



**HAL**  
open science

## Modelling children's inhibitory skills using learning data from an educational app

Guilherme Medeiros Machado, Geoffray Bonnin, Sylvain Castagnos, Lara  
Hoareau, Aude Thomas, Youssef Tazouti

### ► To cite this version:

Guilherme Medeiros Machado, Geoffray Bonnin, Sylvain Castagnos, Lara Hoareau, Aude Thomas, et al.. Modelling children's inhibitory skills using learning data from an educational app. *Journal of Computer Assisted Learning*, 2023, pp.1-13. 10.1111/jcal.12773 . hal-04040321

**HAL Id: hal-04040321**

<https://inria.hal.science/hal-04040321v1>

Submitted on 25 Sep 2024

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# Modelling children's inhibitory skills using learning data from an educational app

Guilherme Medeiros Machado<sup>1</sup>  | Geoffroy Bonnin<sup>1</sup> | Sylvain Castagnos<sup>1</sup> |  
Lara Hoareau<sup>2</sup> | Aude Thomas<sup>2</sup> | Youssef Tazouti<sup>2</sup>

<sup>1</sup>Bird Team, LORIA, University of Lorraine, Nancy, France

<sup>2</sup>LPN (EA 7489), University of Lorraine, Nancy, France

## Correspondence

Guilherme Medeiros Machado, University of Lorraine, Nancy, France.  
Email: [guilherme.medeiros-machado@loria.fr](mailto:guilherme.medeiros-machado@loria.fr)

## Funding information

French Programme of Investments for the Future, Grant/Award Number: e-Fran Linumen

## Abstract

**Background:** Early literacy and numeracy skills are developed during early childhood. Among the many factors that influence the development of such skills, the literature shows that the executive functions, especially the response inhibition (RI)—that is the capability to block out or to tune out what can be considered irrelevant information or action to the learning task—is one of the most essential functions. There are specific tests used to appraise these children's inhibition skills, but these tests are generally time-consuming, and demand specialized human resources.

**Objectives:** We present a computational approach to model children's RI behaviour through the analysis of educational traces left in an educational app. This modelling allows the automatic and instant identification of the RI level of children without the need of a human-conducted test.

**Methods:** Our modelling is based on two definitions of RI found in the literature, from which we derived a mathematical formalism of three variables we used to query the traces dataset and isolate the RI behaviour of each student from the learning traces generated in the app. The sample population is composed of children from diverse socioeconomic backgrounds. The model is then assessed by comparing it to a traditional human-conducted RI test suitable for kindergarten children, the Head-Toes-Knees-Shoulder (HTKS) task.

**Results and conclusions:** The results show that our RI model can explain an important part of the HTKS variance (up to 0.45 according to the adjusted  $R^2$ ) when taking the HTKS results as a dependent variable for a multiple regression model. In practice, our model can be integrated in a learning app and become a powerful tool for instant preliminary identification of dysfunctional RI behaviour, especially in the early stages of children's education. Once students are identified by our model as having a dysfunctional RI behaviour, teachers can rapidly act to help them. Besides, the proposed model requires only very simple data to work, which means it can be easily integrated into different learning apps.

## KEYWORDS

data science applications in education, early childhood education, executive function, response inhibition, student model

## 1 | INTRODUCTION

Research in the area of school readiness emphasizes that, during the preschool years, children experience significant cognitive development and early learning (Duncan et al., 2007). In this context, executive functions are major predictors of later school success (Fuhs et al., 2014; Moffitt et al., 2011). Executive functions are a set of effortful mental processes we use to pay attention and to stay focused (Diamond, 2013). Interestingly, they have been shown to promote the acquisition of early numeracy and early literacy skills (Schmitt et al., 2015), which are themselves major predictors of preschoolers' later success.

In this article, we focus on the relationship between early numeracy and early literacy skills and their relationship with executive functions. More precisely, we argue that early literacy and early numeracy learning requires better assessment and real-time consideration of executive functioning in general, especially response inhibition (RI). We propose to adopt a learning analytics (LA) approach and design a computational model capable to measure the executive functioning of RI through the analysis of e-learning traces. To achieve this goal, we propose and formalize three computational variables that are derived from the analysis of the traces left by the students' interactions with an educational app (shown in Figure 1) and are validated through an executive functioning task.

Developing such a model could be useful by providing an instant assessment of the students' RI skills, and therefore enhance the identification process of students who are lacking these skills. Moreover, the application of such computational models in psychology can help to understand many aspects of human behaviour and cognition, such as memory, perception, reasoning, decision making, and so on. (Farrell & Lewandowsky, 2018). Some similar work from the literature in the domain of LA focuses on modelling students' forgetting (Choffin et al., 2019; Lalwani & Agrawal, 2019), while others like Rajendran et al. (2018), Syed et al. (2020), Marchal, Castagnos & Boyer (2018), and Antón-Méndez et al. (2019) use gaze data to respectively predict learning and working memory (WM). Similarly, the paper of Matayoshi et al. (2019) models and predicts students' knowledge retention, while some other works are related to different types of cognitive flexibility (CF) (Moissa, Bonnin, Castagnos & Boyer, 2019; Naik & Kamat, 2019). As will be shown in this article, our proposed model can for now help to understand, although only to a certain extent, the relationship between RI and early literacy and numeracy.

### 1.1 | Executive functioning

Best and Miller (2010) define executive functioning as the adaptive, goal-oriented control of thoughts, behaviours, and emotions. Executive functioning is also defined as a high-level cognitive system that integrates and controls low-level skills such as processing speed and memory span (Demetriou et al., 2010). Miyake et al. (2000) and Zink et al. (2021) distinguished three related but distinct components of executive functioning: WM, CF, and RI. In the remainder of this article, we will rely on this definition of executive functioning as three different but related components. The WM describes the ability to retain spatial or verbal information in one's short-term memory and to simultaneously eliminate irrelevant information and/or modify the information retained (Willoughby et al., 2012). The CF includes maintaining focus and flexibly adapting to change goals or stimuli (Rueda et al., 2005). Finally, the RI that refers to the ability to override a dominant or prepotent response in favour of a more adapted one (Friedman & Miyake, 2004).

RI mechanisms are distinguished according to the type of motor or cognitive process. Motor RI refers to the ability to control one's behaviour by restraining automatic or predominant motor behaviours (e.g., control of motor impulsivity). The cognitive side of RI refers to the control of the information one is processing (e.g., suppression of automatic or irrelevant information during ongoing cognitive processing). Across the preschool period, children show improvements in the ability to suppress a motor response such as a button press (Carver, Livesey, & Charles, 2001; Simpson & Riggs, 2006) or verbal response (Espy et al., 2006), and the ability to produce a response that conflicts with prepotent tendencies.

Some recent studies show that the components of executive functioning can be dissociated (Lonigan et al., 2016; Miller et al., 2012) and predict different educational outcomes (McClelland et al., 2014). The three components of executive functioning develop in a significant and linear fashion during childhood. The preschool years are considered an important period for the development of executive functioning because of the rapid changes in the prefrontal cortex between the ages of 2 and 5 (Zelazo & Müller, 2011). Several authors suggest that RI and WM develop first, and then allow for the development of CF (Best & Miller, 2010; Diamond et al., 2002; Garon et al., 2008).



**FIGURE 1** Screenshots of the app. The two first pictures correspond to activities of early literacy and the others to activities of early numeracy.

## 1.2 | Early literacy

Early Literacy is the acquisition of the skills, knowledge, and attitudes that are the developmental precursors of conventional forms of reading and writing (Whitehurst & Lonigan, 1998). The four most widely studied aspects of early literacy are alphabet knowledge, phonological awareness, vocabulary, and oral comprehension (Lonigan & Shanahan, 2008). Alphabet knowledge refers to letters' names and sounds, as well as recognizing the visual shape of the letters and their sound values (Noel Foulin, 2005). Phonological awareness is the ability to detect and manipulate the sound structure of oral language (Whitehurst & Lonigan, 2002). Vocabulary is the set of words in a language that a person has or uses. Comprehension is the interpretation of information contained within a text, the use of prior knowledge to interpret this information, and the construction of a coherent representation of the text by the reader (Kendeou et al., 2007). These dimensions can be used to predict future skills (Catts et al., 2015; Duncan et al., 2007).

## 1.3 | Early numeracy

Early mathematics skills cover several domains, including numeracy, geometry, patterning, and problem solving (National Research Council, 2009). Although each of these early skills is distinct, they begin to develop during early childhood (Litkowski et al., 2020) and interact with each other to build more advanced mathematics skills (Purpura et al., 2013). Krajewski and Schneider (2009) argue that children go through several phases to acquire increasingly cognitively complex early numeracy skills. First, children learn to recognize small sets without counting (subitizing), distinguish between small quantities (set comparison), and retain the verbal counting sequence. Second, children apply the counting sequence to fixed sets (one-to-one counting) and make links between each of the number words and their respective quantity (e.g. apply cardinal number knowledge). Third, children combine number words and quantities into new number words and quantities without using physical objects (e.g., story problems). Linking new knowledge to existing knowledge allows children to develop a deeper understanding of mathematics.

## 1.4 | Response inhibition as a predictor of early literacy and early numeracy

The literature shows that the development of executive functioning allows children to be attentive, to persevere in particularly difficult tasks, and above all to avoid distractions. All of these behaviours offer more learning opportunities for children (McClelland et al., 2015). A considerable amount of work has also found links between executive functioning and children's performance in early literacy and early numeracy (Blair & Razza, 2007; McClelland et al., 2014).

These links are usually stronger and more consistent between executive functioning and early numeracy than between executive

functioning and early literacy (Allan et al., 2014). Especially, RI is a good predictor of early numeracy even after statistically controlling for important covariates such as the child's age, verbal and/or general intelligence, and maternal education (Blair & Razza, 2007; Espy et al., 2004). However, the association between RI and early literacy is less clear. While some studies find that RI is an important predictor of some aspects of early literacy (Cameron et al., 2012), other studies find that RI and early literacy are not significantly related (Davidse et al., 2011; Lan et al., 2011). Because of these links between executive functioning, early literacy, and early numeracy, and especially because of the important predictive role of RI in early numeracy performance, we propose in this article a method to model the students' RI.

## 1.5 | Learning analytics

LA first appeared in the early 2010s (Johnson et al., 2012). Historically at the intersection of machine learning, statistics, human-computer interaction, and learning science, they were already erected as a full discipline in 2013 (Siemens, 2013). The general objective is to exploit the large amount of learning traces left by students through their interactions with online educational services (intelligent tutoring systems, learning management systems, MOOCs, and so on) to improve the quality of the teaching (Keshavamurthy & Guruprasad, 2015). These systems have rapidly gained popularity and are now used in nursery school and elementary school (Medeiros Machado et al., 2020; Ebner & Schön, 2013; Martinez & LaLonde, 2020; Triana et al., 2015), high school (Jones, 2019), university (Leitner et al., 2017), and even for lifelong learning in companies (Domenjoz, 2017).

The applications are multiple and include visualization tools such as student and teacher dashboards (Brun et al., 2019), early prediction of at-risk students, intervention techniques to help them (Wise, 2014), and automated recommendation of suitable educational content (Karypis, 2017; Machado et al., 2021).

Year after year, many new objectives and a greater variety of traces emerged, for example integration of skills to be achieved (Lockyer et al., 2013), prediction of memorized items, and adaptation to the speed of learning (Marchal, Castagnos, & Boyer, 2016), adaptation to intellectual disabilities (Cano et al., 2018), use of gaze data (Beuget, Castagnos, Luxemburger, & Boyer, 2019), evaluation of the effort provided by students (Moissa, Bonnin, Castagnos, & Boyer, 2019; Chong et al., 2018), evaluation of the learning quality (Bondareva et al., 2013).

Although large-scale proof of the effectiveness of LA tools and the benefits they bring to learners is still lacking (Ifenthaler & Yau, 2020; Viberg & Grönlund, 2021), first encouraging results do exist. For instance, a between-group study of 482 students (241 students per group), each of them taking seven Moodle courses, showed a 15.11% ( $p < 0.001$ ) gain in final grade in the group that received feedback through LA compared with the control group (Lim et al., 2021). In line with these works, our perspective is to provide teachers

with an additional performance indicator (a measure of each student's progress in RI), to facilitate the detection of students who are struggling. In this article, we make the first step towards that goal by studying to what extent such an indicator can be automatically obtained from the traces left by the learners during their interactions with our application.

## 1.6 | The current study

Given the role that RI plays in early literacy and early numeracy acquisition, this article aims to analyse learning traces left by children during early literacy and early numeracy activities. Our goal is to identify the signs of RI present within these traces and to exploit them to provide some automated assessment of RI.

Traditional (non-automated) tests to measure RI include the Stroop test and the Head-Toes-Knees-Shoulder (HTKS) task, which will be presented in Section 2.3.1. However, they present some limitations such as the need for a human specialist to supervise their execution and compute the results. For children at kindergarten, they also have the drawback of being time-consuming and presenting results with a short lifespan since the children's cognitive capabilities are constantly evolving in this phase.

Being able to automatically assess RI would allow pinpointing in real-time the children who may be facing difficulties in early literacy and early numeracy because of a low level of RI. Such a mechanism could help teachers to be alerted that a specific student is struggling with early literacy or early numeracy acquisition, allowing them to take specific actions to help the student.

The automatic identification of RI from simple students' traces is not a simple task. There are two main challenges to provide such automatic identification: (1) the need to properly process and isolate the specific RI behaviour from the learning traces; and (2) the need to validate the isolated RI behaviour. One of the ways to perform such validation is to compare the automated assessments with traditional RI test results. This article addresses both challenges and shows our findings in explaining children's RI behaviour using learning traces. In Section 2 the method to isolate the RI behaviour is presented, then the results and the discussions of the comparison with a traditional RI task are presented in Section 3. The conclusions and future directions are presented in Section 4.

## 2 | METHOD

### 2.1 | The LINUMEN project

The goal of the LINUMEN project is to create and test an educational tablet app (some screenshots shown in Figure 1) that aims at helping kindergarten students develop their early literacy and early numeracy skills. To ensure the resulting app was geared to the needs of French kindergarten children and was compatible with use in France's kindergartens (*écoles maternelles*), both researchers and education

professionals (teachers, educational advisers, national education inspectors) were included in the design team.

The resulting educational tablet app “appLINOU” is one of only a few French preschool educational apps developed by a university team. It contains 20 activities covering a wide range of early literacy and early numeracy skills (see Appendix 1 and 2). Each activity increases in difficulty in five stages corresponding to children's abilities during five periods of kindergarten, from the start of the second year to the end of the third (final) year. Each stage of an activity includes around 15 items of varying difficulty. A mascot named LINOU introduces each activity by explaining its purpose (thereby making the activity more meaningful to the child) and provides feedback during the exercises. When children answer an item correctly, the mascot gives a positive and motivating feedback. When children answer incorrectly, it encourages them to try again. If the second answer is also incorrect, it provides guidance (scaffolding) on how to find the right answer. If the third answer is still incorrect, it describes the solution in detail (see Appendix 3).

The gestures required to complete the app's activities—dragging and dropping and selecting images—are straightforward and suited to children of kindergarten age. The app includes other functions to enhance its educational impact. For example, rather than receiving automatic feedback on their answers, children must click on a button to check whether a response is correct. This function gives children extra time to think about their answers. Another button allows children to listen to the instructions again at any time.

### 2.2 | Participants

This study was conducted in the Lorraine Region located in northeast France, which is home to approximately 2.4 million people. Participants were 442 kindergarten students (221 girls and 221 boys), their mean age (collected at the beginning of the experiment) was 4.43 years with a standard deviation of 0.28. Only children with no known developmental difficulties were selected to participate in this study. A particular attention was given to select schools from different geographic areas (urban vs. rural), and different socioeconomic status (SES). The SES index of each child's family is calculated by combining scores for five indicators: father's level of education, mother's level of education, father's profession, mother's profession, and household income.

This information was collected by a questionnaire sent to parents. Parents' reported highest education level yielded six categories: 1 = less than high school (Mother = 7%, Father = 3%), 2 = vocational or technical program (M = 7%, F = 12%), 3 = high school diploma (M = 15%, F = 9%), 4 = Associate's Degree (M = 10%, F = 18%), 5 = Master's Degree (M = 35%, F = 24%), and 6 = PhD (M = 26%, F = 34%). The father's and mother's profession have been categorized in six categories (1 = manual worker to 6 = executive manager) according to the classification of the French National Institute for Statistics and Economic Studies (INSEE<sup>1</sup>). The parental monthly income

<sup>1</sup><http://www.insee.fr/>

was measured on a 6-point scale ranging from 1 = less than 1000 € to 6 = more than 4000 € per month. The mean family income was 2550 € (standard deviation = 910 €). The strong correlations between these five indicators (which ranged from 0.53 to 0.78 and were all significant at  $p < 0.01$ ) justified grouping them. The SES was then measured by the average of the five indicators. This average ranges from 1 to 55 (mean = 3.98, standard deviation = 1.17). The children were in 11 classes at three kindergartens which differed in terms of their students' socioeconomic backgrounds. All classes were selected to participate in the experiment in each selected school. For this study, we selected all children who had their parents' permission to participate in the experiment. Students from the first school were mostly from disadvantaged backgrounds; students from the second school were mostly from advantaged backgrounds; students from the third and final school were from a variety of socioeconomic backgrounds.

## 2.3 | Measures

### 2.3.1 | Head-Toes-Knees-Shoulders

To compare the automated assessments with traditional RI test results, we relied on a behavioural task: the HTKS (McClelland et al., 2014). This task requires the children to provide a motor response to a verbal command. In Part 1, two conflicting instructions are presented to the child: Touch your toes and touch your head. When the examiner says, 'Touch your toes', the child must touch his or her head and vice versa. In Part 2, new conflicting instructions are added to the first two instructions: Touch your shoulders and touch your knees. In Part 3, there is a final rule switch: in this section, the head is now associated with the knees and the feet with the shoulders. The scoring system is the same for each of the three parts: two points for a correct response, one point for a self-correct (child corrects his or her response unprompted), and a score of 0 for an incorrect response. In our case, the maximum possible score for the 30 items was 60 points. Higher scores indicate higher levels of self-regulation skills (Cronbach's  $\alpha = 0.92$ ).

### 2.3.2 | Response inhibition variables

As explained in the previous sections, we propose to look at students' traces to identify the RI behaviour. More precisely, we propose three new variables to model RI.

The reasoning used to define these variables is that, by using the app, the students leave behind footprints of RI behaviour mixed in their performance traces. Such footprints are subsets of traces isolated from the students' performance dataset. Each subset is defined by querying the dataset to find a specific combination of traces that can be interpreted as the representation of a RI behaviour. To isolate these footprints, we defined three RI variables. These variables were first proposed in a preliminary short paper (Medeiros Machado et al., 2020), and were then refined and assessed for this study. The three RI

variables are the following: The first represents students who insisted on the same error consecutively (SERIALFAIL); the second represents students who failed very fast (FASTFAIL); the third represents students who committed the most common error of the activity (COMMONFAIL). In the remainder of this article, we will use the following notation to formally define the RI variables.  $S = \{s_1, s_2, \dots, s_{|S|}\}$  will denote the set of students, while  $A = \{a_1, a_2, \dots, a_{|A|}\}$  will represent the collection of all proposed activities. The whole set of responses that a student  $s$  has given to an activity  $a$  will be listed in the sequence  $R_{sa} = \langle r_1, r_2, \dots, r_{|R_{sa}|} \rangle$ . Each response of a specific activity ( $r_i \in R_{sa}$ ) is then assigned to the set of wrong answers  $W_a = \{w_1, w_2, \dots, w_{|W_a|}\}$ , or to the set of correct answers  $C_a = \{c_1, c_2, \dots, c_{|C_a|}\}$ . The time spent by the student to provide the answer is also encoded within the responses  $r$ .

Using this notation, we have  $R_{sa} \cap W_a$  the set of wrong answers given by student  $s$  on activity  $a$  and  $R_{sa} \cap C_a$  the set of correct answers given by student  $s$  on activity  $a$ . We can notice that in our context, this last set can be either a singleton corresponding to the correct answer that the student gave or an empty set when the student gave no correct answer.

#### *Students who insisted in the same error (SERIALFAIL)*

This variable is based on the definition of dysfunctional behaviour of executive functioning given by Henry and Bettenay (2010). The authors define that a dysfunctional executive functioning behaviour is characterized by a person 'trapped in "repetitive cycles" of a previously learned pattern'. To isolate this repetitive behaviour, we took the errors registered by each child in the app and ordered them in temporal sequences. To be considered a sign of a behavioural problem, the same error must happen multiple times in sequence.

A practical question to the definition of students who insisted in the same error (SERIALFAIL) is how many times each specific error must happen in sequence to be considered as a repetitive cycle of previously learned pattern. In the case of our study, each student has up to three chances to correctly answer a question, otherwise, the applINOU gives the correct answer, and the student does not gain the points of the question.

This parameter is key to the definition of the SERIALFAIL variable. To define and formalize it, it must be clear that repetitive errors must appear in a timely sequence without any other registered answers within a SERIALFAIL sequence. Taking as input a timestamp ordered dataset, we can formally define the sequence of recorded errors in each question as:

$$SE_{sa} = \{r_i \in R_{sa} \cap W_a \mid \forall j \in \{1, 2, 3, \dots, k\}, r_i = r_j\}$$

$SE_{sa}$  depicts an error that a student  $s$  perpetrated several times in an activity  $a$ . The reason why  $j$  starts in 1, is because we are interested in the errors the students committed from their first answer to each activity. The value  $k$  represents the parameter of how many times the same wrong answer must appear to be considered as a dysfunctional RI behaviour. Given this definition of  $SE_{sa}$ , we define the SERIALFAIL variable as follows:

$$\text{SERIALFAIL}(s) = \log \left( \sum_a |SE_{sa}| \right). \quad (1)$$

A concrete example of a student's behaviour targeted by this variable is presented in the activity 'The bee is foraging'. This activity demands the student to slide a bee into a targeted flower, a set of flowers is displayed in a queue, and they are described by their colour and position in the queue. The student was requested to put the bee into the first purple flower. However, when performing the activity, the student misplaced the bee twice. For this variable, we are not interested to know where the student put the bee, the important information is that it was misplaced twice in a row.

#### Students who failed very fast (FASTFAIL)

This variable is derived from the definition of RI by Miyake et al. (2000), that is 'one's ability to deliberately inhibit **dominant, automatic, or prepotent** responses when necessary'. We are especially interested in measuring the student's ability to fast inhibit a prepotent wrong answer. To do so, we look at the time the students took to wrongly answer each question.

The key question to ask here is how much time represents a fast failure that can represent an inability to suppress a wrong answer. Differently from the first variable, this time threshold needs to be defined first in the means of each question presented in the app. In other words, we need to know how much time is demanded by each question to be correctly answered. Second, we need to look at the student's behaviour to find traces of dysfunctional behaviour, that is all the time a specific student wrongly answered a question in less time than the threshold specifically identified for this question. This threshold is defined by taking the answer time distribution of each question. Further details about this definition are given in Section 2.4.2. Once the threshold is known, the student dysfunctional traces are isolated and used to compute the students who failed very fast (FASTFAIL) variable.

We can formally note  $AW_a$  as the set of all wrong answers provided by all the students for the exercise  $a$ , ordered by increasing timestamp. A threshold  $\phi_a$  is then used to determine if a fast failure has occurred in activity  $a$ . Once we get  $\phi_a$ , the student's subset of answers representing a very fast failure attempt is easily defined as:

$$VF_{sa} = \{r_i \in R_{sa} \cap W_a \mid \text{time}(r_i) \leq \phi_a\}$$

To exemplify this variable, we take the same activity used before ('The bee is foraging'), but instead of looking for how many times the student misplaced the bee we are interested in how long it takes the student to fail. If a student takes less than 8 s to slide the bee into the wrong flower, we consider such an answer as a dysfunctional RI behaviour. This threshold is different for each activity, and it is learned by our algorithm by looking at the students' traces of each specific activity.

Finally, we have:

$$\text{FASTFAIL}(s) = \log \left( \sum_a |VF_{sa}| \right). \quad (2)$$

#### Students who committed the most common error (COMMONFAIL)

The final variable is also defined by taking the definition of RI of Henry and Bettenay (2010) presented for SERIALFAIL. However, in the case of students who committed the most common error (COMMONFAIL), we are mostly interested in isolating the behaviour that represents wrong 'previously learned patterns'. To identify such wrong learned patterns, we will use the community knowledge. That means such patterns can be shared by a group of students and by the analysis of their behaviour we will be able to identify and isolate them.

A concrete example to clarify this definition is taken from one of the app activities. Imagine that one of the activities demands that the student selects a regular pen, but instead, the student selects a pencil (a very similar object). If most of the students make this same mistake in this activity, then we identify a wrong learned pattern.

The key concept and question here is how many times the same error should happen in an activity to be considered a COMMONFAIL. Further details about how to find the optimal number of times each wrong answer must appear to be isolated by COMMONFAIL are given in Section 2.4.2.

To define COMMONFAIL formally, we first need to isolate the most common error of each activity. Let  $mc_a$  be the most common error of each activity  $a$  defined as  $mc_a = \arg \max_r \sum \delta_{rr'}$ , where  $\{r, r' \in AW_a\}$  and  $\delta$  is the Kronecker delta function. For each activity  $a$  and each student  $s$ , we then define the set of responses that correspond to the most common error as:

$$MC_{sa} = \{r_i \in R_{sa} \mid r_i = mc_a\}$$

Based on this definition, we have:

$$\text{COMMONFAIL}(s) = \log \left( \sum_a |MC_{sa}| \right). \quad (3)$$

Besides these formal definitions, all three variables SERIALFAIL, FASTFAIL, and COMMONFAIL are also represented graphically in Figure 2, which shows a direct overview of the target behaviour each of the proposed variables aims at isolating: SERIALFAIL showing the repetitive cycles, FASTFAIL showing the inability of RI captured in a fast failure, and COMMONFAIL showing the wrong previous learned patterns.

## 2.4 | Procedure

### 2.4.1 | Data collection procedure

The data collected for this study was generated by the use of applINO and was gathered after each kindergarten class finished



**FIGURE 2** A visual representation of the three proposed RI variable SERIALFAIL, FASTFAIL, COMMONFAIL

St. ID	Act. ID	Answer	Correct %	Time	Rec. answer	Timestamp
1	AB1	Right	100%	10 s	1	1 January 2019–0:00 AM
2	AB1	Wrong	25%	30 s	0	1 January 2019–10:01 AM

**TABLE 1** Example of students' traces extracted from appLINO and used to model the response inhibition variables

their daily activities. Each class used the app at least once a week during the period from February 2019 to May 2020 included. This period covers the second and the third year of France's kindergartens. For our experiment, we used only the data from the second year (called 'moyenne section'). The second year is also separated into first and second periods' which we will identify by the acronyms MS1 and MS2, respectively. The data used in our experiments concerns MS1 and MS2 and it was collected from February 2019 to April 2019. To ensure the quality of the gathered data, each teacher received specific training to implement a strict protocol that guided the experiment.

To minimize the bias of the results between each school, all the schools used the app at the same pace, that is the experiments started and finished the same specific weeks in all schools. The students also followed the same schedule to perform the activities. The teachers were responsible to perform a synchronization of the data collected in each tablet at the end of the day.

Because of the difficulty to implement the HTKS task with all children involved in the experiment, we decided to measure students' RI at only two distinct periods. The first was at the beginning of the MS1 period and the second at the end of the MS2 period. The goal was to target two distinct periods in time of the same scholar year to assess how much explanation can the proposed RI variables provide about children's RI behaviour at different periods of time.

## 2.4.2 | Analytic procedure

### Collected dataset

After performing the data preprocessing, the dataset was split into two parts belonging to MS1 and MS2 periods respectively. The identification of each period is also important to later associate each student's respective result in the HTKS test for MS1 and MS2 and build the RI dataset. Each line of the collected dataset at the end looks like the example presented in Table 1. The column 'St. ID' represents the identification of the student in the app. 'Act. ID' is the identification of the question the student answered. 'Answer' stores binary information either *Right* if the student correctly answered the question or *Wrong* otherwise. 'Correct %' is

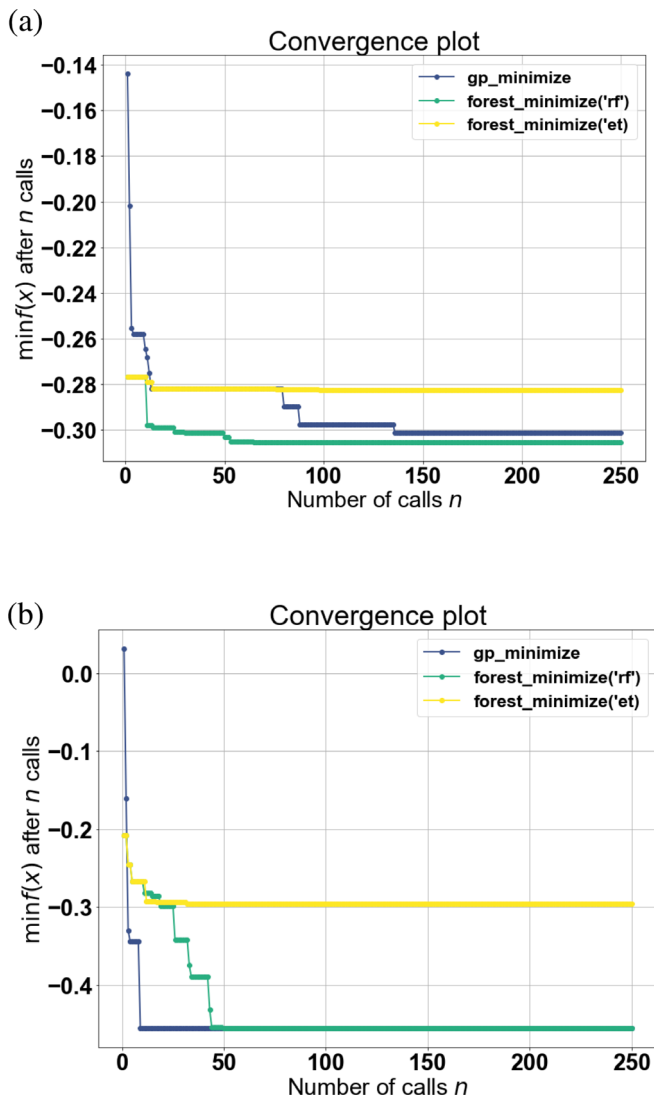
the percentage of the question the student got right; the student needs to get 100% of correctness in a question to register a *Right* code in the column 'Answer'. 'Time' is the information of how many seconds took the student to answer such a question. 'Rec. Answer' is the record of the response the student gave to the question. Finally, 'Timestamp' stores the date and time when the student answered the question.

### Parameters' optimization

As described in Section 2.3.2, each proposed RI variable requires the definition of relevant values to their parameters. For SERIALFAIL, it is important to determine how many times a student should repeat the same error in sequence to consider him/her trapped in a 'repetitive cycle' of a previously learned pattern. For FASTFAIL, the important parameter is the time threshold to consider that the student committed a fast error in an activity. For COMMONFAIL, it is the number of times an error should appear in an activity to be considered as being common. The definition of such parameters is not an easy task because they depend on the student's behaviour, the activities' characteristics, and the characteristics of the group of students. Given the importance of such parameters to the definition of the variables and the correct identification of RI dysfunctional behaviour, we decided to define their values by using an optimization function. This function will look for the parameters' values that correspond to the most acceptable approximation of reality. To optimize the RI variables, we need to define a target function that will receive a set of parameters as input and will give a result that will be analysed to be minimized or maximized. In our case, the target function is a multiple linear regression, and the result to minimize or maximize is the 'adjusted  $R^2$ ' (simply called  $R^2$  in the rest of the text). The function takes the set of our RI variables as independent variables and the HTKS scores as the dependent one to be explained. The optimization process then controls the changing of the RI parameters and analyses the impact of such changes by looking at the  $R^2$ . This process is known as 'hyper-parameter optimization'. Once the best parameters are found, we can retrieve the resulting  $R^2$  and the learning traces that reflect a dysfunctional RI behaviour.

The range used for the variables' optimization is defined as follows:





**FIGURE 3** The RI hyperparameters optimization results for both MS1 and MS2 periods. (a) Optimization for MS1 period. (b) Optimization for MS2 period.

- Number of consecutive errors in `SERIALFAIL`: [2, 3]
- Time threshold for each activity in `FASTFAIL`: [0.01, 0.5]
- Number of times an error must appear to be considered common in `COMMONFAIL`: [10, 60]

The `SERIALFAIL` parameter is limited to up to three errors because this is the limit of attempts a student can make in the appLINO. The `FASTFAIL` parameter range is given by the percentage of the traces belonging to each activity in the dataset when such traces are ordered from the fastest errors to the slowest ones. So, 0.01 represents the top 1% of the fastest errors in the subset of traces belonging to a specific activity. The `COMMONFAIL` parameter range is given by the number of times an error must appear in an activity to be considered a common error. The parameter has an upper limit of 60 because this is the maximum number of times we

observed the same incorrect answer in a specific activity of our dataset.

The optimization is performed through two sequential algorithms that use decision trees, and one algorithm of Bayesian optimization using Gaussian processes. For the decision tree-based algorithms, the surrogate models used are 'Extra trees regression' (et) and 'Random Forest regression' (rf). The results of the hyperparameter optimization are shown in Figure 3, where `gp_minimize` corresponds to the Gaussian-based strategy and `forest_minimize('rf')` and `forest_minimize('et')` corresponds to the tree-based strategies using random forests and extra-trees models respectively. The y-axis presents the values of  $-(R^2)$ .

Based on the results of the optimization process, we selected the `forest_minimize('rf')` algorithm to extract the best parameters for the RI variables, as it was the only one able to minimize the function (i.e., giving the maximum value of  $R^2$ ) for both MS1 and MS2 periods.

### Response inhibition datasets

Once the best parameters for the RI variables are obtained, two new RI datasets are built, which contain the students' id, the mean score in the activities (from now on referred to as activities scores), the RI variables, and the HTKS results, grouping each student in a line of a dataset belonging either to MS1 or to MS2 period. The data about performances in early literacy and early numeracy activities were combined to form the student's activities score, which is given by the mean score in such activities. It is important to stress that the RI variables act as pre-filters to build these datasets. For instance, a student who never presents two or more wrong consecutive answers in any activity will not appear in the RI datasets. For this reason, the students' sample size differs on the regression tests for both periods (MS1 and MS2) when compared with the number of students who used the app. The next sections of this article will present the results and the discussion based on the analysis performed over the RI datasets.

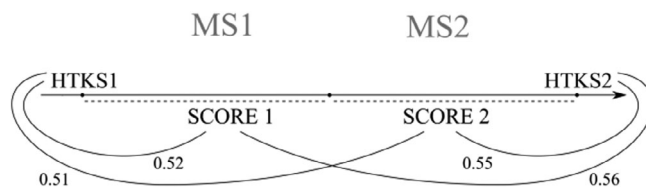
## 3 | RESULTS AND DISCUSSION

We first focus on the descriptive analysis of the students' traces, specifically the students' activities scores, their HTKS scores, and the RI variables. Table 2 presents the statistics where 'MS1' refers to the first period and 'MS2' refers to the second period. The 'SCORE' corresponds to the number of right and wrong answers for each activity. For appLINO, a wrong answer is registered as the number 2 and a correct one as 5. There was a significant enhancement in students' activities scores for the 'MS2' period.

The RI variables `SERIALFAIL`, `FASTFAIL`, and `COMMONFAIL` presented the same behaviour, where we can notice a significant decrease in the students' RI dysfunctional behaviour from MS1 to MS2 period. We can also notice that the students had better activities scores and higher HTKS scores in MS2 than in MS1. Note that the only variable for which we did not find any statistically significant difference is the HTKS result.

Variable	Level	Mean	N	SD	Welch's t	DF	p
SCORE	MS1	3.67	144	0.40	-3.59	54.90	0.00
	MS2	3.97	40	0.48			
SERIALFAIL	MS1	3.71	144	0.70	12.99	69.01	0.00
	MS2	2.22	40	0.62			
FASTFAIL	MS1	1.83	144	0.88	9.02	81.00	0.00
	MS2	0.67	40	0.66			
COMMONFAIL	MS1	3.90	144	0.44	38.32	57.66	0.00
	MS2	0.57	40	0.49			
HTKS	MS1	21.09	144	16.62	-1.11	61.31	0.27
	MS2	24.45	40	16.99			

**TABLE 2** Descriptive statistics between the variables from MS1 and MS2 periods



**FIGURE 4** Spearman's correlations between the students' activities scores and head-toes-knees-shoulder (HTKS) scores for both periods.

**TABLE 3** Spearman's correlation between the variables from MS1 period

Level	Variable	1	2	3	4
MS1	1.SERIALFAIL	1			
	2.FASTFAIL	0.55**	1		
	3.COMMONFAIL	0.76**	0.55**	1	
	4.HTKS	-0.49**	-0.33**	0.55**	1

Note:  $N = 144$ , \*\* $p < 0.01$ .

Next, we examine the correlations between the activities scores and the HTKS scores. Figure 4 shows the Spearman correlations for both MS1 and MS2 periods isolated but also across both periods. Recall that the first HTKS test was performed at the beginning of the MS1 period while the second HTKS test was performed at the end of the MS2 period, while the activities scores were collected throughout each period. Our results show important correlations between the HTKS and the activities scores, in all cases, we found a statistically significant moderate correlation ( $p < 0.05$ ). These results therefore corroborate the correlation between HTKS and students' performance in the context of a learning app.

To determine to which extent our proposed RI variables relate to the HTKS scores, we first looked at their respective Spearman correlations. The results are shown in Table 3 for MS1 and Table 4 for MS2. Note that for MS1, 144 students were considered, while for MS2 40 students were entered in the analysis. The reason for this difference is that the students had a worse performance in MS1, which led to a bigger sample of students that presented one of the target behaviours filtered from the data by the defined RI variables.

**TABLE 4** Spearman's correlation between the variables from MS2 period

Level	Variable	1	2	3	4
MS2	1.SERIALFAIL	1			
	2.FASTFAIL	0.48**	1		
	3.COMMONFAIL	0.36*	0.00	1	
	4.HTKS	-0.64**	-0.13	0.47**	1

Note:  $N = 40$ , \* $p < 0.05$ ; \*\* $p < 0.01$ .

Overall, we can see that, as expected, all RI variables negatively correlate with the HTKS scores. In MS1, all correlations were statistically significant ( $p < 0.05$ ). More precisely, we found a moderate correlation for SERIALFAIL (-0.49), a weak correlation for FASTFAIL (-0.33), and a moderate correlation for COMMONFAIL (-0.55). In MS2, besides FASTFAIL, for which we found no statistically significant correlations, and which had a very weak correlation (-0.13), both other variables were significantly correlated with the HTKS scores ( $p < 0.05$ ). We found a strong correlation for SERIALFAIL (-0.64) and a moderate correlation for COMMONFAIL (-0.47).

The SERIALFAIL and COMMONFAIL variables are also strongly correlated in MS1 (0.76) and provide similar correlations with the HTKS scores. However, SERIALFAIL and COMMONFAIL are weakly correlated in MS2, but SERIALFAIL is highly correlated with HTKS in this period. In general, our proposed variables therefore seem quite relevant indicators of students' RI, especially the SERIALFAIL variable.

To confirm the previous results, we also performed a hierarchical regression. The goal is to assess the effectiveness of each proposed RI variable (independent variables) in explaining the variance of the HTKS scores (dependent variable). To do so, we used each RI variable as input of a multiple linear regression model step-by-step while verifying the impact of inserting these variables by looking at changes in the  $R^2$ .

At each step, a new variable is added and the changes in  $R^2$ , the regression coefficient ( $\beta$ ), the  $T$ -test value ( $t$ ), as well as the  $p$ -value are verified. The first variable inserted is SERIALFAIL, followed by FASTFAIL, and finally COMMONFAIL, the order of input of the variables does not affect the final result. Table 5 shows the obtained statistics.

**TABLE 5** Hierarchical regression from response inhibition variables to head-toes-knees-shoulder (HTKS) result

Variable	$\beta$	$t$	$p$	$R^2$
(A) HTKS for MS1 period (N = 144)				
<b>Model Step 1</b>				0.25
SERIALFAIL	-12.03	-7.03	0.00	
<b>Model Step 2</b>				0.25
SERIALFAIL	-10.91	-5.23	0.00	
FASTFAIL	-1.56	-0.94	0.34	
$F_{\text{for } R^2 \text{ change}} = 0.07$				
<b>Model Step 3</b>				0.30
SERIALFAIL	-5.62	-2.20	0.02	
FASTFAIL	-0.04	-0.02	0.97	
COMMONFAIL	-13.45	-3.35	0.00	
$F_{\text{for } R^2 \text{ change}} = 10.50$				
(B) HTKS for MS2 Period (N = 40)				
<b>Model Step 1</b>				0.41
SERIALFAIL	-17.88	-5.35	0.00	
<b>Model Step 2</b>				0.42
SERIALFAIL	-20.71	-5.35	0.00	
FASTFAIL	5.04	1.40	0.16	
$F_{\text{for } R^2 \text{ change}} = 0.94$				
<b>Model Step 3</b>				0.45
SERIALFAIL	-17.34	-4.04	0.00	
FASTFAIL	3.55	0.98	0.33	
COMMONFAIL	-7.61	-1.65	0.10	
$F_{\text{for } R^2 \text{ change}} = 1.69$				

The table is split into two parts: MS1 at the top and MS2 at the bottom. The first important result is the confirmation of SERIALFAIL as an important variable to explain the variance of HTKS. In MS1 it was responsible for 0.25 and in MS2 for 0.41 of the  $R^2$ . The addition of the other variables was comparably less important to the model explanation, except for the addition of COMMONFAIL in MS1 period which contributed with an extra 0.5 for the  $R^2$  results. When comparing the MS1 and MS2 results, the first difference is that the adjusted  $R^2$  achieved higher values in MS2. This was not a surprise since children still had very few interactions with the app when the HTKS was measured for MS1 because it was the beginning of the year. Another difference is that for MS2 period, the COMMONFAIL variable loses its importance in the model. Overall, all these observations reinforce the importance of SERIALFAIL in explaining HTKS in both periods. These results therefore confirm the relevance of our variables, and especially that of the SERIALFAIL variable, which provided a good explanation power for both periods.

Despite presenting a good explanation power regarding the HTKS scores, the proposed RI variables present some limitations that come either from the context of this study or from the nature of the RI tests themselves. The way we modelled RI reflects the numerous

discussions between the psychology and computer science teams of this project, but it does not mean that such variables can be used as ground truth for modelling RI. Instead, we advise using them as preliminary indicators for dysfunctional RI behaviour, which means they are not able to replace any of the existing RI tests, but they can for instance be used to alert the teachers of a possible dysfunctional RI behaviour. Once our variables have identified some students with a possible dysfunctional behaviour, a RI test (e.g., the HTKS task) can be performed only in the subset of the identified students, saving time while giving the teacher the possibility to propose specific actions to help the students.

To the best of our knowledge, there is no other study that tries to model the complex RI behaviour from simple student traces and assess it with real students. So, this is a novel study that not only proposes to model RI from traces but also assesses the proposition within an important target group, that is kindergarten children. As specified in the introduction, the kindergarten age range is the most important to the development of RI. Moreover, it is also during kindergarten that children will acquire early literacy and early numeracy skills.

Concerning the RI test, the HTKS was selected to be used in this study because it is well suited to kindergarten children and because it is related to the inhibitory control (McClelland et al., 2014). There are fewer tests available to children under 8 years old (Henry & Bettenay, 2010), which makes the HTKS a valuable resource even if the test is also related to other executive functions. An executive functioning test is generally composed of many dimensions of evaluation. In the case of HTKS, visuospatial and verbal abilities are being assessed, which is not always true in the case of appLINOU. For instance, the app cannot assess the motor reaction of the students by using only the traces in the app while the HTKS task can. This makes the task of modelling RI from traces even more challenging, and that is why we suggest the use of our methodology to support the identification of students who will then be followed by a more traditional RI test.

## 4 | CONCLUSIONS

This article presented a novel approach to model RI based on learning traces from a learning app and assessed the extent to which this approach could be used to explain the results of a RI test applied to kindergarten students in France. The assessment was performed in two distinct periods of time while the students were using appLINOU. The results showed a good explanation power of the proposed RI indicators, especially that of SERIALFAIL, as shown by a hierarchical regression analysis.

The theoretical contributions of this article concern the proposition of a methodology to isolate the RI behaviour of kindergarten students from learning traces. Such methodology is sufficiently generic to be used with another learning app that collects simple interaction traces. In practical terms, the RI variables can be used to generate preliminary indicators of RI. Such indicators can be integrated, for instance, into a teacher dashboard to provide real-time measures of

RI. Such measures can alert the teacher about a student who is struggling to learn due to a possible RI dysfunctional behaviour. The teacher can then ask the student to be tested by a traditional RI test like the HTKS.

In future works we intend to extend the data collection to other periods of the kindergarten, and to split the analysis to compare not only periods but also early literacy and early numeracy data. The challenge to such a split is to find a way to relate a traditional RI test with a student's skill in early literacy or early numeracy. Other directions also concern the adoption of multiple RI tests, and a comparison with other datasets. We also intend to assess the generalization power of our method by increasing our sample size.

## ACKNOWLEDGEMENTS

This work has received funding from the French Programme of Investments for the Future within the frame of the eFram LINUMEN project.

## CONFLICT OF INTEREST

The authors declare there is no conflict of interest.

## PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/jcal.12773>.

## DATA AVAILABILITY STATEMENT

Open Practices Statement. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

## ORCID

Guilherme Medeiros Machado  <https://orcid.org/0000-0001-5283-9228>

## REFERENCES

- Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., & Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: A meta-analysis. *Developmental Psychology, 50*(10), 2368–2379.
- Antón-Méndez, I., Talk, A., & Johnston, S. (2019). Gaze direction reveals implicit item and source memory in older adults. *PLoS One, 14*(12), 1–18. <https://doi.org/10.1371/journal.pone.0226018>
- Beuget, M., Castagnos, S., Luxembourger, C., & Boyer, A. (2019). Eye gaze sequence analysis to model memory in E-education. *Artificial Intelligence in Education, 24*–29. [https://doi.org/10.1007/978-3-030-23207-8\\_5](https://doi.org/10.1007/978-3-030-23207-8_5)
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641–1660.
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647–663.
- Bondareva, D., Conati, C., Feyzi-Behnagh, R., Harley, J. M., Azevedo, R., & Bouchet, F. (2013). Inferring learning from gaze data during interaction with an environment to support self-regulated learning. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education* (pp. 229–238). Springer Berlin Heidelberg.
- Brun, A., Bonnin, G., Castagnos, S., Roussanaly, A., & Boyer, A. (2019). Learning analytics made in France: The METAL project. *The International Journal of Information and Learning Technology*.
- Cameron, C. E., Brock, L. L., Murrah, W. M., Bell, L. H., Worzalla, S. L., Grissmer, D., & Morrison, F. J. (2012). Fine motor skills and executive function both contribute to kindergarten achievement. *Child Development, 83*(4), 1229–1244.
- Cano, A. R., Fernández-Manjón, B., & García-Tejedor, A. (2018). Using game learning analytics for validating the design of a learning game for adults with intellectual disabilities. *British Journal of Educational Technology, 49*, 659–672.
- Catts, H. W., Herrera, S., Nielsen, D. C., & Bridges, M. S. (2015). Early prediction of reading comprehension within the simple view framework. *Reading and Writing, 28*(9), 1407–1425. <https://doi.org/10.1007/s11145-015-9576-x>
- Carver, A. C., Livesey, D. J., & Charles, M. (2001). Age related changes in inhibitory control as measured by stop signal task performance. *International Journal of Neuroscience, 107*(1–2), 43–61. <https://doi.org/10.3109/00207450109149756>
- Choffin, B., Popineau, F., Bourda, Y., & Vie, J.-J. (2019). DAS3H: Modeling student learning and forgetting for optimally scheduling distributed practice of skills. *JDSE 2019 - Paris-Saclay Junior Conference on Data Science and Engineering*, Sep 2019, Gif-sur-Yvette, France.
- Chong, T. J., Apps, M. A. J., Giehl, K., Hall, S., Clifton, C., & Husain, M. (2018). Computational modelling reveals distinct patterns of cognitive and physical motivation in elite athletes. *Scientific Reports, 8*, 11888.
- Davidse, N. J., de Jong, M. T., Bus, A. G., Huijbregts, S. C. J., & Swaab, H. (2011). Cognitive and environmental predictors of early literacy skills. *Reading and Writing, 24*(4), 395–412.
- Demetriou, A., Mouyi, A., & Spanoudis, G. (2010). The development of mental processing. In *The handbook of life-span development* (pp. 306–345). John Wiley & Sons, Inc..
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Diamond, A., Kirkham, N., & Amso, D. (2002). Conditions under which young children can hold two rules in mind and inhibit a prepotent response. *Developmental Psychology, 38*(3), 352–362. <https://doi.org/10.1037/0012-1649.38.3.352>
- Domenjoz, J.-C. (2017). *Learning analytics et big data pour la formation: Opportunités et risques*. <https://educationauxmedias.ch/learning-analytics-et-big-data-pour-la-formation-opportunités-et-risques>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446.
- Ebner, M., & Schön, M. (2013). Why learning analytics for primary education matters! *Bulletin of the IEEE Technical Committee on Learning Technology, 15*, 14.
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., & Senn, T. E. (2004). The contribution of executive functions to emergent mathematic skills in preschool children. *Developmental Neuropsychology, 26*(1), 465–486.
- Espy, K. A., Bull, R., Martin, J., & Stroup, W. (2006). Measuring the development of executive control with the shape school. *Psychological Assessment, 18*(4), 373–381. <https://doi.org/10.1037/1040-3590.18.4.373>
- Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316272503>
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General, 133*(1), 101–135. <https://doi.org/10.1037/0096-3445.133.1.101>
- Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across

- content areas. *Developmental Psychology*, 50(6), 1698–1709. <https://doi.org/10.1037/a0036633>
- Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, 134(1), 31–60. <https://doi.org/10.1037/0033-2909.134.1.31>
- Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health*, 15(2), 110–119.
- Ilfenthaler, D., & Yau, J. Y.-K. (2020). Utilising learning analytics to support study success in higher education: A systematic review. *Educational Technology Research and Development*, 68(4), 1961–1990. <https://doi.org/10.1007/s11423-020-09788-z>
- Johnson, L., Adams, S., & Cummins, M. (2012). *The nmc horizon report: 2012 higher education edition*. <https://libros.metabiblioteca.org/handle/001/563>
- Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education*, 16(1), 24.
- Karypis, G. (2017). Improving higher education: Learning analytics & recommender systems research. In *Proceedings of the eleventh acm conference on recommender systems* (p. 2). Association for Computing Machinery.
- Kendeou, P., van den Broek, P., White, M. J., & Lynch, J. (2007). Comprehension in preschool and early elementary children: Skill development and strategy interventions. In *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 27–45). Lawrence Erlbaum Associates Publishers.
- Keshavamurthy, U., & Guruprasad, H. S. (2015). Learning analytics: A survey. *CoRR* <http://arxiv.org/abs/1501.06964>
- Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction*, 19(6), 513–526.
- Lalwani, A., & Agrawal, S. (2019). What does time tell? Tracing the forgetting curve using deep knowledge tracing BT - artificial intelligence in education. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *20th International Conference, AIED 2019, Proceedings, Part II* (pp. 158–162). Springer International Publishing.
- Lan, X., Legare, C. H., Ponitz, C. C., Li, S., & Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: A cross-cultural analysis of Chinese and American preschoolers. *Journal of Experimental Child Psychology*, 108(3), 677–692.
- Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education - a literature review. In *Learning analytics: Fundamentals, applications, and trends: A view of the current state of the art to enhance e-learning* (pp. 1–23). Springer International Publishing.
- Lim, L.-A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2021). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, 72, 101202. <https://doi.org/10.1016/j.learninstruc.2019.04.003>
- Litkowski, E. C., Duncan, R. J., Logan, J. A., & Purpura, D. J. (2020). When do preschoolers learn specific mathematics skills? Mapping the development of early numeracy knowledge. *Journal of Experimental Child Psychology*, 195, 104846.
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439–1459. <https://doi.org/10.1177/0002764213479367>
- Lonigan, C. J., Lerner, M. D., Goodrich, J. M., Farrington, A. L., & Allan, D. M. (2016). Executive function of Spanish-speaking language-minority preschoolers: Structure and relations with early literacy skills and behavioral outcomes. *Journal of Experimental Child Psychology*, 144, 46–65.
- Lonigan, C. J., & Shanahan, T. (2008). Executive summary developing early literacy: Report of the National Early Literacy Panel a Scientific Synthesis of early literacy development and implications for intervention. In *Developing early literacy: Report of the national early literacy panel* (pp. 55–106). National Institute for Literacy.
- Machado, G. M., Maran, V., Lunardi, G. M., Wives, L. K., & de Oliveira, J. P. M. (2021). AwARE: a framework for adaptive recommendation of educational resources. *Computing*, 103(4), 675–705.
- Marchal, F., Castagnos, S., & Boyer, A. (2016). A first step toward recommendations based on the memory of users. In 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 54–61). IEEE.
- Marchal, F., Castagnos, S., & Boyer, A. (2018). First attempt to predict user memory from gaze data. *International Journal on Artificial Intelligence Tools*, 27(6), 1850029.
- Martinez, W., & LaLonde, D. (2020). Data science for everyone starts in kindergarten: Strategies and initiatives from the American Statistical Association. *Harvard Data Science Review*, 2(3). <https://doi.org/10.1162/99608f92.7a9f2fd4>
- Matayoshi, J., Uzun, H., & Cosyn, E. (2019). Deep (un)learning: Using neural networks to model retention and forgetting in an adaptive learning system BT - artificial intelligence in education. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *20th international conference, aied 2019, proceedings* (pp. 258–269). Springer International Publishing.
- Medeiros Machado, G., Bonnin, G., Castagnos, S., Hoareau, L., Thomas, A., & Tazouti, Y. (2020). An approach to model children's inhibition during early literacy and numeracy acquisition. In *International Conference on Artificial Intelligence in Education* (pp. 203–207). Springer.
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., & Pratt, M. E. (2014). *Predictors of early growth in academic achievement: the head-toes-knees-shoulders task* (Vol. 5). <https://www.frontiersin.org/article/10.3389/fpsyg.2014.00599>
- McClelland, M. M., John Geldhof, G., Cameron, C. E., & Wanless, S. B. (2015). *Development and Self-Regulation*. <https://doi.org/10.1002/9781118963418.childpsy114>
- Miller, M. R., Giesbrecht, G. F., Müller, U., McInerney, R. J., & Kerns, K. A. (2012). A latent variable approach to determining the structure of executive function in preschool children. *Journal of Cognition and Development*, 13(3), 395–423. <https://doi.org/10.1080/15248372.2011.585478>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7), 2693–2698. <https://doi.org/10.1073/pnas.1010076108>
- Moissa, B., Bonnin, G., Castagnos, S., & Boyer, A. (2019). Modelling students' effort using behavioral data. In *Technology-Enhanced & Evidence-Based Education & Learning Workshop at LAK (Teel 2019)*.
- Naik, V., & Kamat, V. (2019). Analyzing engagement in an on-line session BT - artificial intelligence in education. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *20th international conference, aied 2019, proceedings* (pp. 359–364). Springer International Publishing.
- National Research Council. (2009). *Mathematics learning in early childhood*. National Academies Press.
- Noel Foulin, J. (2005). Why is letter-name knowledge such a good predictor of learning to read? *Reading and Writing*, 18(2), 129–155. <https://doi.org/10.1007/s11145-004-5892-2>

- Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105*(2), 453–464.
- Rajendran, R., Kumar, A., Carter, K. E., Levin, D. T., & Biswas, G. (2018). Predicting learning by analyzing eye-gaze data of Reading behavior. *International Educational Data Mining Society*.
- Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2005). The development of executive attention: Contributions to the emergence of self-regulation. *Developmental Neuropsychology, 28*(2), 573–594.
- Schmitt, S. A., McClelland, M. M., Tominey, S. L., & Acock, A. C. (2015). Strengthening school readiness for head start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly, 30*, 20–31.
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400.
- Simpson, A., & Riggs, K. J. (2006). Conditions under which children experience inhibitory difficulty with a “button-press” go/no-go task. *Journal of Experimental Child Psychology, 94*(1), 18–26. <https://doi.org/10.1016/j.jecp.2005.10.003>
- Syed, R., Collins-Thompson, K., Bennett, P. N., Teng, M., Williams, S., Tay, D. W. W., & Iqbal, S. (2020). Improving learning outcomes with gaze tracking and automatic question generation. In *Proceedings of the web conference 2020* (pp. 1693–1703). Association for Computing Machinery.
- Triana, M. J. R., Martínez-Monés, A., & Villagrà-Sobrinó, S. (2015). Applying learning analytics to a primary school classroom: Benefits and barriers. In *Workshop on Ethics and Privacy in Learning Analytics at the Learning Analytics & Knowledge Conference*.
- Viberg, O., & Grönlund, Å. (2021). Desperately seeking the impact of learning analytics in education at scale: Marrying data analysis with teaching and learning. In *Online learning analytics* (pp. 19–31). Auerbach Publications.
- Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*(3), 848–872 <https://www.ncbi.nlm.nih.gov/pubmed/9680688>
- Whitehurst, G. J., & Lonigan, C. J. (2002). Emergent literacy: Development from prereaders to readers. In S. Neumann & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11–29). The Guildford Press.
- Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). *The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement*. <http://doi.apa.org/getdoi.cfm?doi=10.1037/a0025361>
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In *Proceedings of the fourth international conference on learning analytics and knowledge* (pp. 203–211). Association for Computing Machinery.
- Zelazo, P. D., & Müller, U. (2011). Executive function in typical and atypical development. In *Blackwell handbook of childhood cognitive development* (pp. 445–469). Blackwell Publishers Ltd.
- Zink, N., Lenartowicz, A., & Markett, S. (2021). A new era for executive function research: On the transition from centralized to distributed executive functioning. *Neuroscience and Biobehavioral Reviews, 124*, 235–244.

### SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Medeiros Machado, G., Bonnin, G., Castagnos, S., Hoareau, L., Thomas, A., & Tazouti, Y. (2023). Modelling children's inhibitory skills using learning data from an educational app. *Journal of Computer Assisted Learning, 39*(3), 856–868. <https://doi.org/10.1111/jcal.12773>