RESEARCH ARTICLE

Developmental Science WILEY

Disentangling age and schooling effects on inhibitory control development: An fNIRS investigation

Courtney McKay[1] | **Sobanawartiny Wijeakumar**[1,2] | **Eva Rafetseder**[1,5] | **Yee Lee Shing**[3,4]

[1] Division of Psychology, University of Stirling, Stirling, Scotland

[2] School of Psychology, University of Nottingham, Nottingham, UK

[3] Department of Psychology, Goethe University Frankfurt, Frankfurt, Germany

[4] Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany

[5] Department of Psychology, University of Konstanz, Konstanz, Germany

Correspondence
Yee Lee Shing, Department of Psychology, Goethe University Frankfurt, Frankfurt am Main, Germany.
Email: shing@psych.uni-frankfurt.de

Abstract

Children show marked improvements in executive functioning (EF) between 4 and 7 years of age. In many societies, this time period coincides with the start of formal school education, in which children are required to follow rules in a structured environment, drawing heavily on EF processes such as inhibitory control. This study aimed to investigate the longitudinal development of two aspects of inhibitory control, namely response inhibition and response monitoring and their neural correlates. Specifically, we examined how their longitudinal development may differ by schooling experience, and their potential significance in predicting academic outcomes. Longitudinal data were collected in two groups of children at their homes. At T1, all children were roughly 4.5 years of age and neither group had attended formal schooling. One year later at T2, one group (P1, $n = 40$) had completed one full year of schooling while the other group (KG, $n = 40$) had stayed in kindergarten. Behavioural and brain activation data (measured with functional near-infrared spectroscopy, fNIRS) in response to a Go/No-Go task and measures of academic achievement were collected. We found that P1 children, compared to KG children, showed a greater change over time in activation related to response monitoring in the bilateral frontal cortex. The change in left frontal activation difference showed a small positive association with math performance. Overall, the school environment is important in shaping the development of the brain functions underlying the monitoring of one own's performance.

KEYWORDS
cut-off design, executive functioning, fNIRS, inhibitory control, response monitoring, schooling

1 | INTRODUCTION

The developmental period of transitioning from kindergarten to formal schooling is characterized by remarkable improvements in cognitive functions. As children prepare for and settle into school and classroom environments, they are increasingly expected to orchestrate and exert control over their own thoughts and behaviors, in accordance to goals and context—a set of skills collectively known as executive functioning (EF; Diamond, 2013). In this study, we investigated the longitudinal development of a key component of EF, namely inhibitory control and its neural correlates, how these differ by schooling experience, and their potential significance in predicting academic outcomes.

There is accumulating evidence to suggest inhibitory control, the capacity to interrupt a prepotent response and enact an alternative less salient response associated with goal attainment, may play a key role in determining school readiness (Müller et al., 2008) as well as predicting future academic achievement (Blair & Razza, 2007; Duckworth et al., 2019; Gawrilow et al., 2014; McClelland et al., 2014; Smith-Donald et al., 2007; Son et al., 2019). For instance, Bierman et al. (2008) found that, in a sample of typically developing preschool children, tasks of working memory and inhibitory control predicted emerging literacy skills. This finding is in agreement with Blair and Razza (2007), who examined the role of self-regulation in relation to emerging academic abilities in 3- to 5-year-old children. While several aspects of self-regulation predicted certain academic outcomes, inhibitory control made independent contributions to all three measures of academic ability (mathematical knowledge, letter knowledge, and phonemic awareness). The authors suggested that the ability to inhibit distracting or irrelevant information while reading or when faced with a numerical problem may be a contributing factor to success, over and above specific knowledge of problem solutions. For example, inhibitory control may allow children to consider multiple dimensions of a problem, rather than focusing on the most salient or recent aspects.

While inhibitory control prior to starting school may play an important role in predicting future academic success, the school environment itself may play an equally important role in shaping these skills. In school, children are required to follow classroom rules, sit still, and pay attention for a large portion of the lessons while suppressing any distractions that may interfere with their learning (Bierman et al., 2008). These demands draw heavily on inhibitory processes. Therefore, it is conceivable that the environment of formal schooling may advance the development of inhibitory control, in comparison to kindergartners that tend to be more play-oriented (Morrison et al., 1997).

1.1 | School cut-off design

To estimate the causal effects of schooling on cognitive development is not trivial, as schooling and development are confounded in time. The cut-off design (for a review, see Morrison et al., 2019) is an effective longitudinal method for examining unique schooling effects by taking advantage of arbitrary school cut-off dates. This method compares children who are similar in age, but due to fixed entry dates, are enrolled into different school years. Previous studies with a cut-off design found causal, beneficial effects of schooling on aspects of literacy (Morrison et al., 1995; Varnhagen et al., 1994) and numeracy (Bisanz et al., 1995; Christian et al., 2000). Recent years have seen a growth in research examining schooling-related effects on more basic cognitive processes, such as EF, given the associations shown between its subcomponents with academic achievement (Morrison et al., 2019). However, the findings here are mixed. For instance, Burrage et al. (2008) assessed inhibition in two groups of 5-year-old children born within 4 months of each other during the fall and spring semesters of the school year. The researchers found no significant difference in performance between children who had attended school and those who had stayed in kinder-

Research Highlights

- Using a modified school cut-off design, we collected longitudinal assessments of two aspects of inhibitory control, namely response inhibition and response monitoring, and their neural correlates.
- For response monitoring, P1 children, compared to kindergarten children, showed a greater difference after 1 year of schooling in activation between correct and incorrect responses in the bilateral frontal cortex.
- The left frontal activation difference in P1 children showed a small association with math performance.
- The school environment plays an important role in shaping the development of brain functions underlying the monitoring of one own's performance.

garten. On the other hand, Kim et al. (2021) used a school cut-off design to examine performance on an inhibitory control task in 4- to 7-year-old children. There was a significant difference between first grade children and kindergarten children, with kindergarteners showing greater improvements across the year. However, this result should be interpreted with caution for several reasons. First, based on the data presented, it appears the first-grade children may have been significantly older than the kindergartners at baseline, which was not controlled for in the analyses. Second, initial differences in performance existed between the two groups at the start of the year, with the first graders outperforming the kindergarteners at baseline. Hence, it is unclear whether the kindergarteners improved more from the experience of kindergarten or were just "catching up" in performance with age.

1.2 | Response inhibition and response monitoring

Despite the growing interest in how schooling may influence various aspects of basic cognition, there have been very few neurodevelopmental investigations. The only longitudinal inquiry into schooling-effects on neural correlates of inhibitory control was conducted by Brod et al. (2017). Using a cut-off design, fMRI data were collected on 5- and 6-year-old children while they completed a go/no-go task. This study sought to uncover schooling-related effects in *response inhibition*, and thus focused on activation for successfully inhibited (no-go) and successfully executed (go) trials. While no group differences in activation were found during correct no-go trials, a larger increase in activation in the right superior posterior parietal cortex (PPC), an area associated with sustained attention, was found for correct go trials, only in children who attended school. The authors concluded the increased engagement of the PPC may reflect a direct effect of the schooling experience, where children are required to pay attention for extended periods of time in classrooms.

Although trials with correct responses have traditionally been the focus of analyses in a go/no-go task, a separate literature have

highlighted a unique pattern of activation in response to errors. First recognized by ERP researchers (Falkenstein et al., 1991; Gehring et al., 1993), the negative and positive components that arise following an incorrect response to a no-go trial are referred to as the error-related negativity (ERN) and error-related positivity (Pe). These components presumably reflect the functioning of a network of structures, including the anterior cingulate cortex (ACC) and lateral prefrontal cortex (LPFC), and are thought to reflect error detection and/or conflict resolution processes associated with *response monitoring* (Grammer et al., 2014; Kim et al., 2016).

Interestingly, response monitoring is one of the components of cognitive control that has been linked to academic success (Denervaud, Knebel, et al., 2020; Kim et al., 2016), and its deficits are associated with developmental disorders including attention deficit hyperactivity disorder (Groom et al., 2013). To be successful in school, children must monitor their own progress, detect errors when they occur, and subsequently adapt their own behaviour. In comparison to kindergarten, teachers in school classrooms also provide more directive feedback on the accuracy of children's schoolwork, possibly shaping their sensitivity to errors (Denervaud, Knebel, et al., 2020). Relating response monitoring and schooling, Grammer et al. (2014) administered a go/no-go task to a sample of 3- to 7-year-old children and found that Pe was sensitive to age-related change during the school transition period, where older children exhibited a larger Pe than younger children. Further, Kim et al. (2016) administered a go/no-go task alongside two measures of academic achievement; math and reading. Using a multiple regression analysis, they found that stronger reading and math skills predicted a larger Pe but did not predict the ERN. Thus, the authors concluded that the Pe, rather than the ERN, may be associated with academic achievement. Most developmental research in response monitoring has been conducted using EEG, with a small number of studies that have used fMRI (Denervaud, Fornari, et al., 2020; Rubia et al., 2007). Specifically, Rubia et al. (2007) compared brain activation between adults and children while they completed a modified stop task. During unsuccessful no-go trials (contrasted with successful go trials), adults and children showed similar activation in the medial prefrontal cortex, anterior, and posterior cingulate gyrus. However, adults showed increased activation compared to children in the ACC. Thus, converging evidence from fMRI and EEG investigations has identified neural signatures of response monitoring after committing errors, and highlights the involvement of a network of frontal regions, which may show developmental and/or schooling differences.

1.3 | Present study

Based on the review above, several questions remain that the current study aimed to address. First, although Brod et al. (2017) reported that 1 year of formal schooling results in increased engagement of the PPC, it is unknown whether this increase predicts academic achievement. Previous studies that have investigated the link between response inhibition and academic achievement have been strictly correlational. Thus, any causal links between the two remain to be demonstrated.

Second, it is unknown whether entering formal education causally impacts the frontal networks underlying response monitoring, as none of the studies that examined response monitoring and schooling utilized a cut-off design. To fill in these knowledge gaps, we conducted a study in Scotland with a modified cut-off design. Rather than comparing children born several months before and after a cut-off date, all children in the current study were born in January and February of 1 year. This was possible because in Scotland, school commencement dates fall in August, with the school-starting cohort consisting of children born between the beginning of March in 1 year (aged 5.5) and the end of February (aged 4.5) of the following year. However, parents of children born in January and February can choose to enroll their child into school or defer their entry until the following year, and these requests are automatically approved. Thus, the current study compared two groups of children across time: one group enrolled into school as soon as they were eligible and completed 1 year of primary school (P1), and the other group deferred their school entry and stayed in kindergarten (KG). At timepoint 1 (T1) children in both groups were 4.5-years-old and in kindergarten. At timepoint 2 (T2), children in both groups were 5.5-years-old, but P1 children had completed one full year of schooling while KG children had completed another year of kindergarten.

Thus, this quasi-experimental design allows for the comparison of two groups of children who are similar in age but differ in their experience in a school context. This is important, given that the kindergarten and schooling environments differ in several ways. Specifically, Sharp (2002) conducted a review of UK and European policy on school starting ages and highlighted four important differences between the schooling and kindergarten environments. First, children in school spend less time on tasks of their own choosing as schoolteachers take on a more instructional and didactic role. Second, children spend less time outside engaging in physical activities and discovering their environment and instead, spend more time in class sitting still. This is reflected in research conducted by Quick et al. (2002), who found almost half of the British school headteachers interviewed felt their outdoor learning facilities were inadequate. Third, the school curriculum places a larger emphasis on subject-related academic material as opposed to learning through play and finally, the adult to child ratio is usually higher in pre-school settings.

Our first question sought to determine whether entering formal schooling leads to increased engagement of the neural networks underlying response inhibition and response monitoring. To answer this question, we employed a portable functional near-infrared spectroscopy (fNIRS) system, which allowed us to collect data on children in their homes (see more details in McKay et al., 2021). This system has several advantages over other imaging modalities as it is non-invasive, cost-effective, portable, and easy to use with young children. Our second question inquired whether schooling-specific improvements in response inhibition and/or monitoring, if any, would be associated with improvements in academic achievement[1]. In line with findings by Brod et al. (2017), we predicted both groups would show improvements in

---

[1] This project was pre-registered on As.Predicted.org (#34866). We initially planned to also examine the relationship between performance/neural activation of the go/no-go task with another behavioural EF task that taps into cognitive flexibility (hearts-and-flowers task).

response inhibition, with P1 children showing a larger increase in parietal activation associated with sustained attention as a result of schooling. Further, based on research suggesting a link between response inhibition and future academic success (Blair & Razza, 2007; Gawrilow et al., 2014; McClelland et al., 2014; Smith-Donald et al., 2007; Son et al., 2019), we predicted the schooling-specific increase in parietal activation in the P1 children would be associated with larger improvements in academic achievement. Next, we predicted P1 children would, over time, show stronger response monitoring after committing error (i.e., acting wrongly based on prepotent response), and thus show a stronger change in activation in the frontal cortex in response to error trials. Lastly, based on work by Grammer et al. (2014), we predicted schooling-specific changes in response monitoring in the P1 children would be associated with improvements in academic achievement.

2 | METHOD

2.1 | Participants

Participants were recruited through gateway organizations such as nurseries and leisure centers. Parents of eligible children contacted the research team to schedule a testing session. All children had normal or corrected to normal vision, no history of colour-blindness, no neurological conditions, and were born full term (> 37 weeks) with an uncomplicated birth. Parents and children provided informed consent prior to testing. The research was approved by the General University Ethics Panel (GUEP 375 and 375A) at the University of Stirling.

Children were tested in their home on two separate occasions, across 2 consecutive years. At T1, 95 4.5-year-olds were recruited for the study. Fifteen children were excluded from all analyses; 12 children (5 P1, 7 KG) interfered with the fNIRS set-up (pulled the cap off) before the completion of the task, and three children provided unusable data (two KG children had thick hair that led to poor signal quality, and data from one P1 child was lost due to experimenter error). Hence, a total of 80 children (39 females, M_{age} at T1 = 53.5 months, SD = 1.2, range = 5 months) provided potentially usable fNIRS data at T1 (see further analysis-specific criteria below). All 80 children agreed to take part at T2 (39 females, M_{age} at T2 = 65.5 months, SD = 1.2, range = 5 months). Of these children, 40 (24 females, M_{age} at T2 = 65.6 months, SD = 1.1, $range$ = 5 months) attended P1 in between the two timepoints, and 40 (15 females, M_{age} at T2 = 65.4 months, SD = 1, $range$ = 4 months) remained in KG.[2]

However, due to an error in task programming, the data from the hearts-and-flowers task could not be interpreted. Therefore, we focused on children's performance and brain response on the go/no-go task, and relate these to measures of academic achievement.

[2] Given the potential associations between task compliance and inhibitory skills, we compared children who refused to participate versus children who agreed to participate on a broad range of questionnaire variables that were collected as part of the larger project. These questionnaires assessed factors such as child temperament, quality of the parent-child relationship, and SES status. We found a significant group difference in two subscores of the Parenting Stress Index (Abidin et al., 2013). The first subscore measured child hyperactivity (p=.015), with those who did not participate scoring higher on this scale. We also found that children who did not participate were scored as less demanding (p=.014). However, it is important to note that these correlations do not survive correction for multiple comparison.

2.2 | fNIRS analysis exclusion (see Figure 1)

2.2.1 | Response inhibition

Two children (1 P1, 1 KG) were excluded from the response inhibition fNIRS analyses for contributing fewer than six usable correct no-go trials across both timepoints and five children were excluded for providing incomplete data (two children (2 KG) refused to complete the task and data from three children (2 P1, 1 KG) was corrupted). Hence, a total of 73 children contributed longitudinal data for the response inhibition fNIRS analyses. Of these children, 37 were in P1 group and 36 were in KG group.

2.2.2 | Response monitoring

Fifteen children (8 P1, 7 KG) were excluded from the response monitoring fNIRS analyses for contributing fewer than six usable incorrect no-go trials across both timepoints, and four children were excluded for providing incomplete data (one KG child refused to complete the task and data from three children (2 P1, 1 KG) was corrupted). A total of 61 children contributed to the response monitoring fNIRS analyses. Of these children, 30 were in P1 group and 31 were in KG group.

2.3 | Behavioural exclusion

2.3.1 | Vocabulary

Three children were excluded from the vocabulary analyses (two children (1 P1, 1 KG) refused to do the task and data from one KG child was lost due to experimenter error). Seventy seven children contributed to the final vocabulary analyses. Of these children, 39 were in P1 group and 38 were in KG group.

2.3.2 | Numeracy

Six children were excluded from the numeracy analyses (five children (3 P1, 2 KG) refused to complete the task and data from one KG child was lost due to experimenter error). Seventy four children contributed to the final numeracy analyses. Of these children, 37 were in P1 group and 37 were in KG group.

2.3.3 | School achievement packs

No children were excluded on either the math or phoneme pack.

2.4 | Experimental task

2.4.1 | Cats-and-dogs task

The cats-and-dogs task (CDT), adapted from Brod et al. (2017), was used to measure response inhibition and response monitoring in
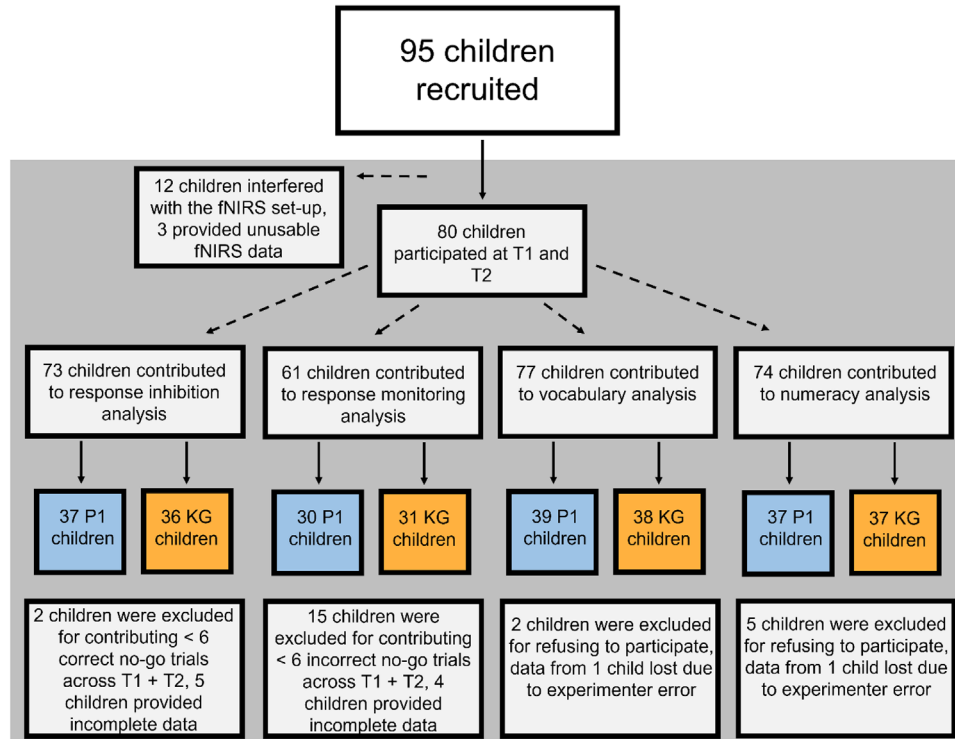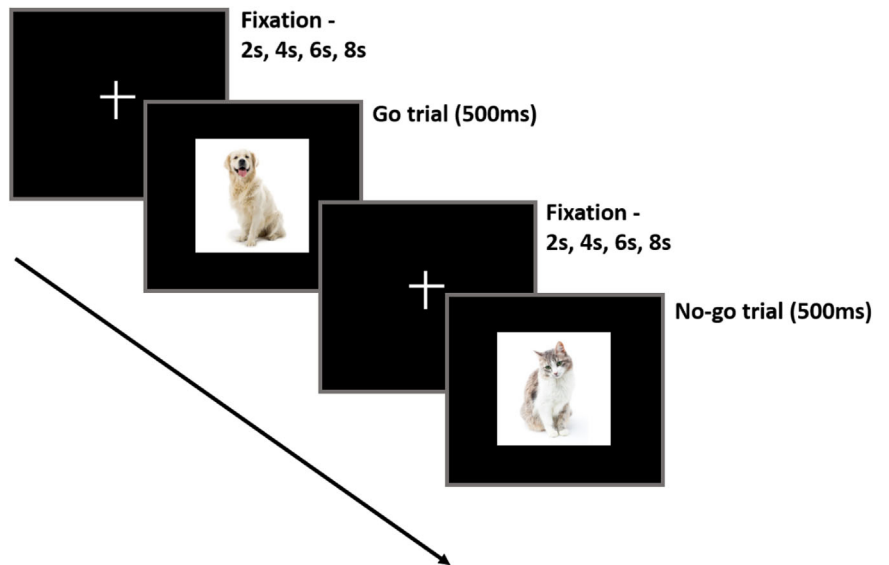
FIGURE 1 Schematic figure for participant recruitment and data exclusion

FIGURE 2 Trial structure of the Cats-and-Dogs Task (CDT)

children—see Figure 2. The task was run in E-prime V.3 software on an HP laptop with a 14-inch screen. During "go" trials, children saw a picture of a dog and were supposed to press a button (spacebar). During "no-go" trials, children saw a picture of a cat and were supposed to withhold pressing a button. To ensure children understood the rules, the session began with three blocks of practice that progressively allowed less time to response. During the practice, children were reminded of the rules if they made a mistake. Performance on the practice runs was not included in final analyses. After children completed all practice blocks, the test session, consisting of two runs began. The first run was comprised of 59 trials: 44 go trials and 15 no-go trials. Run 2 was compromised of 69 trials: 52 go trials and 17 no-go trials. Pictures of cats and dogs were presented for 500 ms, followed by a fixation cross as jitter that ranged in duration from 2 to 8 s. Responses made during stimuli presentation or during the fixation cross period were recorded. The order of presentation of go and no-go trials was pseudorandom, with the constraint that no-go trials were preceded equally often by 1, 2, 4, or 5 go trials.

2.5 | Academic performance measures

2.5.1 | Vocabulary task (administered at T1 and T2)

The vocabulary subset of the Wechsler Preschool and Primary Scale of Intelligence (Warschausky & Raiford, 2018) was used to assess word knowledge. The task included three picture items and 20 verbal items. During the picture items, children were presented with three consecutive pictures of objects (car, scissors, banana) and asked to name each object. If a child incorrectly named the first object (car), they were corrected. Feedback was not provided for the other two picture items. For the verbal items, children were required to provide verbal definitions of words. Corrective feedback was given for the first two verbal items if a child did not receive a perfect score. No feedback was provided for the remaining verbal items. In accordance with the manual, if a child's response was unclear or too vague, the experimenter prompted the child by asking, "What do you mean?", or "Tell me more about it!", or some other neutral query. The test was discontinued if a child gave three consecutive incorrect responses. The picture and verbal items were summed to provide a total vocabulary score (out of 43) at each timepoint.

2.5.2 | Numeracy task (administered at T1 and T2)

The numeracy screener developed by Nosworthy et al. (2013) was used to assess basic numeracy skills. Children were required to compare pairs of magnitudes ranging from one to nine and judge which was larger. Magnitudes were represented symbolically (56 digit pairs) and non-symbolically (56 pairs of dot arrays). In both the symbolic and non-symbolic conditions, numerical magnitude was counterbalanced for the side of presentation. Dot stimuli were also controlled for area and density. Easier items were presented first, followed by more difficult items. Children were given 1 min to complete each condition. The order of the two conditions were counterbalanced across participants. Children received one point for each correct answer. A final score was calculated at each timepoint by subtracting incorrect responses from correct responses.

2.5.3 | School achievement packs (administered at T2 only)

Two measures of achievement were included to assess how much P1 children learned over the course of the first grade in terms of school content. The math pack contained 25 math questions, adapted from the Scottish Curriculum For Excellence teaching resources (twinkl, n.d.). The test was discontinued after three incorrect responses. The phonemes pack contained 20 questions assessing phonetic awareness, adapted from the Heggerty Phonemic Awareness Program (Heggerty, 2019). The pack included 10 items requiring the addition of a phoneme, and 10 items requiring the substitution of a phoneme.

A final score for each pack was calculated by summing the correct responses.

2.5.4 | fNIRS data acquisition

fNIRS data were collected at 7.81 Hz using a NIRSport system 8×8 (8 sources 8 detectors)/release 2.01 with wavelengths of 850 and 760 nm. Fiber optic cables carried light from the machine to a NIRS cap. Probe geometry was designed by collating regions of interest (ROI) from previous fNIRS and fMRI literature (Brod et al., 2017; Wijeakumar et al., 2015). Probe geometry consisted of four channels each on the left and right frontal cortices, and three channels each on the left and right parietal cortices (McKay et al., 2021). Note that short-source-detector channels were not used to regress scalp hemodynamics as all the channels were directed toward maximising coverage of the frontal and parietal cortices. Four cap sizes (50, 52, 54, and 56 cm) were used to accommodate different head sizes. Source-detector separation was scaled according to cap size (50 cm cap: 2.5 cm; 52 cm cap: 2.6 cm; 54 cm cap: 2.7 cm and 56 cm cap: 2.8 cm). To synchronise behavioural and fNIRS data, a McDaq data acquisition device (www.mccdaq.com) was used to send information from the task presentation laptop to the fNIRS system.

2.6 | Procedure

Data were collected in each participant's home. After arrival, the researcher measured the circumference of the child's head and selected an appropriately sized fNIRS cap. Children were given an iPad to watch cartoons during the set-up. Once the cap was fitted to the child's head, measurements were taken from the inion to the nasion and from the two peri-auricular points to make sure that the cap was centered. After the equipment was safely positioned, the instruction and practices for the CDT started, followed by the actual task. During the task, if children indicated that they made an error, the experimenter reassured the child and encouraged them to continue concentrating on the game. To keep children engaged, each test run contained different sets of picture of cats and dogs. To maintain motivation, children were also rewarded with a sticker after each run.

Once children completed the CDT, they were provided the iPad to watch cartoons while the researchers removed the cap. After a short break, testing proceeded with the vocabulary task, followed by the numeracy task. At T2, children were additionally tested on the phonemes pack and math pack. The order of the academic performance tasks presented was counterbalanced across participants. Children were rewarded with stickers after completing each task, regardless of their performance. All children were remunerated with £10 and a toy upon completion of each time point measurement. As part of the overall procedure for the project, children also completed tasks on visual working memory, counterfactual reasoning, and associative memory, while parents filled in questionnaires collecting data on

demographics, child behaviour, and life stress (data not included here, see details in McKay et al., 2021).

3 | DATA ANALYSES

3.1 | Behavioural analyses

Accuracy was calculated separately for each trial type (go and no-go) and test run (run 1 and run 2) and timepoint (T1 and T2). The following formula was used to calculate accuracy and reaction time at each timepoint to account for the different number of trials included in each run.

$$\text{Weighted average:} \frac{((\text{run 1 score} * \text{run 1 number of trials}) + (\text{run 2 score} * \text{run 2 number of trials}))}{\text{Total number of trials}}$$

After computing the weighted averages, a corrected measure of accuracy (against response biases) was calculated for each subject, by subtracting no-go incorrect responses from go correct responses ($\text{Go}_{correct} - \text{NoGo}_{incorrect}$), separately at each timepoint.

3.2 | Outlier correction

All behavioural data were screened for outliers. To correct for longitudinal outliers, we used the Mahalanobis distance (MD) method. Further, we screened for outliers that were ± 3 SDs from the mean at each timepoint. Three outliers were identified: two P1 children were removed from the phonemes pack analyses and one KG child was removed from the math pack analyses. No other outliers were identified.

3.2.1 | fNIRS preprocessing

fNIRS data were pre-processed using the Homer2 package (https://www.nitrc.org/projects/homer2/). Raw data were pruned using the *enPrunechannels* function (SNRthresh = 2, SDrange = 0.0–45). Signals were converted from intensity values to optical density (OD) units using the *Intensity2OD* function. Data were corrected for motion using the *hmrMotionCorrectPCArecurse* function, (tMotion = 1, tMask = 1, STDEVthresh = 50, AMPthresh = 0.5, nSV = 0.97, maxIter = 5, turnon = 1). Data were scanned for motion artifacts using *hmrMotionArtifactByChannel* function (tMotion = 1, tMask = 1, STDEVthresh = 50, AMPthresh = 0.5). Then, the function *enStimRejection* (tRange = −1 to 3) was used to turn off stimulus triggers that contained motion artifacts. The data were band-pass filtered using *hmrBandpassFilt* to include frequencies between 0.016 and 0.5 Hz. Using the function *hmrOD2Conc*, the OD units were converted to concentration units. To find trials that were outliers with respect to the aver-

age HRF, we used the function *hmrFindHrfOutlier* (tRange = −1 to 3, STDEVthresh = 3, minNtrials = 3). Lastly, the HRF was estimated using the ordinary least squares method with a modified gamma function with a square wave (*hmrDeconvHRF_DriftSS* function [tRange = −1 to 3, paramsBasis = 0.1,0.5,0.5, rhoSD_ssThresh = 0, flagSSmethod = 0, driftOrder = 3, flagMotionCorrect = 0]).

3.2.2 | fNIRS group analyses

Oxygenated haemoglobin (HbO) and deoxygenated haemoglobin (HbR) beta values were extracted for each run (run 1 and run 2) and each condition (cue onset of correct go trials, cue onset of correct no-go trials, cue onset of incorrect go trials, cue onset of incorrect no-go trials, response at go trials, response at no-go trials). A weighted average was then calculated to account for the different number of trials included in each test run to produce one beta estimate per subject, per condition, per chromophore, and per timepoint.

3.3 | Response inhibition analyses

For the response inhibition analyses, we focused on HbO and HbR beta estimates for cue onset on correct no-go trials versus correct go trials. These beta values captured activation right after the onset of the stimulus. At T1, the mean number of correct trials included for P1s were 60 ± 4 go trials and 16 ± 1 no-go trials. The mean number of correct trials included for KGs were 66 ± 3 go trials and 18 ± 2 no-go trials. At T2, the mean number of correct trials included for P1s were 69 ± 4 go trials and 19 ± 1 no-go trials. The mean number of correct trials included for KGs were 75 ± 4 go trials and 18 ± 1 no-go trials.

3.4 | Response monitoring analyses

In the pre-registration, we initially only planned for analysis of response inhibition, focusing on correct responses on no go trials. However, based on consideration from the literature, we also investigated activation relating to response monitoring, namely contrasting erroneous responses on no-go trials against correct response on go trials. In both trial types a motor response was conducted, followed by no explicit feedback. Therefore, the post-processing of the erroneous response in the case of no-go trials is assumed to involve the self-detection of error and conflict, which should lead to more monitoring and careful responding in subsequent trials, consequently overall better performance on the task. Thus, for the response monitoring
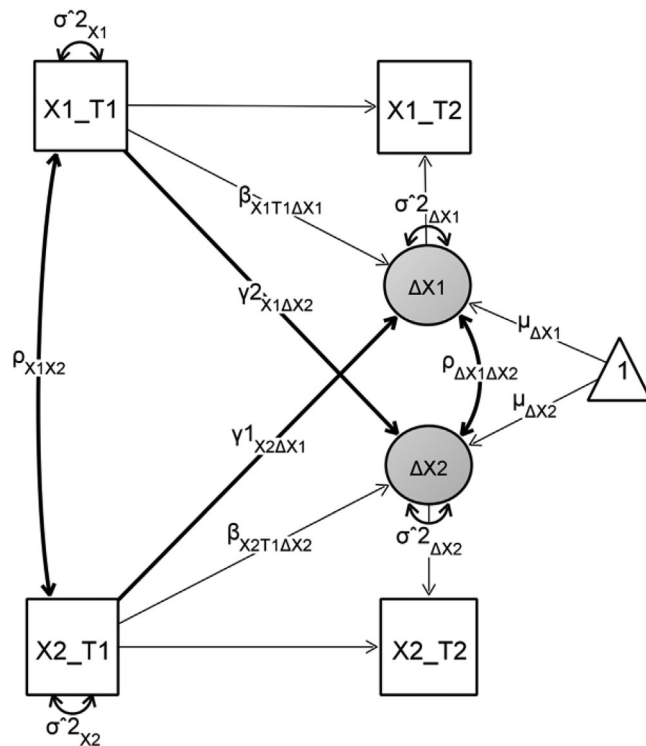
FIGURE 3 Graphical illustration of a bivariate latent change score model. Observed variables are depicted as squares and latent variables as circles. Variances are shown by two-headed arrows self, covariances are shown by two-headed arrows across variables, and regressions are shown by one-headed arrows. Figure created in Onyx (http://onyx.brandmaier.de)

analyses, we focused on HbO and HbR betas estimates for response on go trials and response on no-go trials. These beta values captured activation at the onset of the child's button press. At T1, the mean number of trials included for P1s were 61 ± 4 correct go trials and 12 ± 1 incorrect no-go trials. The mean number of trials included for KGs were 65 ± 4 correct go trials and 10 ± 1 incorrect no-go trials. At T2, the mean number of trials included for P1s were 69 ± 4 correct go trials and 9 ± 1 incorrect no-go trials. The mean number of trials included for KGs were 74 ± 4 correct go trials and 12 ± 1 incorrect no-go trials.

3.5 | Modelling framework

Univariate latent change score (LCS) models (Kievit et al., 2018; McArdle & Hamagami, 2004) were used to investigate the degree of longitudinal change in behavioral performance and brain activation. All univariate models were set up as multi-group models, allowing the same model to be fitted for each group (P1 vs. KG) and later on parameter comparisons. Individual growth is captured by T1 (i.e., the intercept of *X1_T1*–Figure 3) and the latent change score factor ($\Delta X1$), modelled as the difference between the initial observation and subsequent observation. Average group change across time is captured by the mean of the latent change score factor ($\mu_{\Delta X1}$), and between-person

differences in change are captured by the variance ($\sigma 2_{\Delta X1}$). Lastly, the covariance or regression parameter ($\beta_{XT1\Delta X1}$) determines to what extent the amount of change depends on scores at T1.

At the next step, with the inclusion of an extra domain, a univariate LCS model can be extended into a bivariate LCS model, allowing for testing of cross-domain coupling (see Figure 3). To determine whether scores at T1 in one domain (*X1*) are associated with scores at T1 in a second domain (*X2*), the intercept covariance ($\rho_{X1 \times 2}$) is estimated. To examine whether the change in *X1* is associated with the change in *X2*, the change covariance is estimated ($\rho_{\Delta X1\Delta X2}$). Further, the coupling effect ($\gamma 2_{X1\Delta X2}$) determines whether the change in *X1* is a function of the starting point of *X2*, and vice versa ($\gamma 1_{X2\Delta X1}$). For the bivariate LCS model, as motivated by our second research question, only measures that showed schooling-specific effects, from response inhibition/monitoring on the one hand, and academic performance, on the other hand, were included.

3.6 | Model fit indices

Models were estimated in the lavaan software package in R (version 3.6.2, 2019; Rosseel, 2012). Full information maximum likelihood was used for model estimation and to handle missing data. To formally test for significance of parameters of interest, equality constraint was made on the parameter and significance of change in model fit (compared to the just-identified free model) was assessed using the chi-square difference test (at $p < 0.05$). To account for any age and gender effects, these variables were added as covariates into all models.

4 | RESULTS

4.1 | Behavioural results

4.1.1 | Univariate LCS modelling

Four separate univariate models were fitted to each group (P1 and KG) with (1) corrected accuracy on CDT ($Go_{correct}$-$NoGo_{incorrect}$) (2) vocabulary scores (3) symbolic numeracy scores (4) non-symbolic numeracy scores. Raw mean performance levels are illustrated in Figure 4. Parameter estimates are shown in Table 1.

CDT
P1 children showed a significant increase in corrected accuracy between T1 and T2, while KG children did not. However, when the change in corrected accuracy was constrained to be equal across groups, model fit was not significantly worse, $\Delta x^2 = 1.237$, $\Delta df = 1$, $p = 0.266$. There was also no significant worsening in model fit when the baseline scores at T1 were constrained to be equal across groups $\Delta x^2 = 0.189$, $\Delta df = 1$, $p = 0.664$. This suggests that P1 children and KG children started out with similar accuracy and changed comparably across the two timepoints, contrary to our hypothesis.
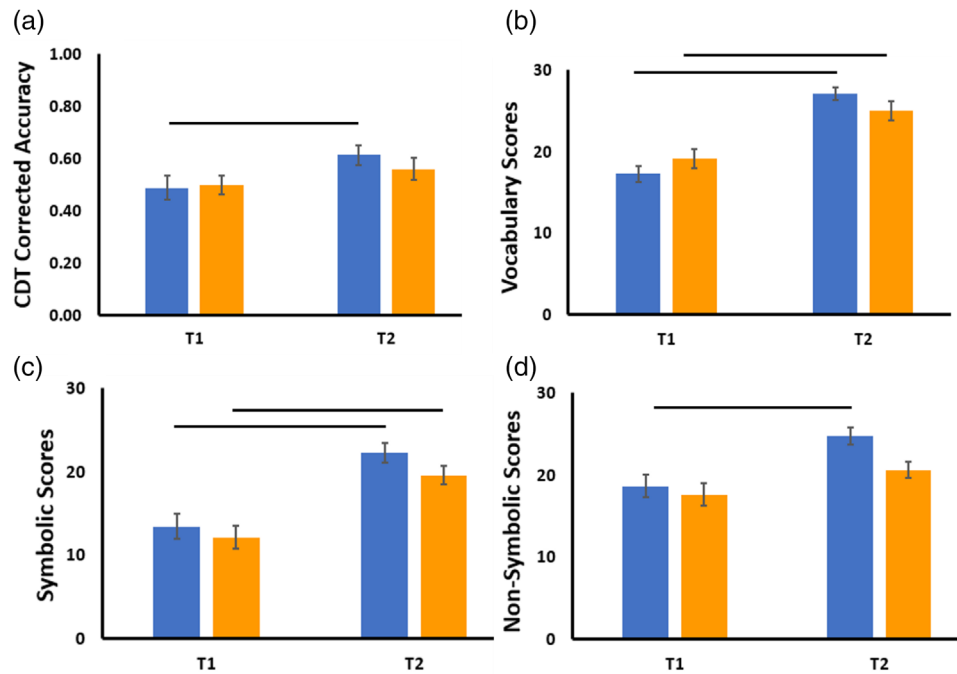
FIGURE 4 Behavioural estimates for the **(a)** CDT task (corrected accuracy based on $Go_{correct}$-$NoGo_{incorrect}$) **(b)** vocabulary task **(c)** numeracy task (symbolic) **(d)** numeracy task (non-symbolic). P1 children are shown in blue and KG children are shown in orange. "-" denotes significance at $p < 0.05$ level (see text for the results of formal model comparison). Error bars show SEM

Vocabulary

Both P1 children and KG children showed a significant increase in vocabulary scores between T1 and T2. Constraining the change to be equal across groups led to a significant drop in model fit $\Delta x^2 = 5.001$, $\Delta df = 1$, $p = 0.025$, suggesting P1 children increased significantly more than KG children. No significant differences at T1 were found $\Delta x^2 = 0.084$, $\Delta df = 1$, $p = 0.772$. Therefore, P1 children and KG children started out with similar accuracy, but the improvement in P1 children on vocabulary knowledge was greater than the improvement in KG children.

Numeracy

For the symbolic condition, both P1 children and KG children showed a significant increase in scores between T1 and T2. No significant drop in model fit was found when the change was constrained to be equal across groups $\Delta x^2 = 0.413$, $\Delta df = 1$, $p = 0.520$. Further, no significant baseline difference was found when the scores at T1 were constrained to be equal across groups $\Delta x^2 = 3$, $\Delta df = 1$, $p = 0.083$. For the non-symbolic condition, P1 children significantly improved between the two timepoints while KG children did not. However, when the change was constrained to be equal across groups, no significant drop in model fit was observed $\Delta x^2 = 2.037$, $\Delta df = 1$, $p = 0.154$. Further, no significant drop in model fit was found after constraining T1 estimates to be equal across groups $\Delta x^2 = 0.002$, $\Delta df = 1$, $p = 0.969$. Thus, for both conditions of the task, P1 children and KG children started out with similar scores and they changed comparably between the two timepoints.

School achievement packs

Univariate models could not be fitted to the school achievement packs as they were only administered at T2. Thus, simple *t*-tests were conducted to compare performance between P1 and KG children on these measures. As expected, we found that P1 children (Math: $M = 30.1$, $SD = 6.6$; Phonemes: $M = 6.4$, $SD = 4$) performed significantly better than KG children (Math: $M = 23.9$, $SD = 6.5$; Phonemes: $M = 2.5$, $SD = 2.7$) on both math and phonemes, respectively ($t[77] = 4.233$, $p < 0.001$; $t[76] = 5.067$, $p < 0.001$).

4.1.2 | fNIRS results

fNIRS data were comprised of HbO and HbR beta values for each of the 14 channels. To reduce data dimension and focus subsequent analyses on effects that had a difference between HbO and HbR, an initial repeated measure ANOVA including chromophore (HbO, HbR) as a factor was run for each channel, using the Benjamini-Hochberg method to correct for multiple comparisons. For the response inhibition analyses, a repeated measures ANOVA with a within-subject factor of trial type (go correct, no-go correct) and chromophore (HbO, HbR) and a between-subjects factor of group (P1, KG) was run for each of the 14 channels. For the response monitoring analyses, a repeated measures ANOVA with a within-subject factor of trial type (go correct, no-go incorrect) and chromophore (HbO, HbR) and a between-subjects factor of group (P1, KG) was run for each of the 14 channels. We focused on significant interactions involving chromophore as a

TABLE 1 Parameter estimates for P1 children and KG children from four separate univariate models on the behavioral tasks

	$Go_{correct}$-$NoGo_{incorrect}$		Vocabulary		Numeracy (Symbolic)		Numeracy (Non-Symbolic)	
	P1	KG	P1	KG	P1	KG	P1	KG
Mean change $\mu_{\Delta X}$	0.13* (0.05)	.06 (.05)	9.79* (1.15)	5.72* (1.31)	8.88* (1.53)	7.47* (1.55)	6.18* (1.45)	2.96 (1.66)
Change variance $\sigma^2_{\Delta X}$	0.09* (0.02)	.08* (.02)	52.59* (13.23)	64.65* (20.55)	91.24* (20.61)	88.9* (17.31)	78.56* (19.06)	105.5* (22.56)
Intercept X_T1	−1.97 (1.58)	−1 (1.57)	−96* (31.63)	−80.03 (44.05)	69.42 (47.81)	−67.68 (63.16)	−10.44 (50.94)	−14.03 (76.52)
Intercept variance σ^2_X	0.07* (0.01)	.05* (.01)	31.08* (6.41)	46.8* (7.16)	76.26* (24.08)	64.38* (13.33)	65.37* (14.02)	67.13* (14.2)
Intercept-change regression $\beta_{XT1\Delta X}$	−0.05* (0.01)	−.03* (.01)	−26.96* (8.23)	−34.18* (10.24)	−56.59* (20.81)	−50.93* (11.86)	−50.22* (12.9)	−67.39* (16.3)
Age[a] onto intercept	0.15 (0.1)	.09 (.1)	6.88* (1.93)	6.07* (2.68)	−3.41 (2.91)	4.88 (3.87)	1.76 (3.1)	1.94 (4.7)
Age-change covariance	−0.03 (0.016)	−.02 (.02)	−.92* (.44)	.11 (.35)	.387 (.528)	−.89* (.44)	−.46 (.54)	.12 (.57)
Gender[b] onto intercept	0.07 (0.04)	.06 (.03)	.9 (.9)	.57 (1.16)	.67 (1.52)	−.09 (1.55)	.48 (1.46)	.38 (1.81)
Gender-change covariance	0.01 (0.05)	−.02 (.04)	−.87 (1.07)	−.97 (1.33)	−1.9 (1.48)	2.19 (1.62)	−1.54 (1.52)	.84 (1.68)

Standard errors are in parentheses.
*Asterisks denote significance at p < 0.05 level.
[a] Age = age in days/100.
[b] Gender coded as 1 = girls, −1 = boys.

factor, and followed up with post-hoc analyses conducted on the HbO estimates.

Response inhibition analyses

Only channels that showed a significant interaction involving chromophore and that survived the Benjamini-Hochberg correction are reported. The interaction between trial type and chromophore was significant in channels overlying the right middle frontal gyrus ($F[1,71] = 12.052$, $p = 0.001$), the right inferior frontal gyrus ($F[1,71] = 8.241$, $p = 0.005$), the right supramarginal gyrus ($F[1,70] = 7.932$, $p = 0.006$), and the left supramarginal gyrus ($F[1,71] = 11.876$, $p = 0.001$). Following up on the interaction, post-hoc tests revealed that HbO activation for go correct trials was greater than activation for no-go correct trials (see Table 2). The remaining three-and four-way interactions between group, trial type, time, and chromophore were either not significant or did not survive the Benjamini-Hochberg correction.

Response monitoring analyses

Only channels that showed a significant interaction with chromophore and that survived the Benjamini-Hochberg correction are reported. The interaction between trial type and chromophore was significant in channels overlying the right middle frontal gyrus ($F[1,57] = 21.134$, $p < 0.001$; $F[1,57] = 15.341$, $p < 0.001$), the right inferior frontal gyrus ($F[1,57] = 19.023$, $p < 0.001$), the left middle frontal gyrus ($F[1,57] = 40.548$, $p < 0.001$), the left inferior frontal gyrus ($F[1,57] = 18.279$, $p < 0.001$; $F[1,57] = 10.769$, $p = 0.002$), and the right supramarginal gyrus ($F[1,56] = 6.773$, $p = 0.012$). Following up on the interaction, post-hoc tests revealed that HbO activation for (erroneous) response at no-go trials was more negative than for (correct) response at go trials (see Table 3).

A significant four-way interaction between group, time, trial, and chromophore was observed in channels overlying the right middle frontal gyrus ($F[1,57] = 10.198$, $p = 0.002$; $F[1,57] = 5.671$, $p = 0.021$), the right inferior frontal gyrus ($F[1,57] = 7.402$, $p = 0.009$), the left middle frontal gyrus ($F[1,57] = 9.912$, $p = 0.003$), the left inferior frontal gyrus ($F[1,57] = 5.897$, $p = 0.018$), and the right superior occipital gyrus ($F[1,56] = 5.976$, $p = 0.018$). All post-hoc tests are shown in Table 3. Importantly, in the bilateral middle frontal gyrus and bilateral inferior frontal gyrus, P1 children showed greater negative activation for response at incorrect no-go trials than for correct go trials at both T1 and at T2. This was not the case for KG children, who only showed a difference in activation between these trials at T1. Therefore, the ANOVA revealed that the difference in activation between correct go trials and incorrect no-go trials across time differentiated P1 children from KG children. To relate these neural differences in response monitoring to behavior using the bivariate LCS models, an average difference in activation (go correct activation–no-go incorrect activation) was computed across channels of nearby regions that showed the significant four-way interaction with similar patterns. Specifically, this led to two clusters covering the right frontal cortex (averaging channels 1, 2, and 3; see Figure 5a) and the left frontal cortex (averaging channels 5 and 7; see Figure 5b).

TABLE 2 Response inhibition analysis: Channels showing significant interactions between trial type (go correct and no-go correct) and chromophore

Channel No.	Brain areas (MNI coordinates)	Trial x Chromophore (HbO)
Channel 1	Right middle frontal gyrus	Go > No-go ($p = 0.007$)
Channel 2	Right middle frontal gyrus	
Channel 3	Right inferior frontal gyrus	Go > No-go ($p = 0.006$)
Channel 4	Right inferior frontal gyrus	
Channel 5	Left middle frontal gyrus	
Channel 6	Left middle frontal gyrus	
Channel 7	Left inferior frontal gyrus	
Channel 8	Left inferior frontal gyrus	
Channel 9	Right angular gyrus	
Channel 10	Right superior occipital gyrus	
Channel 11	Right supramarginal gyrus	Go > No-go ($p = 0.008$)
Channel 12	Left inferior parietal lobule	
Channel 13	Left angular gyrus	
Channel 14	Left supramarginal gyrus	Go > No-go ($p = 0.004$)

Significant post-hoc results are shown for HbO estimates.

TABLE 3 Response monitoring analysis: Channels showing significant two-way interaction between trial type (go correct vs. no-go incorrect) and chromophore, and four-way interaction between group, trial type, time, and chromophore

Channel No.	Brain areas (MNI coordinates)	Trial × Chromophore (HbO)	Group × Trial × Time × Chromophore (HbO)
Channel 1	Right middle frontal gyrus	Go > No-go ($p < 0.001$)	P1 T1: Go > No-go ($p = 0.048$); P1 T2: Go > No-go ($p = 0.001$); KG T1: Go > No-go ($p = 0.036$)
Channel 2	Right middle frontal gyrus	Go > No-go ($p < 0.001$)	P1 T1: Go > No-go ($p = 0.013$); P1 T2: Go > No-go ($p < 0.001$)
Channel 3	Right inferior frontal gyrus	Go > No-go ($p < 0.001$)	P1 T1: Go > No-go ($p = 0.004$); P1 T2: Go > No-go ($p < 0.001$); KG T1: Go > No-go ($p = 0.005$)
Channel 4	Right inferior frontal gyrus		
Channel 5	Left middle frontal gyrus	Go > No-go ($p < 0.001$)	P1 T1: Go > No-go ($p = 0.001$); P1 T2: Go > No-go ($p = 0.002$)
Channel 6	Left middle frontal gyrus		
Channel 7	Left inferior frontal gyrus	Go > No-go ($p < 0.001$)	P1 T1: Go > No-go ($p = 0.012$); P1 T2: Go > No-go ($p = 0.001$); KG T1: Go > No-go ($p < 0.001$)
Channel 8	Left inferior frontal gyrus	Go > No-go ($p = 0.025$)	
Channel 9	Right angular gyrus		
Channel 10	Right superior occipital gyrus		P1 T2: Go > No-go ($p = 0.05$)
Channel 11	Right supramarginal gyrus	Go > No-go ($p = 0.042$)	
Channel 12	Left inferior parietal lobule		
Channel 13	Left angular gyrus		
Channel 14	Left supramarginal gyrus		

Significant post-hoc results are shown for HbO estimates.
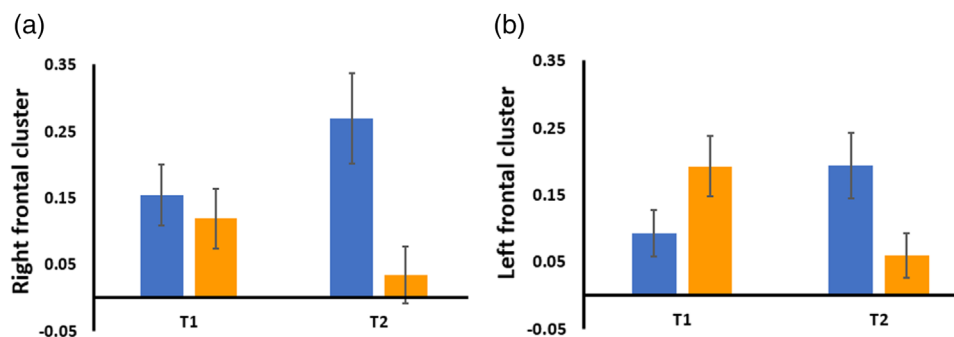
(a)

(b)

FIGURE 5 The difference in activation between go correct and no-go incorrect trials (response monitoring contrast) in the **(a)** right frontal cluster **(b)** left frontal cluster. P1 children are shown in blue and KG children are shown in orange. Error bars show SEM

TABLE 4 CDT bivariate couplings between (a) right frontal cluster and corrected accuracy ($Go_{correct}$-$NoGo_{incorrect}$)and (b) left frontal cluster and corrected accuracy ($Go_{correct}$-$NoGo_{incorrect}$), separately for P1 children and KG children

	a. Right frontal cluster		b. Left frontal cluster	
	P1	KG	P1	KG
Intercept covariance $\rho_{X1 \times 2}$	.65 (1.3)	2.97* (.78)	3.32* (1)	1.83* (.82)
Right frontal onto corrected accuracy change $Y1_{X2\Delta X1}$	0 (0)	−.01 (0)	0 (0)	0 (0)
Corrected accuracy onto right frontal cluster change $Y2_{X1\Delta X2}$	−53.5* (22.37)	3.03 (20.73)	−10.1 (19.92)	31.91* (14.6)
Change-change covariance $\rho_{\Delta X1\Delta X2}$	.98 (1.41)	−.47 (1.11)	1.35 (1.31)	1.16 (.73)

Standard errors are in parentheses.
*Asterisks denote significance at $p < 0.05$ level.

4.2 | Bivariate LCS modelling

As the first step, we tested the longitudinal coupling between activation difference in the two frontal clusters and corrected accuracy on the CDT task for both groups. This is mainly to verify the functional relevance of the two frontal clusters of response monitoring activation for overall task performance.

4.2.1 | Right frontal cluster and CDT corrected accuracy

Parameter estimates are shown in Table 4. For KG Children, corrected accuracy at T1 was positively correlated with the difference in activation in the right frontal cluster at T1. Namely, children who showed more difference in activation related to response monitoring had better performance. Constraining the baseline correlation at T1 to be 0 in KG children led to a significant drop in model fit, $\Delta x^2 = 10.707$, $\Delta df = 1$, $p = 0.001$. No other cross-domain parameters were significant.

For P1 children, corrected accuracy at T1 negatively predicted the change in the difference in activation in the right frontal cluster from T1 to T2. Thus, children with better performance at T1 showed less change in activation over time. However, constraining the coupling pathway to be 0 did not lead to a significant drop in model fit

$\Delta x^2 = 3.776$, $\Delta df = 1$, $p = 0.052$. No other cross-domain parameters were significant.

4.2.2 | Left frontal cluster and CDT corrected accuracy

For KG children, better corrected accuracy at T1 was correlated with higher difference in activation in the left frontal at T1. Constraining the baseline correlation at T1 to be 0 in KG children led to a significant drop in model fit, $\Delta x^2 = 5.028$, $\Delta df = 1$, $p = 0.025$. Furthermore, higher corrected accuracy at T1 predicted more change in the difference in activation in the left frontal. To follow-up on this, the coupling pathway was constrained to be 0 in KG children, which led to a significant drop in model fit $\Delta x^2 = 4.492$, $\Delta df = 1$, $p = 0.034$.

For P1 children, similar to KG children, better corrected accuracy at T1 was correlated with higher difference in activation in the left frontal at T1. Constraining the baseline correlation at T1 to be 0 in P1 children lead to a significant drop in model fit, $\Delta x^2 = 5.536$, $\Delta df = 1$, $p = 0.019$. No other cross-domain pathways were significant.

Taken together, in KG and P1 children, higher response monitoring activation difference in the left frontal cluster (additionally right frontal cluster for KG) was related to better overall performance in the inhibitory control task.

4.3 | Academic achievement in P1 children

To address our second research question, we tested to what extent the schooling-specific response monitoring activation changes in the two frontal clusters could predict academic achievement. The longitudinal coupling between the activation difference with performance on the academic tasks was examined. Here, we focused on bivariate relationships of P1 children (since they were the only group that attended school and showed a greater response monitoring activation difference across time). Bivariate relationships for KG children are shown in Supplementary Material (Table S1).

4.3.1 | Right/left frontal cluster and academic achievement

Bivariate longitudinal models were fitted for the response monitoring activation in the right frontal (or left frontal, respectively) and (1) vocabulary scores (2) math pack, and (3) phonemes pack. The longitudinal change in activation in the left frontal cluster was positively correlated with math pack scores at T2 ($p = 0.04$). To follow-up on this finding, the coupling pathway was constrained to be 0, which led to a trend in drop of model fit $\Delta x^2 = 3.488$, $\Delta df = 1$, p = 0.062. No other cross-domain parameters were found to be significant in all other models.

5 | DISCUSSION

The present study sought to examine to what extent 1 year of formal schooling shapes the development of neural processes underlying response inhibition and response monitoring, as well as establish whether these effects, if significant, were related to academic achievement. First, we found that P1 children and KG children started out with similar corrected accuracy on the go/no-go task. Although P1 children, but not KG children, showed significant improvement on task accuracy over time, the magnitude of change between the two groups was statistically comparable. In terms of brain activation, no significant differences in response inhibition were found between the two groups of children. However, for response monitoring, after 1 year of schooling P1 children showed a greater activation difference than KG children. Functionally, this activation difference was associated with better performance on the go/no-go task. When relating to broader measures of academic achievement, we found a small association between response monitoring and math performance. Each of these aspects of results are discussed in the following.

While we hypothesized that P1 children would show greater improvement than KG children across the year, our findings are in line with Brod et al. (2017) who also reported no group differences in response inhibition behaviour across the year. However, unlike Brod et al. (2017) and in contrary to our hypothesis, we also did not find any group difference in neural activation related to response inhibition (or parietal activation during go trials as in Brod et al., 2017). Several methodological differences exist that may account for this

inconsistency. First, children in the current study were between 1 to 2 years younger than the children in the Brod et al. (2017) study, due to national differences in school entry age. It is conceivable that the first year of schooling may be set up to be less demanding and formally structured where children start school at a younger age. Thus, the increase in parietal activation resulting from a schooling environment may only appear if there is a sufficiently large change in terms of demand and expectations transitioning from kindergartens to classrooms. Second, the current study and that of Brod et al. (2017) employed different modalities to record brain activation. The fNIRS channel-based analyses employed here may not have been as sensitive as the fMRI analyses conducted by Brod et al. (2017) to detect changes in activation in small clusters of voxels (as reported in that study). A potential way to improve upon this would be to conduct more targeted analyses. For instance, novel image reconstruction uses a head model to generate functional images of the fNIRS data, transforming surface level channel-based data into a volumetric representation within the brain (Forbes et al., 2021). This would allow for greater comparability with fMRI investigations.

Another limitation of our research may be related to the longitudinal nature of the study. Longitudinal research with fNIRS (and all neuroimaging modalities in general) faces the challenge that the recorded areas may not remain consistent over time, particularly in development when children's brains are actively developing and growing. However, Collins-Jones et al. (2021) recently used image reconstruction to investigate the effects of variation in array position and head size in channel-space analysis of longitudinal fNIRS infant data (when head growth is most rapid). Specifically, they investigated the effect of variation in head size and array position on inferences drawn from both individual and group-level data. They found differing inferences between the individual and group-level data were primarily due to variability in array position, however, this effect decreased as group size increased. Specifically, this study included a sample size of 53 at 5-months, 40 at 8-months, and 45 at 12-months. Thus, the authors concluded that at these group sizes, the inferences drawn from group level channel-based analysis are unlikely to be significantly affected by variability in array position and shifting head sizes.

Finally, Brod et al. (2017) employed a traditional school cut-off design, in which children whose birthdates fell shortly before and shortly after an arbitrary cut-off date were compared, resulting in group assignment that is to some extent random (parents may request deferral of school entry). On the other hand, the current study took advantage of school commencement regulations in Scotland, where parents of children born in January and February each year can autonomously chose to enrol or defer their child's entry to school. Thus, it is conceivable that parents make this decision based on certain child characteristics, leading to fundamental differences between children who are enrolled versus those who are deferred. While this is possible, the evidence we have in the study suggests that this was likely not the case. First, P1 and KG children showed no differences in performance at the first timepoint on any of the cognitive and academic measures included here. Second, the current study is part of a larger project and thus, we could compare performance between P1 and KG

children at the first timepoint (prior to starting school) on other measures. This included a colour change-detection task (assessing visual working memory) and a grid memory task (assessing associative memory). No group differences in performance were found on any of these measures. Finally, and most interestingly, we also administered a battery of questionnaire to the parents of the children assess here measuring child temperament, quality of the parent-child relationship, and a range of environmental factors including SES, level of disorganisation in the home, and number of daily hassles experienced by parents. Critically, no differences between P1 and KG children emerged in any of these parent-reported measures. In sum, parental belief concerning the child's school readiness is not systematically reflected in quantitative measures of cognitive function, academic performance, and temperament. Descriptive statistics and statistical test results for these comparisons can be found in Supplementary Materials (Table S2 and S3).

Our second major finding concerns the activation related to response monitoring, in which we found that P1 children, but not KG children, showed a greater difference after 1 year of schooling. As the response monitoring contrast was not part of the study pre-registration, it was important for us to first establish the functional relevance of the two frontal clusters (left and right middle/inferior frontal gyrus) that emerged from this contrast. Therefore, we tested the coupling between the difference in activation with performance on the CDT task, and found that a greater response monitoring activation difference in the left frontal cluster was related to better performance in the inhibitory control task in both groups. For KG children, a similar relationship was also found for the right frontal cluster. This is in line with previous research reporting that a greater difference in activation between correct go and incorrect no-go trials reflects more efficient response monitoring (Grammer et al., 2014; Torpey et al., 2012), which may support better task performance. Previous adult fMRI studies have implicated a broader network of frontal regions subserving response monitoring. For example, Chevrier et al. (2007) administered a stop-signal task and found error-related activity in frontal regions including the right middle frontal gyrus and dorsal ACC. Furthermore, Edwards et al. (2012) administered a go/no-go task and combined ERP time courses and fMRI spatial maps allowing for the identification of brain regions that are associated with portions of the time course in the ERP data. They identified two components associated with significant activation in the bilateral middle frontal gyrus and caudal ACC, demonstrating that both regions are engaged during error processing. The authors argued the simultaneous involvement of both areas may reflect a post-error cognitive response, where conflict between the executed and supposedly correct response occurs via the caudal ACC and LPFC. Based on experimenter observations in the current study, this interpretation seems likely as children occasionally showed a reaction reflecting conflict after making an incorrect button press in a no-go trial. Children would either verbally indicated that they made a mistake (e.g., saying "oh no") or show behavior of having committed an error (e.g., clasping hands over mouth, pulling hand away from keyboard).

Most importantly, in the two frontal clusters identified from the response monitoring contrast, P1 children showed a greater difference

in activation across time than KG children. We posit that, across the first school year, P1 children show stronger response monitoring due to the nature of the schooling environment. In school, emphasis is placed on instructional learning where children are provided with opportunities to engage in schoolwork and gain insights into their own performance based on teacher feedback (Denervaud, Knebel, et al., 2020). As this instructional learning takes hold, children learn to value correct answers and avoid errors (Denervaud, Knebel, et al., 2020). In contrast, the kindergarten environment introduces learning through more play-initiated activities (Morrison et al., 1997). While free play orientation may benefit children in many ways, it likely does not encourage the identification of errors on academic tasks as effectively as formal schooling (Denervaud, Knebel, et al., 2020).

Third, as the next step to determine whether the larger activation difference in response monitoring in the P1 children could predict academic performance, we investigated the longitudinal coupling between these variables. We found borderline significant positive correlations between the change in activation in the left frontal cluster with performance on the math pack. This is in line with Kim et al. (2016) who found that stronger math skills (as well as reading skills) predicted stronger ERP component related to response monitoring. Further support for our finding stem from previous adult EEG research that found a larger ERN was significantly correlated with better academic performance (Hirsh & Inzlicht, 2010). Given that monitoring one's own performance is a key aspect of self-regulation, the authors interpreted that individuals with a greater ability to monitor tend to engage in self-regulatory behaviours that are important for academic success (Pintrich & De Groot, 1990). It is however important to note that the change-change association between the left frontal cluster activation and math performance did not survive the formal model comparison. Therefore, the result needs to be interpreted with caution and stands for replication test. Future studies need to be better powered in terms of sample size. Hertzog et al. (2006) evaluated the statistical power of latent change score models and found even with large sample sizes and multiple measurement occasions, statistical power to detect covariance in change remains low. Given the modest sample size of the present study coupled with the inclusion of only two measurement occasions, we likely did not have sufficient statistical power to detect meaningful relationships, even when present.

Fourth, in terms of more general schooling effects on academic measures, we found that P1 children showed greater improvements than KG children in vocabulary. The existing literature into whether and why schooling might improve vocabulary has been somewhat mixed. Morrison et al. (2019) conducted a review of the literature into schooling effects on vocabulary and found that three out of the five studies failed to find a positive effect of schooling. Further, the two studies that did find a positive effect either had higher program standards or a curriculum that emphasised greater vocabulary instruction. A potential reason for this discrepancy in findings might be related to SES. Wright and Neuman (2014) found that children from lower income schools (i.e., schools where more than 50% of students received free and reduced lunch) encounter less opportunity for vocabulary learning than children from higher income schools

(i.e., schools where fewer than 25% of students received free and reduced lunch). Specifically, the authors found that teachers from economically disadvantaged schools were less likely to discuss word meanings with children and also explained fewer challenging words. Consequently, children from lower income schools received only 60% of the vocabulary instruction provided to their more economically advantaged peers. This is important, as several studies have demonstrated that instruction aids children's vocabulary acquisition (Biemiller & Boote, 2006; Kim, 2017). In the current study, parental education and income was above the national average, and thus, it is likely children were enrolled into economically advantaged schools that provided a greater degree of instructional learning, leading to an improvement in vocabulary beyond children attending kindergarten.

Finally, it is worth mentioning that, for KG children only, those who began the study with better performance on the CDT task showed a greater increase in response monitoring activation across the year. We did not predict this result but it seems interesting, given that the KG children, at the mean level, did not show a significant change in activation difference across time. One interpretation for this finding relates to the interplay between children's individual characteristics and the schooling/kindergarten environment. We posit that the schooling environment may have facilitated all school children, regardless of their starting point, to become more sensitive to task accuracy and error, leading to a mean change in brain activation across the year associated with stronger response monitoring. On the other hand, for the reasons highlighted above, kindergarten children may encounter less explicit instruction. Only those who are already advanced at the start, presumably by eliciting more advanced interaction with adult caregivers, show a change in brain activation associated with more efficient response monitoring. Future studies should test this postulation by getting more direct measurement of social/instructional environment of children.

The current study uniquely contributes to the current special issue on the development of self-regulation, cognitive control, and executive function, by being the first study to use a cut-off design to assess the impact of 1 year of schooling on both response inhibition and response monitoring and to relate these differences to measures of academic achievement. Our findings highlight the causal roles of the school environment in shaping the development of brain functions underlying EF, particularly in the monitoring of one own's error. Such schooling-specific neurocognitive changes can predict specific aspect of academic performance across the first school year and may indicate how adaptive children are in adjusting to the new formal schooling environment. At the methodological level, our study demonstrated the feasibility of collecting good quality neural data using fNIRS from children in their homes. Future studies may explore to what extent such neural measures can be utilized to identify children who are potentially struggling during the critical transitional period from kindergarten to first grade.

CONFLICT OF INTEREST

No potential conflict of interest is reported by the authors.

ETHICAL APPROVAL STATEMENT

This research was approved by the General University Ethics Panel (GUEP 375 and 375A) at the University of Stirling.

DATA AVAILABILITY STATEMENT

All group-level data and analysis scripts have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/nyjqe/. Further information regarding the fNIRS pre-processing and analysis are available upon request.

ORCID

Sobanawartiny Wijeakumar https://orcid.org/0000-0002-1008-3407

Eva Rafetseder https://orcid.org/0000-0002-9816-8607

Yee Lee Shing https://orcid.org/0000-0001-8922-7292

REFERENCES

Abidin, R., Flens, J. R., & Austin, W. G. (2013). The parenting stress index. *Forensic Uses of Clinical Assessment Instruments*, https://doi.org/10.4324/9780203726587

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology*, 98(1), 44–62. https://doi.org/10.1037/0022-0663.98.1.44

Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, 20(3), 821–843. https://doi.org/10.1017/S0954579408000394

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. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Brod, G., Bunge, S. A., & Shing, Y. L. (2017). Does one year of schooling improve children's cognitive control and alter associated brain activation? *Psychological Science*, 28(7), 967–978. https://doi.org/10.1177/0956797617699838

Burrage, M. S., Ponitz, C. C., McCready, E. A., Shah, P., Sims, B. C., Jewkes, A. M., & Morrison, F. J. (2008). Age- and schooling-related effects on executive functions in young children: A natural experiment.

Child Neuropsychology, *14*(6), 510–524. https://doi.org/10.1080/09297040701756917

Bisanz, J., Morrison, F. J., & Dunn, M. (1995). Effects of age and schooling on the acquisition of elementary quantitative skills. *Developmental Psychology*, *31*(2), 221–236. https://doi.org/10.1037/0012-1649.31.2.221

Chevrier, A. D., Noseworthy, M. D., & Schachar, R. (2007). Dissociation of response inhibition and performance monitoring in the stop signal task using event-related fMRI. *Human Brain Mapping*, *28*(12), 1347–1358. https://doi.org/10.1002/hbm.20355

Collins-Jones, L. H., Cooper, R. J., Bulgarelli, C., Blasi, A., Katus, L., McCann, S., Mason, L., Mbye, E., Touray, E., Ceesay, M., Moore, S. E., Lloyd-Fox, S., & Elwell, C. E. (2021). Longitudinal infant fNIRS channel-space analyses are robust to variability parameters at the group-level: An image reconstruction investigation. *Neuroimage*, *237*, 118068. https://doi.org/10.1016/j.neuroimage.2021.118068

Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development*, *1*(4), 429–448. https://doi.org/10.1207/S15327647JCD0104_04

Denervaud, S., Fornari, E., Yang, X.-F., Hagmann, P., Immordino-Yang, M. H., & Sander, D. (2020). An fMRI study of error monitoring in Montessori and traditionally-schooled children. *Npj Science of Learning*, *5*, 11. https://doi.org/10.1038/s41539-020-0069-6

Denervaud, S., Knebel, J. F., Immordino-Yang, M. H., & Hagmann, P. (2020). Effects of traditional versus montessori schooling on 4- to 15-year old children's performance monitoring. *Mind, Brain, and Education*, *14*(2), 167–175. https://doi.org/10.1111/mbe.12233

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Duckworth, A. L., Taxer, J. L., Eskreis-Winkler, L., Galla, B. M., & Gross, J. J. (2019). Self-control and academic achievement. *Annual Review of Psychology*, *70*, 373–399. https://doi.org/10.1146/annurev-psych-010418-103230

Edwards, B. G., Calhoun, V. D., & Kiehl, K. A. (2012). Joint ICA of ERP and fMRI during error-monitoring. *Neuroimage*, *59*(2), 1896–1903. https://doi.org/10.1016/j.neuroimage.2011.08.088

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9

Forbes, S. H., Wijeakumar, S., Eggebrecht, A. T., Magnotta, V. A. & Spencer, J. P. (2021). Processing pipeline for image reconstructed fNIRS analysis using both MRI templates and individual anatomy. *Neurophotonics*, *8*(2), 025020. https://doi.org/10.1117/1.NPh.8.2.025010

Gawrilow, C., Fäsche, A., Guderjahn, L., Gunzenhauser, C., Merkt, J., & von Suchodoletz, A. (2014). The impact of self-regulation on preschool mathematical achievement. *Child Indicators Research*, *7*, 805–820. https://doi.org/10.1007/s12187-013-9201-y

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*(6),. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x

Grammer, J. K., Carrasco, M., Gehring, W. J., & Morrison, F. J. (2014). Age-related changes in error processing in young children: A school-based investigation. *Developmental Cognitive Neuroscience*, *9*, 93–105. https://doi.org/10.1016/j.dcn.2014.02.001

Groom, M. J., Liddle, E. B., Scerif, G., Liddle, P. F., Batty, M. J., Liotti, M., & Hollis, C. P. (2013). Motivational incentives and methylphenidate enhance electrophysiological correlates of error monitoring in children with attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *54*(8), 836–845. https://doi.org/10.1111/jcpp.12069

Heggerty, M. (2019). *Literacy resources, Inc.* Heggerty Phonemic Awareness Curriculum. https://heggerty.org/

Hertzog, C., Lindenberger, U., Ghisletta, P., & Von Oertzen, T. (2006). On the power of multivariate latent growth curve models to detect correlated change. *Psychological Methods*, *11*(3), 244–252. https://doi.org/10.1037/1082-989X.11.3.244

Hirsh, J. B., & Inzlicht, M. (2010). Error-related negativity predicts academic performance. *Psychophysiology*, *47*(1), 192–196. https://doi.org/10.1111/j.1469-8986.2009.00877.x

Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A. L., de Mooij, S. M. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., Fonagy, P., Lindenberger, U., & Dolan, R. J. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, *33*(February 2017), 99–117. https://doi.org/10.1016/j.dcn.2017.11.007

Kim, M. H., Ahmed, S. F., & Morrison, F. J. (2021). The effects of kindergarten and first grade schooling on executive function and academic skill development: evidence from a school cutoff design. *Frontiers in Psychology*, *11*, 607973. https://doi.org/10.3389/fpsyg.2020.607973

Kim, M. H., Grammer, J. K., Marulis, L. M., Carrasco, M., Morrison, F. J., & Gehring, W. J. (2016). Early math and reading achievement are associated with the error positivity. *Developmental Cognitive Neuroscience*, *22*, 18–26. https://doi.org/10.1016/j.dcn.2016.09.002

Kim, Y. S. G. (2017). Multicomponent view of vocabulary acquisition: An investigation with primary grade children. *Journal of Experimental Child Psychology*, *162*, 120–133. https://doi.org/10.1016/j.jecp.2017.05.004

McArdle, J. J., & Hamagami, F. (2004). Latent difference score structural models for linear dynamic analyses with incomplete longitudinal data. *New methods for the analysis of change*. https://doi.org/10.1037/10409-005

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. *Frontiers in Psychology*, *5*, 599. https://doi.org/10.3389/fpsyg.2014.00599

McKay, C. A., Shing, Y. L., Rafetseder, E., & Wijeakumar, S. (2021). Home assessment of visual working memory in pre-schoolers reveals associations between behaviour, brain activation and parent reports of life stress. *Developmental Science*, *24*(4), e13094. https://doi.org/10.1111/desc.13094

Morrison, F. J., Griffith, E. M., & Alberts, D. M. (1997). Nature-nurture in the classroom: Entrance age, school readiness, and learning in children. *Developmental Psychology*, *33*(2), 254–262. https://doi.org/10.1037/0012-1649.33.2.254

Morrison, F. J., Kim, M. H., Connor, C. M., & Grammer, J. K. (2019). The causal impact of schooling on children's development: Lessons for developmental science. *Current Directions in Psychological Science*, *28*(5), 441–449. https://doi.org/10.1177/0963721419855661

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Developmental Psychology*, *31*(5), 789–799. https://doi.org/10.1037/0012-1649.31.5.789

Müller, U., Lieberman, D., Frye, D., & Zelazo, P. D. (2008). Executive function, school readiness, and school achievement. In S. K. Thurman & C. A. Fiorello, *Applied cognitive research in K–3 classrooms. BT – Applied cognitive research in K–3 classrooms* (pp. 41–83). Routledge.

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *Plos One*, *8*(7), e67918. https://doi.org/10.1371/journal.pone.0067918

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, *82*(1), 33–40. https://doi.org/10.1037/0022-0663.82.1.33

Quick, S., Lambley, C., Newcombe, E., & Aubrey, C. (2002). *Implementing the foundation stage in reception classes*. Research Report 350, London: Department for Education and Skills.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rubia, K., Smith, A. B., Taylor, E., & Brammer, M. (2007). Linear age-correlated functional development of right inferior fronto-striato-cerebellar networks during response inhibition and anterior cingulate during error-related processes. *Human Brain Mapping*, 28(11), 1163–1177. https://doi.org/10.1002/hbm.20347

Sharp, C. (2002). School starting age : European Policy and Recent Research. *LGA seminar "When should our children start school?"*.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, 22(2), 173–187. https://doi.org/10.1016/j.ecresq.2007.01.002

Son, S. H. C., Choi, J. Y., & Kwon, K. A. (2019). Reciprocal associations between inhibitory control and early academic skills: Evidence from a Nationally representative sample of head start children. *Early Education and Development*, 30(4), 456–477. https://doi.org/10.1080/10409289.2019.1572382

Torpey, D. C., Hajcak, G., Kim, J., Kujawa, A., & Klein, D. N. (2012). Electrocortical and behavioral measures of response monitoring in young children during a Go/No-Go task. *Developmental Psychobiology*, 54(2), 139–150. https://doi.org/10.1002/dev.20590

twinkl. (n.d.). https://www.twinkl.co.uk/

Varnhagen, C. K., Morrison, F. J., & Everall, R. (1994). Age and schooling effects in story recall and story production. *Developmental Psychology*, 30(6), 969–979. https://doi.org/10.1037/0012-1649.30.6.969

Warschausky, S., & Raiford, S. E. (2018). Wechsler preschool and primary scale of intelligence. In *Encyclopedia of clinical neuropsychology*. https://doi.org/10.1007/978-3-319-57111-9_1606

Wijeakumar, S., Spencer, J. P, Bohache, K., Boas, D. A., & Magnotta, V. A. (2015). Validating a new methodology for optical probe design and image registration in fNIRS studies. *Neuroimage*, 106, 86–100. https://doi.org/10.1016/j.neuroimage.2014.11.022

Wright, T. S., & Neuman, S. B. (2014). Paucity and disparity in kindergarten oral vocabulary instruction. *Journal of Literacy Research*, 46(3). https://doi.org/10.1177/1086296/14551474

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.