Received: 5 February 2021 | Revised: 22 November 2021 | Accepted: 4 February 2022

DOI: 10.1111/desc.13252

Developmental Science WILEY

PAPER

Intra-individual variability in task performance after cognitive training is associated with long-term outcomes in children

Ana Cubillo[1,2] | Henning Hermes[3] | Eva Berger[4] | Kirsten Winkel[5] | Daniel Schunk[4] | Ernst Fehr[1] | Todd A. Hare[1]

[1]Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Zurich, Switzerland

[2]Department of Child and Adolescent Psychiatry, University Psychiatric Clinics Basel, Basel, Switzerland

[3]DICE, Heinrich Heine University of Dusseldorf, Dusseldorf, Germany

[4]Chair of Public and Behavioral Economics, Johannes Gutenberg University of Mainz, Mainz, Germany

[5]Chair of Statistics and Econometrics, Johannes Gutenberg University of Mainz, Mainz, Germany

Correspondence
Ana Cubillo, Zurich Center for Neuroeconomics, Department of Economics, University of Zurich, Zurich, Switzerland.
Email: ana.cubillo@econ.uzh.ch

Funding information
National Center of Competence in Research Affective sciences, Grant/Award Number: 51NF40-104897

Abstract

The potential benefits and mechanistic effects of working memory training (WMT) in children are the subject of much research and debate. We show that after five weeks of school-based, adaptive WMT 6–9 year-old primary school children had greater activity in prefrontal and striatal brain regions, higher task accuracy, and reduced intra-individual variability in response times compared to controls. Using a sequential sampling decision model, we demonstrate that this reduction in intra-individual variability can be explained by changes to the evidence accumulation rates and thresholds. Critically, intra-individual variability is useful in quantifying the immediate impact of cognitive training interventions, being a better predictor of academic skills and well-being 6–12 months after the end of training than task accuracy. Taken together, our results suggest that attention control is the initial mechanism that leads to the long-run benefits from adaptive WMT. Selective and sustained attention abilities may serve as a scaffold for subsequent changes in higher cognitive processes, academic skills, and general well-being. Furthermore, these results highlight that the selection of outcome measures and the timing of the assessments play a crucial role in detecting training efficacy. Thus, evaluating intra-individual variability, during or directly after training could allow for the early tailoring of training interventions in terms of duration or content to maximise their impact.

KEYWORDS

attention control, children, cognitive training, fMRI, intra-individual variability, working memory

1 | INTRODUCTION

Cognitive training programs have received considerable attention over the years given their potential to improve cognitive abilities in healthy and clinical populations. However, the effectiveness and persistence of benefits from cognitive training programs are still being closely examined and vigorously debated (Au et al., 2015; Bogg & Lasecki, 2014; Cortese et al., 2015; Karbach & Verhaeghen, 2014; Melby-Lervag et al., 2016; Sala & Gobet, 2020; Schwaighofer et al., 2015; Smid et al., 2020; Wass et al., 2012). Although cognitive training programs have been shown to improve performance on similar untrained tasks (near-transfer), the evidence for transfer to cognitive skills in other domains (far-transfer) remains more sparse and controversial (Au et al., 2015; Cortese et al., 2015; Delalande et al., 2020; Gilligan et al., 2020;

Jones et al., 2020; Karbach & Verhaeghen, 2014; Melby-Lervag et al., 2016; Sala & Gobet, 2020; Schwaighofer et al., 2015; Smid et al., 2020). We still lack sufficient understanding of the types of cognitive skills and abilities that are most beneficial to train, types of training methods and dosages that work best for particular skills, and the types of individuals that can reap sufficient benefits to justify the time and monetary costs of cognitive training interventions.

As with many aspects of cognitive training, the extent of far-transfer effects to academic achievements is subject to intense debate. Improvements in academic performance seem to be stronger for the language and reading domain and less consistent in mathematics, although this varies depending on the type of training regime and study sample (see Sala & Gobet, 2020 and Titz & Karbach, 2014 for review and meta-analysis). For example, there are reports that initial transfer effects to mathematics do not persist three months later (Jones et al., 2020) and that children with low working memory ability show worse math skills than a normal classroom instruction control group 2 years after working memory training (WMT) (Roberts et al., 2016). On the other hand, a recent study found that the effects of training can emerge and increase over time in a cohort of over 500 first-grade children that were not preselected based on working memory abilities (Berger et al., 2020). This study found that the far-transfer benefits from adaptive WMT to academic skills were only evident 6–12 months after the end of training. Moreover, this work showed that five weeks of adaptive WMT during the first-grade year led to an increased probability of entering the highest academic track of the German secondary school system 3–4 years later. Given the aggregate results across multiple studies, it is clear that longitudinal study designs that include follow-up measures over multiple years will be important for determining the potential effectiveness of different types and/or doses of cognitive training, especially for children.

Research Highlights

- Response time variability in traditional working memory and selective attention tasks successfully detects immediate training effects
- Measures of intra-individual response time variability are closely associated with current and future academic performance and well-being.
- Attention control abilities may serve as a mechanism underlying working memory cognitive training effects, supporting the development of later benefits in other areas of cognition.
- The types of outcome measures investigated and the timing of assessments relative to the training period are critical aspects in determining training efficacy.

It is important to understand the cognitive and neurobiological changes that take place during or just after training. Presumably, these proximal effects allow for the eventual emergence of wide-ranging benefits in the future. The level of attention can determine how well information is remembered (Gazzaley & Nobre, 2012). Working memory processes, defined as the temporary storage and manipulation of information that allows for the performance of complex cognitive tasks (Baddeley, 1996; Baddeley, 2010), are therefore closely inter-related with attention control. We refer to the capacity to coordinate and allocate attention to the relevant stimuli in the environment regardless of distractions and fatigue as *attention control* (Cartwright, 2012; Corbetta & Shulman, 2002; Norman & Shallice, 1983). Working memory requires attention control to maintain and reassess task-relevant information while blocking interference from task-irrelevant information (Engle, 2018; Fukuda & Vogel, 2011; Kane et al., 2008; Mcnab & Klingberg, 2008). Both working memory and attention control processes rely on fronto-parietal and striatal brain networks (Klingberg, 2010). Cognitive training has been reported to alter brain structure and function, with induced changes often observed in prefrontal, parietal, and striatal regions (Astle et al., 2015; Buschkuehl et al., 2012; Flegal et al., 2019; Klingberg, 2010; McNab et al., 2009; Salmi et al., 2018; Schneiders et al., 2012). These are crucial regions supporting executive functions such as working memory and attention control (D'Esposito & Postle, 2015; Frank et al., 2001; Mcnab & Klingberg, 2008; Owen et al., 2005; Wager & Smith, 2003). Brain imaging studies suggest that successful transfer from trained to untrained skills requires that both cognitive processes engage at least partially overlapping structural and functional brain systems (Dahlin et al., 2008; Morrison & Chein, 2011). Thus, to be most beneficial cognitive training programs should facilitate neural developments that allow for more effective and efficient engagement of such shared neural systems.

Sensitive and reliable measures of changes in mental and neural functions are necessary to detect the immediate effects of training interventions and to forecast long-term benefits of training. In the current work, we test the hypothesis that intra-individual variability in response times may be useful in this regard. Intra-individual variability measures based on either accuracy or response times have been shown to be more informative than averaged accuracy measures when trying to understand the mechanisms by which beneficial effects of cognitive trainings might transfer to academic skills (Karbach & Unger, 2014; Könen & Karbach, 2015), to anticipate long-term benefits in academic performance (Judd et al., 2021), and to facilitate the identification of those individuals that would benefit most from the training intervention (Karbach & Unger, 2014; Könen & Karbach, 2015; MacDonald et al., 2009; Saville et al., 2011).

Several methods have been used to quantify and distinguish between different cognitive processes that may give rise to intra-individual variability in response times. The individual coefficient of variation (ICV) is a common measure, computed as a straightforward ratio of the standard deviation relative to the mean. In addition, the shape of the response time distribution can be parameterized by fitting ex-gaussian models (Geurts et al., 2008; Hervey et al., 2006; van Belle et al., 2015), and potential sources of variability can be distinguished by fitting Diffusion Decision Models (DDM) (Forstmann et al., 2016; Karalunas & Huang-Pollock, 2013; Ratcliff et al., 2016; Schmiedek et al., 2009). Furthermore, DDMs can also be used to measure and understand the effects of attention on task performance and decision making (Cavanagh et al., 2014; Krajbich & Rangel, 2011; Krajbich et al., 2015).

Intra-individual variability in performance is associated with the cognitive abilities and brain functions targeted by cognitive training interventions (Castellanos et al., 2005; Geurts et al., 2008; Judd et al., 2021; Kofler et al., 2013; MacDonald et al., 2006). Intra-individual variability is also associated with prefrontal brain function and dopaminergic neuromodulation (Ilg et al., 2018; Johnson et al., 2015; MacDonald et al., 2006, 2009; Papenberg et al., 2013; Tamnes et al., 2012; van Belle et al., 2015), especially the inhibitory and cognitive control abilities mediated by frontal and subcortical dopaminergic brain systems (Bellgrove et al., 2004; Isbell et al., 2018; Montez et al., 2017; van Belle et al., 2015). Measures of intra-individual variability may be especially useful when comparing across heterogeneous groups, such as children and adolescents whose cognitive development is still ongoing or populations with cognitive difficulties such as ADHD or Autism (Castellanos et al., 2005; Dirk & Schmiedek, 2016; Geurts et al., 2008; Karalunas et al., 2014; Könen & Karbach, 2015). Consistent with the well-established pattern of brain and cognitive development across the lifespan, intra-individual variability shows an inverted-U shaped association with age, decreasing from childhood through adolescence until young adulthood, and increasing again in old age (Montez et al., 2017; Papenberg et al., 2013; Williams et al., 2005).

An advantage of intra-individual variability measures is that they capture not just the outcome, but also the efficiency of cognitive processes. Increased variability in response times is associated with difficulties in attention control or the ability to maintain attention and goals (Unsworth, 2015). Improved cognitive capacity, enhanced efficiency, or stability of cognitive processes have all been hypothesized as potential mechanisms underlying the impact of training interventions (von Bastian & Oberauer, 2014). Failures of attention during task performance might indicate inconsistent implementation rather than reduced capacity or abnormal function. Such inconsistency in implementing the relevant cognitive systems may not be associated with reduced accuracy if the task or test is not difficult enough and/or yields only coarse measures of accuracy, but may still significantly impair academic performance in children (Judd et al., 2021). Inconsistent implementation of behaviorally relevant cognitive processes can change the response time distribution. These effects will not necessarily lead to differences in average response times, but can be quantified through various metrics of intra-individual response time variability (Ali et al., 2019; Geurts et al., 2008; van Belle et al., 2015).

Intra-individual variability has indeed been associated with attentional lapses. This has been shown in children with ADHD who often show abnormally long RTs on a subset of trials (Hervey et al., 2006; van Belle et al., 2015). Compared to measures of central tendency (e.g., mean or median), intra-individual variability measures have shown to be more sensitive to fatigue in young adults (Wang et al., 2014), or to externally rated attention problems (Gómez-Guerrero et al., 2011), and to correctly classify patients with ADHD (Klein et al., 2006). Therefore, intra-individual variability metrics could plausibly detect changes caused by training interventions that cannot (yet) be captured by mean performance measures because performance variability measures are more sensitive to the efficiency of cognitive processes.

Recent work has shown that performance variability is related to working memory abilities, training, and transfer to academic skills. Intra-individual variability in accuracy within and between sessions in working memory tasks has been associated with academic performance in 3rd and 4th grade school children (Dirk & Schmiedek, 2016), and a recent study in 6-year-old children showed that intra-individual variability after working memory cognitive training was associated with performance in mathematics 3 years later (Judd et al., 2021). Given the predictive association between academic performance at school and well-being in adulthood (Tomasik et al., 2019), it is important to investigate the impact of any cognitive training on academic performance.

In summary, there is sufficient reason to hypothesize that intra-individual response time variability metrics can detect short-term training effects and may be useful in predicting the degree of long-term benefits. Here, we test the hypothesis that intra-individual variability in task performance—quantified via response times—can be used to assess training efficacy in the short term and is correlated with future far-transfer effects. We use a combination of cognitive tasks (N-Back and Flanker), functional magnetic resonance imaging (fMRI), and Diffusion Decision Modelling of individual performance to examine the effects of five weeks of adaptive WMT on brain and cognitive function in first-grade children. Our specific hypotheses were that WMT would benefit performance on the N-Back and Flanker tasks in the short term, increasing accuracy and reducing intra-individual variability in response times. We also hypothesized that decreased response time variability following WMT would be associated with brain activation in key working memory regions such as dorsolateral prefrontal cortex and the striatum. Furthermore, we hypothesized that, if WMT influenced the ability or motivation to selectively attend to task relevant information, then we should see differences in the estimated drift rates for the DDM between the training and control groups. Lastly, given the existing evidence for associations between intra-individual variability and cognitive function across different psychiatric and ageing populations, we hypothesised that intra-individual variability measures would be indicative of future outcomes at the subsequent 6 and 12-month follow-up assessments. We test these hypotheses using three independent data sets.

2 | METHOD

2.1 | Participants

For this paper, we analysed data from three separate samples of children (*N* = 28, 572, and 11,878). We describe the participants and tasks used in the two larger conceptual replication samples in subsection 2.7 below. The initial fMRI sample included 28 typically developing 7–9 year-old primary school children (mean age = 93 months, SD = 5 months, 14 females, working memory training group [**WMT**] = 16, comparison group [CMP] = 12). These children were recruited out of an ongoing intervention study of over 500 children and 29 different classrooms. The local ethics committee (Kantonale Ethikkommission Zürich) approved all procedures and methods used during this study.

2.2 | Cognitive training program

The training procedures consisted of a five-week intervention and four assessment waves, one pre-intervention (baseline), one immediately after the end of the five-week intervention, and two follow-up waves at 6 and 12–13 months, respectively. The assessment battery included tests of working memory and IQ (digit span, location span, object span, Raven's test), educational outcomes (math numeracy and math geometry, reading abilities) and concentration tests (Go/NoGo and bp task).

The working memory training program implemented was Cogmed's RoboMemo[1]. It is a computerized program, highly adaptive to individual performance, implemented via notebook computers including headphones for the spoken instructions and an external mouse. The intervention consisted of a daily WMT session per day (duration ~30 min), over a period of 5 weeks (25 sessions). Each training session included six adaptive modules (working memory tasks), including each 12 trials (75 trials in total). During the intervention, there was one specifically trained student coach in each class.

We compare the WMT group to children that either received standard classroom instruction ($N = 3$) or a self-regulation training over six school lessons ($N = 9$). In these lessons, the teacher taught a version of the mental contrasting with implementation intentions (MCII) technique (Duckworth et al., 2013) that was adapted to the relevant age group and the classroom context.

2.3 | Post-training cognitive and decision tasks

Working memory (N-Back) task: The 11-min block design working memory task consists of four conditions (Figure S1a). In the '0-Back' condition, they have to respond whenever they see the picture of a sun on the screen. In the '1-Back', '2-Back', and '3-Back' conditions, they have to respond whenever the picture on the screen is the same as 1, 2, or 3 before it, respectively. Performance data were recorded during scanning. The main performance variables are the d-prime index (d' = z(HitRate) – z(FalseAlarms)) and the intra-individual coefficient of variation (ICV = SDRT/MRT) for each working memory condition.

Flanker task: The 11-min event-related task was designed based on Rueda et al. (2004). Participants were presented with 240 trials (Figure S1b). Each trial consisted in a central row of five yellow fishes over a blue background. They were instructed to 'feed' the fish located in the centre of the screen. To do so, the child had to press the right/left button on the button box, depending on the direction of the central fish and ignoring the direction of the flankers. The main performance variables are the % of correct responses, and intra-individual coefficient of variation in response times. Furthermore, we fit a decision diffusion model to the response outcomes and times to determine the source(s) of differential variability in performance (see Section 2.4.2). Lastly, for comparison with previously published papers, we also conducted a post-hoc analysis of the RT data using the ex-gaussian approach. Based on previous studies (Geurts et al., 2008; Hervey et al., 2006; van Belle et al., 2015), we expected differences between the groups in the parameters of the exponential component of the ex-gaussian distribution of RTs.

For completeness, we note that the children also completed an intertemporal choice task similar to one in Steinbeis et al. (2014) while in the scanner, which was administered to test a separate hypothesis from the one that we focus on in this paper.

Further details on the three tasks can be found in the Supplementary Methods.

2.4 | Behaviour data analyses

Two of the 28 children from the fMRI sample withdrew from the study after the first task (in both cases, the intertemporal choice task). For three participants there were technical failures collecting the performance data during the Flanker task, which resulted in one participant being excluded due to the complete loss of performance data, and for two participants only one run of the task could be used in the analysis.

2.4.1 | Regressions on behavioural performance

Statistical analyses were conducted using RStudio (Version 1.1.442) (RStudio Team 2020).

We investigated differences between the trained and untrained groups on the main performance variables of the tasks right after the WMT. To do so, we conducted a general linear model (GLM) for the N-Back and Flanker tasks with training group (WMT vs. CMP) as fixed-effects factor and task condition (N-Back: high vs. low working memory; Flanker: congruent vs. incongruent) as random-effects factors.

2.4.2 | Decision Diffusion Modelling analyses

We used a Bayesian hierarchical approach to fit the parameters of the decision diffusion model (DDM) to the Flanker task using JAGS (Plummer, 2003) and the JAGS-Wiener module (Wabersich & Vandekerckhove, 2014) together with the rjags package (Plummer, 2018) in R. We used the priors recommend for hierarchical diffusion decision modelling in Wiecki et al. (2013). The fitting was run with three chains, 100,000 burn-in samples, and 10,000 posterior samples with a thinning rate of 10 samples. Convergence was assessed using visual inspection, and by ensuring psrf measures were below 1.05 for all parameters. Drift rates were calculated as a weighted linear combination of the target and non-target stimuli to distinguish the relative contribution of each to the evidence accumulation rate. Here, we fit the DDM to children's behaviour in the Flanker task. We did not fit data from the N-Back task because it required responses only on target trials, which were a small minority (25%) of all trials.

In our specification of the DDM, the magnitude of the drift rate coefficients informs us about how strongly each stimulus influences the evidence accumulation processes. In the flanker task, children should focus on the target fish because it alone provides evidence for the correct response in each trial. The direction the flanking distractor fish are facing is irrelevant and should be ignored. Thus, we specified the drift rate according to equation (1) below. We hypothesized that β_1–β_2 (i.e., the weight on relevant minus irrelevant information) should be greater in the WMT than the CMP group.

$$\text{Drift} = \beta_1 * \text{Target\_Direction} + \beta_2 * \text{Distractor\_Direction} \quad (1)$$

2.5 | fMRI data collection and analysis

Detailed descriptions of the fMRI preprocessing, scanning parameters and fMRI GLM can be found in the Supplementary Methods.

2.6 | Associations between post-training cognitive task performance and follow-up measures in the fMRI sample

We investigated if intra-individual variability measures computed immediately after the training could serve as indicators of relevant future outcomes at the subsequent follow-up assessments. Specifically, we examined the total score in the SDQ (Woerner et al., 2002), a behaviour and psychological well-being screening measure typically administered in clinical settings to identify potential problematic areas in a child that merit further assessment by a specialist. The SDQ was filled out by parents 6 months after training. We also examined tests of academic performance in reading and two mathematics subscales (geometry and arithmetic) conducted 12 months after training. We focused on these specific academic skills because of the results from the independent sample in Berger et al. (2020), which show that WMT improved geometry and reading scores, but not arithmetic.

In order to investigate whether intra-individual variability measures could be indicative of future outcomes at the subsequent follow-up assessments across all intervention groups, we conducted Bayesian robust linear regression analyses. These analyses tested whether changes in SDQ and academic skills (i.e., controlling for baseline scores) could be explained by children's accuracy (d-prime) or response time variability in cognitive tasks performed at the end of the training period. Specifically, we used the coefficient of variation in response times and d-prime scores from the N-Back task and the estimates of DDM drift rates from the flanker task to explain future outcomes. Certain follow-up or intra-individual variability measures were missing for some children (maximum number of missing values for any measure was 4). In order to use as much of the data from our small fMRI sample as possible, we imputed the missing values using the 'mice' package (van Buuren & Groothuis-Oudshoorn, 2011) in R. We generated 10 different imputed datasets and fit Bayesian linear regressions to each of them using the R package, 'brms' (Bürkner, 2018) as an interface to STAN (Stan Development Team, 2020). We drew our final inferences from the combined posterior distributions of all ten robust regression models to reduce the influence of any one set of imputed values on our results. Each of the 10 models used z-scored dependent and independent variables and student-t priors for all fixed effects (mean = 0, SD = 1, degrees of freedom = 10). Each model used 6000 MCMC samples across three independent chains after 1000 warmup samples for each chain and a thinning step = 5. The full set of regressor variables and results from these regressions are reported in Table S7. All regressions controlled for the baseline performance for each dependent variable, which effectively estimates changes in the outcomes.

2.7 | Conceptual replication and generalization using independent samples

We tested whether the core aspects would replicate or generalize in two independent, larger samples. The conceptual replication involved novel analyses of the data from Berger et al. (2020), henceforth referred to as the BFHSW study. The BFHSW study was conducted in a separate sample of 1st-grade children (age = 6–7 years, mean = 6.8 years, SD = 4.3 months) than the one from which our fMRI sample was drawn. However, the two studies used the same WMT procedures, as well as many of the same assessment instruments at baseline and follow up. These overlaps allow us to test the association between changes in response-time ICVs after the training and their association with academic skills at the 12-month follow-up in a manner similar to our fMRI sample, although computing response-time ICVs from a response inhibition rather than a working memory task (see 2.7.1 for details). We also tested whether the association between intra-individual variability in task performance and measures of current and future well-being we found in our small fMRI sample would generalize to the much larger set of children taking part in the Adolescent Brain and Cognitive Development (ABCD) study (Casey et al., 2018).

2.7.1 | BFHSW sample

Berger et al. (2020) implemented the same WMT intervention and the same pre- and post-training assessments at the same follow up time points (6 and 12 months after the end of the training) used in our fMRI sample in a separate set of 572 children (6–7 years of age). In this sample, the WMT group included 279 participants who performed the same WMT as in our initial sample. The control groups received either standard school instruction, the self-regulation training similar to our initial sample, or a learning software training. Following the procedures established in Berger et al. (2020), we compared the change in ICV in response times after WMT to the 101 children that received standard school instruction. However, our primary tests of the association between changes in ICV over the five-week intervention period and improvements in reading or geometry score 1 year after training were conducted across the 521–565 children for whom we have all of the relevant measures at each time point (see Tables 1 and 2 for details).

TABLE 1 Regression analyses testing for an association between intra-individual variability and academic performance in the BFHSW study

A	Dependent variable: Geometry Scores at 12-month follow up			
	Coefficient	Robust Std error	*t*-Value	*p*-Value
Go/NoGo ICV after training	−0.1351	0.0619	−2.18	0.0185
Go/NoGo ICV baseline	−0.1254	0.0393	−3.19	0.0015
Baseline Geometry score	0.3429	0.5478	6.26	<0.001
WMT group	0.3569	0.1068	3.34	0.001
B	**Dependent variable: Reading Scores at 12-month follow up**			
	Coefficient	**Robust Std error**	***t*-Value**	***p*-Value**
Go/NoGo ICV after training	−0.0847	0.0441	−1.92	0.032
Go/NoGo ICV baseline	−0.0301	0.0367	−0.82	0.209
Baseline Reading score	0.5532	0.0528	10.48	<0.001
WMT group	0.2288	0.1064	2.15	0.020

Note: This table shows the relationship between the change in the intra-individual coefficient of variation in response times during the Go/NoGo task (ICV) after the five-week intervention period and improvements in academic performance at the 12-month follow up assessment in BFHSW sample. The results are derived from linear least squares regression models including fixed effect control covariates for the child's age, sex, school site, and other treatment groups (omitted here for brevity and clarity). All outcome scores are standardized to mean = 0 and SD = 1. The robust standard errors are based on clustering at the classroom level. The coefficient for Go/NoGo ICV after training tests for the relationship between post-training changes in ICV and improvement in the academic skills at the 12-month follow up relative to baseline in both regressions. **A**) Changes in Go/Nogo ICV after training are significantly associated with improvements in geometry 12 months later (N = 521). **B**) Changes in Go/Nogo ICV after training show a similar relationship with improvements in reading 12 months later ($N = 522$), but the significance of this effect does not survive correction for multiple comparisons. Note that the *p*-values in the 4th column represent uncorrected one-sided tests of the a priori hypothesis that improvements (decreases) in Go/Nogo ICV would be associated with increased academic performance and that the threshold for Bonferroni correction 0.05/2 = is 0.025.

We computed intra-individual variability in response times from a response inhibition task instead of a working memory task in the BFHSW study. That study did not include a working memory task with enough speeded responses from each child to reliably compute intra-individual variability in response times. It did, however, include a Go/Nogo task, which measures response inhibition, at all assessment waves that we could use to compute intra-individual variability in response times. Therefore, to test the association between improvements in intra-individual variability after training and long-term, far transfer to academic skills, we computed the ICV as the standard deviation of go-trial RTs divided by the mean of the go-trial RTs. Note that we compute the ICV from correct trials only just as we did for all tasks in the initial sample and consistent with the standard procedure in the literature (Bellgrove et al., 2004; Bos et al., 2020; Fagot et al., 2018; Marciano & Yeshurun, 2017). We refer to these analyses as a conceptual replication because we use the Go/Nogo task instead of the N-back. Lastly, we only tested academic skills in the BFHSW sample because many of the parents did not complete the SDQ for their children in that study.

We fit linear regression models using Stata (StataCorp 2015). Specifically, we followed the methods reported in Berger et al. (2020) and estimated ordinary least squares regressions with robust standard errors clustered at the classroom level. All regression models included control covariates for treatment type, school, sex, age, and baseline performance in both the dependent variables (geometry or reading scores) and ICV from the Go/Nogo task.

2.7.2 | Adolescent Brain and Cognitive Development (ABCD) study

The ABCD study (https://abcdstudy.org, data release version 3.0) is a longitudinal, multicentre study of children's cognitive and neurobiological development starting from age 10, with data on 11,878 children. This study includes a wide range of standardized questionnaires and interviews covering both general well-being and clinical measures. In addition, participants perform several cognitive tasks, including an N-back task (Casey et al., 2018). We tested if intra-individual variability, once again quantified as the coefficient of variation in response times, during the N-Back task was associated with scores on the Child Behavioural Checklist (CBCL, Achenbach, 1991). The CBCL is a measure of current behavioural and psychopathological symptoms with high correspondence to the SDQ when both scales are applied to the same individuals. The ABCD study includes the CBCL, but not SDQ. We also tested for potential relationships between intra-individual variability and body mass index (BMI) scores. We chose BMI as an additional translational measure because it is robustly associated with physical, cognitive, and socioeconomic well-being. We used only those participants whose performance in the N-Back task was deemed adequate by the ABCD study's established QA procedure (overall response accuracy for 0-back or 2-back >60%) at all three currently available time points, and whose BMI fell between the 1st and 99th percentile (BMI percentiles = 13.3 and 35.0, respectively). We restricted the range for BMI because there were a small minority (<2%) of children

TABLE 2 Regression analyses testing for a decrease in Go/Nogo ICV after working memory training in the BFHSW sample

A	Dependent variable: Go/NoGo ICV right after training			
	Coefficient	Robust Std Error	*t*-Value	*p*-Value
Go/NoGo ICV baseline	0.3775	0.0293	12.90	<0.001
WMT group	0.0115	0.1134	0.10	0.460
B	**Dependent variable: Go/NoGo ICV at 12-month follow up**			
	Coefficient	Robust Std Error	*t*-Value	*p*-Value
Go/NoGo ICV baseline	0.2989	0.0386	7.75	<0.001
WMT group	−0.2960	0.1007	−2.94	0.003

Note: This table shows effects of 5 weeks of working memory training on the intra-individual coefficient of variation in response times (ICV) in the Go/NoGo task in the BFHSW sample. The results are derived from linear least squares regression models including fixed effect control covariates for the child's age, sex, school site, and other treatment groups (omitted here for brevity and clarity). All outcome scores are standardized to mean = 0 and SD = 1. The robust standard errors are based on clustering at the classroom level. **A**) Go/NoGo ICV measures do not yet show a significant decrease just after the working memory training (N = 565). **B)** Improvements (i.e., decreases) in Go/NoGo ICV in the working memory training group (WMT group) have emerged by the 12-month follow up assessment (N = 527). Note that the p-values in the 4th column represent uncorrected one-sided tests of the a priori hypothesis that working memory training would decrease response-time ICV and that the threshold for Bonferroni correction is 0.05/2 = 0.025.

with extremely low or high BMI values; however, robustness checks including all BMI values yielded similar results. The final sample size for our analyses was 8,522 children.

We computed Hierarchical Bayesian regression models estimating association between ICV and concurrent as well as future BMI and CBCL scores using the R package brms and MCMC sampling with STAN (Bürkner, 2018; Bürkner, 2017; Core Team, 2018; Stan Development Team, 2020). We fit and compared regression models that 1) assumed the association between ICV measured at baseline and baseline scores on CBCL or BMI would remain constant for CBCL or BMI values measured at the 1 or 2-year follow-up assessments, or 2) allowed for the explanatory power of baseline ICV to decrease in future assessment waves (i.e., these models included an interaction between ICV and assessment wave). All models controlled for potential effects of age, N-back accuracy, sex, race and parental education. We used a linear model that assumed a gaussian distribution for CBCL scores. However, the BMI scores did not follow a gaussian distribution. Therefore, we use a modified link function for the BMI regressions based on a combination of a gaussian and exponential distributions. To facilitate interpretability and comparison across dependent variables, we restricted the influence of ICV and all control variables to the mean parameter of the gaussian portion of the distribution in the BMI regressions. All models used z-scored dependent and independent variables as well as weakly regularizing priors for all fixed effects (gaussians with mean = 0 and SD = 1). Posterior distributions were estimated using 3000 MCMC samples across 3 independent chains after 1000 warmup samples for each chain and thinning step = 1. Convergence was assessed using visual inspection, and by ensuring psrf measures were below 1.05 for all parameters.

3 | RESULTS

3.1 | Behavioural results

Children in the WMT did not differ from those in the CMP at baseline. Before the start of any training program, all participants were assessed using a battery of tests that included general intelligence (a modified version of the Raven Matrices), working memory (visual, spatial), inhibition (Go-NoGo task), school performance (including reading, arithmetic, geometry, etc) and psychological well-being screening measures (SDQ). Statistical comparisons between the two groups show that they did not differ in any of these baseline measures (Table S1). While the two groups showed no differences in any measure at baseline, including cognitive tasks measuring working memory and attention skills, we do not have baseline measures on the N-Back and Flanker tasks reported in the following section.

3.2 | Accuracy and response-time variability findings

Overall, the group of children randomly assigned to undergo adaptive WMT performed more accurately and with less trial-to-trial variability in response times during the N-Back and Flanker tasks than those in the CMP (see Supplementary Tables S2 and S3 for the full set of descriptive statistics and results). The WMT group responded more accurately in the Flanker task across both the congruent and incongruent trials. In the N-Back task, children in the WMT group were more accurate than those in the CMP on low working memory trials (0-1 back), but the two groups did not significantly differ on high working memory trials (2-3 back).

In addition to better accuracy, children that received adaptive WMT also showed less intra-individual variability in response times than children in the CMP group (Supplementary Tables S2 and S3). We computed the intra-ICV as intra-individual RT standard deviation/intra-individual RT mean. Children in the WMT group used external mouses during the training intervention rather than the MRI-compatible button-box used during both the Flanker and N-back tasks, and thus familiarity with the response device cannot be a source for differences in ICV across groups.

3.3 | Diffusion Decision Model analyses

We fit a DDM model to the children's behaviour in the Flanker task to determine the mechanisms leading to differences in response time

TABLE 3 Diffusion Decision Model parameters for the Flanker task

	WMT			CMP			WMT – CMP		
DDM parameter	***m***	***HDI***		***m***	***HDI***		***m***	***HDI***	
Target drift coef.	2.84	2.30	3.38	1.99	1.65	2.37	0.84*	0.20	1.49
Distractor drift coef.	0.20	0.03	0.38	0.25	0.00	0.51	–0.05	–0.36	0.25
Target – Distractor	2.64	2.06	3.20	1.75	1.31	2.19	0.89*	0.14	1.57
Boundary	2.53	2.13	2.96	2.09	1.75	2.42	0.45*	– 0.08	0.98
Non-decision time	0.30	0.22	0.37	0.24	0.14	0.35	0.05	– 0.07	0.18

Note: This table lists the mean (m) as well as the lower and upper bounds of the 95% highest density interval (HDI) of the posterior distributions for parameters or parameter differences from the decision diffusion model (DDM) fit to the Flanker task. The DDM was fit to the Flanker task data separately for the children in the group that received working memory training (WMT) and those in the comparison group (CMP). The asterisks next to the mean differences between WMT and CMP denote those means that are significantly different based on a one-sided test of posterior probability of the mean for WMT being greater than the CMP group.

variability between the treatment groups. Using the DDM, we can determine if response time variability is driven by (1) differences in the non-decision time; (2) differences in the boundary or threshold determining when there is sufficient evidence to make one response versus the other (often interpreted as response caution); and/or (3) the drift rates (i.e., how quickly and robustly evidence is accumulated). We separated the drift rate into two components to measure children's sensitivity to the relevant information from the target compared to the irrelevant information from the flankers. The fits are summarized in Table 3 and show that children in the WMT group were more sensitive to the information carried by the target fish (i.e., its direction) relative to distractor fish (posterior probability = 0.992) and utilized a higher response threshold (posterior probability = 0.952) than children in the CMP. These DDM results are consistent with better attention to task-relevant features following WMT.

We also simulated responses from the fitted diffusion decision model to test if it could reproduce the patterns of response time variability observed in the Flanker task. To generate simulated responses in the Flanker task, we used each participant's best-fitting DDM parameters. We then compared the simulated response times across groups and found that the RTs were less variable for simulated agents using parameters from the WMT participants than for simulated agents based on CMP children's parameters (Table S4). Thus, the fitted DDM can both explain and generate different levels of response time variability in the two groups.

In addition to fitting the DDM, we also conducted a post-hoc test fitting the ex-gaussian distribution to each child's response times. Although the ex-gaussian model does not allow for the same type of mechanistic inferences as the DDM, we fit and report it to facilitate comparison with previously published papers using this method to quantify intra-individual variability in RT. These ex-gaussian results are consistent with the ICV and DDM results and indicated that the standard deviation (sigma) and exponential (tau) parameters differed between the working memory trained and CMP, but there was no significant difference in the means (mu) of the response time distributions (Table S5). In other words, more variable individuals were not reliably faster or slower to respond overall. Rather they were more inconsistent in the way they executed their responses. Intra-individual variability was highly correlated across the N-Back and Flanker tasks ($r = 0.65$, $p = 0.0008$, 95% CI [0.32, 0.84]).

Together, the pattern of behavioural results across both cognitive tasks and several complementary forms of analysis suggest that the adaptive WMT intervention may have increased children's ability to engage and maintain attention on task-relevant information in a domain-general manner soon after the five weeks of training were complete.

3.4 | fMRI results

Along with better accuracy, the WMT group showed increased activation compared to the CMP in brain regions that are part of attention and control networks during the low working memory trials. These included portions of fronto-striatal-thalamic systems such as the right caudate, putamen, pallidum, thalamus, inferior middle and superior frontal gyri, the dorsal anterior cingulate and the supplementary motor cortex (Table S6, Figure 1). Consistent with the behavioural findings of similar accuracy in the high working memory condition, there were no significant differences in the BOLD signal across groups during the high working memory trials. We did not detect any significant differences in activity as a function of WMT during the Flanker task.

In addition, we found that task-related BOLD signal levels in regions that showed greater activity in the WMT group (see Figure S2) also correlated with the intra-individual coefficients of variation and/or accuracy on the N-back task across all participants (Figure 1, bottom row). Some relation to accuracy and the intra-ICV in these regions is to be expected given that there are group differences in intra-individual variability. However, activity in the dorsal striatal functional ROI, encompassing dorsal caudate and putamen, was significantly associated with intra-individual variability even after accounting for the effects of WMT condition (coef = –0.25, $p = 0.004$; Table 4). There were similar, though not significant, trends in the dorsolateral prefrontal cortex (dlPFC) for intra-individual variability, and in the anterior cingulate cortex/supplementary motor area for accuracy (Table 4).

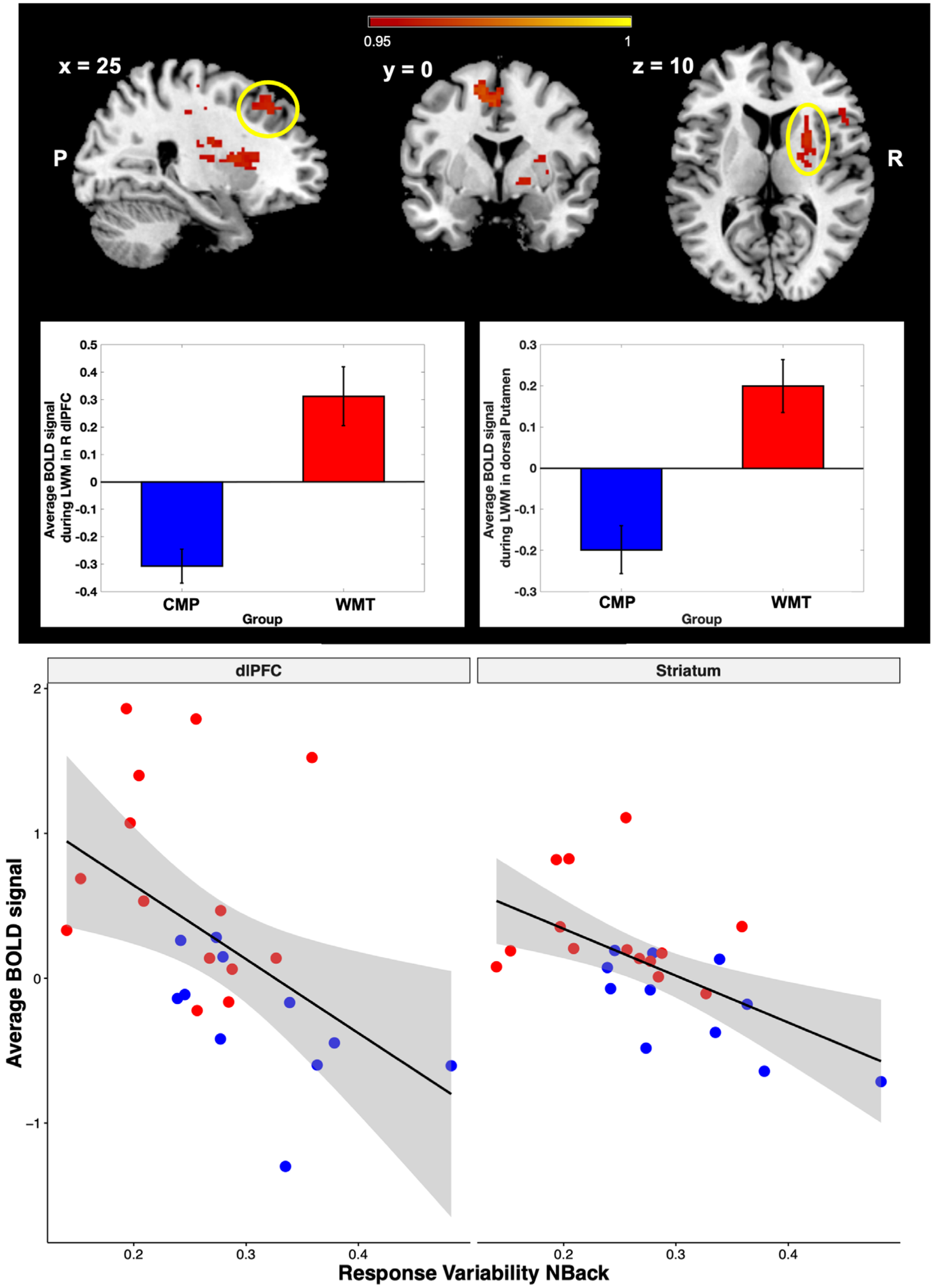

FIGURE 1 Group differences on brain activation. The WMT group had greater activity in frontostrial regions than the CMP group in N-Back tasks. Consistent with the behavioural results, these differences were specific to the low working memory condition. The bar graphs show the average BOLD signal in each group, in the two clusters circled in yellow, a) right dorsolateral prefrontal cortex (left panel) and b) right putamen (right panel). The scatter plots in the bottom row show the association between BOLD signal and individual differences in the coefficient of variation across all trials. Children in the WMT group are shown in red while those in the CMP group are shown in blue.

TABLE 4 Associations between BOLD signal in ROIs where group differences were identified in the LWM contrasts and intra-individual coefficient of variation and d-prime across all trials

	Estimate (se)	*t*-Value	*p*-Value
Dorsal striatum			
Group	0.17 (0.15)	1.16	–
ICV	−0.25 (0.08)	−3.26	0.004
d-prime	0.12 (0.06)	2.10	0.049
dlPFC			
Group	0.56 (0.31)	1.782	–
ICV	−0.32 (0.16)	−1.978	0.0619
d-prime	0.09 (0.13)	0.689	0.4989
dACC-SMA			
Group	0.50 (0.20)	2.497	–
ICV	−0.13 (0.10)	−1.249	0.2260
d-prime	0.14 (0.08)	1.794	0.0879

Note: The table reports the results from three linear regression models testing the association between the intra-individual coefficient of variation and d-prime during the N-Back task and BOLD signal after accounting for the effects of training condition (Group). The linear model in each region was computed as BOLD signal ~ Group + ICV + d-prime + Delay, where Delay is a control variable that accounts for the delay between the end of the intervention and the scanning session (omitted from the table for conciseness). We list the T-values for the binary regressor indicating training group for comparison with the linear regressor for ICV. However, we do not report p-values for the Group regressor because the functional ROIs were originally identified with this contrast, making the analysis circular. The point of these regressions is to test if differences in ICV provide additional explanatory power in the region where activity differed between training groups. se = standard error; ICV = Individual Coefficient of Variation; dlPFC = dorsolateral prefrontal cortex; dACC = dorsal anterior cingulate; SMA = supplementary motor area. Note that the fourth column reports uncorrected, two-tailed *p*-values and that the threshold for Bonferroni correction across the three ROIs is 0.0167.

3.5 | Associations between post-training cognitive task performance and follow-up measures in the fMRI sample

Response time variability explained significant additional variance in future SDQ scores (standardized coef. = 0.32 ± 0.14), geometry (standardized coef. = −0.66 ± 0.23), and reading (standardized coef. = −0.32 ± 0.14), after accounting for baseline scores in those measures and IQ (Table S7). Thus, response time variability after 5 weeks of training was associated with future improvement in SDQ scores and academic skills that were not yet evident in direct tests of those skills at the same time point. In contrast, no post-training cognitive task performance or baseline measures were significantly associated with future arithmetic scores. Thus, in our sample, the intra-individual variability in response times measured right after the intervention correlated with future performance in the same academic domains that Berger and colleagues previously found to be improved 1 year after WMT in an independent sample.

3.6 | Conceptual replication of the association between post-training ICV and follow-up measures in the BFHSW sample

Consistent with our results in the fMRI sample, changes in ICV (post- versus pre-intervention) during the Go/Nogo task were also associated with improvement in geometry and reading skills 1 year after WMT in the independent BFHSW sample (Tables 1 and 2), although the effect for reading does not survive Bonferroni correction for multiple comparisons. Note that our regression specification includes regressors for both baseline (W1) and post-training (W2) Go/Nogo ICV, and in this specification, the coefficient for W2 Go/Nogo ICV represents the effect of the change in performance between W1 and W2. The same holds for the use of the future academic skill scores as the dependent variable when including the baseline score as an independent variable in the regression.

Interestingly, while N-back ICV was significantly different between WMT and control groups just after training, significant differences in Go/Nogo ICV did not appear until 12 months after training (Table 2). This delayed emergence of significant improvements in Go/Nogo ICV is consistent with the delayed emergence of Go/Nogo accuracy improvements reported in Berger et al. (2020). Even though they did not yet significantly differ across treatments, changes in Go/Nogo ICV from baseline to post-training still predicted future improvements in performance. A relevant question for future research is to determine which types of cognitive tasks (e.g., working memory, response inhibition, etc.) are best suited to evaluate intra-individual variability in response time or accuracy in order to forecast the emergence of far transfer benefits following working memory or other training regimes.

3.7 | The association between ICV and measures of well-being generalizes to the ABCD study

We used data from the first three waves of the longitudinal ABCD study to test whether the relationship we observed in our fMRI sample between response-time variability during the N-back task and measures of well-being generalized to an independent and larger set of children. We used the data on BMI and scores on the Child Behavioral Checklist (CBCL) as our measures of well-being in the ABCD study. Unlike our fMRI and the BFHSW studies, the ABCD study does not include a WMT intervention. Therefore, we used the ABCD data to test if there was a significant association between N-back ICV and current well-being and, if so, whether this relationship holds over the first- and second-year follow-ups in this longitudinal study.

We fit and compared Hierarchical Bayesian regression models that assumed the association between N-back ICV at baseline and BMI, or CBCL, at baseline, 1- and 2-year follow-up was either stable or decreased over time. Concretely, we tested whether regression

TABLE 5 Results of the regression analyses testing the association between Go/Nogo ICV and BMI or CBCL scores in the ABCD data

Dependent variable: BMI					
	Mean	SD	HDI		Post. Prob.
Go/Nogo ICV	0.024	0.004	0.016	0.032	0.999
Go/Nogo Accuracy	−0.011	0.005	−0.02	−0.002	0.992
Year 1	0.04	0.011	0.019	0.061	0.999
Year 2	0.083	0.016	0.05	0.115	0.999
Sex	−0.011	0.008	−0.027	0.005	0.901
Age	0.085	0.006	0.073	0.098	0.999
Dependent variable: CBCL					
	Mean	SD	HDI		Post. Prob.
Go/Nogo ICV	0.028	0.007	0.014	0.042	0.999
Go/Nogo Accuracy	−0.092	0.008	−0.106	−0.077	0.999
Year 1	−0.026	0.019	−0.063	0.012	0.908
Year 2	−0.074	0.028	−0.129	−0.018	0.996
Sex	0.175	0.013	0.149	0.202	0.999
Age	0.003	0.011	−0.018	0.024	0.616

Note: This table reports hierarchical Bayesian linear regressions testing for relationships between N-Back ICV and body mass index (BMI) or behavioural difficulties (CBCL) in 8,522 children from the ABCD study. In addition to the fixed effects reported here, the models included grouping or random effects for race and parental education level. The columns labelled mean and SD list the mean and standard deviation of the posterior distribution for each coefficient. All independent variables in these models were standardized using the *z*-score transformation. The two columns labelled HDI give the lower and upper bounds of the 95% highest density interval for the posterior distributions of each coefficient. Post. Prob. = posterior probability that the estimated regression coefficient is greater or less than zero. The coefficient for Go/Nogo ICV measures the average association between response time variability during the N-back task at the initial baseline visit (age 10 years) and BMI or CBCL at ages 10, 11, and 12 years. Model comparisons showed that the relationship between baseline Go/Nogo ICV and BMI or CBCL is stable over this time period (see text for details).

models allowing for an interaction between baseline N-back ICV values and assessment wave (i.e., the explanatory power of ICV could decrease or increase) were better than models assuming a fixed association between baseline N-back ICV values and well-being at all waves. The baseline coefficients were the same in both the fixed and interaction models (Table 5) and indicated that greater variability in response times during the N-back at baseline was associated with decreased well-being in terms of both baseline BMI (standardized coefficient = 0.02, posterior probability >0.999) and CBCL (standardized coefficient = 0.03, posterior probability >0.999) scores, consistent with the findings in the fMRI sample.

We compared the fixed and interaction models using leave-one-out cross-validation with pareto-smoothed importance sampling (PSIS-LOO, Vehtari et al., 2017). The model comparisons modestly favoured the simpler fixed model without follow-up wave interactions when explaining both BMI (difference in expected log pointwise predictive density (elpd) for the interaction model = −2.3, standard error (SE) of the difference = 1.6) and CBCL (elpd difference = −2.6, SE = 1.9). Moreover, neither interaction model suggested a decrease in the explanatory power of baseline N-back ICV for well-being at 1- or 2-year follow-up visits relative to baseline. If anything, there was a slight increase in the regression coefficient for ICV between baseline and year 2 when seeking to explain BMI. Thus, the results from the ABCD data show that the relationship between N-back ICV and children's well-being generalizes across measures of well-being (SDQ, CBCL, BMI), and the explanatory power of N-back ICV persists through at least 2 years of experience and development in the absence of any cognitive training intervention.

4 | DISCUSSION

The present study examined how the neurocognitive mechanisms underlying the short-term impact of adaptive WMT in primary school children relate to training benefits that emerge months or years after training. Overall, our results suggest that in addition to working memory itself, there may be concurrent benefits to selective and sustained attention during or directly after five weeks of training. We show that intra-individual variability in response times during several different cognitive tasks can be used to detect short-term training effects in children, and that such measures may be indicative of the persistence and/or emergence of far-transfer benefits months to years after training is completed.

Our findings indicate that better attention is among the immediate results of adaptive WMT. Working memory and attention processes are thought to be closely linked and interdependent (Astle & Scerif, 2011; D'Esposito & Postle, 2015; Engle, 2018; Eriksson et al., 2015; Gazzaley & Nobre, 2012; Unsworth & Robison, 2017; Wass et al., 2012). Although they have different primary targets, the Flanker, Go/Nogo, and N-Back tasks require the ability to maintain attentional focus throughout the duration of the task (sustained attention), and to identify the target stimuli and filter out or inhibit responses to non-target stimuli (selective attention). At the neural level, differences between the WMT and CMP groups were found in striatum as well as the lateral and medial prefrontal cortices, which are brain regions that, among other things, support selective and sustained attention functions (Frank et al., 2001; Mcnab & Klingberg, 2008; Zanto et al., 2011). These neural differences were accompanied by better signal detection performance (i.e., higher d-prime), reduced intra-individual variability in response times, and more efficient accumulation of relevant information (i.e., higher DDM drift rates) in children that received adaptive WMT. All of these behavioural measures are related to and dependent on attention. Therefore, taken together, our neural and behavioural results suggest that the benefits of the WMT program used in this study are at least partially mediated by more effective attention processes leading to consistent and effective responses to task-relevant information and reduced processing of irrelevant, distracting stimuli.

These results lend further support to theories of the mechanisms underlying training benefits. A meta-analysis of previous training studies concluded that the Cogmed-RM adaptive WMT program has effects on attention in daily life (Spencer-Smith & Klingberg, 2015). The more effective attention processes we detected at the end of the training are consistent with previous results and theories about the basis of far-transfer effects following cognitive training as well (Dahlin et al., 2008; Greenwood & Parasuraman, 2016; Morrison & Chein, 2011). Specifically, these far-transfer benefits occur when the trained and transfer skills share common fundamental cognitive processes. Given the important role of attention as a prerequisite to many cognitive processes, it could serve as a basis for far-transfer effects following WMT.

Recently, benefits of the adaptive WMT in school-age children, relative to standard classroom instruction, have been shown to emerge over 6–12 months (Berger et al., 2020). Initial improvements in attention may serve as a scaffold for later changes in higher cognitive processes that facilitate better school performance. Our current results suggest that attention functions might be among the first to improve from this type of training, and that later emerging benefits to academic skills and general well-being are associated with immediate improvements in attention processes. It is not surprising that WMT would also influence attention control (e.g., selective attention, sustained attention, or goal-directed attention reallocation) given that these processes are postulated to be pre-requisites for the successful implementation of working memory (Astle & Scerif, 2011; D'Esposito & Postle, 2015; Eriksson et al., 2015; Gazzaley & Nobre, 2012; Unsworth & Robison, 2017; Wass et al., 2012). There is also evidence that the associations between working memory capacity and various cognitive and academic skills are partially mediated by a common reliance on attention control (Engle, 2018; Fukuda & Vogel, 2011; Unsworth & Robison, 2017). Given the apparent role of attention processes in mediating the far-transfer of training effects, it is important to measure these processes when assessing the efficacy of WMT and other forms of cognitive training. It will also be important for future studies to directly compare adaptive WMT versus interventions that use adaptive mechanisms to train attention, inhibition, or other cognitive abilities. Our current results indicate that metrics quantifying intra-individual variability in response time will be useful in determining the relative short and long-term efficacy of different training regimes.

The ability of intra-individual variability metrics to detect individual differences in attention control could explain the association we find between ICV and the future emergence of benefits to academic skills and general well-being after WMT. Intra-individual response time variability metrics are sensitive and reliable measures of individual differences in attention control processes (MacDonald et al., 2009; Saville et al., 2011). They are often used as an index of an individual's attention allocation efficiency or degree of fluctuation in attention control during task performance (Bellgrove et al., 2004; Isbell et al., 2018; Kelly et al., 2008; Stuss et al., 2003; Unsworth, 2015). Intra-individual variability has been linked with cognitive control measures in healthy children and adults, and the variability in response times measured in one task is correlated with working and long-term memory or intelligence measured in separate tasks (Bellgrove et al., 2004; Isbell et al., 2018; Larson & Saccuzzo, 1989; Montez et al., 2017; van Belle et al., 2015). It also differs between healthy individuals and those with attention deficit hyperactivity disorder (ADHD) (Castellanos et al., 2005; Geurts et al., 2008; Karalunas et al., 2014; Kofler et al., 2013; van Belle et al., 2015). However, increased response time variability is not unique to ADHD and is seen in various psychiatric and neurological disorders (e.g., traumatic brain injury, dementia, and schizophrenia), in which attention deficits may play an important, though less prominent, role (Geurts et al., 2008; Haynes et al., 2017; Ilg et al., 2018; Kofler et al., 2013; MacDonald et al., 2006). Increased intra-individual variability is commonly observed in non-affected relatives as well as patients, indicating that it may capture shared genetic or environmental risk factors for current and future psychopathologies (Adleman et al., 2014; Ilg et al., 2018; Karalunas et al., 2014; Kuntsi et al., 2010; Stuss et al., 2003). In fact, a recent review by Haynes et al. highlights several longitudinal studies in older adults that have shown that the intra-individual variability in response times is associated with future levels of cognitive impairment and mortality (Haynes et al., 2017). Thus, our current results, together with the existing body of work indicate that intra-individual variability measures are sensitive to not only to current cognitive and neurological function, but also associated with the future stability, improvement, or decline of those functions.

We found that intra-individual variability metrics can detect the short-term efficacy and are indicative of the emergence of longer-term benefits of working memory interventions aimed at improving cognitive skills and academic performance in children. We could detect significant differences between trained and untrained groups in intra-individual response time variability during cognitive tasks probing working memory and attention (N-Back and Flanker) directly after five weeks of WMT, while significant improvements in variability during a response inhibition task (Go/Nogo) did not emerge until months later. Nevertheless, consistent with their ability to forecast cognitive decline in the elderly, we found that measures of the intra-individual variability in the N-back and Go/Nogo tasks computed at the end of training were associated with improvements in academic skills and general well-being in children up to 1 year after training. Across both tasks, lower post-training variability was related to better future scores on tests of academic skills and strengths/weaknesses in classroom and social behaviour. The results from the ABCD data are also consistent with the idea that measures of performance variability are associated with both current and future well-being, specifically behavioural problems and BMI.

Our results suggest that measures of intra-individual variability are useful in evaluating intervention efficacy. However, there are several important questions that still need to be addressed. For example, can we use intra-individual variability metrics to determine when an individual has received a sufficient dose of the training intervention? If so, then we could tailor the amount of training to each person in order to improve the cost benefit trade-offs inherent in any training program. Another key question our findings raise is what types of tasks (e.g., those targeting working memory, attention, task-switching, etc) and measures of intra-individual variability are best suited to assessing the short and long-term outcomes of cognitive training. Previous work has quantified intra-individual variability in response times in several different ways (Geurts et al., 2008; Karalunas et al., 2014;

van Ravenzwaaij et al., 2011). We found significant differences in response time variability between training groups in selective attention (Flanker), working memory (N-Back) tasks, and response inhibition tasks (Go/Nogo) using several complementary measures of variability. However, there may be differences in how well the different measures of variability and/or task designs predict the emergence of benefits to specific areas of academic performance or general well-being in the longer term. This question will be important to address in future studies that collect and compute multiple longitudinal measures in large samples of participants.

We note a few potential limitations of this work. One potential limitation is that familiarity with the use of computer devices may underlie the differences in response time variability between control and training groups. However, we think this is unlikely given that in the fMRI study the training group implemented the responses during the training with a mouse whereas inside the scanner children responded using an MRI-adapted button box. Moreover, significant improvements in performance variability in the Go/Nogo task did not emerge until months after training suggesting the difference was due to further development in cognitive skills rather than simple action familiarity or motor skills. Secondly, our initial fMRI study did not include the N-Back or Flanker at baseline so we could not control for baseline performance in those exact tasks. The lack of between-group differences in any other pre-intervention measures of working memory or attention suggests that the probability of randomization failures leading to training-independent differences in working memory or attention is very low. The fact that we can replicate our results from the fMRI sample in the BFHSW sample using a Go/Nogo task measured at baseline and post-intervention further indicates that randomization failure is an unlikely cause for our original results. Lastly, an important limitation is that our current data cannot tell us whether these effects are specific to adaptive WMT per se or if other forms of cognitive training might lead to similar benefits. Our results on RT variability suggest that some of the initial training benefits are mediated by improvements in attention control. While attention control and working memory are interrelated, it should be possible to train attention control using cognitive tasks that make limited demands on working memory in order to better distinguish between the two skills. Determining the best types and forms of cognitive training, and potentially how to customize the training for individuals of different ages or abilities is an important goal for future research.

5 | CONCLUSION

Effective means of enhancing cognitive abilities have been a longstanding goal in many disciplines. Our current work adds to the existing evidence that adaptive WMT can significantly benefit school-aged children (Berger et al., 2020; Jones et al., 2020; Karbach et al., 2015; Titz & Karbach, 2014; Wass et al., 2012). Moreover, it provides additional insights into the mechanisms underlying these benefits. Together with the recent findings of Berger et al. (2020), it also highlights the importance of including long-term follow-ups in any evaluation of training efficacy. In addition to long-term follow-up data, we demonstrate the utility of using response time variability metrics as an immediate indicator of intervention success. The practical relevance of such an immediate assessment tool should not be overlooked, as it could potentially allow for tailoring training interventions in terms of duration or content without needing to wait for years for follow-up data to determine whether long-term benefits will emerge.

ACKNOWLEDGEMENTS

This study was supported by the National Center of Competence in Research Affective sciences hosted by the University of Geneva (SNSF grant number 51NF40-104897).

Open access funding provided by Universitat Zurich.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

ETHICS APPROVAL STATEMENT

The local ethics committee (Kantonale Ethikkommission Zürich) approved all procedures and methods used during this study.

AUTHOR CONTRIBUTIONS

AC and TH designed the task-based fMRI study and analysis plan with input from HH. AC collected the combined fMRI plus Flanker, Intertemporal choice, and N-Back task data. EB, HH, EF, DS and KW designed and implemented the in-school training interventions and follow-up measures. AC and TH analyzed the data. AC and TH wrote the manuscript with input from EB, HH, EF, DS and KW.

DATA AVAILABILITY STATEMENT

Code for the data analysis and behavioural data collected during the imaging study is openly available at https://osf.io/35f7g/. Behavioural data collected during the school-based assessments might be shared subject to reasonable request to the authors.

ORCID

Ana Cubillo https://orcid.org/0000-0001-5657-7341
Henning Hermes https://orcid.org/0000-0002-9339-4779
Ernst Fehr https://orcid.org/0000-0001-6325-7821
Todd A. Hare https://orcid.org/0000-0002-0260-2772

ENDNOTE

[1] Cogmed and Cogmed Working Memory Training are trademarks, in the U.S. and/or other countries, of Cogmed Inc. (www.cogmed.com).

REFERENCES

Achenbach, T. M. (1991) *Manual for the child behavior checklist/4-18 and 1991 profile*. Department of Psychiatry, University of Vermont.

Adleman, N. E., Yi, J. Y., Deveney, C. M., Guyer, A. E., Leibenluft, E., & Brotman, M. A. (2014). Increased intrasubject variability in response time in unaffected preschoolers at familial risk for bipolar disorder. *Psychiatry Research*, 30, *219*(3), 687–689.

Ali, S., Macoun, S. J., Bedir, B., & MacDonald, S. W. S. (2019). Intraindividual variability in children is related to informant ratings of attention and executive function. *Journal of Clinical and Experimental Neuropsychology*, *41*(7), 740–748.

Astle, D. E., Barnes, J. J., Baker, K., Colclough, G. L., & Woolrich, M. W. (2015). Cognitive training enhances intrinsic brain connectivity in childhood. *The Journal of Neuroscience : The official journal of the Society for Neuroscience*, *35*(16), 6277–6283.

Astle, D. E., & Scerif, G. (2011). Interactions between attention and visual short-term memory (VSTM): What can be learnt from individual and developmental differences? *Neuropsychologia*, *49*(6), 1435–1445.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: A meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377.

Baddeley, A. (1996). The fractionation of working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(24), 13468–13472.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140.

Bellgrove, M. A., Hester, R., & Garavan, H. (2004). The functional neuroanatomical correlates of response variability: Evidence from a response inhibition task. *Neuropsychologia*, *42*(14), 1910–1916.

Berger, E. M., Fehr, E., Hermes, H., Schunk, D., & Winkel, K. The impact of working memory training on children's cognitive and noncognitive skills. *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3622985

Bogg, T., & Lasecki, L. (2014). Reliable gains? Evidence for substantially underpowered designs in studies of working memory training transfer to fluid intelligence. *Frontiers in Psychology*, *5*, 1589.

Bos, D. J., Dreyfuss, M., Tottenham, N., Hare, T. A., Galván, A., Casey, B. J., & Jones, R. M. (2020). Distinct and similar patterns of emotional development in adolescents and young adults. *Developmental Psychobiology*, *62*(5), 591–599.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R Package brms. *The R Journal*, *10*(1), 395.

Buschkuehl, M., Jaeggi, S. M., & Jonides, J. (2012). Neuronal effects following working memory training. *Developmental Cognitive Neuroscience*, *2*(Suppl 1), S167–79.

Cartwright, K. B. (2012). Insights from cognitive neuroscience: The importance of executive function for early reading development and education. *Early Education & Development*, *23*(1), 24–36.

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. M., … Dale, A. M. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54.

Castellanos, F. X., Sonuga-Barke, E. J., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biological Psychiatry*, *57*(11), 1416–1423.

Cavanagh, J. F., Wiecki, T. V., Kochar, A., & Frank, M. J. (2014). Eye tracking and pupillometry are indicators of dissociable latent decision processes. *Journal of Experimental Psychology General*, *143*(4), 1476–1488.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201–215.

Core Team, R. (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing.

Cortese, S., Ferrin, M., Brandeis, D., Buitelaar, J., Daley, D., Dittmann, R. W., Holtmann, M., Santosh, P., Stevenson, J., Stringar, A., Zudda, A., Sonuga-Bark, E. J. S., & European ADHD Guidelines Group (EAGG). (2015). Cognitive training for attention-deficit/hyperactivity disorder: Meta-analysis of clinical and neuropsychological outcomes from randomized controlled trials. *Journal of the American Academy of Child and Adolescent Psychiatry*, *54*(3), 164–174.

Dahlin, E., Neely, A. S., Larsson, A., Backman, L., & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science*, *320*(5882), 1510–1512.

Delalande, L., Moyon, M., Tissier, C., Dorriere, V., Guillois, B., Mevell, K., Charron, S., Salvia, E., Poirel, N., Vidal, J., Lion, S., Oppenheim, C., Houdé, O., Cachia, A., & Borst, G. (2020). Complex and subtle structural changes in prefrontal cortex induced by inhibitory control training from childhood to adolescence. *Developmental Science*, *23*(4), https://doi.org/10.1111/desc.12898

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142.

Dirk, J., & Schmiedek, F. (2016). Fluctuations in elementary school children's working memory performance in the school context. *Journal of Educational Psychology*, *108*(5), 722–739. https://doi.org/10.1037/edu0000076

Duckworth, A. L., Kirby, T. A., Gollwitzer, A., & Oettingen, G. (2013). From fantasy to action: Mental contrasting with implementation intentions (MCII) improves academic performance in children. *Social Psychological and Personality Science*, *4*(6), 745–753.

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190–193.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, *88*(1), 33–46. https://doi.org/10.1016/j.neuron.2015.09.020

Fagot, D., Mella, N., Borella, E., Ghisletta, P., Lecerf, T., & De Ribaupierre, A. (2018). Intra-individual variability from a lifespan perspective: A comparison of latency and accuracy measures. *Journal of Intelligence*, *6*(1), 16.

Flegal, K. E., Ragland, J. D., & Ranganath, C. (2019). Adaptive task difficulty influences neural plasticity and transfer of training. *Neuroimage*, *188*, 111–121.

Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, *67*, 641–666.

Frank, M. J., Loughry, B., & O'Reilly, R. C. (2001). Interactions between frontal cortex and basal ganglia in working memory: A computational model. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(2), 137–160.

Fukuda, K., & Vogel, E. K. (2011). Individual differences in recovery time from attentional capture. *Psychological Science*, *22*(3), 361–368.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135.

Geurts, H. M., Grasman, R. P., Verte, S., Oosterlaan, J., Roeyers, H., van Kammen, S. M., & Sergeant, J. A. (2008). Intra-individual variability in ADHD, autism spectrum disorders and Tourette's syndrome. *Neuropsychologia*, *46*(13), 3030–3041.

Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (2020). First demonstration of effective spatial training for near transfer to spatial performance and far transfer to a range of mathematics skills at 8 years. *Developmental Science*, *23*(4), https://doi.org/10.1111/desc.12909

Gómez-Guerrero, L., Martín, C. D., Mairena, M. A., Di Martino, A., Wang, J., Mendelsohn, A. L., Dreyer, B. P., Isquith, P. K., Gioia, G., Petkova, E., & Castellanos, F. X. (2011). Response-time variability is related to parent ratings of inattention, hyperactivity, and executive function. *Journal of Attention Disorders*, *15*(7), 572–582.

Greenwood, P. M., & Parasuraman, R. (2016). The mechanisms of far transfer from cognitive training: Review and hypothesis. *Neuropsychology*, *30*(6), 742–755.

Haynes, B. I., Bauermeister, S., & Bunce, D. (2017). A systematic review of longitudinal associations between reaction time intraindividual variability and age-related cognitive decline or impairment, dementia, and mortality. *Journal of the International Neuropsychological Society*, *23*(5), 431–445.

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Keith Conners, C., Swanson, J. M., & Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology*, *12*(2), 125–140.

Ilg, L., Klados, M., Alexander, N., Kirschbaum, C., & Li, S.-C. (2018). Long-term impacts of prenatal synthetic glucocorticoids exposure on

functional brain correlates of cognitive monitoring in adolescence. *Science Reports*, *8*(1), 7715.
Isbell, E., Calkins, S. D., Swingler, M. M., & Leerkes, E. M. (2018). Attentional fluctuations in preschoolers: Direct and indirect relations with task accuracy, academic readiness, and school performance. *Journal of Experimental Child Psychology*, *167*, 388–403.
Johnson, B. P., Pinar, A., Fornito, A., Nandam, L. S., Hester, R., & Bellgrove, M. A. (2015). Left anterior cingulate activity predicts intra-individual reaction time variability in healthy adults. *Neuropsychologia*, *72*, 22–26.
Jones, J. S., Milton, F., Mostazir, M., & Adlam, A. R. (2020). The academic outcomes of working memory and metacognitive strategy training in children: A double-blind randomized controlled trial. *Developmental Science*, *23*(4), https://doi.org/10.1111/desc.12870
Judd, N., Klingberg, T., & Sjöwall, D. (2021). Working memory capacity, variability, and response to intervention at age 6 and its association to inattention and mathematics age 9. *Cognitive Development*, *58*, 101013.
Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2008). Variation in working memory capacity as variation in executive attention and control. In: A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse eds. *Variation in working memory*. Oxford University Press; p. 21–48.
Karalunas, S. L., Geurts, H. M., Konrad, K., Bender, S., & Nigg, J. T. (2014). Annual research review: Reaction time variability in ADHD and autism spectrum disorders: Measurement and mechanisms of a proposed transdiagnostic phenotype. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(6), 685–710.
Karalunas, S. L., & Huang-Pollock, C. L. (2013). Integrating impairments in reaction time and executive function using a diffusion model framework. *Journal of Abnormal Child Psychology*, *41*(5), 837–850.
Karbach, J., Strobach, T., & Schubert, T. (2015). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychology*, *21*(3), 285–301.
Karbach, J., & Unger, K. (2014). Executive control training from middle childhood to adolescence. *Frontiers in Psychology*, *5*, https://doi.org/10.3389/fpsyg.2014.00390
Karbach, J., & Verhaeghen, P. (2014). Making working memory work: A meta-analysis of executive-control and working memory training in older adults. *Psychological science*, *25*(11), 2027–2037.
Kelly, A. M., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *Neuroimage*, *39*(1), 527–537.
Klein, C., Wendling, K., Huettner, P., Ruder, H., & Peper, M. (2006). Intra-subject variability in attention-deficit hyperactivity disorder. *Biological Psychiatry*, *60*(10), 1088–1097.
Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317–324.
Kofler, M. J., Rapport, M. D., Sarver, D. E., Raiker, J. S., Orban, S. A., Friedman, L. M., & Kolomeyer, E. G. (2013). Reaction time variability in ADHD: A meta-analytic review of 319 studies. *Clinical Psychology Review*, *33*(6), 795–811.
Könen, T., & Karbach, J. (2015). The benefits of looking at intraindividual dynamics in cognitive training data. *Frontiers in Psychology*, *6*, https://doi.org/10.3389/fpsyg.2015.00615
Krajbich, I., Bartling, B., Hare, T., & Fehr, E. (2015). Rethinking fast and slow based on a critique of reaction-time reverse inference. *Nature Communications*, *6*(1), https://doi.org/10.1038/ncomms8455
Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857.
Kuntsi, J., Wood, A. C., Rijsdijk, F., Johnson, K. A., Andreou, P., Albrecht, B., Arias-Vasquez, A., Buitelaar, J. K., McLoughlin, G., Rommelse, N. N. J., Sergeant, J. A., Sonuga-Barke, E. J., Uebel, H., van der Meere, J. J., Banaschewski, T., Gill, M., Manor, I., Miranda, A., Mulas, F., …, Asherson, P. (2010). Separation of cognitive impairments in attention-deficit/hyperactivity disorder into 2 familial factors. *Archives of General Psychiatry*, *67*(11), 1159–1167.
Larson, G. E., & Saccuzzo, D. P. (1989). Cognitive correlates of general intelligence: Toward a process theory of g. *Intelligence*, *13*(1), 5–31. https://doi.org/10.1016/0160-2896(89)90003-2
MacDonald, S. W. S., Li, S.-C., & Bäckman, L. (2009). Neural underpinnings of within-person variability in cognitive functioning. *Psychology and Aging*, *24*(4), 792–808. https://doi.org/10.1037/a0017798
MacDonald, S. W. S., Nyberg, L., & Bäckman, L. (2006). Intra-individual variability in behavior: Links to brain structure, neurotransmission and neuronal activity. *Trends in Neurosciences*, *29*(8), 474–480. https://doi.org/10.1016/j.tins.2006.06.011
Marciano, H., & Yeshurun, Y. (2017). Large inter-individual and intra-individual variability in the effect of perceptual load. *PLOS ONE*, *12*(4), e0175060. https://doi.org/10.1371/journal.pone.0175060
McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*(1), 103–107.
McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., & Forssberg, H. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science*, *323*(5915), 800–802.
Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "Far Transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612
Montez, D. F., Calabro, F. J., & Luna, B. (2017). The expression of established cognitive brain states stabilizes with working memory development. *eLife*, *6*, e25606.
Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46–60.
Norman, D. A., & Shallice, T. (1983). Attention to action—Willed and automatic-control of behavior. *Bulletin of the Psychonomic Society*, *21*(5), 354–354.
Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.
Papenberg, G., Hämmerer, D., Müller, V., Lindenberger, U., & Li, S.-C. (2013). Lower theta inter-trial phase coherence during performance monitoring is related to higher reaction time variability: A lifespan study. *Neuroimage*, *83*, 912–920.
Plummer, M. M. (2018). Stukalov A. Package "rjags' [Software update for JAGS].
Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In: Proceedings of the 3rd international workshop on distributed statistical computing. p. 1–10. (124; vol. 124).
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281.
Roberts, G., Quach, J., Spencer-Smith, M., Anderson, P. J., Gathercole, S., & Gold, L. (2016). Academic outcomes 2 years after working memory training for children with low working memory: A randomized clinical trial. *JAMA Pediatrics*, *170*(5), e154568.
RStudio Team. *RStudio: Integrated development for R*. RStudio, PBC; (2020).
Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040.
Sala, G., & Gobet, F. (2020). Working memory training in typically developing children: A multilevel meta-analysis. *Psychonomic Bulletin & Review*, *27*(3), 423–434.
Salmi, J., Nyberg, L., & Laine, M. (2018). Working memory training mostly engages general-purpose large-scale networks for learning. *Neuroscience and Biobehavioral Reviews*, *93*, 108–122.
Saville, C. W. N., Pawling, R., Trullinger, M., Daley, D., Intriligator, J., & Klein, C. (2011). On the stability of instability: Optimising the reliability of

intra-subject variability of reaction times. *Personality and Individual Differences*, *51*(2), 148–153. https://doi.org/10.1016/j.paid.2011.03.034

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2009). On the relation of mean reaction time and intraindividual reaction time variability. *Psychology and Aging*, *24*(4), 841–857.

Schneiders, J. A., Opitz, B., Tang, H., Deng, Y., Xie, C., Li, H., & Mecklinger, A. (2012). The impact of auditory working memory training on the fronto-parietal working memory network. *Frontiers in Human Neuroscience*, *6*, 173.

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does working memory training transfer? A meta-analysis including training conditions as moderators. *Educational Psychologist*, *50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a science of effective cognitive training. *Current Directions in Psychological Science*, *29*(6), 531–537.

Spencer-Smith, M., & Klingberg, T. (2015). Benefits of a working memory training program for inattention in daily life: A systematic review and meta-analysis. *PLOS ONE*, *10*(3), e0119522. https://doi.org/10.1371/journal.pone.0119522

Stan Development Team "RStan: The R interface to Stan." R package version 2.21.2. 2020. Available from: http://mc-stan.org/

StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; (2015).

Steinbeis, N., Haushofer, J., Fehr, E., & Singer, T. (2016). Development of behavioral control and associated vmPFC–DLPFC connectivity explains children's increased resistance to temptation in intertemporal choice. *Cerebral Cortex*, *26*(1), 32–42. https://doi.org/10.1093/cercor/bhu167

Stuss, D. T., Murphy, K. J., Binns, M. A., & Alexander, M. P. (2003). Staying on the job: The frontal lobes control individual performance variability. *Brain: A Journal of Neurology*, *126*(Pt 11), 2363–2380.

Tamnes, C. K., Fjell, A. M., Westlye, L. T., Ostby, Y., & Walhovd, K. B. (2012). Becoming consistent: Developmental reductions in intraindividual variability in reaction time are related to white matter integrity. *The Journal of Neuroscience: The official journal of the Society for Neuroscience*, *32*(3), 972–982.

Titz, C., & Karbach, J. (2014). Working memory and executive functions: Effects of training on academic achievement. *Psychological Research*, *78*(6), 852–368.

Tomasik, M. J., Napolitano, C. M., & Moser, U. (2019). Trajectories of academic performance across compulsory schooling and thriving in young adulthood. *Child Development*, *90*(6), Available from: https://onlinelibrary.wiley.com/doi/10.1111/cdev.13150

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49*, 110–128.

Unsworth, N., & Robison, M. K. (2017). A locus coeruleus-norepinephrine account of individual differences in working memory capacity and attention control. *Psychonomic Bulletin & Review*, *24*(4), 1282–1311.

van Belle, J., van Raalten, T., Bos, D. J., Zandbelt, B. B., Oranje, B., & Durston, S. (2015). Capturing the dynamics of response variability in the brain in ADHD. *NeuroImage: Clinical*, *7*, 132–141.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), https://doi.org/10.18637/jss.v045.i03

van Ravenzwaaij, D., Brown, S., & Wagenmakers, E. J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition*, *119*(3), 381–393.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

von Bastian, C. C., & Oberauer, K. (2014). Effects and mechanisms of working memory training: A review. *Psychological Research*, *78*(6), 803–820.

Wabersich, D., & Vandekerckhove, J. (2014). Extending JAGS: A tutorial on adding custom distributions to JAGS (with a diffusion model example). *Behavior Research Methods*, *46*(1), 15–28.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective & Behavioral Neuroscience*, *3*(4), 255–274.

Wang, C., Ding, M., & Kluger, B. M. (2014). Change in intraindividual variability over time as a key metric for defining performance-based cognitive fatigability. *Brain and Cognition*, *85*, 251–258.

Wass, S. V., Scerif, G., & Johnson, M. H. (2012). Training attentional control and working memory – Is younger, better? *Developmental Review*, *32*(4), 360–387. https://doi.org/10.1016/j.dr.2012.07.001

Wiecki, T. V., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the drift-diffusion model in Python. *Frontiers in Neuroinformatics*, *7*, 14.

Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A., & Tannock, R. (2005). Inconsistency in reaction time across the life span. *Neuropsychology*, *19*(1), 88–96.

Woerner, W., Becker, A., Friedrich, C., Rothenberger, A., Klasen, H., & Goodman, R. (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, *30*(2), 105–112. https://doi.org/10.1024//1422-4917.30.2.105

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, *14*(5), 656–661.

SUPPORTING INFORMATION

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

How to cite this article: Cubillo, A., Hermes, H., Berger, E., Winkel, K., Schunk, D., Fehr, E., & Hare, T. A. (2023). Intra-individual variability in task performance after cognitive training is associated with long-term outcomes in children. *Developmental Science*, *26*, e13252. https://doi.org/10.1111/desc.13252