C-C)	6	préside -> prédise
Non-adjacent including vowel migration (C-V, V-V)	14	foule -> floue
Exterior migrations		
Adjacent consonant-vowel (CV) migration	68	lier -> lire
Non-adjacent consonant-consonant migration (C-C)	66	coude -> douce
Non-adjacent including vowel migration (C-V, V-V)	14	pela -> pale
No letter transpositions possible	102	maison

Statistical Analysis

Crawford and Howell's significance t-test was used to compare the performance of each of the experimental participants with performance of the control group (Crawford & Garthwaite, 2002; Crawford & Howell, 1998). In the descriptions of each participants' responses in a test, within-participant comparisons between two conditions were conducted using chi-squared tests (two-tailed comparison). When applicable, comparisons between conditions at the group level were conducted using a Wilcoxon signed-rank test (results reported with T, the minimum sum of ranks).

Results

Double dissociation between letter transpositions within words and letter migrations between words

We first sought to confirm the existence of a double dissociation, suggested by the initial Malabi screener, between attentional dyslexia (causing letter migrations between words, as observed in participants LD, TR and LP) and letter-position dyslexia (causing letter transpositions within words, as observed in participant JB). To test this, we had all four students read both the word pair test with 468 migratable words and the single word test with 304 letter transposition opportunities. The results of the four participants and controls are in Table 7, together with statistics. The findings were clear-cut. First, all of the attentional deficit cases made a significant greater number of migration errors on the word pair test than transposition errors on the single word test. They also made a greater number of errors than the control group and the participant screened for letter position dyslexia. Second, conversely, the opposite picture surfaced for patient JB: he was the only participant to make a significantly greater number of transposition errors than of migration errors and differed from controls only in his abnormal percentage of transpositions.

In the case of the migration errors in the attentional test, letter swaps were always possible between both words (e.g., *passee tendu* --> *tasse pendu*), however, we were also aware that the reader may only make a migration of one letter to the other word (e.g., *passee tendu* -> *passee pendu*), possibly due to a migration to the more frequent word. We counted each migration made, meaning that in the first example, two errors were counted and in the second example, one error was counted. Double migrations represented 2% of LD's errors, 4% of LP's errors and 15% of TR's errors.

Table 7, Percentage of errors in dyslexic participants and the six controls					
	LD Attentional Dyslexia	LP Attentional dyslexia	TR Attentional dyslexia	JB Letter-position dyslexia	Controls
Between word migrations in the Attentional Dyslexia test	11	15	16	3	2.2 (sd=1.6)
Within word transpositions in the Letter Position Dyslexia test	1	3	1	10	1.7 (sd=1.6)
migrations errors made in the attentional dyslexia test vs. transpositions made in letter-position test	$\chi^2=7.79$ $p<0.01^{**}$	$\chi^2=8.80$ $p0.01^{**}$	$\chi^2=14.47$ $p<0.001^{***}$	$\chi^2=4.03$ $p<0.05^*$	$\chi^2=0.07$ $p=0.798$
migrations errors made in the attentional dyslexia test compared to control	$t=4.20,$ $p<0.01^{**}$	$t=6.11,$ $p<0.001^{***}$	$t=6.59,$ $p<0.001^{***}$	$t=0.86,$ $p=0.22$	
letter transposition errors made in the letter-position dyslexia test compared to control	$t=-0.20,$ $p=0.42$	$t=1.43,$ $p=0.12$	$t=-0.41,$ $p=0.35$	$t=6.33,$ $p<0.001^{***}$	

Detailed analysis of migration errors

Our data afforded a detailed analysis of the nature and causes of migration errors. We first noted that the deficit was highly selective. On the word pair test, in the three attentional dyslexics, migration errors far outnumbered any other types of errors that they made. LD made only 1 error out of 51 (2%) that could not be explained as a migration, while LP made 8 errors (10%) and TR 10 errors (12%) that could not be explained by a migration. Also, all migration errors resulted in a real word, suggesting the presence of a lexical filter: participants struggled to produce real words, such that letter migrations that would result in a pseudoword were eliminated.

In order to understand what factors may modulate attentional dyslexia, we next examined which stimuli were most likely to induce migration errors.

Word length. We first looked at the impact of word length on the rate of between-word migrations in developmental attentional dyslexia. Previous research on attentional dyslexia in Hebrew have shown an effect of longer words leading to more between word migrations (Friedmann et al., 2010), a possible explanation being that more letters in the visual field may lead to more possibility for error. To assess this question, we always used word pairs that included two words of the same length. As described in Table 5, we used a systematic design that included 156 word-pairs for each of the lengths four, five and six letters. For each length, we calculated the number of between-word attentional errors (see Table 8). Unlike previous

results reported in Hebrew, we did not observe a length effect. Children were just as likely to produce a migration error if the words comprised 4, 5 or 6 letters.

Table 8, Percentage of between-word migrations in words of different lengths.						
	4- letters	5- letters	6- letters	4 vs. 5 letters	5 vs. 6 letters	4 vs. 6 letters
LD	9	6	11	$\chi^2 = 0.64$ $p = 0.42$	$\chi^2 = 1.61$ $p = 0.20$	$\chi^2 = 0.22$ $p = 0.64$
LP	16	9	15	$\chi^2 = 2.24$ $p = 0.13$	$\chi^2 = 1.71$ $p = 0.19$	$\chi^2 = 0.04$ $p = 0.85$
TR	13	10	14	$\chi^2 = 0.44$ $p = 0.51$	$\chi^2 = 0.75$ $p = 0.38$	$\chi^2 = 0.04$ $p = 0.84$
average	12.67%	8.33%	13.33%	$\chi^2 = 0.28$ $p = 0.87$	$\chi^2 = 0.22$ $p = 0.90$	$\chi^2 = 0.19$ $p = 0.91$

Migration position. It has been previously observed that Hebrew readers, who read right to left, make more attentional errors on the final (left-most) letter of words (Friedmann & Rahamim, 2007). Furthermore, there is evidence that other orthographic visual-analyzer dyslexia types are more sensitive to making errors on certain positions. For example, neglect dyslexia is generally categorized as omissions or substitution systematically on one-side or the other of words. In the literature, reading errors typically occur on the left side (Ellis et al., 1987; Reznick & Friedmann, 2015; Haywood and Coltheart 2001). Letter-position errors, meanwhile, occur more frequently in middle positions (Friedmann & Rahamim, 2007; Güven & Friedmann, 2019; Kohnen et al., 2012). In this context, we were interested in seeing if French children with attentional dyslexia would also make more errors on the last letter. The 528-word pairs include 192 pairs with a potential for first letter migration, 192 pairs with a potential for middle letter migration, and 192 pairs with a potential for final letter migration. In each of these cases (first, middle, last) 120 pairs only permitted a migration in one position, while 72 had multiple migration positions. Our results (Table 9) showed that the final letter (the rightmost letter), like in Hebrew, was more likely to be affected by a migration. All three participants made significantly more errors on the last letter compared to the first letter, but no difference when comparing the first and the middle letter. Two readers produced significantly more errors on the last letter than on a middle letter.

Table 9, Percentage of between-word migrations of first, middle, and final letters.						
	first	middle	final	first vs middle	middle vs final	first vs final
LD	4	9	17	$\chi^2 = 1.41$ $p = 0.23$	$\chi^2 = 2.41$ $p = 0.12$	$\chi^2 = 7.04$ $p < 0.01^{**}$
LP	7	10	27	$\chi^2 = 0.31$ $p = 0.58$	$\chi^2 = 8.59$ $p < 0.01^{**}$	$\chi^2 = 11.66$ $p < 0.001^{***}$
TR	7	11	25	$\chi^2 = 0.68$ $p = 0.41$	$\chi^2 = 5.98$ $p = 0.02^*$	$\chi^2 = 10.21$ $p = 0.01^{**}$
average	6%	10%	23%	$\chi^2 = 1.09$ $p = 0.30$	$\chi^2 = 6.13$ $p = 0.01$	$\chi^2 = 11.66$ $p < 0.001$

Migration direction. We next looked at the issue of the directionality of the migrations. We were interested in knowing if the migration was more likely to result from a carry-over from the first word onto the next (perseveration), as previously found in developmental attentional dyslexia in Hebrew (Friedmann et al., 2010), or in the converse direction (anticipation). We looked at the number of times a migration appeared to travel in one of these directions. In a previous report in Hebrew (Friedmann et al., 2010), a letter was more likely to travel from one word to the next (e.g. ‘cat hub’ was more commonly misread as ‘cat hut’ than as ‘cab hub’). This provides a possible argument against attentional dyslexia as a purely visual problem since the migration could stem from a perseveration in a memory buffer that carries information from one word to the next (Cohen & Dehaene, 1998), and which could affect an orthographic stage, but also a later phonological stage. If the migrations stemmed from a visual-attentional deficit at the orthographic analyzer stage, however, they might conceivably occur in the opposite direction, i.e., as encroachments from letters that have not been read, but are already in the peripheral visual field and therefore compete within the current attentional window, while the visual system is attempting to bind letters to the currently fixated word. For our participants, we scored the direction of movement for each migration error. All word pairs permitted both migrations from a previous or from a following word. As both a pair and its inverse were presented, the direction of the migration would not be affected by a movement towards a migration that would lead to reading a more frequent word. Summarized in Table 10, indeed showed that LD and TR made significantly more right-to-left errors, i.e., anticipations of letters from the forthcoming word. For example, ‘ceux pela’ was read ‘cela pela’, ‘réparé soutes’ as ‘séparé soutes’ and ‘dons cric’ as ‘donc cric’. No significant difference between the two directions was found for LP.

The observation of anticipations is surprising, particularly since previous work in Hebrew found that participants with migration errors made twice as many perseveration errors than anticipation errors (Friedmann et al., 2010). Note, however, that the next word appears to the left of the page in Hebrew; thus, both findings might be putatively reconciled as an effect, onto the currently fixated word, of the word that lies on the right side of the page (and which therefore has the most direct projection to the left-hemispheric visual word form area; see Cohen et al., 2000; Molko et al., 2002). Whether or not this interpretation is correct, the present findings in French patients are important in that they highlight attentional dyslexia as a truly visual-attentional deficit as opposed to a mere consequence of perseveration.

Table 10, Number of migrations exclusively explained by			
	perseverance	anticipation	perseverance vs. anticipation
LD	33%	67%	$\chi^2 = 23.12$ $p=0.001^{***}$
LP	43%	57%	$\chi^2 = 3.92$ $p=0.05^*$
TR	35%	65%	$\chi^2 = 18$ $p=0.001^{***}$
average	37%	63%	$\chi^2 = 13.52$ $p=0.001^{***}$

Number of shared letters and opportunities for errors. Our list of word pairs also systematically manipulated the similarity of the two words: it comprised pairs of words differing in only one, two or all letters. 180-word pairs existed for the category of 1- or all-different letters, and 108-word pairs with 2- different letters (see Table 5 for examples). This variable could impact reading in several different ways. On the one hand, shared letters increase the similarity between words and may therefore increase the chances that they are confused. This hypothesis finds support in experiments that flash two words simultaneously in the visual field. In these cases, an ‘illusory word’ made from the two words combined is more likely to appear the more similar the word pair (Davis & Bowers, 2006). On the other hand, when two words share more letters in the same position, migrations between them may go undetected, thus reducing the apparent number of errors. In the final analysis, what may count is simply the number of opportunities for errors that respect the lexicality filter. In our stimuli, the word pairs differing by 1 letter (e.g., ‘douce douze’) offered, by definition, a

single opportunity for migration, but so did the majority word pairs differing by all letters (e.g. *passee tendu* → *tasse pendu*; for all other letter swaps, e.g. *tassee pensee*, at least one of the resulting strings was not a word). The word pairs differing by 2 letters, however, offered two opportunities for errors (e.g., in ‘*belle salle*’, swapping the initial consonants *b* and *s* results in ‘*selle balle*’, and swapping the vowels *e* and *a* result in ‘*balle selle*’). Thus, if this was the dominant factor, then word pairs differing by two letters should yield more migration errors.

Since we did not find a length effect earlier, we collapsed the data across word lengths to look at differences created by the number of different letters. The results provided modest evidence in favor of the “opportunities for error” account (see Table 11). LD was not affected by the similarity between words. LP tended to make migrations when there were 2 different letter different, and significantly fewer errors when only 1 letter was different. Similarly, TR made significantly more errors when 2 letters were different, and significantly fewer errors when only one letter was different. All-and-all, patients made more errors when the letters differed by two letters, and when averaged over patients, those word pairs yielded approximately twice as many migrations as the other two categories.

Table 11, Percentage of between-word migrations for word pairs with 1, 2 or all different letters						
	1 different letter (canal banal)	2 different letters (belle salle)	all different letters (passee tendu)	1 vs. others	2 vs. others	all vs. others
LD	8%	15%	8%	$\chi^2 = 0.70$ $p < 0.40$	$\chi^2 = 2.4$ $p = 0.12$	$\chi^2 = 0.70$ $p < 0.40$
LP	14%	21%	7%	$\chi^2 = 0$ $p = 1$	$\chi^2 = 3.03$ $p = 0.08$	$\chi^2 = 5.13$ $p < 0.02^*$
TR	8%	24%	13%	$\chi^2 = 4.79$ $p < 0.03^*$	$\chi^2 = 6.38$ $p = 0.01^*$	$\chi^2 = 0.36$ $p = 0.55$
average	10.33%	20%	9.33%	$\chi^2 = 86$ $p < 0.35$	$\chi^2 = 4.08$ $p = 0.04^*$	$\chi^2 = 1.59$ $p = 0.21$

Lexical status. Taken together, those analyses provided a strong case for LD, LP and TR exhibiting an attentional deficit originating in the visual orthographic stage and causing letter migrations. Since this stage is prelexical, migration errors should show up in both lexical and sub-lexical reading. Other studies of orthographic visual deficits, such as letter position dyslexia, have also observed an equal sensitivity to pseudoword and word reading (Friedmann

& Rahamim, 2007; Güven & Friedmann, 2019; Kohnen et al., 2012). To see if we could capture a similar result in sub-lexical reading, we asked all three participants to read a list of 72 pseudoword pairs, the presentation of which was the same as the real word test. 54 of the pseudo word pairs had a migratable letter (e.g., 'sogne timbe' could be read as the words 'signe tombe'), while 18 did not. Pairs were matched in length and ranged in 3, 4 or 5 letters.

Detailed analysis of transposition errors

As with the participants with attentional dyslexia, we also took a deeper look into the factors that impacted the likelihood of a transposition error in participant JB, and that had already been assessed in other languages.

Out of the 48 errors JB made on the single word reading test, 33 of them could be explained by letter transpositions, 13 could be explained as the regularization of non-frequent sounds and 2 as other errors. All in all, this meant that 69% of JB's errors could be explained by letter transpositions, thus revealing the great selectivity of this deficit. Furthermore, 88% of the letter-transposition errors were lexical.

Our next goal was to see what factors in the stimuli of the single word test were more likely to spur an error. The list was designed to probe several factors: exteriority (are transpositions more frequent between letters in the middle of the word, or between letters including at least one letter on the exterior), adjacency (are transpositions more likely to occur when two letters are next to each other), frequency (is the reader more likely to transpose letters when they yield a more frequent word) and consonant-vowel status.

For the following analysis, we focus on the transposition errors that were made on designated trap words, thus excluding two of JB's non-lexical transpositions that occurred on words that were not specifically put in the test as traps. Words with 3-letters (N=14) were removed from the analysis because no errors were made on these small words, and they did not allow for an analysis of exterior versus middle errors.

Middle versus exterior letters. It has also been observed that transpositions are more likely to involve middle letters (Friedmann & Gvion, 2001; Friedmann & Haddad-Hanna, 2012; Friedmann & Rahamim, 2007; Güven & Friedmann, 2019; Kohnen et al., 2012). The theory is that the position of these letters may be less precisely coded, causing greater potential for transposition than the exterior letters. We examined if this was also the case with JB as a French speaker. We compared the rates of letter transposition for stimuli affording a

transposition of middle letters (N=156 stimuli) versus those affording a transposition including at least one exterior letter (N=134). Our results showed that, like previous cases, JB made significantly more migrations of middle-letters (17%) than exterior letters (3%), $\chi^2 = 10.88$, $p < 0.001$.

Adjacent versus non-adjacent letters. Another factor that had been previously shown to modulate letter transpositions is whether or not the letters transposed are adjacent (e.g., plier-piler) or not (e.g., paveur-vapeur): adjacent letters are more frequently transposed (Friedmann & Gvion, 2001; Güven & Friedmann, 2019). Our list comprised 196 words whereby the potential transposition was between adjacent letters, and 94 non-adjacent possibilities. Our results showed that 100% of JB's letter transposition concerned adjacent letters.

Word frequency. It has been previously noted that letter transpositions errors are far more likely to result in a more frequent word than the original stimulus. The hypothesis is that if the positions of letters are ambiguously encoded, and such partial information is combined with a Bayesian prior based on word frequency (e.g. Norris, 2006), then the best-matching lexical entry is likely to be the more frequent word. Since in our list, all words and their transposition were included (N=145 each), we could easily test this theory by looking at lexical errors and assessing if JB was more likely to respond with the more frequent word. Our results indicated that JB was more likely to convert a low frequency word to its higher frequency anagram (14%) than vice versa (3%), $\chi^2 = 7.78$, $p = 0.005$. This finding replicates prior observations in Hebrew (Friedmann & Rahamim, 2007), Turkish (Güven & Friedmann, 2019) and in English (Kohnen et al., 2012), and further emphasizes the importance of choosing appropriate error-eliciting stimuli in a screener for different subtypes of dyslexia.

Consonants versus vowels. Transpositions can affect two consonants (e.g., linge->ligne), two vowels (e.g., lion->loin) or a consonant and a vowel (e.g., porche->proche). Previous research with developmental dyslexia in school age children in English did not find a difference between transpositions with only consonants versus the inclusion of at least one vowel. In another study with adults, errors were more likely to be made with consonant-consonant transpositions (Perea & Lupker, 2004). In Hebrew, researchers also failed to find a "consistent tendency" (Friedmann & Rahamim, 2007). This is important, because it rules the possibility of a specific vowel deficit, as reported for instance by Khentov-Kraus & Friedmann (2011). In our own list, 92 words allowed for a consonant-consonant transposition and 198-word allowed for transposition that included a vowel (either a vowel-vowel or consonant-vowel

transposition; note that only 28 words allowed for VV transpositions, so this category was considered too small for separate analysis). Similar to the results in other languages, JB showed an equal tendency in transposing consonants and vowels transpositions, making an equal percentage of errors in both categories (9%).

Lexicality. If the transposition of letters occurs during the orthographic-visual analysis stage, then even if the reader knows that they are reading pseudowords, the competition created from the presence of nearby words stored in the lexicon should give rise to a bias towards words, i.e., a significantly greater number of words read correctly relative to pseudowords, compared to an unimpaired normal reader. Based on this hypothesis, we presented a list of 72 written items to the participant for lexical decision. Half of the items in the list were existing words, the other half were pseudowords matched for length. Of the 36 pseudo words, 18 pseudowords offered a possible middle transposition that yielded a real word, while the other 18 pseudowords did not. The participants were given the list (presented in 14pt Calibri font) and asked to silently read the stimuli and circle all of the ‘real words’. JB, and the controls, were all left alone to read silently and finish the task at their own pace.

Control participants made very few errors on this test. Four out of six controls did not make any errors at all. One student failed to circle one ‘word’ and another student incorrectly circled one ‘transposition’ pseudoword. JB’s results provided a stark contrast. Similar to the control group, JB responded correctly to the non-transposition pseudowords, but he incorrectly circled six of the pseudowords with a possible transposition. These pseudowords (with their possible anagram) were: *reprati* (*reparti*), *pocile* (*police*), *blacon* (*balcon*), *voclan* (*volcan*), *sorucil* (*sourcil*) and *matreau* (*marteau*). In other words, JB incorrectly swapped letters on 33% of the pseudowords in which a transposition was possible. Conversely, two real words were missed. Table 12 presents JB’s results in comparison to the control group.

Table 12 Percentage of errors in a silent lexical decision task			
Lexical decision (N=72)	JB	Control	patient vs. control
<i>words, N=36</i>	6	0.46 (1.13)	t=4.00, p=0.08
<i>migratable pseudowords, N=18</i>	33	0.93 (2.27)	t=2.07, p=0.05*
<i>non-migratable pseudowords, N=18</i>	0	0	-

Discussion

In summary, we developed a novel screener capable of eliciting different types of reading errors in French, the Malabi, and we used it, as well as further tests, to obtain evidence for the existence of dissociable deficits corresponding to developmental attentional dyslexia (3 patients) and developmental letter-position dyslexia (1 patient) in French. The four cases we report resemble other previous cases of selective developmental deficit in surface (Castles, 1996; Friedmann & Lukov, 2008; Zoccolotti et al., 1999) and phonological dyslexia (Campbell & Butterworth, 1985) in that the readers are impaired only in specific reading contexts (Sotiropoulos & Hanley, 2017). However, whereas selective deficits have historically focused on central reading processes as described by the Dual Route Model, the present research provides additional evidence supporting the locus of deficit in the visual orthographic analyzer. For example, we found that our attentional dyslexics were more likely to make errors of letter encroachment from the right visual field, than of perseveration from previous stimuli. Our findings help specify the conditions that are most likely to yield errors. For attentional dyslexics, this is the presence, on the right side of fixation, of another competing word whose final letters may be switched with those of the current target word. For letter-position dyslexics, it is the existence, within the target word, of two nearby letters that, if transposed, yield another word of higher frequency.

This work, importantly, highlights the need to examine several key factors in orthographic analysis of reading models. In the case of attentional dyslexia, this means taking into account the attentional window which selectively amplifies information from a target word and filters out the letter information arising from other words, particularly on the right side of fixation. In this respect, our results concur with considerable prior research suggesting the existence of a non-conscious processing of parafoveal words (McConkie & Rayner, 1975; Rayner, 1975) and the importance of selective attention during reading (Facoetti et al., 2006, 2008; Franceschini et al., 2012; Peyrin et al., 2011; Vidyasagar & Pammer, 2010). In the case of migration errors, it may be that the attentional window that relegates specific letters to a word is less certain than in normal readers. In other word, causing letters to hold their position but jump between words. Considering letter position dyslexia, models should account for results that could arise from an abnormal uncertainty about letter position, which would result in an ambiguous coding of the incoming letter string, particularly for letters inside the word.

Note that even normal readers make transposition errors, particularly under conditions of impoverished inputs (e.g. with flashed and masked priming word) and more likely for middle letter primes (Perea & Lupker, 2003). These results are compatible with spatial coding models of reading and especially models of the orthographic stage, that differentiate encoding for letter identity and the relative position of letters, demonstrated by a cumulative benefit in masked primes that share a letter and position, share a letter in a position once removed and share neither letter or position (Davis & Bowers, 2006).

Transposition errors are also compatible with the existence of neurons tuned to intermediate-size units such as bigrams (letter pairs) (Perea & Lupker, 2003), as long as these units have large and therefore relatively imprecise receptive fields and therefore tend to also respond to misordered bigrams. However, recent evidence from both behavioral and brain-imaging studies is that a bigram stage may not be required to account for fluent word reading – rather, the information flow may proceed directly from letter position coding to lexical access (Agrawal et al., 2019, 2020; Woolnough et al., 2020). Most interestingly, the dominant effect of reading acquisition seems to be to refine the positional accuracy with which nearby letters are encoded (Agrawal et al., 2019). Letter-position dyslexia would occur at precisely that stage where each letter must be precisely bound to a position, and which is strictly needed in order to avoid confusing anagrams.

In both subtypes of dyslexia, the presence of a lexical bias (almost all of the errors are plausible words) and of a frequency bias (errors tend to be of higher frequency than the original stimuli) suggests that visual dyslexics partially compensate their deficit by over-relying on their input lexicon, using a lexical prior based on the lexicality and frequency of different strings in the target language, exactly as would be optimal for a “Bayesian reader” (Norris, 2006). This finding, as well as the double dissociation between transposition and migration errors, exemplify the fact that developmental deficits can be very “pure”, selectively affecting a particular stage of the orthographic identification process while leaving subsequent lexical stages intact.

Future Directions

There are two primary future directions. First and foremost, we present very few case studies in this paper. This makes the evidence for the factors that influence errors disputable. More cases should be found to support the selective deficit theory. Second, we need to assess

if providing targeted training can reduce sensitivity to the deficit. In the majority of cases, dyslexia does not make reading impossible, only difficult. The difficulty faced then has an impact on whether or not the child spends time reading or avoiding it as much as possible. Previous research, although scant has shown that remedial training, or tactics that alleviate the deficit, have been shown to work and may provide a way to get the learner to engage in reading. For example, auditory processing and language training has been shown to improve cases of dyslexia that stem from phonological processing deficits (Alexander et al., 1991; Temple et al., 2003), including through video games that support grapheme-phoneme associating (Lyytinen et al., 2007). Providing a cut-out window that can be passed from word to word while reading stops migrations from neighboring words and improves fluency (Friedmann et al., 2010; Rayner et al., 1989). Crowding of nearby letters is a frequent source of errors in beginner readers and struggling readers (Martelli et al., 2009; Pelli et al., 2007), who may benefit from the s p a c i n g of letters (Zorzi et al., 2012). This simple adjustment to text also seems to be a factor in alleviating letter transpositions (Friedmann & Rahamim, 2014).

Another avenue to pursue will also be the consequences for brain imaging and genetics for cases of specific deficits. The existence of very different subtypes of dyslexia, with doubly-dissociated performances on specific tests, should not be surprising, neither from the perspective of classical adult cognitive neuropsychology where many such dissociations have been attested (Beauvois & Derouesne, 1979; Campbell & Butterworth, 1985) nor from the perspective of multiple-route models of reading (Coltheart et al., 2001; Coltheart, 2005), which clearly point to the possibility that many different impairments may result in a reading deficit. However, it is ripe of consequences for large-scale studies of the genetic or neurological basis of those deficits, which often treat “dyslexia” as a single entity. By doing so, they run the risk of identifying very broad genetic risk factors, for instance related to intelligence or education (Gialluisi et al., 2020) or, in the case of brain imaging, the generic consequences of non-proficient reading rather than the specific causes of the reading deficit (Dehaene et al., 2015b; Maisog et al., 2008; Martin et al., 2015; Rueckl et al., 2015). The present results suggest that a careful analysis of behavior, based on the existence of distinct types of errors, should precede, guide and facilitate the mechanistic understanding of the various causes of reading impairments, as well as improve the efficiency of their rehabilitation.

One place to start would be the testing for shared selective deficits in cases where dyslexia runs in families, as there is strong evidence for the familial transmission (Cardon et al., 1994; DeFries et al., 1978; Gialluisi et al., 2020; Pennington et al., 1991; Plomin et al., 1997).

Conclusion

The results of this current study support the existence of selective deficits stemming from the orthographic visual analysis stage of reading by showing a double dissociation in errors made by dyslexic participants on the Malabi screener and on follow-up tests, specifically designed with word traps to expose selective deficits. Importantly, this work in French adds to the growing body of languages presenting evidence for specific dyslexia types with a possible origins in the orthographic analysis stage (see, for instance, in Hebrew Friedmann et al., 2010; Friedmann & Gvion, 2001; Friedmann & Rahamim, 2007; Arabic : Friedmann & Haddad-Hanna, 2014; English: Ellis et al., 1987; Kohnen et al., 2012; Turkish: Güven & Friedmann, 2019; Italian: Lavelli et al., 2019). It is possible that the majority of dyslexia stems from underlying language processing deficits or combinations of deficits (only 43% of the dyslexic participants were noted as having a selective deficit as measured by the Malabi). This work also highlights the importance of developing dyslexic screeners that include specific traps to detect selective deficits, as opposed to just screening for the number of correct responses and reading speed, as is generally done in the majority of brain-imaging or genetic studies. None of the patients in our sample of highly trained dyslexic students were considered dyslexic on the traditional French test for dyslexia. Understanding dyslexia through word traps and error types will in turn help researchers and practitioners to provide improved remediation tactics that are tailored to helping the individual get around their impairment.

References

- Agrawal, A., Hari, K., & Arun, S. (2020). A compositional neural code in high-level visual cortex can explain jumbled word reading. *eLife*, 9, e54846. <https://doi.org/10.7554/eLife.54846>
- Agrawal, A., Hari, K. V. S., & Arun, S. P. (2019). Reading Increases the Compositionality of Visual Word Representations. *Psychological Science*, 0956797619881134. <https://doi.org/10.1177/0956797619881134>
- Alexander, A. W., Andersen, H. G., Heilman, P. C., Voeller, K. K. S., & Torgesen, J. K. (1991). Phonological awareness training and remediation of analytic decoding deficits in a group of severe dyslexics. *Annals of Dyslexia*, 41(1), 193-206. <https://doi.org/10.1007/BF02648086>
- Araújo, S., Faisca, L., Bramão, I., Petersson, K. M., & Reis, A. (2014). Lexical and Phonological Processes in Dyslexic Readers : Evidence from a Visual Lexical Decision Task: Lexical and Phonological Processes in Dyslexia. *Dyslexia*, 20(1), 38-53. <https://doi.org/10.1002/dys.1461>
- Beauvois, M. F., & Derouesne, J. (1979). Phonological alexia : Three dissociations. *Journal of Neurology, Neurosurgery & Psychiatry*, 42(12), 1115-1124. <https://doi.org/10.1136/jnnp.42.12.1115>
- Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—A causal connection. *Nature*, 301(5899), 419-421. <https://doi.org/10.1038/301419a0>
- Bruck, M., Genesee, F., & Caravolas, M. (1997). A cross-linguistic study of early literacy acquisition. In *Foundations of reading acquisition and dyslexia : Implications for early intervention*.
- Campbell, R., & Butterworth, B. (1985). Phonological Dyslexia and Dysgraphia in a Highly Literate Subject : A Developmental Case with Associated Deficits of Phonemic Processing and Awareness. *The Quarterly Journal of Experimental Psychology Section A*, 37(3), 435-475. <https://doi.org/10.1080/14640748508400944>
- Cardon, L., Smith, S., Fulker, D., Kimberling, W., Pennington, B., & DeFries, J. (1994). Quantitative trait locus for reading disability on chromosome 6. *Science*, 266(5183), 276-279. <https://doi.org/10.1126/science.7939663>

- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, 91(1), 77-111.
- Castles, Anne. (1996). Cognitive Correlates of Developmental Surface Dyslexia : A Single Case Study. *Cognitive Neuropsychology*, 13(1), 25-50.
<https://doi.org/10.1080/026432996382051>
- Castles, Anne, & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47(2), 149-180. [https://doi.org/10.1016/0010-0277\(93\)90003-E](https://doi.org/10.1016/0010-0277(93)90003-E)
- Castles, Anne, Coltheart, M., Larsen, L., Jones, P., Saunders, S., & McArthur, G. (2009). Assessing the basic components of reading : A revision of the Castles and Coltheart test with new norms. *Australian Journal of Learning Difficulties*, 14(1), 67-88.
<https://doi.org/10.1080/19404150902783435>
- Cohen, L., & Dehaene, S. (1998). Competition between past and present. Assessing and explaining verbal perseverations. *Brain*, 121, 1641-1659.
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M. A., & Michel, F. (2000). The visual word form area : Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123, 291-307.
- Cohen, L., Martinaud, O., Lemer, C., Lehéricy, S., Samson, Y., Obadia, M., Slachevsky, A., & Dehaene, S. (2003). Visual Word Recognition in the Left and Right Hemispheres : Anatomical and Functional Correlates of Peripheral Alexias. *Cerebral Cortex*, 13(12), 1313-1333. <https://doi.org/10.1093/cercor/bhg079>
- Cohen, Laurent, Henry, C., Dehaene, S., Martinaud, O., Lehéricy, S., Lemer, C., & Ferrieux, S. (2004). The pathophysiology of letter-by-letter reading. *Neuropsychologia*, 42(13), 1768-1780. <https://doi.org/10.1016/j.neuropsychologia.2004.04.018>
- Coltheart, M, Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC : A dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1), 204-256.
- Coltheart, Max. (2005). Modeling Reading : The Dual-Route Approach. In M. J. Snowling & C. Hulme (Éds.), *The science of reading : A handbook* (p. 6-23). Blackwell Publishing.
<https://doi.org/10.1002/9780470757642.ch1>

- Coslett, H. B., & Turkeltaub, P. (2016). Chapter 63—Acquired Dyslexia. In G. Hickok & S. L. Small (Éds.), *Neurobiology of Language* (p. 791-803). Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00063-8>
- Cuetos, F., & Ellis, A. W. (1999). Visual Paralexias in a Spanish-Speaking Patient with Acquired Dyslexia : A Consequence of Visual and Semantic Impairments? *Cortex*, 35(5), 661-674. [https://doi.org/10.1016/S0010-9452\(08\)70826-8](https://doi.org/10.1016/S0010-9452(08)70826-8)
- Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, 117(3), 713-758. <https://doi.org/10.1037/a0019738>
- Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding : Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3), 535-557. <https://doi.org/10.1037/0096-1523.32.3.535>
- DeFries, J. C., Singer, S. M., Foch, T. T., & Lewitter, F. I. (1978). Familial nature of reading disability. *The British Journal of Psychiatry: The Journal of Mental Science*, 132, 361-367. <https://doi.org/10.1192/bjp.132.4.361>
- Dehaene, S., Cohen, L., Sigman, M., & Vinckier, F. (2005). The neural code for written words : A proposal. *Trends Cogn Sci*, 9(7), 335-341.
- Dehaene, S., Jobert, A., Naccache, L., Ciuciu, P., Poline, J. B., Le Bihan, D., & Cohen, L. (2004). Letter binding and invariant recognition of masked words : Behavioral and neuroimaging evidence. *Psychol Sci*, 15(5), 307-313.
- Dehaene, Stanislas, Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate : Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, 16(4), 234-244. <https://doi.org/10.1038/nrn3924>
- Dehaene, Stanislas, Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359-1364. <https://doi.org/10.1126/science.1194140>
- Dotan, D., & Friedmann, N. (2015). Steps towards understanding the phonological output buffer and its role in the production of numbers, morphemes, and function words. *Cortex*, 63, 317-351. <https://doi.org/10.1016/j.cortex.2014.08.014>

- Elbro, C., & Jensen, M. N. (2005). Quality of phonological representations, verbal learning, and phoneme awareness in dyslexic and normal readers. *Scandinavian Journal of Psychology*, 46(4), 375-384. <https://doi.org/10.1111/j.1467-9450.2005.00468.x>
- Elbro, C., & Petersen, D. K. (2004). Long-Term Effects of Phoneme Awareness and Letter Sound Training : An Intervention Study With Children at Risk for Dyslexia. *Journal of Educational Psychology*, 96(4), 660-670. <https://doi.org/10.1037/0022-0663.96.4.660>
- Ellis, A. W. (2016). *Reading, Writing and Dyslexia (Classic Edition) : A Cognitive Analysis*. Psychology Press. <https://doi.org/10.4324/9781315669991>
- Ellis, A. W., Flude, B. M., & Young, A. W. (1987). "Neglect dyslexia" and the early visual processing of letters in words and nonwords. *Cognitive Neuropsychology*, 4(4), 439-464. <https://doi.org/10.1080/02643298708252047>
- Evans, B. J. W., & Allen, P. M. (2016). A systematic review of controlled trials on visual stress using Intuitive Overlays or the Intuitive Colorimeter. *Journal of Optometry*, 9(4), 205-218. <https://doi.org/10.1016/j.optom.2016.04.002>
- Facoetti, A., Ruffino, M., Peru, A., Paganoni, P., & Chelazzi, L. (2008). Sluggish engagement and disengagement of non-spatial attention in dyslexic children. *Cortex*, 44(9), 1221-1233.
- Facoetti, A., Zorzi, M., Cestnick, L., Lorusso, M. L., Molteni, M., Paganoni, P., Umiltà, C., & Mascetti, G. G. (2006). The relationship between visuo-spatial attention and nonword reading in developmental dyslexia. *Cognitive Neuropsychology*, 23(6), 841-855. <https://doi.org/10.1080/02643290500483090>
- Ferrand, L., & Grainger, J. (1992). Phonology and Orthography in Visual Word Recognition : Evidence from Masked Non-Word Priming. *The Quarterly Journal of Experimental Psychology Section A*, 45(3), 353-372. <https://doi.org/10.1080/02724989208250619>
- Ferrand, L., & Grainger, J. (1993). The time course of orthographic and phonological code activation in the early phases of visual word recognition. *Bulletin of the Psychonomic Society*, 31(2), 119-122. <https://doi.org/10.3758/BF03334157>
- Finkbeiner, M., Almeida, J., & Caramazza, A. (2006). Letter identification processes in reading : Distractor interference reveals an automatically engaged, domain-specific mechanism. *Cognitive Neuropsychology*, 23(8), 1083-1103. <https://doi.org/10.1080/02643290600665778>

- Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facchetti, A. (2012). A Causal Link between Visual Spatial Attention and Reading Acquisition. *Current Biology*, 22(9), 814-819. <https://doi.org/10.1016/j.cub.2012.03.013>
- Friedmann, N., Biran, M., & Gvion, A. (2012). Patterns of visual dyslexia : Patterns of visual dyslexia. *Journal of Neuropsychology*, 6(1), 1-30. <https://doi.org/10.1111/j.1748-6653.2011.02000.x>
- Friedmann, N., & Coltheart, M. (2018). 35 .Types of developmental dyslexia. In E. Dattner & D. Ravid (Éds.), *Handbook of Communication Disorders* (p. 721-752). De Gruyter. <https://doi.org/10.1515/9781614514909-036>
- Friedmann, N., & Gvion, A. (2001). Letter Position Dyslexia. *Cognitive Neuropsychology*, 18(8), 673-696. <https://doi.org/10.1080/02643290143000051>
- Friedmann, N., & Haddad-Hanna, M. (2012). Letter Position Dyslexia in Arabic : From Form to Position. *Behavioural Neurology; Hindawi*. <https://doi.org/10.3233/BEN-2012-119004>
- Friedmann, N., & Haddad-Hanna, M. (2014a). Types of Developmental Dyslexia in Arabic. In E. Saiegh-Haddad & R. M. Joshi (Éds.), *Handbook of Arabic Literacy* (Vol. 9, p. 119-151). Springer Netherlands. https://doi.org/10.1007/978-94-017-8545-7_6
- Friedmann, N., & Haddad-Hanna, M. (2014b). Types of Developmental Dyslexia in Arabic. In E. Saiegh-Haddad & R. M. Joshi (Éds.), *Handbook of Arabic Literacy* (Vol. 9, p. 119-151). Springer Netherlands. https://doi.org/10.1007/978-94-017-8545-7_6
- Friedmann, N., Kerbel, N., & Shvimer, L. (2010). Developmental attentional dyslexia. *Cortex*, 46(10), 1216-1237. <https://doi.org/10.1016/j.cortex.2010.06.012>
- Friedmann, N., & Lukov, L. (2008). Developmental surface dyslexias. *Cortex*, 44(9), 1146-1160. <https://doi.org/10.1016/j.cortex.2007.09.005>
- Friedmann, N., & Nachman-Katz, I. (2004). Developmental Neglect Dyslexia in a Hebrew-Reading Child. *Cortex*, 40(2), 301-313. [https://doi.org/10.1016/S0010-9452\(08\)70125-4](https://doi.org/10.1016/S0010-9452(08)70125-4)
- Friedmann, N., & Rahamim, E. (2007). Developmental letter position dyslexia. *Journal of Neuropsychology*, 1(2), 201-236. <https://doi.org/10.1348/174866407X204227>
- Friedmann, N., & Rahamim, E. (2014). What can reduce letter migrations in letter position dyslexia? : TREATMENT OF LETTER POSITION DYSLEXIA. *Journal of Research in Reading*, 37(3), 297-315. <https://doi.org/10.1111/j.1467-9817.2011.01525.x>

- Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2008). The Long Road to Automation : Neurocognitive Development of Letter–Speech Sound Processing. *Journal of Cognitive Neuroscience*, 21(3), 567-580. <https://doi.org/10.1162/jocn.2009.21061>
- Gialluisi, A., Andlauer, T. F. M., Mirza-Schreiber, N., Moll, K., Becker, J., Hoffmann, P., Ludwig, K. U., Czamara, D., Pourcain, B. S., Honbolygó, F., Tóth, D., Csépe, V., Huguet, G., Chaix, Y., Iannuzzi, S., Demonet, J.-F., Morris, A. P., Hulstender, J., Willcutt, E. G., ... Schulte-Körne, G. (2020). Genome-wide association study reveals new insights into the heritability and genetic correlates of developmental dyslexia. *Molecular Psychiatry*, 1-14. <https://doi.org/10.1038/s41380-020-00898-x>
- Güven, S., & Friedmann, N. (2019). Developmental Letter Position Dyslexia in Turkish, a Morphologically Rich and Orthographically Transparent Language. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02401>
- Jacquier-Roux, M., Valdois, S., & Zorman, M. (2002). ODEDYS: Outil de dépistage des dyslexies. Laboratoire Cogni-Sciences.
- Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When ‘slime’ becomes ‘smile’ : Developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681-3692. <https://doi.org/10.1016/j.neuropsychologia.2012.07.016>
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O’Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J.-F., ... Schulte-Körne, G. (2013a). Predictors of developmental dyslexia in European orthographies with varying complexity : Cross-linguistic predictors of dyslexia. *Journal of Child Psychology and Psychiatry*, 54(6), 686-694. <https://doi.org/10.1111/jcpp.12029>
- Landerl, K., Ramus, F., Moll, K., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., O’Donovan, M., Williams, J., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Chaix, Y., Démonet, J.-F., ... Schulte-Körne, G. (2013b). Predictors of developmental dyslexia in European orthographies with varying complexity : Cross-linguistic predictors of dyslexia. *Journal of Child Psychology and Psychiatry*, 54(6), 686-694. <https://doi.org/10.1111/jcpp.12029>
- Lavelli, M., Barachetti, C., Majorano, M., Florit, E., Brotto, C., & Miottello, P. (2019). Impacts of a shared book-reading intervention for Italian-speaking children with developmental

- language disorder. *International Journal of Language & Communication Disorders*, 54(4), 565-579. <https://doi.org/10.1111/1460-6984.12460>
- Lefavrais, P. (2005). Alouette-R, Test d'analyse de la lecture et de la dyslexie.
- Lyytinen, H., Ronimus, M., Alanko, A., Poikkeus, A.-M., & Taanila, M. (2007). Early identification of dyslexia and the use of computer game-based practice to support reading acquisition. *Nordic Psychology*, 59(2), 109-126. <https://doi.org/10.1027/1901-2276.59.2.109>
- Maisog, J. M., Einbinder, E. R., Flowers, D. L., Turkeltaub, P. E., & Eden, G. F. (2008). A meta-analysis of functional neuroimaging studies of dyslexia. *Ann N Y Acad Sci*, 1145, 237-259. <https://doi.org/10.1196/annals.1416.024>
- Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. *Journal of Vision*, 9(4), 14.1-18. <https://doi.org/10.1167/9.4.14>
- Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults : A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping*, 36(5), 1963-1981. <https://doi.org/10.1002/hbm.22749>
- McCloskey, M., & Rapp, B. (2000). A Visually Based Developmental Reading Deficit. *Journal of Memory and Language*, 43(2), 157-181. <https://doi.org/10.1006/jmla.2000.2724>
- McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, 17(6), 578-586. <https://doi.org/10.3758/BF03203972>
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read : A meta-analytic review. *Psychological Bulletin*, 138(2), 322-352. <https://doi.org/10.1037/a0026744>
- Molko, N., Cohen, L., Mangin, J. F., Chochon, F., Lehéricy, S., Le Bihan, D., & Dehaene, S. (2002). Visualizing the neural bases of a disconnection syndrome with diffusion tensor imaging. *Journal of Cognitive Neuroscience*, 14, 629-636.
- Monzalvo, K., & Dehaene-Lambertz, G. (2013). How reading acquisition changes children's spoken language network. *Brain and Language*, 127(3), 356-365. <https://doi.org/10.1016/j.bandl.2013.10.009>
- Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition*, 24, 45-64.

- Norris, D. (2006). The Bayesian reader : Explaining word recognition as an optimal Bayesian decision process. *Psychol Rev*, 113(2), 327-357.
- Parkin, J. R. (2018). Wechsler Individual Achievement Test–Third Edition Oral Language and Reading Measures Effects on Reading Comprehension in a Referred Sample. *Journal of Psychoeducational Assessment*, 36(3), 203-218. <https://doi.org/10.1177/0734282916677500>
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *J Vis*, 7(2), 20 1-36. <https://doi.org/10.1167/7.2.20>
- Pennington, B. F., Gilger, J. W., Pauls, D., Smith, S. A., Smith, S. D., & DeFries, J. C. (1991). Evidence for Major Gene Transmission of Developmental Dyslexia. *JAMA*, 266(11), 1527-1534. <https://doi.org/10.1001/jama.1991.03470110073036>
- Perea, M., & Lupker, S. J. (2003). Transposed-letter confusability effects in masked form priming. In *Masked Priming : State of the art* (1re éd., p. 97-120). Psychology Press. <https://doi.org/10.4324/9780203502846>
- Peterson, R. L., & Pennington, B. F. (2015). Developmental Dyslexia. *Annual Review of Clinical Psychology*, 11(1), 283-307. <https://doi.org/10.1146/annurev-clinpsy-032814-112842>
- Peyrin, C., Demonet, J. F., N’Guyen-Morel, M. A., Le Bas, J. F., & Valdois, S. (2011). Superior parietal lobule dysfunction in a homogeneous group of dyslexic children with a visual attention span disorder. *Brain and Language*, 118(3), 128-138. <https://doi.org/10.1016/j.bandl.2010.06.005>
- Piquard-Kipffer, A., & Sprenger-Charolles, L. (2013). Early predictors of future reading skills : A follow-up of French-speaking children from the beginning of kindergarten to the end of the second grade (age 5 to 8). *L’Année psychologique*, 113(04), 491-521.
- Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading : Computational principles in quasi-regular domains. *Psychol Rev*, 103(1), 56-115.
- Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, Nurture, and Cognitive Development from 1 to 16 Years : A Parent-Offspring Adoption Study. *Psychological Science*, 8(6), 442-447. <https://doi.org/10.1111/j.1467-9280.1997.tb00458.x>
- Ramus, F. (2003). Developmental dyslexia : Specific phonological deficit or general sensorimotor dysfunction? *Current Opinion in Neurobiology*, 13(2), 212-218. [https://doi.org/10.1016/S0959-4388\(03\)00035-7](https://doi.org/10.1016/S0959-4388(03)00035-7)

- Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., & Frith, U. (2003). Theories of developmental dyslexia : Insights from a multiple case study of dyslexic adults. *Brain*, 126(4), 841-865. <https://doi.org/10.1093/brain/awg076>
- Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, 25(2), 482-503. <https://doi.org/10.1037/0096-1523.25.2.482>
- Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, 7(1), 65-81. [https://doi.org/10.1016/0010-0285\(75\)90005-5](https://doi.org/10.1016/0010-0285(75)90005-5)
- Rayner, K., Murphy, L. A., Henderson, J. M., & Pollatsek, A. (1989). Selective attentional dyslexia. *Cognitive Neuropsychology*, 6(4), 357-378.
- Rueckl, J. G., Paz-Alonso, P. M., Molfese, P. J., Kuo, W.-J., Bick, A., Frost, S. J., Hancock, R., Wu, D. H., Mencl, W. E., Duñabeitia, J. A., Lee, J.-R., Oliver, M., Zevin, J. D., Hoeft, F., Carreiras, M., Tzeng, O. J. L., Pugh, K. R., & Frost, R. (2015). Universal brain signature of proficient reading : Evidence from four contrasting languages. *Proceedings of the National Academy of Sciences of the United States of America*, 112(50), 15510-15515. <https://doi.org/10.1073/pnas.1509321112>
- Saksida, A., Iannuzzi, S., Bogliotti, C., Chaix, Y., Démonet, J.-F., Bricout, L., Billard, C., Nguyen-Morel, M.-A., Le Heuzey, M.-F., Soares-Boucaud, I., George, F., Ziegler, J. C., & Ramus, F. (2016). Phonological skills, visual attention span, and visual stress in developmental dyslexia. *Developmental Psychology*, 52(10), 1503-1516. <https://doi.org/10.1037/dev0000184>
- Schaadt, G., Pannekamp, A., & van der Meer, E. (2013). Auditory phoneme discrimination in illiterates : Mismatch negativity—A question of literacy? *Developmental Psychology*, 49(11), 2179-2190. <https://doi.org/10.1037/a0031765>
- Shallice, T., & Warrington, E. K. (1977). The possible role of selective attention in acquired dyslexia. *Neuropsychologia*, 15(1), 31-41. [https://doi.org/10.1016/0028-3932\(77\)90112-9](https://doi.org/10.1016/0028-3932(77)90112-9)
- Snell, J., & Grainger, J. (2019). Readers Are Parallel Processors. *Trends in Cognitive Sciences*, 23(7), 537-546. <https://doi.org/10.1016/j.tics.2019.04.006>
- Sotiropoulos, A., & Hanley, J. R. (2017). Developmental surface and phonological dyslexia in both Greek and English. *Cognition*, 168, 205-216. <https://doi.org/10.1016/j.cognition.2017.06.024>

- Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends in Neurosciences*, 20(4), 147-152. [https://doi.org/10.1016/s0166-2236\(96\)01005-3](https://doi.org/10.1016/s0166-2236(96)01005-3)
- Temple, C. M. (2003). Deep dyslexia in Williams syndrome. *Journal of Neurolinguistics*, 16(6), 457-488. [https://doi.org/10.1016/S0911-6044\(01\)00044-6](https://doi.org/10.1016/S0911-6044(01)00044-6)
- Temple, C. M. (2006). Developmental and Acquired Dyslexias. *Cortex*, 42(6), 898-910. [https://doi.org/10.1016/S0010-9452\(08\)70434-9](https://doi.org/10.1016/S0010-9452(08)70434-9)
- Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., & Gabrieli, J. D. E. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation : Evidence from functional MRI. *Proceedings of the National Academy of Sciences*, 100(5), 2860-2865. <https://doi.org/10.1073/pnas.0030098100>
- Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of Phonological Awareness and Rapid Automatic Naming Ability to the Growth of Word-Reading Skills in Second-to Fifth-Grade Children. *Scientific Studies of Reading*, 1(2), 161-185. https://doi.org/10.1207/s1532799xssr0102_4
- Valdois, S., Bosse, M.-L., Ans, B., Carbonnel, S., Zorman, M., David, D., & Pellat, J. (2003). Phonological and visual processing deficits can dissociate in developmental dyslexia : Evidence from two case studies. *Reading and Writing*, 16(6), 541-572. <https://doi.org/10.1023/A:1025501406971>
- Valdois, S., Bosse, M.-L., & Tainturier, M.-J. (2004). The cognitive deficits responsible for developmental dyslexia : Review of evidence for a selective visual attentional disorder. *Dyslexia*, 10(4), 339-363. <https://doi.org/10.1002/dys.284>
- Vallar, G., Burani, C., & Arduino, L. S. (2010). Neglect dyslexia : A review of the neuropsychological literature. *Experimental Brain Research*, 206(2), 219-235. <https://doi.org/10.1007/s00221-010-2386-0>
- Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia) : What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, 45(1), 2-40. <https://doi.org/10.1046/j.0021-9630.2003.00305.x>
- Vidyasagar, T. R., & Pammer, K. (2010). Dyslexia : A deficit in visuo-spatial attention, not in phonological processing. *Trends Cogn Sci*, 14(2), 57-63. <https://doi.org/10.1016/j.tics.2009.12.003>

- Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192-212. <https://doi.org/10.1037/0033-2909.101.2.192>
- Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2020). Spatiotemporal Dynamics of Orthographic and Lexical Processing in the Ventral Visual Pathway. *BioRxiv*, 2020.02.18.955039. <https://doi.org/10.1101/2020.02.18.955039>
- World Health Organization. (2020). International statistical classification of diseases and related health problems (11th ed.). <https://icd.who.int/en>
- Yamada, J. (1995). Developmental Deep Dyslexia in Japanese : A Case Study. *Brain and Language*, 51(3), 444-457. <https://doi.org/10.1006/brln.1995.1070>
- Ziegler, J. C., Castel, C., Pech-Georgel, C., George, F., Alario, F. X., & Perry, C. (2008). Developmental dyslexia and the dual route model of reading : Simulating individual differences and subtypes. *Cognition*, 107(1), 151-178.
- Zoccolotti, P., De Luca, M., Di Pace, E., Judica, A., Orlandi, M., & Spinelli, D. (1999). Markers of developmental surface dyslexia in a language (Italian) with highgrapheme–phoneme correspondence. *Applied Psycholinguistics*, 20(2), 191-216. <https://doi.org/10.1017/S0142716499002027>
- Zorzi, M., Barbiero, C., Facchetti, A., Lonciari, I., Carrozzi, M., Montico, M., Bravar, L., George, F., Pech-Georgel, C., & Ziegler, J. C. (2012). Extra-large letter spacing improves reading in dyslexia. *Proc Natl Acad Sci U S A*, 109(28), 11455-11459. <https://doi.org/10.1073/pnas.1205566109>
- Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud ? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1131-1161.

Appendix

Table of number of errors for each error type made by students previously diagnosed as dyslexic

	single words									pseudowords									wordpairs								
	attentional	letter position	Neglect	Visual	phonological	vowel	Surface	deep	attentional	letter position	Neglect	visual	phonological	vowel	Surface	deep	attentional	letter position	Neglect	visual	phonological	vowel	Surface	deep			
7th grade																											
GR	1	0	0	2	1	3	7	1	0	4	0	5	0	0	0	0	4	0	0	2	6*	1	1	1			
AD	0	0	0	0	0	1	6	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0			
TF	0	0	0	0	0	1	2	0	0	4	1*	1	0	0	1	0	1	0	0	0	1	0	1	1			
CC	5*	20*	3*	15*	5*	23*	9	5*	1	17*	2	10*	2	17*	6*	0	6*	0	1	2	5*	2	1	1			
TF	2	6*	2	8*	0	5	13*	0	0	1	1*	3	2	3	3	0	3	0	1	0	3	1	0	0			
HW	0	1	1	3	1	5	9	1	0	8*	0	2	3*	3	2	0	2	0	0	2	2	1	0	0			
LD	3*	1	1	4	1	4	7	1	1	4	1*	3	0	2	9*	0	1	2	3	2	3	2	0	0			
TR	1	2	2	2	2	2	6	2	2*	6	0	6	0	4	11*	0	0	2	4	3*	2	1	0	2			
PO	0	12*	2	5	1	12*	12	1	0	10*	0	1	1	5*	3	0	1	2	3	0	0	0	0	0			
6th grade																											
DH	0	1	1	4	0	6	4	1	0	1	0	5	0	1	0	0	0	0	0	0	0	0	0	0			
LL	9*	13*	12*	26*	7*	30*	24*	9*	1	6	4*	10*	1	5*	12*	2*	3*	12*	10*	4*	4*	2*	2*	2*			
LB	7*	7*	1	15*	3	22*	15*	3	1	4	0	2	3	4	1	0	1	2	3	0	0	1	1	1			
FC	7*	14*	2	16*	3	17*	22*	3	0	2	1*	5	1	2	5	1*	2*	5*	5*	4*	1	1	1	1			
JC	1	2	0	6	2	7	19*	2	1	1	0	2	1	3	2	0	0	3	4	2	1	1	1	1			
JB	1	8*	1	11	3	11	8	3	1	8*	0	4	3	8*	2	0	2*	2	2	1	1	0	0	0			
UP	1	2	0	3	0	5	13	0	1	3	0	2	2	1	6*	0	0	3	1	1	1	1	1	0			

Chapter 6, General discussion

Our contribution

I began the general introduction of this thesis describing the bridge between cognitive science and education in elucidating the reading brain. On the one hand, functional imaging of the brain before and after learning to read (Dehaene-Lambertz et al., 2018), and for illiterate populations (Dehaene et al., 2015) has established the cortical areas dedicated to reading and the connection created between visual and language areas that develops during the learning process (Dehaene et al., 2010). Scientists have also played a key role in providing a protocol for randomized control trials, national testing procedures and comparative cross-language meta-analyses with key insights into what works (Groupe de travail du Csen, 2019; National Reading Panel, 2000; Rose, 2006). Members of the education community, through their curiosity and desire to provide quality education have been valuable partners in developing protocols, curriculums, questions and feedback from the class, that cannot be captured in lab experimentation. The main outcome of this collaboration has led to the overwhelming support of phonics curriculum as the best initial means to literacy instruction in the class (National Reading Panel, 2000; Rose, 2006), to boost the automatization of sensitivity to print (Brem et al., 2010) and access to meaning in previously established cortical areas for language.

With the end of the reading wars (Castles et al., 2018), research in reading can now turn to the detailed work of assessing what types of phonics practices works best, and when they should be applied to be the most conducive to learning. This has been the central theme of the present research, with our main question being, can technology provide a possible support to phonics education through the use of a tablet-based software for explicit systematic phonics instruction. To this end we developed and tested two different software packages for reading education in 1st grade and kindergarten. The results of both of these interventions have supported the use of our game as a means to early literacy, but with several caveats that help us to better understand how and when phonics is best taught in the class. For example, in Chapter 2 we show that our software for phonics instruction helped children learn to read faster and with better comprehension. Our results supported the Simple View of reading, which states that comprehension is the product of vocabulary and decoding (Hoover & Gough, 1990). In other words, by improving decoding skills through the game, working memory was freed up from slow decoding processes to allow students to access meaning from print. Importantly, this intervention also built on our knowledge of when

phonics education is important. Children that participated in the intervention only made improvements when they used the software in the first half of the year. These results support phonics training as an important element in literacy education early in the learning phase. Results were not found for students that used the software in the second part. From this, we can hypothesize that the explicit phonics instruction provide by Kalulu was beneficial when combined with the learning phase of reading but not when seen as review, or at the minimum, when the review was not tailored to the child's needs, as children had to complete each level in the game in a linear order. These results contribute to similar research pointing to an important window for phonics intervention that starts with teaching many grapheme-phoneme correspondences at a rapid tempo at the onset of reading education (Goigoux, 2016; Sunde et al., 2020).

The outcomes of our 1st-grade intervention study, however, came with their own limitations: positive results were only obtained for children during the early learning period; the software for our active control game was not well adapted for play lasting several months and not a well-matched control; children in this experiment were also learning to read in the class, thus limiting our ability to measure the benefits of the game separate from teacher methods. In response to these limitations, we decided to re-test our game with kindergarten students, the school year before formal literacy education in France. The game was adapted to this age level, following a pilot and teacher feed-back. We also developed a matched math game that was housed in the same game environment. The goal of both reading and math games was to teach children about symbols (graphemes and symbolic numbers) and their meaning (phonemes and quantity).

Test results for the intervention showed that children's symbol to meaning understanding improved both in grapheme-phoneme comprehension (when children played with the reading games) and number to quantity associations (when they played with the math games). For kindergarteners, these results were found independent of whether the software was used during the first or the second half of the year. In other words, we had accomplished our first two goals of teaching children irrespective of session in the cross-over training and providing a comparative quality active control. In follow-up tests, at the beginning of the school year after the intervention, students that used the reading software continued to outperform the control group in two important predictors of reading at school: letter knowledge and phoneme awareness. As these tests were done on paper, with teachers that

were unaware of the intervention, we can report that the games were successful in helping students not only respond to tablet-based tests but also to paper-and-pen tests. These initial positive results in phonics comprehension, however, did not help children spontaneously learn to read aloud more words than the passive control, nor did playing the game provide continued benefits once formal literacy education began. Our conclusion was that the tablet games were not sufficient enough to teach reading in the absence of real reading aloud practice.

Combining the results from the two interventions improves our knowledge of what works in phonics education. It seems that tablet-based phonics games can boost literacy and help children enter into reading through better grapheme-phoneme knowledge, but only when used early in literacy education and accompanied by a larger program that also includes reading aloud exercises. This tentative conclusion should be further tested in future randomized control studies.

Finally, the goal of this thesis has been to develop support for teachers by the creation and research of applicable tools for learning. Along with making our tablet-based game available for teachers, as well as the source code, in chapter 3 we present work that was done to provide teachers with a phonics progression based on the frequency and the consistency of the grapheme-phoneme correspondences for any alphabetic language using the neural network Transformer model (Vaswani et al., 2017). Importantly, our results made a slight improvement over previous methods reported in the literature on word error rate and phoneme error rate for grapheme-to-phoneme transcription (Potier Watkins et al., 2019). These improved rates of transcription can be used to analyze the consistency and frequency of grapheme-phoneme correspondences in any alphabetic language and build a phonics progression based on these variables, with accompanying 100% decodable word lists. These stimuli can be printed and used by teachers or, as we did, used to develop phonics games.

Future avenues in phonics instruction

Considering our key finding, that our games worked best when used early in learning and that grapheme-phoneme understanding only transfers to reading ability when combined with classroom activities, we hope that future research can explore the possibility that the game may boost reading if combined with reading aloud. This could be done through the integration

of voice recognition mini-games ‘plugged’ into the game engine (all mini-games in Kalulu are designed to be optional. Any developer can ‘plug’ or ‘unplug’ one of the mini-games we have already included in the suite, or design new games that use the phonics progression). This could also be done by integrating Kalulu into a larger teacher-led phonics curriculum whereby children learn and automatize knowledge of the grapheme-phoneme correspondences using Kalulu, then participate in teacher assessed reading aloud exercises. In this way, teachers could follow the child’s progression in the game, which would also provide the teacher with immediate feed-back and decodable text to be read aloud, adapted to the student’s progress and difficulties. To this end, the Kalulu software comes with a printable teacher-oriented manual, which is included as an appendix to this PhD thesis, that explains its logic and day-to-day progression, and can be used to accompany it in the classroom. It is currently our goal to test just such an approach by comparing kindergarten and 1st grade learning with the same methodology during the 2021-2022 school year.

Improving the game environment

Along with improving and testing the pedagogy provided in Kalulu, we believe that the game environment itself is also ripe for several different experiments in order to improve how the game supports learning. In both ELAN (chapter 2) and Kalulu (chapter 4) we tried to recreate some of the visual spatial attention elements of first-person shooter games to help children automatize learning (Bavelier et al., 2012; Green & Bavelier, 2003). For example, our games required quickly selecting the correct response amongst distractors. Children therefore must emit speeded responses while keeping track of multiple positions and avoiding distractors. Whether or not these elements help children learn, provide unnecessary stress or only make learning more like a game, is unclear. Future lab experimentation on the effects of game strategies to help children learn are needed.

Another element of the game environment worth improving is the strength of the storyline and reward system to motivate learning. As mentioned in chapter 4, the Kalulu game story of ‘learning to make the garden of one’s mind flourish’ was designed as an improvement from ELAN where there was no clear learning goal described to the player. This adaptation took its cue from elements of growth mindset theory that aim to teach children that effort is a sign of learning and improving (Rattan et al., 2015) by clearly explaining the learning goal as

the primary purpose of the game and providing children with comments from the game mascot that emphasize effort as a sign of progress. However, Kalulu does not fully exploit a reward system based on principles from growth mindset theory, such as rewarding players that re-try a difficult game and avoid re-playing easy games (Andersen & Nielsen, 2016; O'Rourke et al., 2014). Any new version of the game should implement and compare if rewarding 'perseverance' as opposed to 'winning' motivates students to spend more time working on difficult elements and progressing faster.

Contributions and improvements to dyslexia

Another contribution of this project came from our research into different types of developmental dyslexia. We developed a screener that contains specific words to test for deficits in different stages in orthographic analyses and central reading stages according to the Dual Route Model of reading (Coltheart, 2005; Friedmann & Coltheart, 2018) and coded responses representative of these error types. We presented the first cases of French attentional dyslexia (a deficit demonstrated by the migration of letters between words, reading "bar cat" as "bat car") and letter-position dyslexia (a deficit demonstrated by the transposition of letters read with-in words, reading 'diary' as 'dairy'). This work builds on the cases of sub-types of dyslexia already found in other languages (Friedmann et al., 2010; Friedmann & Gvion, 2001; Friedmann & Haddad-Hanna, 2014; Güven & Friedmann, 2019, 2019; Kohnen et al., 2012), while also providing support to models of orthographic analysis that take into account different stages for letter identification and letter position encoding in normal readers (Davis & Bowers, 2004, 2006).

Our results were encouraging not only because they help us to understand the many cognitive processes involved in reading, but because they will help us to improve remediation for dyslexia. The current limitation to this work is that we only report four cases with a selective deficit. To support our findings concerning the different factors that modulate the different deficits, it will be important to find many more cases. Another critical future exploration in the case of our analysis of dyslexia will be to find specific interventions to alleviate the types of errors produced. For example, it has already been reported that extra word spacing may reduce crowding effects when reading, and therefore help children with letter position dyslexia read faster (Friedmann & Rahamim, 2014). In the case of attentional

dyslexia, using a movable window that shifts from word to word has been reported as alleviating the nuisance of letter migrations (Friedmann et al., 2010). Finding more cases of selective dyslexia, designing and testing remediation practices, then developing applicable solutions for speech therapists and teachers to help readers with selective deficits are all necessary future goals.

Helping the children the most at risk for poor literacy skills.

An adaptive solution for a global crisis

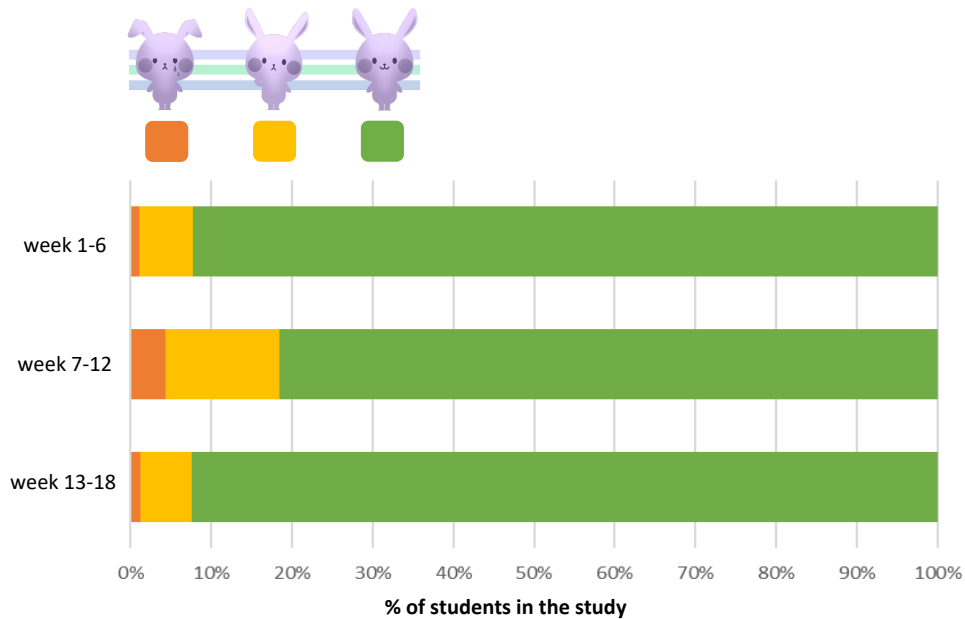
As presented in the introduction of this paper, phonics instruction has proven to be the best method for literacy success for all alphabetic languages, independent of the language's orthographic transparency (Landerl, 2000; Vousden et al., 2011). The focus and implementation of phonics methods in France and other developed nations has become more and more prevalent as nations seek to provide quality evidence-based pedagogy (Groupe de travail du Csen, 2019; National Reading Panel, 2000; Rose, 2006; Skibbe, 2016). However, in many parts of the world, phonics education is still not the norm. It is for this reason that Kalulu was designed to flexibly adapt to any language with the goal of creating an easy-to-use mobile solution for phonics instruction. A pedagogical progression of grapheme-phoneme correspondences to be taught, with accompanying stimuli, can be developed and implemented in the game using the GPA4.0 module described in chapter 3. Mini-games can be 'plugged' or 'unplugged' to the game engine depending on learning goals and language specific lessons to be taught (i.e., Kalulu contains a specific mini-game for silent-letters for French).

We were recently given the chance to test our goal for a language-adaptable game with colleagues Olalla and Lukasova from the Universidade Federal do ABC in Sao Paulo, Brazil. According to national literacy assessments from 2016 in Brazil, over half of students read and spell below grade level, 55% and 58% respectively (Temer et al., 2016). The textbook for reading provided to all public schools in Sao Paulo, shared with us by our colleagues, focuses on whole-word reading methods formerly criticized as inferior to phonics instruction (National Reading Panel, 2000). Working together, we were able develop a phonics progression and stimuli from a Portuguese word database (Estivalet & Meunier, 2015) that was quickly implemented in Kalulu for a Brazilian Portuguese version of the game. Our colleagues piloted

it with 19 children (the small group was due to the limited number of tablets available for the project) from two local public kindergarten classes. Students came from low to medium income neighborhoods. The intervention period lasted six weeks, with each child playing Kalulu for three 15-to-20-minute sessions a week. A control group of students that did not use the game was also tested.

Measures taken from the intervention included children's reported enjoyment of the game, collected at the end of each 3-week session. This was done by showing children a picture of a sad, neutral and happy rabbit, and asking them to point to which rabbit described how they felt about playing the game. Children's alphabetic knowledge was collected at four test periods, two before the intervention, immediately after the intervention and during the following school year on the same month as the first test. The majority of children reported enjoying playing the game (Fig. 1a). This was a welcome result as we had not collected data on student's appreciation of the games in the interventions described in chapter 2 and chapter 4 of this work. Our colleagues also reported that students that used Kalulu made greater progress in their alphabetic knowledge immediately after the intervention. In the longitudinal follow-up, when the students were in a new class with a teacher that did not know about the intervention, children that used Kalulu continued to outperform the 'business-as-usual' control (Fig. 1b).

a, Continuous evaluation of children’s enjoyment of the game Kalulu



b, Children’s alphabetic knowledge before, after and in the following school year

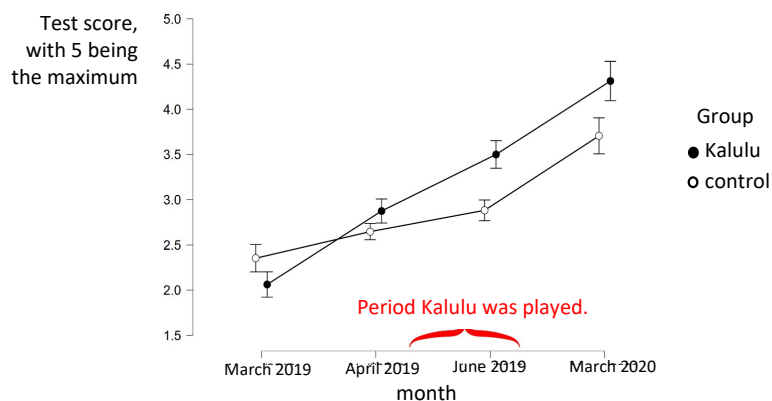


Figure 1. This data comes from a preliminary pilot of 19 children from a kindergarten in Sao Paulo that played 18 sessions of Kalulu over a six-week period. Graph **a)** shows children’s ratings of enjoyability in playing Kalulu at the end of three 3-week periods. Children gave their response by choosing either the sad, neutral or happy rabbit. Graph **b)** show children’s results to an alphabetic knowledge test taken at four different time periods (beginning of the year, pre-intervention, post intervention and at the start of the following year). On the March 2020 test, teachers for the current school year were unaware that children had participated in the intervention.

Developing an easy to adapt and deployable evidence-based curriculum across multiple-languages and environments is an opportunity for teachers and students, but also for researchers. This type of research allows for cross-language testing of the general ideas we have developed in this paper for phonics instruction. In the future, we hope that it will also provide an opportunity for rapid A/B testing of small factors in pedagogy (i.e., testing the ideal number of grapheme-phoneme correspondences to be taught; measuring reading speed as a

function of language transparency; assessing the possibility that voice recognition may be a viable way to have children practice reading) and the game environment (i.e., benefits of timed games and varying distractor levels; rewards based on principles from growth mid-set) that may help children with different orthographies and different learning environments.

The end of this thesis comes on the tail of another global crisis, the spread of COVID-19, which saw the closing of schools around the world to slow down the spread of the virus. This period of global school closures varied from region to region. In some cases, schools were closed for almost the entire school year. In France, where our research takes place, schools were closed for two months. During this period, school administrators, teachers, parents and students were forced to rapidly find hybrid solutions that would keep students connected to classroom learning while working from home, and often on one's own. For our part, we distributed an alpha version of the game Kalulu through our lab website (www.moncerveaualecole.com) along with a manual for learning how to read (I have attached the manual, in French, to the end of this thesis).

When children returned to school in September 2020, all students in 1st and 2nd grade took the National Evaluations described in chapter 4 (see the chapter 4 appendix for test questions). The results collected from the nation's students ($N > 80,000$ students) confirmed a slight trend to lower scores compared to tests taken by the previous 1st grade and 2nd grade populations (scores showed a slight upward trend in 2019, up from 2018). Concurrent with much research in education, socio-economic status was a determining factor in outcome, with children from low-SES classes disproportionately affected by school closures (Sirin, 2005) (see Fig., 2). In spite of efforts, the crisis revealed a lack of hybrid solutions that allow students to follow a school curriculum that could easily adapt to the situation. Future goals for Kalulu include testing it as a stand-alone curriculum that allows it to be used in school and at home.

**COVID-19 Effects, Comparing equivalent tests from 2020 to 2019,
Percentage point increase gap in the proportion of students from low-SES compared
to all other students that met expected outcomes on the National Evaluations**

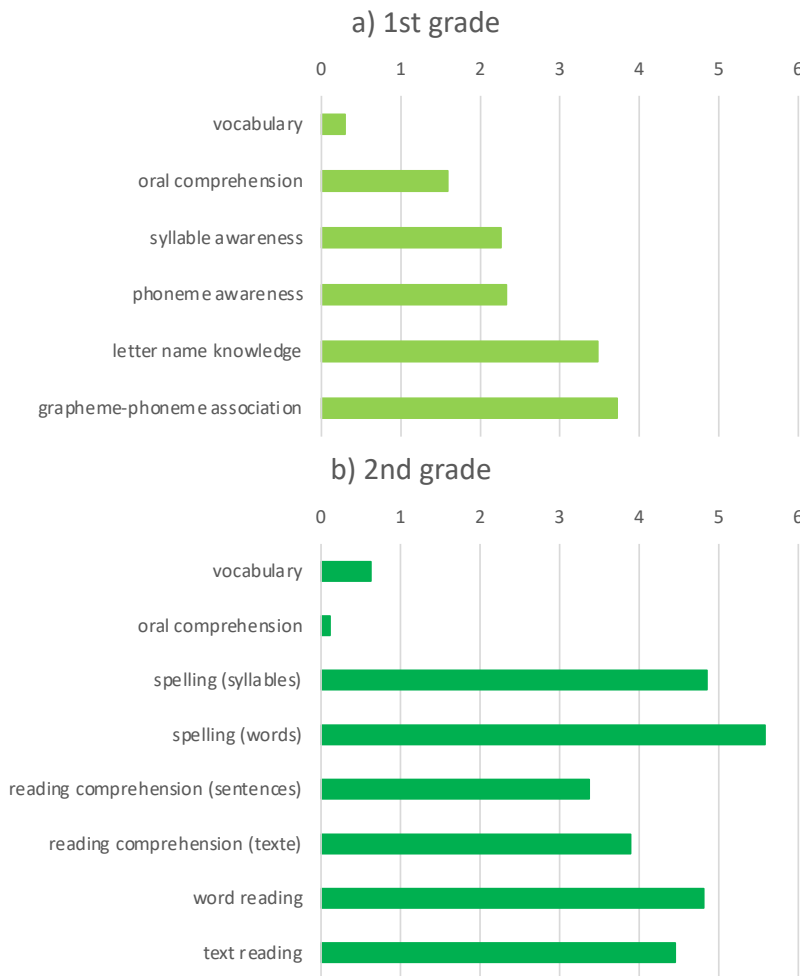


Figure 2, Percentage point increase (from 2019 to 2020) in the gap between the proportion of students from low-SES schools compared to ‘all other students’ that did not meet expected outcomes. Children tested at the beginning of the 2020-2021 school year would have not attended school from March to June during the preceding 2019-2020 school year, followed by summer break after less than a month of school reopening. Graph **a)** presents measures for the beginning of 1st grade and graph **b)** for the beginning of 2nd grade. The gap between the proportion of students obtaining expected outcomes between low-SES schools and ‘others’ increased on all measures, but especially in reading and spelling skills for students entering 2nd grade.

Source: MENJS-DEPP, Repères CP-CE1 copyright DEPP.

Developing low-cost paper-based games for learning

While the basis of this work has been to test the use of technology to boost school learning, we also recognize that technology is not always accessible to the classroom. In these cases,

board and card games may provide the solution for learning through games. They might also provide different benefits due to the fact that board and card games can offer different kinesthetic clues (the difference between tapping on a flat screen and actually moving a game piece with one's hand) and more opportunity for social cooperation (i.e., taking turns, dealing cards to others, providing oral responses to an opponent). The research benefits for math skills from board games is already well documented. For example, it was observed that children who play board games (e.g., Chutes and Ladders) are more likely to have a linear representation of numbers when responding to a number line task (Ramani & Siegler, 2008). The number line task requires the viewer to place a number on its position on a line according to the space provided between two endpoints, generally 0 and 100. Children's linear understanding correlates with school math ability (Booth & Siegler, 2006). This finding spurred a series of board game and math ability experiments by Booth, Siegler and colleagues (Laski & Siegler, 2014; Ramani & Siegler, 2008; Siegler & Ramani, 2008, 2009). These experiments showed that playing on linear board games, combined with a simple 'counting-on' procedure improved children's linear representation of numbers. Counting-on refers to counting each space landed on in a board (i.e., the player's piece is on the number 5 and a two is rolled, the player advances the piece while counting 'six, seven') as opposed to the typical counting-from-0 (i.e., in the same situation, the player advances the piece while counting 'one, two'). The take home message from these papers for scientists, teachers and parents was that board games provide a combined visual, kinesthetic, auditory and temporal feedback to the linear order of numbers that helps cement the child's mental representation of their linear number line.

With a similar goal of improving early math and geometry intuitions and school skills, colleagues Dillon, Kannan, Dean, Spelke and Duflo developed and tested number and shape games with children from Indian slums (Dillon et al., 2017). The goal of the games was to develop early intuitions about math concepts that children from non- low SES families often gain from playing games at home, and test if these early intuitions would later garner improved math skills in school. The games focused on non-symbolic properties of numbers and geometrical representation (i.e., quantity comparison, differences in angles) but not their symbolic components (i.e., indo-arabic numerals, number names, shape names). Compared to an active control group, gains were initially made in non-symbolic and symbolic math. Gains

in symbolic math, however, disappeared once formal schooling began. The control group played games for social awareness.

Today, we are currently working with this team to improve and adapt these games for children in India, the USA and France. Improvements have included making connections between symbolic and non-symbolic components of numbers and geometry. We are also using our experience with phonics instruction to develop matched card and boards games for reading. The games we have developed for reading are designed to support the factors described by the Simple View of Reading (Hoover & Gough, 1990) as essential to literacy: decoding and vocabulary (see appendix for visual game posters from the *Bien Joué !* project, in French, describing game play for each of the 3 reading and 3 math games). Importantly, these games, played in small groups, require children to read aloud and make comprehension choices based on their reading. The games all follow the pedagogical progression designed for our tablet software, Kalulu.

In early 2020, we started an intervention in French classrooms to test the benefits of these games with 1st grade students. Alas, this experiment was interrupted due to school closures during the COVID-19 pandemic. It has restarted this year, focusing on schools in low-SES neighborhoods, with the principal goal of correcting the gap in learning deficits detected in this category of schools as reported above. Participating classes have been randomized into either playing reading games or math games for the entire school year. We will follow students' progress on the National Evaluations through 1st and 2nd grade. Students play the games three times a week for a minimum of 20 minutes per session. Classes will remain in the same game group for the entire year. In the case of positive results, the games will be made available as a resource for teachers (as part of the Covid19 reaction, we also made them freely available on MonCerveauALecole.com). It is also our goal for these games to become another part of the Kalulu pedagogy as they all require reading aloud, a critical element missing in our tablet games. These games, if successful, will also provide teachers with an activity that can be distributed to classes for just the cost of printing, as most schools in France do not have access to tablets.

Conclusion

If we had to offer one take-home message from this thesis, it would be that the development of education initiatives based on findings from the cognitive sciences and in-the-field testing are critical to improving education. This body of this work highlights two clarifications that can be made to general beliefs about education: phonics instruction is a critical part of learning how to read, but only when taught early in literacy and accompanied by reading aloud practice; technology, as seen in previous research (OECD, 2015; Warschauer et al., 2011), will not revolutionize education, but it can boost learning when used as a support for quality evidence-based software. In conclusion, developing hypotheses about how the brain learns and adapts to the cultural tools of our education heritage is an important element in building new curriculum and technologies for schools. However, these technologies should in turn be tested using school measures of learning to ensure that expected outcomes are met. Boosting learning in school is a win-win goal that helps children succeed while providing researchers with real world evidence on how our brains flourish at school.

References

- Andersen, S. C., & Nielsen, H. S. (2016). Reading intervention with a growth mindset approach improves children's skills. *Proceedings of the National Academy of Sciences*, 113(43), 12111-12113. <https://doi.org/10.1073/pnas.1607946113>
- Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain Plasticity Through the Life Span : Learning to Learn and Action Video Games. *Annual Review of Neuroscience*, 35(1), 391-416. <https://doi.org/10.1146/annurev-neuro-060909-152832>
- Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, 42(1), 189-201. <https://doi.org/10.1037/0012-1649.41.6.189>
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., & Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17), 7939-7944. <https://doi.org/10.1073/pnas.0904402107>
- Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading Wars : Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest*, 19(1), 5-51. <https://doi.org/10.1177/1529100618772271>
- Coltheart, M. (2005). Modeling Reading : The Dual-Route Approach. In M. J. Snowling & C. Hulme (Éds.), *The science of reading : A handbook* (p. 6-23). Blackwell Publishing. <https://doi.org/10.1002/9780470757642.ch1>
- Davis, C. J., & Bowers, J. S. (2004). What Do Letter Migration Errors Reveal About Letter Position Coding in Visual Word Recognition? *Journal of Experimental Psychology: Human Perception and Performance*, 30(5), 923-941. <https://doi.org/10.1037/0096-1523.30.5.923>
- Davis, C. J., & Bowers, J. S. (2006). Contrasting five different theories of letter position coding : Evidence from orthographic similarity effects. *Journal of Experimental Psychology: Human Perception and Performance*, 32(3), 535-557. <https://doi.org/10.1037/0096-1523.32.3.535>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate : Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews. Neuroscience*, 16(4), 234-244. <https://doi.org/10.1038/nrn3924>

- Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., & Cohen, L. (2010). How Learning to Read Changes the Cortical Networks for Vision and Language. *Science*, 330(6009), 1359-1364. <https://doi.org/10.1126/science.1194140>
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form : Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLoS Biology*, 16(3), e2004103. <https://doi.org/10.1371/journal.pbio.2004103>
- Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., & Duflo, E. (2017). Cognitive science in the field : A preschool intervention durably enhances intuitive but not formal mathematics. *Science*, 357(6346), 47-55. <https://doi.org/10.1126/science.aal4724>
- Estivalet, G., & Meunier, F. (2015). Brazilian Portuguese Lexicon Manual. www.lexicodoportugues.com
- Friedmann, N., & Coltheart, M. (2018). 35 .Types of developmental dyslexia. In E. Dattner & D. Ravid (Éds.), *Handbook of Communication Disorders* (p. 721-752). De Gruyter. <https://doi.org/10.1515/9781614514909-036>
- Friedmann, N., & Gvion, A. (2001). Letter Position Dyslexia. *Cognitive Neuropsychology*, 18(8), 673-696. <https://doi.org/10.1080/02643290143000051>
- Friedmann, N., & Haddad-Hanna, M. (2014). Types of Developmental Dyslexia in Arabic. In E. Saiegh-Haddad & R. M. Joshi (Éds.), *Handbook of Arabic Literacy* (Vol. 9, p. 119-151). Springer Netherlands. https://doi.org/10.1007/978-94-017-8545-7_6
- Friedmann, N., Kerbel, N., & Shvimer, L. (2010). Developmental attentional dyslexia. *Cortex*, 46(10), 1216-1237. <https://doi.org/10.1016/j.cortex.2010.06.012>
- Friedmann, N., & Rahamim, E. (2014). What can reduce letter migrations in letter position dyslexia? : TREATMENT OF LETTER POSITION DYSLEXIA. *Journal of Research in Reading*, 37(3), 297-315. <https://doi.org/10.1111/j.1467-9817.2011.01525.x>
- Goigoux, R. (2016). Apprendre à lire et à écrire au cours préparatoire: enseignements d'une recherche collective. *Revue Française de pédagogie*, (3), 5-6.
- Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, 423(6939), 534-537. <https://doi.org/10.1038/nature01647>
- Groupe de travail du Csen. (2019). ÉvalAide, un dispositif scientifique de prévention des difficultés en lecture et en mathématiques en CP et en CE1. Canopé.

- https://www.reseau-canope.fr/fileadmin/user_upload/Projets/conference_role_experimentation_domaine_educatif/EvalAide_CSEN_Definitif_Mai2019.pdf
- Güven, S., & Friedmann, N. (2019). Developmental Letter Position Dyslexia in Turkish, a Morphologically Rich and Orthographically Transparent Language. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02401>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2(2), 127-160. <https://doi.org/10.1007/BF00401799>
- Kohnen, S., Nickels, L., Castles, A., Friedmann, N., & McArthur, G. (2012). When 'slime' becomes 'smile': Developmental letter position dyslexia in English. *Neuropsychologia*, 50(14), 3681-3692. <https://doi.org/10.1016/j.neuropsychologia.2012.07.016>
- Landerl, K. (2000). Influences of orthographic consistency and reading instruction on the development of nonword reading skills. *European Journal of Psychology of Education*, 15(3), 239. <https://doi.org/10.1007/BF03173177>
- Laski, E. V., & Siegler, R. S. (2014). Learning from number board games : You learn what you encode. *Developmental Psychology*, 50(3), 853-864. <https://doi.org/10.1037/a0034321>
- Michel Temer, José Mendonça Bezerra Filho, Maria Helena Guimarães de Castro, & Maria Inês Fini. (2016). Overview of the Brazilian Education System. INEP/MEC-Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. http://download.inep.gov.br/acoes_internacionais/pisa/documentos/2016/pisa_overview_of_the_brazilian_education_system.pdf
- National Reading Panel. (2000). Teaching children to read : An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction (NIH Publication No. 00-4769). National Institute of Child Health & Development. <https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf>
- OECD. (2015). Students, Computers and Learning. OECD Publishing. <https://doi.org/10.1787/9789264239555-en>

- O'Rourke, E., Haimovitz, K., Ballweber, C., Dweck, C., & Popović, Z. (2014). Brain Points : A Growth Mindset Incentive Structure Boosts Persistence in an Educational Game. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3339–3348. <https://doi.org/10.1145/2556288.2557157>
- Potier Watkins, C., Dehaene, O., & Dehaene, S. (2019). Automatic Construction of a Phonics Curriculum for Reading Education Using the Transformer Neural Network. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Éds.), *Artificial Intelligence in Education* (p. 226-231). Springer International Publishing. https://doi.org/10.1007/978-3-030-23207-8_42
- Ramani, G. B., & Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child development*, 79(2), 375–394.
- Rattan, A., Savani, K., Chugh, D., & Dweck, C. S. (2015). Leveraging Mindsets to Promote Academic Achievement : Policy Recommendations. *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691615599383>
- Rose, J. (2006). Independent review of the teaching of early reading : Final report. DfES. <http://dera.ioe.ac.uk/5551/3/5d970d28fc535dc54eb4dee9995bef36.pdf>
- Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, 11(5), 655-661. <https://doi.org/10.1111/j.1467-7687.2008.00714.x>
- Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—But not circular ones—Improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, 101(3), 545-560. <https://doi.org/10.1037/a0014239>
- Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement : A Meta-Analytic Review of Research. *Review of Educational Research*, 75(3), 417-453. <https://doi.org/10.3102/00346543075003417>
- Skibbe, L. E. (2016). A Content Analysis of Phonological Awareness and Phonics in Commonly Used Head Start Curricula. 9.
- Sunde, K., Furnes, B., & Lundetræ, K. (2020). Does Introducing the Letters Faster Boost the Development of Children's Letter Knowledge, Word Reading and Spelling in the First Year of School? *Scientific Studies of Reading*, 24(2), 141-158. <https://doi.org/10.1080/10888438.2019.1615491>

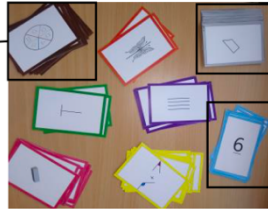
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. ArXiv:1706.03762 [Cs]. <http://arxiv.org/abs/1706.03762>
- Vousden, J. I., Ellefson, M. R., Solity, J., & Chater, N. (2011). Simplifying Reading : Applying the Simplicity Principle to Reading. *Cognitive Science*, 35(1), 34-78. <https://doi.org/10.1111/j.1551-6709.2010.01134.x>
- Warschauer, M., Cotten, S. R., & Ames, M. G. (2011). One Laptop per Child Birmingham : Case Study of a Radical Experiment. *International Journal of Learning and Media*, 3(2), 61-76. https://doi.org/10.1162/ijlm_a_00069

Appendix – Games in the project Bien Joué !

EUCLIDO

Matériel :

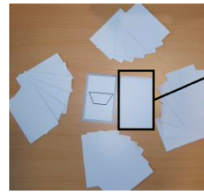
Des cartes joker
(marrons)



Des cartes forme (grises)

Des cartes propriété
(colorées)

1. Distribue 7 cartes à chaque joueur, forme une pioche avec les autres et retourne la première carte de la pioche

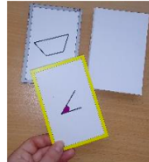


Pioche

2. Sur une carte forme, tu as le droit de poser :

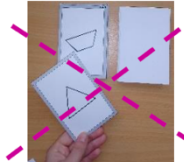


Une carte forme identique

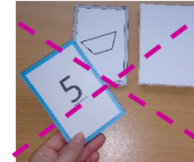


Une carte propriété que la forme possède

3. Sur une carte forme, tu n'as pas le droit de poser :



Une carte forme différente



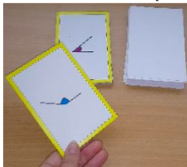
Une carte propriété que la forme ne possède pas



Un joker

Un carré !

4. Sur une carte propriété, tu as le droit de poser :

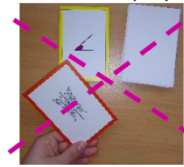


Une carte propriété de même type (même couleur)

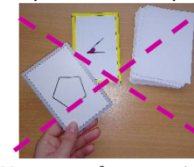


Une carte forme qui possède cette propriété

5. Sur une carte propriété, tu n'as pas le droit de poser :

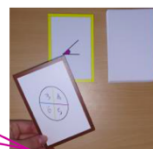


Une carte propriété d'une autre couleur



Une carte forme qui ne possède pas cette propriété

4 côtés !



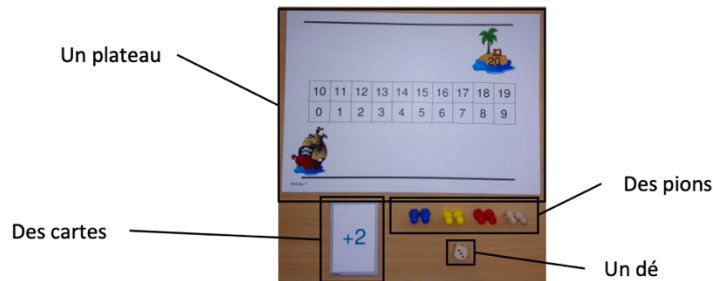
Un joker

Si tu ne peux pas jouer, pioche une carte. Si tu peux la poser, alors pose-là, et sinon, garde la dans ta main.
Tu as gagné si tu es le premier à avoir posé toutes tes cartes !

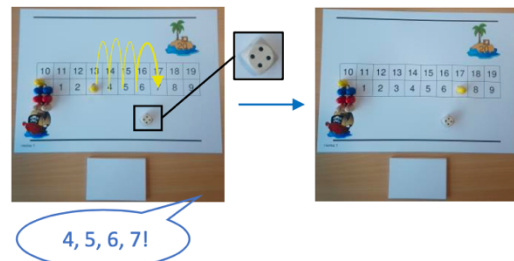
20 PAS AU TRESOR

Règles du jeu

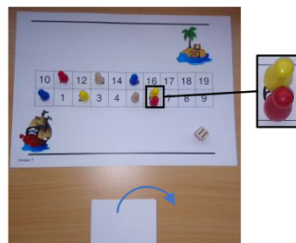
Matériel:



1. Choisi une couleur et installe tes pions sur la case '0'.
Fait une pioche avec les cartes
2. Lancer le dé et avancer



3. Si un autre pion se trouve déjà sur la case où tu arrives, même s'il est de ta couleur, pioche une carte
4. Déplace ton pion du nombre de cases indiquées sur la carte



5. Si tu es le premier à amener tes deux pions sur l'île au trésor, tu as gagné !



LA BATAILLE

Règles du jeu

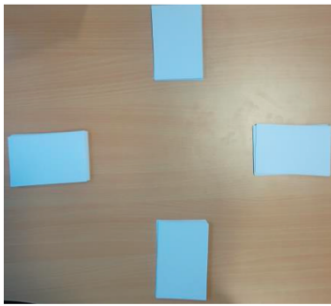
Matériel:

Des cartes avec des points

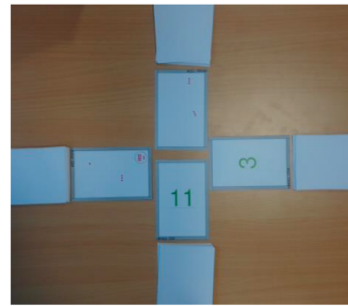


Des cartes avec des nombres

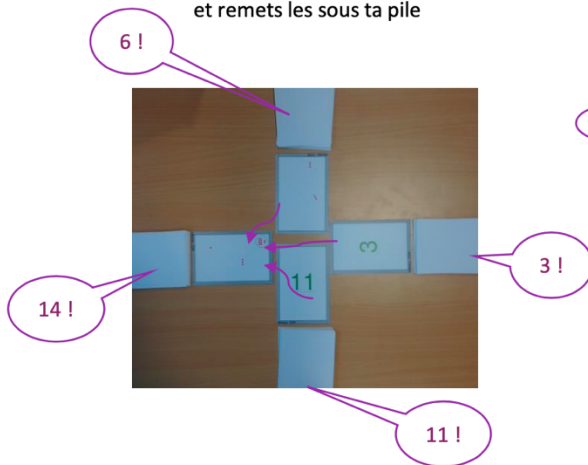
1. Distribue les cartes à tous les joueurs



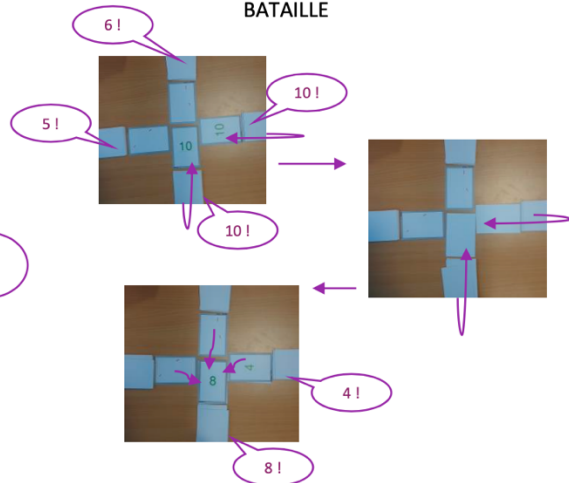
2. Retourne la carte du dessus de ta pile



3. Si tu as le nombre le plus grand, ramasse les cartes et remets les sous ta pile



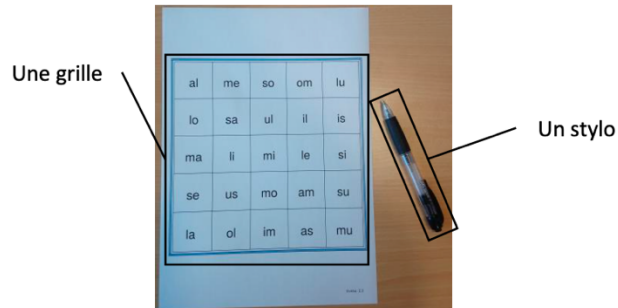
4. Si deux personnes ont le nombre le plus grand :
BATAILLE



5. Lorsqu'un joueur n'a plus de cartes, compte les cartes que tu as dans ta pile : si tu en as le plus, tu as gagné !

LE BINGO

Matériel :



Ecoute la syllabe qui va être dite par ton professeur.
Cherche-la sur ta grille et coche la case

Si tu es le premier à compléter une ligne, une colonne
ou une diagonale du carré, tu as gagné !



mu	is	us	mo	la
om	li	so	sa	lo
as	il	si	im	lu
me	le	se	am	al
ul	mi	su	ol	ma

mu	is	us	mo	la
om	li	so	sa	lo
as	il	si	im	lu
me	le	se	am	al
ul	mi	su	ol	ma

mu	is	us	mo	la
om	li	so	sa	lo
as	il	si	im	lu
me	le	se	am	al
ul	mi	su	ol	ma

mu	is	us	mo	la
om	li	so	sa	lo
as	il	si	im	lu
me	le	se	am	al
ul	mi	su	ol	ma

LIT CHERCHE ET TROUVE

Règles du jeu

Matériel:



1. Place le plateau au centre de la table et distribue les cartes et les jetons à tous les joueurs

2. Retourne les 3 premières cartes du dessus de ta pile et place-les devant toi



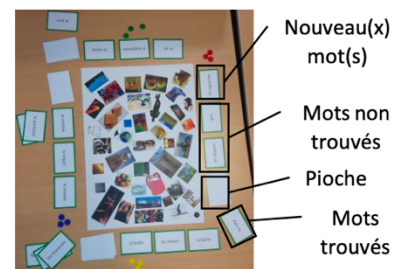
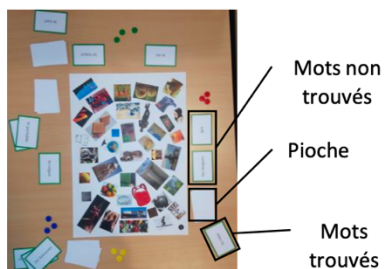
3. Lorsque l'un des joueurs dit « TOP ! », place tes jetons rapidement sur les images associées à tes mots

4. Lorsque tu as trouvé tes trois images, dit « STOP ! » et tout le monde arrête de jouer



5. Mets de côté tous les mots que tu as trouvés et regroupe tes jetons

6. Retourne tes cartes pour en avoir de nouveau 3 devant toi



7. Le premier à avoir trouvé toutes ses cartes a gagné !

LA BATAILLE DES MOTS

Règles du jeu

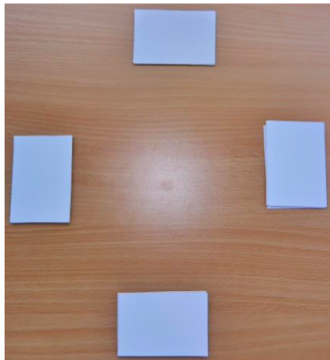
Matériel:

Des cartes avec des mots



Des cartes avec des images

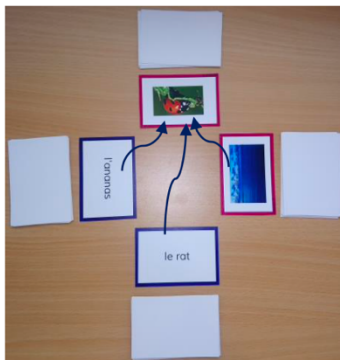
Distribue les cartes à tous les joueurs



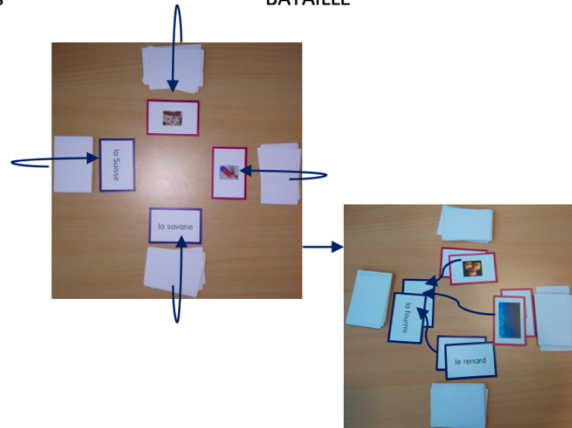
Retourne la carte du dessus de ta pile



Si le mot que tu lis ou l'image que tu vois représente la chose la plus petite, ramasse les cartes et remets les sous ta pile



Si les deux plus petits objets sont de tailles proches :
BATAILLE



Lorsque l'un des joueurs n'a plus de cartes, compte les cartes que tu as dans ta pile : si tu en as le plus, tu as gagné !

