Are facet-specific task trainings efficient in improving children's executive functions, and why (they might not be)? A multi-facet latent change score approach

Sascha Zuber[1,2,3], Emilie Joly-Burra[1,3,4], Caitlin E.V. Mahy[5], Vanessa Loaiza[6], & Matthias Kliegel[1,3,4]

[1] Center for the Interdisciplinary Study of Gerontology and Vulnerability, University of Geneva, Geneva, Switzerland
[2] Institute on Aging & Lifelong Health, University of Victoria, BC, Canada
[3] Swiss National Centre of Competences in Research LIVES–Overcoming vulnerability: life course perspectives, Lausanne and Geneva, Switzerland
[4] Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland
[5] Department of Psychology, Brock University, St Catharines, Ontario, Canada
[6] Department of Psychology, University of Essex, United Kingdom

Correspondence:
Sascha Zuber, University of Geneva, Centre for the Interdisciplinary Study of Gerontology and Vulnerabilities (CIGEV), Boulevard du Pont d'Arve 28, CH- 1205 Genève, Office 103

Contact: E-Mail: Sascha.Zuber@unige.ch; Phone: +41 22 379 83 81

This is an Accepted Manuscript (AM) of an article accepted for publication by Elsevier in *Journal of Experimental Child Psychology* on November 21, 2022. The final authenticated version will be available online at: https://doi.org/10.1016/j.jecp.2022.105602

Reference:

Zuber, S., Joly-Burra, E., Mahy, C.E.V., Loaiza, V., & Kliegel, M. (in press). Are facet-specific task trainings efficient in improving children's executive functions, and why (they might not be)? A multi-facet latent change score approach. *Journal of Experimental Child Psychology*. https://doi.org/10.1016/j.jecp.2022.105602

Abstract

It currently remains unclear how facet-specific trainings of three core modules of executive function (EF; updating, switching, inhibition) directly compare regarding efficacy, whether improvements on trained tasks transfer to non-trained EF tasks, and which factors predict children's improvements. The present study systematically investigated three separate EF trainings in 6- to 11-year-old children (N = 229), using EF-specific trainings that were similar in structure, design, and intensity. Children participated in pre- and posttest assessments of the three EFs and were randomly allocated to one of three EF trainings, or to an active or passive control group.

Multivariate latent change score models revealed that only the updating group showed training-specific improvements in task performance that were larger compared with active as well as passive controls. In contrast, there were no training-specific benefits of training switching or inhibition. Latent changes in the three EF tasks were largely independent, and there was no evidence for transfer effects to non-trained EF tasks. Lower baseline performance and older age predicted larger changes in EF performance. These seemingly opposing effects support compensation accounts as well as developmental theories of EF, and highlight the importance of simultaneously accounting for multiple predictors within one model.

In line with recent theoretical proposals of EF development, we provide new systematic evidence that questions whether modular task trainings represent an efficient approach to improve performance in narrow or broader indicators of EF. Thereby, this evidence ultimately highlights the need for more comprehensive assessments of EF and, subsequently, the development of new training approaches.

Keywords: executive functions; school age; children; cognitive training; plasticity; latent change score modelling;

Highlights

- First systematic comparison training the three core EF across the entire school age
- Latent change score models reveal that updating is the most beneficial EF training
- Seemingly opposing effects support compensation accounts and developmental theories
- Lower baseline performance but older age predict larger training benefits
- Improvements across EFs are largely independent and transfer to other EF is lacking

Are facet-specific task trainings efficient in improving children's executive functions, and why (they may not be)? A multi-facet latent change score approach

Executive function (EF) represents a group of higher order cognitive processes that enable individuals to be attentive, to solve problems, to pursue goals, and to regulate behaviors, thoughts, and emotions (Diamond, 2013; Zelazo et al., 2008). From a developmental perspective, EF is crucial, for example, because it contributes to children's attainment of autonomy, socioemotional functioning, and academic performance (Best et al., 2011; Dawson & Guare, 2018; Denham et al., 2015; Diamond, 2016; Liew, 2012; Riggs et al., 2006). Motivated by the high everyday relevance of broader, more ecological indicators of EF, extensive research has aimed to improve children's performance on more narrow laboratory measures of EFs via cognitive training (Strobach & Karbach, 2021). Such efforts are based on the concept of *transfer*, which postulates that repeatedly performing a cognitive task will improve performance on tasks that deploy similar cognitive processes (often labelled *near transfer*), and that this might even improve performance on structurally more distant tasks deploying similar cognitive processes (often labelled *far transfer*; e.g., Kliegel et al., 2017; Strobach & Karbach, 2021).

In this context, the three most widely studied and trained EF components are: *updating* of information in working memory, *switching* attention between different task sets, and *inhibition* of automatic or predominant responses or of irrelevant distractors (Miyake et al., 2000). It is crucial to highlight here that this three-partite view of EF has received increasing criticism (e.g., Doebel, 2020; Perone et al., 2021). Importantly, EF is still developing during childhood and certain components may only fully mature later in life (e.g., switching; Garon et al., 2008; Karr et al., 2018; Müller & Kerns, 2015). Thus, it is unlikely that EF is best understood with a three- (compared to uni- or two-) dimensional model across all ages (see Karr et al., 2018; Miyake & Friedman, 2012). Related to this, more and more research questions such modular views of different EF components and whether training performance on these rather narrow indicators of EF can actually transfer to broader outcomes (Diamond & Ling, 2016; Kassai et al., 2019; Perone et al., 2021). Recent theoretical contributions therefore urge that EF should be conceived more comprehensively as using control "in the service of particular goals that activate and are influenced by diverse mental content such as knowledge, beliefs, and values" (Doebel, 2020, p. 11). Building on this view, rather than static modules that are activated one at a time, EF may be better understood as a dynamic system that allows momentary behavior by assembling multiple components (physiological, cognitive, emotional, and motor processes, together with the social and physical forces) of prior experiences and abilities to pursue a goal (Perone et al., 2021).

Despite such proposals and calls for more comprehensive views of EF, the three component model is currently still a persisting conceptualization of EF, with updating, shifting, and inhibition representing the most widely studied and trained EF components in childhood (for an in-depth review, see Müller & Kerns, 2015). Thus, the goal of the present study was to systematically compare cognitive trainings of these three EFs. Although our study focused on these narrow measures of EF, we will discuss findings within the broader context of the current EF literature, which will ultimately lead to new insights that align with these more comprehensive views.

1.1. Cognitive training and EF

A significant body of literature suggests that computerized cognitive process trainings can enhance EF performance on laboratory tasks (for reviews, see e.g., Diamond & Ling, 2016; Diamond & Ling, 2020; Kliegel et al., 2017; for meta-analyses, see e.g., Cao et al., 2020; Sala & Gobet, 2017; Scionti et al., 2020; Takacs & Kassai, 2019). However, most previous studies have either focused on training a single EF per study or all three core EFs simultaneously. So far, only one study has applied specific, separate trainings for each EF component within a sample of typically developing children (Johann & Karbach, 2020). Johann and Karbach (2020) compared standard with game-based trainings of the three EFs and examined potential transfer effects to mathematical and reading abilities in 153 8- to 11-year-old children. They found that both trainings improved EF performance. They also found transfer effects to reading abilities. EF improvements were greater in children that participated in a game-based switching or inhibition training compared to the passive control group, and these improvements persisted at a three-month follow-up.

Johann and Karbach's study thus provided the first integrative insights into the broader, long-lasting benefits of training EF to improve non-EF domains. However, although the authors reported no transfer between the three EFs, they did not examine in detail how improvements in each EF directly compared to the others nor whether improvements may be interrelated. The authors suggested that transfer between EFs may not have occurred because each training program consisted of a set of multiple tasks and that training only on a single EF task might facilitate transfer to untrained EF tasks. Despite these relevant suggestions, systematic comparisons of the three EF trainings are currently still lacking, leaving a series of conceptually important questions unanswered – such as how different EFs directly compare in terms of how easily task performance can be improved, whether improved performance in one EF relates to improved performance in the other EFs, or which factors predict training benefits in children. Building on and extending recent work such as that of Johann and Karbach (2020), the present study set out to tackle these pressing questions.

1.2 Performance improvements in the trained EF

Currently, it is unclear how updating, switching, and inhibition directly compare in terms of specific performance improvements in the trained EF. Performance improvements are most frequently observed in studies that train updating, whereas results are less consistent when training switching or inhibition (e.g., Kassai et al., 2019; Rapport et al., 2013; Takacs & Kassai, 2019). Previous findings have to be interpreted with caution, however, because updating also represents the most studied EF component and thus is most likely to produce a larger number of positive findings (Takacs & Kassai, 2019). Further, studies typically train only one specific EF and contrast benefits either to an active or to a passive control group. Yet, studies largely vary in terms of target population, design, and training intensity, which further contributes to the inconsistent pattern of EF training benefits (Klingberg, 2010), making it difficult to investigate whether performance is more likely to improve on certain EF tasks.

1.3. Transfer to performance improvements in untrained EF tasks

Even more debated is the extent to which modular task trainings translate into performance improvements on untrained EF tasks (Diamond & Ling, 2020; Smid et al., 2020). For each EF, transfer effects have been inconsistent (for reviews and meta-analyses, see Kliegel et al., 2017; Klingberg, 2010; Melby-Lervåg et al., 2016; Morrison & Chein, 2011; Sala & Gobet, 2017, 2020), potentially again because of studies examining one EF training at a time and trainings being heterogeneous across studies. Importantly, there is currently no systematic examination of whether improvement in one EF task directly translates into improvements in other EF tasks.

1.4. Theoretical accounts and predictors of training benefits

Although research consistently shows that there is important variance in how much individuals benefit from EF trainings (e.g., Cao et al., 2020; Smid et al., 2020; Traut et al., 2021), there is currently no consensus on which factors predict training benefits in children nor on which theoretical and developmental accounts best describe training mechanisms in children. From a theoretical perspective, *compensation accounts* suggest that training benefits are largest for individuals that have an initial disadvantage in performance (e.g., individuals with low baseline performance, atypically developing children, children at higher developmental risk or from lower socioeconomic conditions; Karbach et al., 2017; Smid et al., 2020; Strobach & Karbach, 2021; Traut et al., 2021). Accordingly, children with lower initial performance have more room to improve and engaging in new cognitive activities might be more beneficial for them. In contrast, *magnification accounts* suggest that children with initial advantages in performance benefit more, because they are more able to fully engage in the intervention program and build on already existing skills (e.g., Foster et al., 2017; Lövdén et al., 2012; Swanson, 2014, 2015). Although both accounts have been supported by empirical studies (see Katz et al., 2021; Traut et al., 2021), they assign a central, but opposing role to baseline performance.

Similarly, from a developmental perspective, there currently is no agreement on the role of age. On one hand, younger children may benefit more, because of greater neuroplasticity and because their EFs are being differentiated into distinct abilities (Best & Miller, 2010; Huizinga et al., 2006). On the other hand, older children may benefit more because the neural underpinnings of EF – prefrontal networks – continue to undergo important structural and synaptic changes in late childhood and throughout adolescence (Best & Miller, 2010; Diamond, 2013). So far, the literature has mostly compared rather distant age groups, typically with a focus on older versus younger adults (Katz et al., 2021). It therefore remains unclear whether early school age children versus pre-adolescents may benefit more from EF trainings. Certain meta-analyses suggest larger training benefits for younger children (e.g., Cao et al., 2020; Cepeda et al., 2001; Wass et al., 2012), whereas others do not find age effects (Kassai et al., 2019; Scionti et al., 2020). Furthermore, other demographic variables that could interact with age, namely gender, remain largely unstudied, even though boys and girls may respond differently to computerized tasks (e.g., Delalande et al., 2020; Martinovic et al., 2016).

1.5. The present study

The present study aimed to provide the first systematic and comprehensive examination of how facet-specific single-task trainings directly compare to each other across the entire middle childhood and whether they translate into benefits in untrained EFs. We aimed to extend Johann and Karbach's (2020) study by: (a) systematically disentangling whether benefits translate to performance improvements in untrained EF tasks; (b) extending the age range to cover the entire primary school age (i.e., 6 to 11 years); (c) more directly examining the role of multiple predictors (i.e., baseline performance and age, controlling for gender) within a single, latent change score model (LCSM); and (d) including an active control group, for which the activity was closely matched to the training interventions regarding task design, difficulty, adaptability of difficulty, intensity, and duration.

The present study thereby aimed to answer the following research questions: (1) How do the different EFs directly compare in terms of how easily task performance can be improved?, (2) Does improved performance on one EF task relate to improved performance on tasks deploying other EFs?, (3) Which factors predict training benefits?, (4) Is there support for compensation versus magnification accounts of cognitive training when all predictors are considered simultaneously within one model?, and (5) Would the efficacy of the training be different when comparing training groups to active versus passive controls?

2. Method

2.1. Participants

Two hundred and thirty-nine school-aged children initially participated in the study. They were recruited through advertisements at publication locations, schools and the experimenters' network. In view of the important differences in the training literature between typically versus atypically developing children as well as the large age range of our sample, we excluded children whose indices of general cognitive functioning were outliers, in order to the render the sample more homogenous in terms of overall development of cognitive functioning. Therefore, nine children were excluded from subsequent analyses because they scored below two and a half standard deviations of their age-group norms on fluid and/or crystallized intelligence measures (assessed via the subtests “Matrices” and “Vocabulary” of the WISC-IV; Wechsler, 2004). Of these nine children, one child came from the updating group, four from the inhibition group, three from the active control group, and one from the passive control group. It is important to highlight that the exclusion occurred after data collection and that the cut-off was adapted during revisions of this manuscript (i.e., we initially planned to exclude scores below two standard deviations, which would exclude one additional participant). Note that pattern of results of our findings would remain the same if analyses were performed on data including all participants as well as if the cut-off was 2 standard deviations. The remaining children did not report any history of (neuro-)psychopathology (as indicated by children’s caregivers in questionnaires) and were either native French speakers or had fluent proficiency in French. All children as well as their caregivers gave informed consent.

The final sample consisted of 230 children (M_{age} = 8 years; 4 months, *SD* = 1 year; 5 months), 121 of which were female (53%; there were no significant differences in age between genders, $p = .502$). Children’s ethnicity was not collected, as this is not common practice in the country of data collection. Before pretest assessment, children were randomly allocated to one of the five groups (updating training, switching training, inhibition training, active control, or passive control group). Table 1 displays number of children, percentage of girls, and age per experimental group. ANOVAs and subsequent Tukey HSD tests showed that there were no significant differences between any of the experimental groups regarding percentage of girls or age (all *p*s > .05). A chi square test of homogeneity indicated that the number of children did not significantly vary between groups, $\chi^2(4) = 1.15$, $p = .89$.

Table 1

Number of children, percentage of girls, and age, per experimental group.

Training group	*N*	Gender	Age			
		% girls	*M*	*SD*	*min*	*max*
Updating	48	50%	8;8	1;5	5;10	11;5
Switching	47	53%	8;0	1;6	5;11	10;8
Inhibition	42	59%	8;4	1;4	6;5	11;2
Active control	43	49%	8;6	1;6	5;10	10;11
Passive control	50	54%	8;1	1;5	6;4	10;7
ANOVAs		p = .91	p = .15			

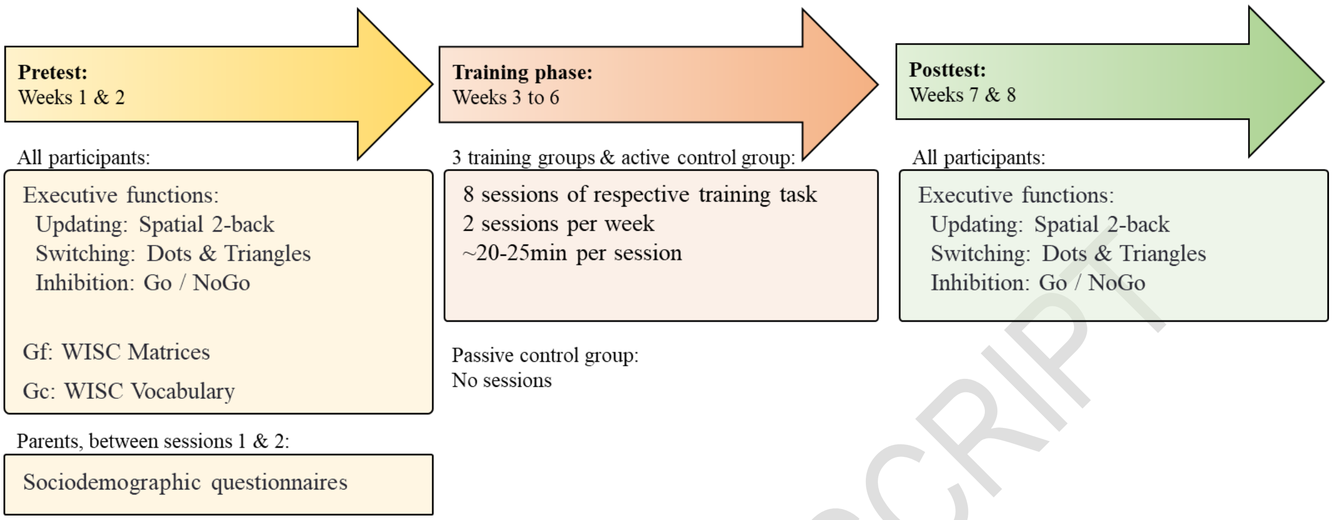
Note. Age in years;months.

2.2. Procedure

Figure 1 illustrates the procedure of the study for the five experimental groups, separated by study phase. Pre- and posttest assessments consisted of two sessions each (~45 min per session) during which children performed different EF tasks as well as other cognitive tasks (e.g., measuring fluid and crystalized intelligence) in a pseudo-randomized order. Socio-demographic questionnaires were filled out by children's parents between the two pretest assessments. Pre- and posttest assessments were separated by four weeks on average. During this period, the three EF training groups and the active control group participated in eight sessions on a computer (~20-25min each; two sessions per week), whereas the passive control group did not participate in any activities. All sessions (pre, post, and trainings) were conducted by two experimenters (one leading the experiment, one being present due to ethical requirements for testing children) in a quiet environment where children were not distracted.

Figure 1

Study procedure, separated per experimental group and study phase.



Note. Gf = fluid intelligence; Gc = crystallized intelligence.

2.3. Measures

2.3.1 Pre-training assessment of fluid and crystallized intelligence

Fluid intelligence: Matrices of WISC-IV (Wechsler, 2004). For each trial, children were shown a 2 by 2 grid of four boxes with the bottom-right box displaying a question mark and the others displaying images. Below this grid, six images were displayed and children were instructed to select the image that would complete the series above (e.g., selecting a green lightbulb among bulbs of other colors). The task consisted of 32 grids in total but was ended earlier if children selected four incorrect answers in five consecutive trials. The outcome measure was the number of correct responses (note that raw scores were age-standardized).

Crystallized intelligence: Vocabulary of WISC-IV (Wechsler, 2004). Children were asked to explain the meaning of words (e.g., "what is an umbrella?") and received two points for correct answers (e.g., "to protect you from rain"), one point for partial, vague answers (e.g., "you hold it above your head") and no points for incorrect answers. The task consisted of 31 words in total but was ended earlier if children gave five consecutive incorrect answers. The outcome measure was the sum of points (note that raw scores were age-standardized).

2.3.2. Pre- and post-training assessment of EF performance

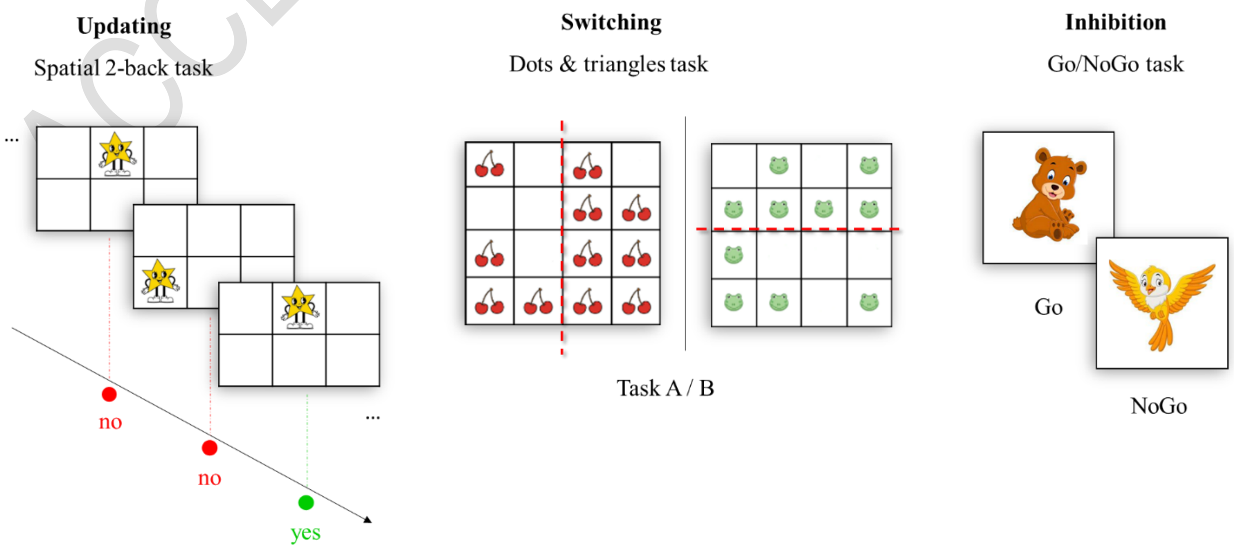
Updating: Spatial 2-back task (adapted from Jaeggi et al., 2011). For each trial, children were shown a 3 by 2 grid of six boxes, and they had to indicate whether a cartoon character was displayed in the same box as two trials before (by pushing the green button stuck on the right arrow key) or not (by pushing the red button stuck on the left arrow key; see Figure 2). Children first performed a practice block of 17 trials (which was repeated if accuracy was below 60%). This was followed by five test blocks of 17 trials, each containing five hit trials (for which the character was in the same location as two trials before) and 12 non-hit trials. For the parallel version of the posttest assessment, a different cartoon character was used, and hit/non-hit trials appeared in a different order. In both assessments the order of hit/non-hit trials was the same for all participants. The outcome measure was the proportion of correctly detected hits minus the proportion of false alarms on non-hit trials.

Switching: Dots & triangles task (adapted from Huizinga et al., 2006). This task consisted of two single task blocks (task A and task B) and a mixed-task block (task A/B). For each trial, children saw a grid of 4x4 boxes, and, using the four arrow keys, they had to answer whether there were more dots (i.e., frog faces) on the left or the right half of the grid (task A), or whether there were more triangles (i.e., cherries) in the top or bottom half of the grid (task B; see Figure 2). Children first worked on both single task blocks (in counter-balanced order), which consisted of 10 practice trials (and an additional 10 practice trials if accuracy was below 60%) and 40 experimental trials. Children then worked on the mixed-task block, which consisted of 21 practice trials (and 21 possible re-practice trials) and 81 experimental trials. The mixed-task block shifted between task A and task B every four trials. For the parallel version of the posttest assessment, stimuli were inversed (i.e., dots were used for task B and triangles for task A). The outcome measure was switching costs (i.e., mean reaction time on switch trials minus mean reaction time on non-switch trials, both on trials with correct responses only). Note that reaction times were initially recorded in milliseconds, but were rescaled to seconds to avoid that variances were much larger than on the other outcome measures.

Inhibition: Go/NoGo task (adapted from Schulz et al., 2007). Children were shown a series of animal pictures and had to push the space bar as fast as possible as soon as a new picture appeared (Go stimuli, 75% of all trials) except for birds (for which no response had to be made; NoGo stimuli, 25% of all trials; see Figure 2). Go and NoGo trials were presented in pseudo-randomized order. Children practiced this first for 16 trials (and for another 16 practice trials if performance was below 60%) and then worked on a block of 96 trials. For the parallel version of the posttest assessment, monkeys were used as NoGo stimuli (note that monkeys were used as Go stimuli at pretest, but birds were not used as Go stimuli at posttest). The outcome measure was inhibition accuracy (i.e., the proportion of correctly inhibited NoGo trials).

Figure 2

Example stimulus for updating, switching, and inhibition tasks



2.3.3. Training programs

The three EF and the active control programs were similar in terms of training design and intensity. Each program consisted of eight sessions lasting for 20-25min each. For the three EFs, training tasks resembled the pre-post assessment of the same EF (see section 2.3.2). Specifically, the updating trainings consisted of a spatial 2-back paradigm, for which children had to indicate for each trial whether a cartoon character was displayed in the same location as two trials before. Each session consisted of 170 trials (50 hit trials). The switching training tasks consisted of a task A/task B switching paradigm, for which children either had to indicate whether stimuli belonged to one vs. to another category (task A), or whether one vs. two objects were displayed (task B). The paradigm switched between tasks on every third trial and each training consisted of 410 trials (200 switching trials). The inhibition training consisted of a Go/NoGo response inhibition paradigm, for which children had to push the spacebar as fast as possible after stimuli appeared on the screen (Go-trials) except for when the stimulus corresponded to a specific category (NoGo-trials). Each session consisted of 320 trials (80 NoGo trials). The active control training children had to categorize images (similar to the categorization tasks of the switching paradigm, but without having to switch between different task sets, thus not particularly tapping into EF). Each training session consisted of 410 trials.

For all four training programs, task difficulty was individually adapted to children's performance throughout the program. During each training session, there were a total of 10 difficulty levels, to which children could advance or revert depending on their performance. Between levels, difficulty was increased by decreasing how long stimulus and fixation cross were presented and, for certain levels, by presenting more complex stimuli (i.e., more spatial locations and more challenging maps for the updating training; categories that are more difficult to distinguish for the inhibition training). To be able to compare training progress between groups, one outcome measure was the highest level achieved in each training session. In addition, group specific outcome measures were: the proportion of correctly detected hits minus the proportion of false alarms on non-hit trials for updating; switching costs on trials with correct responses only for switching; inhibition accuracy for inhibition; mean reaction time on correctly categorized trials for the active control training. More detailed descriptions of the different trainings can be found in supplementary material S1. The passive control only participated in pre- and posttest assessments without receiving any activities between the two assessment times (i.e., 'business as usual' control group).

2.4. Statistical analyses

First, to assess how performance on the training tasks changed throughout the training, we compared the highest task level achieved as well as task specific outcome measures on the first versus on the last training session by conducting a paired-sample t-test for each training group separately (including the active control group). These t-tests as well as descriptive statistics, correlations, ANOVAs, and Tukey HSD test were conducted with the jamovi software.

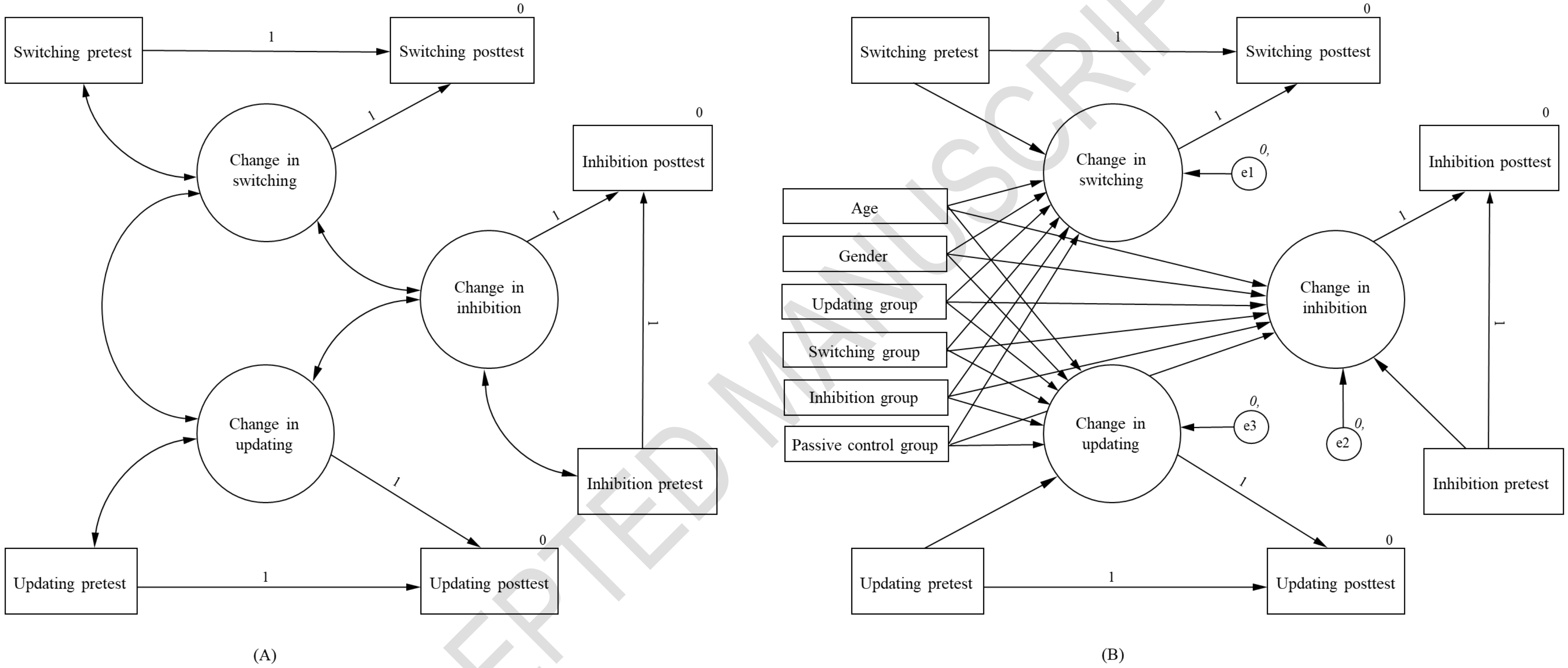
Second, to assess predictors of pre-post change and potential transfer effects to untrained EFs, we conducted factorial latent change score modelling (LCSM). To examine means and variances in change of each EF task as well as how these changes correlated, in a first LCSM (Model 1, see Panel A of Figure 3), we computed latent variables of change as the difference between pre- and posttest performance for each of the three EF tasks. To examine whether change in one EF task was related to change in the other EF tasks, latent change variables were allowed to covary. Performances at pretest for the three EF tasks were also allowed to covary.

Third, to investigate predictors of changes in EF performance, in a second LCSM (Model 2, see Panel B of Figure 3) the following variables were added as predictors of the three latent change scores: (1) baseline performance, (2) children's age, (3) gender, and (4) the specific training group. Four dummy variables were created to investigate the effect of each training group compared to the active control group (i.e., the updating group has the value '1' on the variable 'updating training' and values '0' on the three remaining dummy variables; the active control group has '0' on all four variables). A significant effect of a dummy variable means that this group displayed larger EF changes than the active control group. Residual variances of latent change variables were allowed to covary.

Fourth, to examine whether the efficacy of the training would be judged differently when comparing training groups to active versus passive controls, we computed a third model (Model 3), in which we used the passive control group as a reference group. Note that although changing the reference group can result in different parameter estimates, this model is mathematically equivalent to Model 2 in terms of model fit, which therefore will only be reported for Model 2. LCSMs models were estimated in IBM SPSS AMOS (version 26) using Maximum Likelihood estimation. As indicated by the Little MCAR test (computed in IBM SPSS version 26), missing data (which were less than 1% of all data points) were missing completely at random ($\chi^2(41) = 48.54$, $p = .20$) and therefore subsequently were imputed using Full Information Maximum Likelihood (Little, 1988). We assessed the goodness-of-fit for the two LCSMs using the χ^2/df ratio, the root mean square error of approximation (RMSEA), and the comparative fit index (CFI). Model fits were considered as good when the χ^2/df ratio was smaller than 3, when the RMSEA was comprised between 0 and 0.06, and when the CFI was greater than 0.95 (Hu & Bentler, 1999).

Figure 3

Factorial latent change score models



Note. Panel A represents latent score change Model 1. The latent change variables have estimated means and variances. Single-headed arrows represent regression coefficients while two-headed arrows represent covariances. For the purpose of readability, correlations between pre-tests for the three EF tasks are not depicted. Panel B represents Model 2 where predictors of change are included in the model and allowed to covary. The error terms (e1, e2, and e3) indicate residual variances from the latent change scores. For the purpose of readability, covariances between the different predictors and covariances between residual variances of change are not depicted.

Table 2

Descriptive statistics and correlations between age, pre- and posttest performances on the three EF tasks (across all groups).

	M (SD)	1	2	3	4	5	6
1. Age	8.29 (1.43)	—					
2. Updating pretest	.35 (.25)	.33***	—				
3. Updating posttest	.46 (.29)	.28***	.46***	—			
4. Switching pretest	0.38 (0.32)	-.21**	-.16*	-.11	—		
5. Switching posttest	0.27 (0.27)	-.26***	-.20**	-.19**	.37***	—	
6. Inhibition pretest	.81 (.14)	.20**	.19**	.26***	-.06	-.05	—
7. Inhibition posttest	.82 (.15)	-.01	.07	.15*	-.07	-.08	.33***

Note. *M* = mean. *SD* = standard deviation. * *p < .05*, ** *p < .01*, *** *p < .001*. Updating scores = proportion of correctly detected hits minus the proportion of false alarms on non-hit trials; switching scores = switching cost in seconds (= mean reaction time on shift trials minus mean reaction time on non-shift trials on trials with correct responses only); inhibition scores = inhibition accuracy (proportion of correctly inhibited NoGo trials); age = children's age in years.

3. Results

3.1. Descriptive statistics

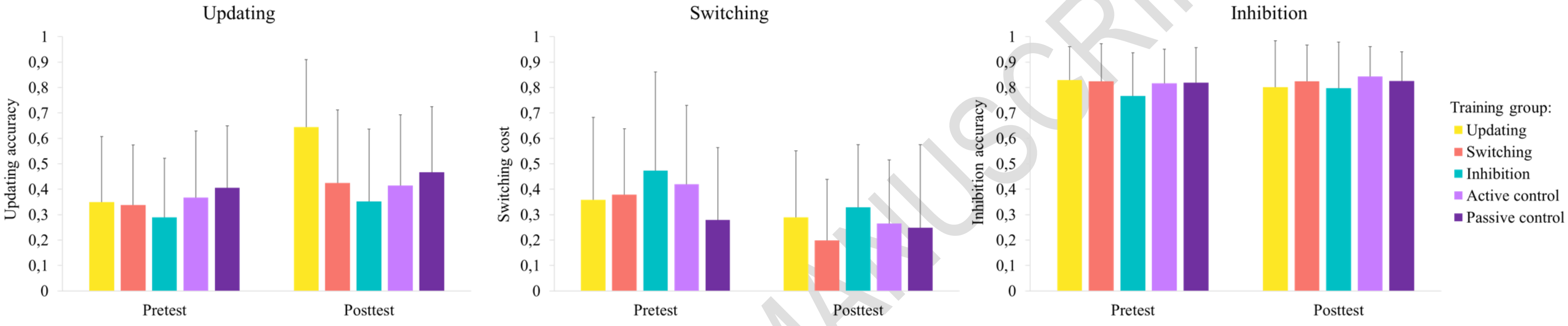
Table 2 presents means, standard deviations, and correlations between age, pre- and posttest performance on the three EFs tasks across all groups. Figure 4 displays pre- and posttest performances on the three EF tasks separated by group.

3.2. Changes in performance across training sessions

Figure 5 depicts trajectories of the highest level achieved on each training session (i.e., group mean) for the four training groups separately. It also displays performance trajectories on each training session on the four training tasks. Regarding the highest levels achieved, paired-sample t-tests between the first and the eighth training session showed that there were significant differences with large improvements for the updating, $t(45) = 12.95$, $p < .001$; $d = 1.91$, and for the switching group, $t(40) = 6.83$, $p < .001$; $d = 1.07$. In contrast, performance of the inhibition and of the active control group did not significantly improve on the highest level achieved, $t(37) = 0.45$, $p = .65$; $d = 0.07$, and $t(42) = - 0.84$, $p = .41$; $d = -0.13$, respectively. Regarding performance on the trained tasks, paired-sample t-tests between the first and the eighth training session showed that there were significant improvements for the updating, $t(45) = 5.13$, $p < .001$; $d = 0.76$, and for the switching group, $t(40) = -5.10$, $p < .001$; $d = -0.80$. In contrast, performance of the inhibition group did not significantly improve, $t(37) = 0.88$, $p = .39$; $d = 0.14$, whereas the active control group became significantly slower across sessions, $t(42) = 3.91$, $p < .001$; $d = 0.60$.

Figure 4

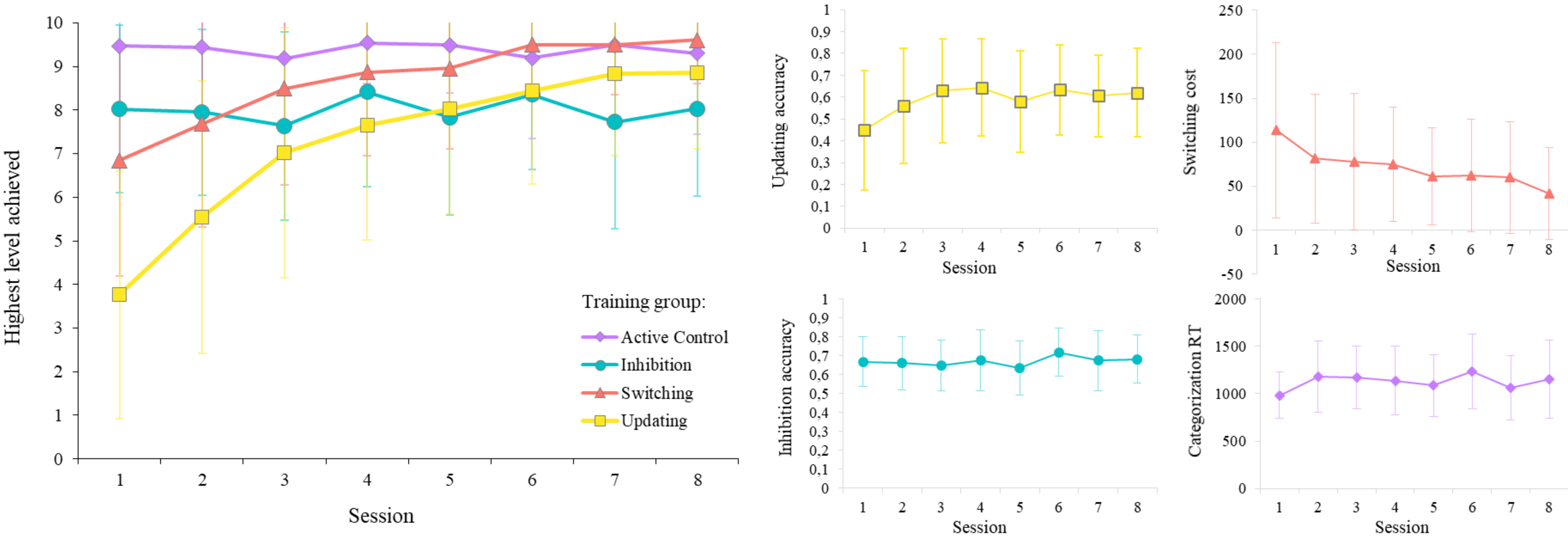
Pre- and posttest performances on the three EF tasks separated by training group.



Note. Updating accuracy = proportion of correctly detected hits minus the proportion of false alarms on non-hit trials; Switching cost = switching cost in seconds (= mean reaction time on switch trials minus mean reaction time on non-switch trials on trials with correct responses only); Inhibition accuracy = proportion of correctly inhibited NoGo trials. Error bars represent standard deviations.

Figure 5

Progression of highest level achieved (large figure left) and performance (small figures right) across the eight training sessions, separated per training group



Note. Updating accuracy = proportion of correctly detected hits minus the proportion of false alarms on non-hit trials; Switching cost = switching cost in milliseconds (= mean reaction time on switch trials minus mean reaction time on non-switch trials on trials with correct responses only); Inhibition accuracy = proportion of correctly inhibited NoGo trials. Categorization RT = mean reaction time (in milliseconds) on correctly categorized trials of active control training task. Error bars represent standard deviations.

Table 3

Estimated means, variances and covariances for the latent score change variables in Model 1.

Parameter	Estimate	S.E.	*p*
Mean change in updating	**.11**	.02	< .001
Mean change in switching	**-.11**	.02	< .001
Mean change in inhibition	.01	.01	.60
Variance of change in updating	**.08**	.01	< .001
Variance of change in switching	**.11**	.01	< .001
Variance of change in inhibition	**.03**	.01	< .001
Change in updating ↔ Change in switching	-.01	.01	.50
Change in updating ↔ Change in inhibition	.01	.01	.82
Change in switching ↔ Change in inhibition	-.01	.01	.56

Note. Double-headed arrows denote covariances. S.E. = standard error of estimation. Significant estimates in **bold**.

Table 4

Changes in performance per EF and per group

	Updating			Switching			Inhibition		
Improvement	Increase in performance accuracy	% of pretest	SD of pretest	Reduction of cost (in ms)	% of pretest	SD of pretest	Increase in performance accuracy	% of pretest	SD of pretest
Across all groups	.11	32.85	0.47	115	30.30	0.36	.01	0.81	0.05
Updating group	.29	84.76	1.14	69	19.05	0.21	-.03	-2.31	-0.20
Switching group	.09	25.86	0.37	181	47.19	0.70	-.01	-0.12	-0.01
Inhibition group	.06	21.17	0.27	144	30.80	0.37	.03	3.96	0.18
Active controls	.05	14.27	0.20	165	39.15	0.53	.03	3.31	0.20
Passive controls	.06	15.28	0.25	30	10.48	0.10	.01	0.83	0.05

Table 5

Regression weights, standard error of estimation, and p-values for predictors of latent changes in EFs in Model 2.

	Predicting change in updating				Predicting change in switching				Predicting change in inhibition			
Predictors	*b*	β	S.E.	*p*	*b*	β	S.E.	*p*	*b*	β	S.E.	*p*
Updating pretest	**-.52**	**-.46**	.07	< .001	-	-	-	-	-	-	-	-
Switching pretest	-	-	-	-	**-.73**	**-.70**	.05	< .001	-	-	-	-
Inhibition pretest	-	-	-	-	-	-	-	-	**-.66**	**-.55**	.07	<.001
Age	**.03**	**.13**	.01	.03	**-.04**	**-.17**	.01	< .001	-.01	-.06	.01	.33
Gender	.01	.02	.03	.73	-.02	-.03	.03	.54	.01	.03	.02	.65
Updating group	**.24**	**.34**	.05	< .001	.06	.08	.05	.23	-.04	-.10	.03	.14
Switching group	.04	.05	.05	.45	-.07	-.08	.05	.20	-.03	-.07	.03	.36
Inhibition group	-.02	-.03	.05	.66	.05	.06	.05	.34	-.03	-.07	.03	.30
Passive control group	.04	.06	.05	.40	.01	.02	.05	.78	-.03	-.07	.03	.31
R^2	.30				.47				.33			

Note. *b* = raw regression weights. β = standardized regression weights. S.E. = standard error of estimation. Updating pretest = proportion of correctly detected hits minus the proportion of false alarms on non-hit trials; switching pre-test = switching cost in seconds (= mean reaction time on shift trials minus mean reaction time on non-shift trials on trials with correct responses only); inhibition scores = inhibition accuracy (proportion of correctly inhibited NoGo trials); age = children's age in years; gender = coded *0* for girls and *1* for boys. To code for group, four dummy variables were computed, with the active control group as a reference. R^2 = Squared multiple correlation. Significant estimates in **bold**.

3.3. Baseline differences

To assure that training improvements were not confounded with potential differences in performance at baseline, we compared pretest performance of the five groups for each EF. Tukey HSD analyses indicated that were no significant differences between any of the five groups in baseline performance on updating, switching, or inhibition, except for one: the inhibition group performed significantly worse than the passive control group on the switching task at baseline ($t(220) = -2.80$, $p = .043$; all other *p*s > .05).

3.4. Variability in change and transfer effects to other EF tasks

Model 1 showed excellent fit ($\chi^2(6) = 7.06$, $\chi^2/df = 1.18$, $p = .32$; CFI = .99, RMSEA = .03). Parameter estimates are reported in Table 3. Mean changes between pre- and posttest were significant in updating and in switching, but not in inhibition. Skewness of the factor scores for the mean changes in updating, in switching, and in inhibition was .04, .36, and -.11 respectively, whereas kurtosis was .19, 1.64, and .88 respectively. Table 4 shows changes in performance per EF and per group. In addition, variances in change (i.e., variances of the latent change scores) were significant in the three EFs, indicating that for each EF there was interindividual variability in change. However, changes in the three EFs did not correlate (all *p*s > .05), indicating that changes in EFs were independent from each other.

3.5 Predictors of change and transfer to untrained EF tasks

As there was significant variance in change on all three EFs, Model 2 examined predictors of change for each EF. Model 2 showed excellent fit ($\chi^2(6) = 10.58$, $\chi^2/df = 1.76$, $p = .10$; CFI = .99, RMSEA = .06). When all predictors were considered simultaneously, they predicted a substantial portion of variance in changes in updating, switching, and inhibition (30%, 47%, and 33%, respectively). Raw and standardized estimates for each EF are reported in Table 5. In sum, results show that change in updating was predicted negatively by updating performance at pretest but positively by age. In addition, only the updating group showed larger changes than the active control group. Change in switching[1] was predicted negatively by switching costs at pretest and by age. None of the groups showed larger changes in switching than the active control group. Change in inhibition was predicted negatively by baseline inhibition performance, indicating that children with lower initial performance showed greater improvements. There were no effects of age nor of group on change in inhibition. Gender did not predict change in any of the three EFs. As for correlations between the latent change variables in Model 1, residual errors for the latent change score variables in the three EFs did not correlate in Model 2, again indicating that changes in EFs were independent from each other. Further, results show that none of the training group variables significantly predicted changes in untrained EFs, again indicating no evidence for transfer effects to untrained EFs.

3.6 Active versus passive controls as reference group

Finally, using the passive controls as reference group in Model 3 revealed a similar pattern of results as in Model 2. Regression weights, standard error of estimation, and p-values for predictors of latent changes in EFs in Model 3 are therefore displayed in supplementary material S2.

4. Discussion

The present study set out to perform the first systematic comparison of facet-specific single-task EF trainings across the entire middle childhood. It aimed to: (1) directly compare whether updating, switching, and inhibition performance improve when these three EF tasks are trained under similar conditions, (2) examine whether training certain EF tasks improves performance on the others, (3) investigate which factors predict training benefits when accounting for all predictors simultaneously within a single model, including whether

[1] For the interpretation of switching results, it is important to keep in mind a) that pre- and posttest scores represent switching costs (hence, larger scores indicate worse performance), and b) that mean change in switching was negative (representing a reduction of switching costs; hence, larger negative regression weights of predictors indicate a larger reduction of switching costs).

predictors support compensation versus magnification accounts of cognitive training, and finally (4) to study whether the efficacy of the training would be judged differently when comparing training groups to active versus passive controls.

4.1. Does performance improve when training on EF tasks?

Overall, our findings show that updating training is the most likely to produce performance improvements when the three EFs are trained under similar conditions using a single-task training paradigm. In contrast, we find no evidence that the other trainings benefit children's EF performance beyond retest effects, general learning, and short-term developmental changes. In more detail, the updating group improved performance across training sessions, and only children in this group displayed larger pre-to-post improvements in updating performance than the active as well as the passive control group. On average, children of the updating group increased their accuracy by 85% whereas the switching, inhibition, active and passive control groups only improved by 19% on average. Although the switching group improved switching performance throughout the training, children did not show significantly larger improvements than the active nor the passive control groups from pre- to post-test. Further, there were no improvements across inhibition training sessions, and there were no group differences in changes in inhibition performance from pre- to post-test.

At first glance our findings might seem to diverge from Johann and Karbach's (2020), who also reported benefits of training switching and inhibition. However, different conclusions seem to have mainly resulted from how the specific outcome measures were interpreted. Johann and Karbach reported benefits of training inhibition because children responded faster and more often on Go trials. However, as in our study, they also did not find improvements in how often children successfully inhibited responses on NoGo trials (children actually responded *more* often on NoGo trials at posttest), which we consider to be the core indicator of inhibitory control. Overall, both studies suggest no benefits of training inhibition.

Similarly, Johann and Karbach report benefits of training switching because they observed that switching costs decreased in the switching training groups. However, as in our study, there were no significant differences compared to the passive control group. This aligns with our findings and raises the question whether multiple groups slightly improve on switching tasks (i.e., they become faster at posttest), but that there may be no EF-specific training benefits beyond mere learning, re-test effects, and generally faster processing.

Taken together, findings of our and Johann and Karbach's study dovetail with previous findings, where significant performance benefits of training updating have more consistently been shown than of training switching or inhibition (e.g., Kassai et al., 2019; Rapport et al., 2013; Takacs & Kassai, 2019). Importantly, the current results confirm this pattern in one systematic overall randomized controlled trial that applied comparable single-task training regimes to these three EF components.

4.2. Does training on one EF task improve performance on the others?

Regarding transfer effects to untrained EF tasks, our findings show that training either on updating, switching, or inhibition tasks does not improve performance on the others. Specifically, LCSM show that the different training groups do not predict the magnitude of change in other EFs and that latent changes in the three EFs are largely independent (i.e., do not correlate). These findings are in line with Johann and Karbach (2020) and several other studies (for reviews, see e.g., Diamond & Ling, 2016; Diamond & Ling, 2020). Johann and Karbach (2020) argued that the lack of transfer effects may be due to the fact that they applied a multi-task training (i.e., trainings consisted of multiple tasks of the same EF component). They suggest that variability in tasks may hinder transfer effects in children and that transfer would be more likely to occur if only one type of tasks was used during the training. Our study provides additional insights in this regard, as we applied a single-task training for each EF component, but – in contrast to these suggestions – also did not observe any transfer effects. Indeed, the few studies that found transfer effects have applied both single- and multi-task trainings (e.g., Klingberg et al., 2005; Klingberg et al., 2002; Kray et al., 2012). We argue that the uniformity versus variability in tasks does not seem to be the main driving mechanism for between-EF transfer effects.

Interestingly, studies that have reported transfer to untrained EFs were largely conducted on atypically developing populations (e.g., children with ADHD; Klingberg et al., 2005; Klingberg et al., 2002; Kray et al., 2012). Thus, transfer effects between EFs may depend on specific cognitive characteristics of the target population rather than on the training-task design. Taking together the systematic evidence of our study and of Johann and Karbach (2020) as well as meta-analytical evidence of Kassai et al. (2019), there is currently no evidence of transfer effects between the three EFs in typically developing children.

4.3. Predictors of training benefits and theoretical implications

When baseline performance, age, gender, and training group were considered simultaneously, they predicted substantial 30%, 47%, and 33% of variance in updating, switching, and inhibition changes. Disentangling the specific role of each predictor while accounting for the others, results show that children with lower baseline performance displayed larger performance improvements on all three EF tasks, whereas older children showed larger improvements in updating and in switching tasks, but there was no effect of age on inhibition performance. There was no effect of gender on change for any of the three EF tasks.

As older children typically display better baseline performance, opposing effects of age and baseline performance may have cancelled out or blurred findings in previous studies that applied more classical analyses examining one predictor at a time. Our findings thereby illustrate an important conceptual implication of directly contrasting multiple predictors within one latent change model (Karbach et al., 2017) and suggest two seemingly opposite, yet complementary mechanisms that drive training effects in children. On the one hand, our findings align with previous studies reporting larger benefits for those that have the most room for improvement (see Karbach et al., 2017; Smid et al., 2020; Strobach & Karbach, 2021; Traut et al., 2021). From a theoretical perspective, we thereby provide more systematic evidence for compensation (rather than magnification accounts) of process-based cognitive trainings.

On the other hand, our findings are in contrast with previous studies reporting larger benefits for younger children or no effects of age (e.g., Cao et al., 2020; Cepeda et al., 2001; Kassai et al., 2019; Scionti et al., 2020; Wass et al., 2012), and suggest that, after accounting for baseline performance, training benefits are larger in older children. This aligns with previous research showing that the neural underpinnings of EFs – most importantly, the prefrontal cortex – continue to undergo structural and synaptic changes in late childhood and subsequent developmental stages (Best & Miller, 2010; Davidson et al., 2006; Diamond, 2013). Similarly, EF continues to develop and become more distinct across childhood (Karr et al., 2018) and certain components (e.g., switching) may only fully mature later in life (Garon et al., 2008; Müller & Kerns, 2015). Relatedly, previous research also shows that other key cognitive abilities – such as metacognitive skills – develop incrementally with schooling and are more developed by the end of middle childhood (Schneider & Lockl, 2008; Schneider & Löffler, 2016). Increases in metacognition facilitate learning across different school subjects (e.g., Dimmitt & McCormick, 2012; McCormick; Schneider, 2008; Smortchkova & Shea, 2020) and it is possible that they also bolster training benefits in older children. From a developmental perspective, our findings thereby suggest that training benefits are maximized when children's cognitive abilities are malleable, the underlying cognitive and neural systems are sufficiently developed, and the training occurs during an appropriate developmental stage that favors improvements.

4.4. Are training effects interpreted differently when compared to active versus passive controls?

To control for potential benefits of engaging in cognitively stimulating activities, participant's expectations, and other placebo effects, including an active control group has become the gold standard in cognitive training research. Yet, including active controls is also more resource- and time-consuming compared to passive controls, and it currently remains debated whether the type of control group actually affects results or the interpretation of training efficacy (e.g., Au et al., 2020). So far, this issue has mostly been examined with meta-analytical approaches between studies, with those focusing on updating reporting that benefits seem larger when comparing training effects to passive controls (Melby-Lervåg & Hulme, 2013; Sala & Gobet, 2017; Schwaighofer et al., 2015), whereas meta-analyses targeting multiple EFs did not find differences between the two control types (Au et al., 2020; Scionti et al., 2020).

With the present study, we provide the first directly comparable within-study evidence that aligns with latter meta-analyses, demonstrating similar patterns of results between active and passive control groups (Au et al., 2020; Scionti et al., 2020). Although active controls present many methodological advantages, such findings can be highly relevant for the efficient allocation of resources and time in future training studies. They suggest that interpretation of training benefits may not differ when training groups are compared to a passive control group versus to an active control group that participated in a cognitively rather low-demanding activity. Depending on the specific study focus, for many studies it may therefore be more ethical either to compare cognitive trainings to passive controls that are on waiting lists and can participate in the training at a later time; or to compare trainings to active control interventions that are more engaging and may better promote children's development while allowing to disentangle specific effects of the different interventions.

4.5. Implications of the present findings

Looking at the important number of cognitive training studies that have been published over the last decades suggests that, in general, researchers have been rather optimistic that these interventions benefit task performance and children's development in a broader context. Yet, more recent literature has questioned the efficacy of how EFs currently are assessed and trained (e.g., Doebel, 2020; Perone et al., 2021). Overall, the present data aligns with such skepticism, showing that when performance on three EF components was trained under very similar and comparable conditions, only one training (i.e., updating) led to specific improvements. Importantly, even this training did not improve performance on supposedly related EF tasks (i.e., switching and inhibition), which aligns with current research showing little evidence for far transfer of EF task training (e.g., Diamond & Ling, 2016; Kassai et al., 2019; Perone et al., 2021). Thus, it is questionable how and why such training approaches should lead to broader improvements and generalize to everyday relevant outcomes, such as school achievement, behavioral regulation, attentional control, and more. Although our study does not allow to conclude whether task-specific improvements would persist over time or affect any everyday outcomes, current findings strongly dampen the enthusiasm towards repeated single-task cognitive training interventions and urge for new, more efficient, and more naturalistic approaches.

This is important, because for children, training interventions are typically carried out in school or childcare. Therefore, they take away from crucial time spend on the educational curriculum and other valuable activities that may benefit their development, such as engaging in physical exercise, artistic activities, mindfulness (for meta-analyses, see Takacs & Kassai, 2019), or in programs providing new strategies to bolster self-regulation, social and other skills (e.g., McClelland & Tominey, 2015; Petersen, 1995). This seems even more relevant for atypically developing or at-risk populations, which may need support the most and, unfortunately, can least afford to spend time and resources on interventions that currently lack systematic evidence to promote children's development. Taken together, we argue that it is crucial for future research to thoroughly examine if and how different intervention approaches can lead to broader, long-lasting improvements in children's everyday outcomes.

In this context, Doebel (2020) and Perone et al. (2021) present inspiring new conceptual models of EF and how its development might be fostered. They question the validity of the current modular conceptualization of EF and whether it is useful to train performance on these modules. Instead, they suggest, future interventions should target children's specific goals by considering children's prior knowledge, beliefs, values, and more, as a dynamic ensemble that allows for momentary behavior to unfold. For example, if a child should learn not to hit another child that took their toy, modular task training (e.g., of inhibition) may be rather inefficient. Instead, it may be more useful to build on the child's previous experiences, such as expecting that hitting will lead to punishment, having experienced how it feels to be hit by someone, preferring to maintain the friendship with the other child, and more (see Doebel et al., 2020). Similarly, providing children with contextual, multi-level information may help them to link specific goals to environmental cues (e.g., asking how the child feels at the moment, explaining why the other child may have taken the toy, providing a context that allows for the conflict to be solved; see Perone et al., 2021). Together, by strengthening the association between goals, cues, and contextual information rather than training modular task performance, more comprehensive goal-oriented interventions may be better suited for helping children to learn and reproduce the target behavior and thereby ultimately bolster development in real-life contexts.

4.6. Limitations and outlook

Although the present study provides important systematic insights on the (in)efficacy of EF trainings and thereby may guide future research towards more integrative and more ecological assessment and training of EF, it also is important to highlight its limitations. A first, methodological shortcoming is that the present study did not include any follow-up nor ecological measures of EF. Besides examining whether training benefits lasted across time, a follow-up would further allow evaluating training-specific versus general effects of the interventions. Yet, as we only found training-specific performance improvements at posttest for one EF task (i.e., updating), we are skeptical that this or similar training approaches would benefit children across longer periods of time nor that they would improve performance on more ecological EF measures or in real-life contexts (see Diamond & Ling, 2016; Kassai et al., 2019; Perone et al., 2021).

Another limitation is that baseline performance levels varied between EFs. For example, updating and switching performance were rather low. Although this may partially be due to our choice of outcome measures, children's performance may also have been affected by the rather low number of trials to assess each EF. In contrast, baseline performance was rather high for inhibition. Although it was sufficiently low at the first training session to leave room for potential improvements, high baseline performance may be particularly challenging for training studies, as supported by our finding that training benefits were largest when baseline performance was lowest.

An important conceptual limitation of the present as well as other training studies is that they typically do not allow us to conclude *why* one EF may show larger performance benefits. It could be that certain EF simply are more trainable than others and allow to increase *existing* resources. However, it could also be that because of how EFs are typically assessed, it may be easier to add *new* resources when performing certain tasks, (e.g., discovering strategies such as rehearsing spatial locations of previous stimuli before onset of the next stimulus in updating tasks) whereas this may be more difficult for other tasks (e.g., inhibiting the impulse to respond after stimulus onset in inhibition tasks).

A final limitation is that correlations between pre- and posttest measures of the same EF were rather low. One possible explanation for this may be that, whereas all children were relatively comparable at pre-test, at post-test children had participated in different interventions, and thus may approach certain tasks differently. This also suggests that, besides training-specific effects, other mechanisms – such as task-novelty, familiarity, motivation, or fatigue – may have affected posttest performance, which further illustrates issues with current approaches to assess and train EFs.

4.6. Conclusion

Our study represents the first systematic and comprehensive examination that directly compared the trainability of the three core EF components across the entire primary school age range using latent change score modelling. It demonstrates that when performance on an updating, switching, or inhibition task are trained under similar conditions within a single group of children, only updating performance showed training-specific improvements. In terms of potential transfer effects, it underlines that there is currently no systematic evidence for transfer of improvements to non-trained EF tasks. Such findings question how likely it is that classical task-paradigm trainings can lead to improvements in even broader outcomes of EF in everyday contexts. Further, taken together with other studies, findings illustrate that the current conceptualization of EF and how it develops is still incomplete (for similar views, see Doebel, 2020; Perone et al., 2021). A better understanding of how EF should be assessed and conceptualized throughout childhood is necessary before future research will be able to explore new interventions supporting children's real-life behaviors that rely on EF.

In terms of predictors of improvements, our results show that performance improvements are largest for children that are older but have still have relatively low EF performance. This provides evidence for opposite, yet complementary mechanisms of baseline performance and age, thereby supporting both compensation accounts as well as developmental theories of EF and underlining the importance of accounting for multiple predictors simultaneously. Finally, our findings show that the efficacy of a cognitive training is similar when comparing training groups to active or to passive controls. In certain situations, it may therefore be more ethical to use passive controls or other interventions that are more likely to benefit children's development than the typical active control paradigms.

Declaration of interest

The authors report no conflict of interest.

Data availability statement

The datasets generated and analyzed for the current study are not publicly available because participant consent forms did not include authorization for public data sharing. However, data will be made available by the corresponding author [SZ] upon reasonable request.

Acknowledgements

The authors thank Laura Abdili, Niloufar Anjileli, Chiara Bernasconi, Giuliano Bizon, Marie-France Bonny, Morgane Budry, Marie Bugnon, Solange Gautier, Maëlle Gousset, Ali Hedhili, Lea Maréchal, Elsa Martinelli, Cécile Mueller, Mélissa Pegat-Toquet, Margaux Pimont, Félix Rambicur, Mélanie Silva Damil, Lysiane Sublet and Monika Toellner for their assistance in study preparation and data collection, and Adeline Moulin for her assistance with literature search. The authors acknowledge funding from the Swiss National Science Foundation (SNSF; grant numbers: 100014_152841 & P400PS_199283) and the Swiss National Centre of Competence in Research LIVES – Overcoming vulnerability: Life course perspectives (NCCR LIVES; grant number: 51NF40-185901).

References

Au, J., Gibson, B. C., Bunarjo, K., Buschkuehl, M., & Jaeggi, S. M. (2020). Quantifying the Difference Between Active and Passive Control Groups in Cognitive Interventions Using Two Meta-analytical Approaches. *Journal of Cognitive Enhancement*, *4*(2), 192-210. https://doi.org/10.1007/s41465-020-00164-6

Best, J. R., & Miller, P. H. (2010). A Developmental Perspective on Executive Function. *Child Development*, *81*(6), 1641-1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327-336. https://doi.org/10.1016/j.lindif.2011.01.007

Cao, Y., Huang, T., Huang, J., Xie, X., & Wang, Y. (2020). Effects and Moderators of Computer-Based Training on Children's Executive Functions: A Systematic Review and Meta-Analysis. *Frontiers in Psychology*, *11*(2937). https://doi.org/10.3389/fpsyg.2020.580329

Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, *37*(5), 715-730. https://doi.org/10.1037/0012-1649.37.5.715

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia*, *44*(11), 2037-2078. https://doi.org/10.1016/j.neuropsychologia.2006.02.006

Dawson, P., & Guare, R. (2018). *Executive skills in children and adolescents: A practical guide to assessment and intervention*. Guilford.

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), e12898. https://doi.org/10.1111/desc.12898

Denham, S. A., Bassett, H. H., Sirotkin, Y. S., Brown, C., & Morris, C. S. (2015). "No-o-o-o peeking": Preschoolers' executive control, social competence, and classroom adjustment. *Journal of Research in Childhood Education, 29*(2), 212-225. https://doi.org/10.1080/02568543.2015.1008659

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

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research.* (pp. 11-43). American Psychological Association. https://doi.org/10.1037/14797-002

Diamond, A., & Ling, D. S. (2016). Conclusions about interventions, programs, and approaches for improving executive functions that appear justified and those that, despite much hype, do not. *Developmental Cognitive Neuroscience, 18,* 34-48. https://doi.org/https://doi.org/10.1016/j.dcn.2015.11.005

Diamond, A., & Ling, D. S. (2020). Review of the Evidence on, and Fundamental Questions About, Efforts to Improve Executive Functions, Including Working Memory. In J. M. Novick, M. F. Bunting, M. R. Dougherty, & R. W. Engle (Eds.), *Cognitive and Working Memory Training: Perspectives From Psychology, Neuroscience, and Human Development*. Oxford University Press. https://doi.org/10.1093/oso/9780199974467.003.0008

Dimmitt, C., & McCormick, C. B. (2012). Metacognition in education. In *APA educational psychology handbook, Vol 1: Theories, constructs, and critical issues.* (pp. 157-187). American Psychological Association. https://doi.org/10.1037/13273-007

Doebel, S. (2020). Rethinking Executive Function and Its Development. *Perspectives on Psychological Science*, *15*(4), 942-956. https://doi.org/10.1177/1745691620904771

Foster, J. L., Harrison, T. L., Hicks, K. L., Draheim, C., Redick, T. S., & Engle, R. W. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1677-1689. https://doi.org/10.1037/xlm0000426

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31-60. https://doi.org/10.1037/0033-2909.134.1.31

Hu, L. T., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling-a Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017-2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(25), 10081-10086. https://doi.org/10.1073/pnas.1103228108

Johann, V. E., & Karbach, J. (2020). Effects of game-based and standard executive control training on cognitive and academic abilities in elementary school children.

Developmental Science, *23*(4), e12866. https://doi.org/https://doi.org/10.1111/desc.12866

Karbach, J., Könen, T., & Spengler, M. (2017). Who Benefits the Most? Individual Differences in the Transfer of Executive Control Training Across the Lifespan. *Journal of Cognitive Enhancement*, *1*(4), 394-405. https://doi.org/10.1007/s41465-017-0054-z

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psycholical Bulletin*, *144*(11), 1147-1185. https://doi.org/10.1037/bul0000160

Kassai, R., Futo, J., Demetrovics, Z., & Takacs, Z. K. (2019). A meta-analysis of the experimental evidence on the near- and far-transfer effects among children's executive function skills. *Psychological Bulletin*, *145*(2), 165-188. https://doi.org/10.1037/bul0000180

Katz, B., Jones, M. R., Shah, P., Buschkuehl, M., & Jaeggi, S. M. (2021). Individual Differences in Cognitive Training Research. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 107-123). Springer International Publishing. https://doi.org/10.1007/978-3-030-39292-5_8

Kliegel, M., Hering, A., Ihle, A., & Zuber, S. (2017). Cognitive Training to Promote Executive Functions. In S. A. Wiebe & J. Karbach (Eds.), *Executive Function: Development Across the Life Span* (pp. 200-213). Taylor & Francis. https://doi.org/10.4324/9781315160719

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*(7), 317-324. https://doi.org/10.1016/j.tics.2010.05.002

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., Gillberg, C. G., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD - A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 177-186. https://doi.org/10.1097/00004583-200502000-00010

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*(6), 781-791. https://doi.org/10.1076/Jcen.24.6.781.8395

Kray, J., Karbach, J., Haenig, S., & Freitag, C. (2012). Can task-switching training enhance executive control functioning in children with attention deficit/-hyperactivity disorder? *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00180

Liew, J. (2012). Effortful Control, Executive Functions, and Education: Bringing Self-Regulatory and Social-Emotional Competencies to the Table. *Child Development Perspectives*, *6*(2), 105-111. https://doi.org/https://doi.org/10.1111/j.1750-8606.2011.00196.x

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198-1202. https://doi.org/10.1080/01621459.1988.10478722

Lövdén, M., Brehmer, Y., Li, S.-C., & Lindenberger, U. (2012). Training-induced compensation versus magnification of individual differences in memory performance. *Frontiers in Human Neuroscience*, *6*(141). https://doi.org/10.3389/fnhum.2012.00141

Martinovic, D., Burgess, G. H., Pomerleau, C. M., & Marin, C. (2016). Computer games that exercise cognitive skills: What makes them engaging for children? *Computers in Human Behavior*, *60*, 451-462. https://doi.org/10.1016/j.chb.2016.02.063

McClelland, M. M., & Tominey, S. L. (2015). *Stop, Think, Act - Integrating Self-Regulation in the Early Childhood Classroom*. Routledge. https://doi.org/10.4324/9781315798059

McCormick, C. B. Metacognition and Learning. In I. B. Weiner (Ed.), *Handbook of Psychology* (pp. 79-102). https://doi.org/10.1002/0471264385.wei0705

Melby-Lervåg, M., & Hulme, C. (2013). Is Working Memory Training Effective? A Meta-Analytic Review. *Developmental Psychology*, *49*(2), 270-291. https://doi.org/10.1037/A0028228

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

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current directions in psychological science*, *21*(1), 8-14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

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. https://doi.org/10.3758/S13423-010-0034-0

Müller, U., & Kerns, K. (2015). The development of executive function. In *Handbook of child psychology and developmental science: Cognitive processes* (7th ed., Vol. 2, pp. 571-623). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118963418.childpsy214

Perone, S., Simmering, V., R., & Buss, A. T. (2021). A Dynamical Reconceptualization of Executive-Function Development. *Perspectives on Psychological Science*, *16*(6), 1198-1208. https://doi.org/10.1177/1745691620966792

Petersen, L. (1995). Stop Think Do. In H. P. J. G. van Bilsen, P. C. Kendall, & J. H. Slavenburg (Eds.), *Behavioral Approaches for Children and Adolescents: Challenges for the Next Century* (pp. 103-111). Springer US. https://doi.org/10.1007/978-1-4757-9406-9_9

Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review*, *33*(8), 1237-1252. https://doi.org/10.1016/j.cpr.2013.08.005

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of applied developmental psychology*, *27*(4), 300-309. https://doi.org/10.1016/j.appdev.2006.04.002

Sala, G., & Gobet, F. (2017). Working memory training in typically developing children: A meta-analysis of the available evidence. *Developmental Psychology*, *53*(4), 671-685. https://doi.org/10.1037/dev0000265

Sala, G., & Gobet, F. (2020). Working memory training in typically developing children: A multilevel meta-analysis. *Psychonomic Bulletin & Review*, *27*(3), 423-434. https://doi.org/10.3758/s13423-019-01681-y

Schneider, W. (2008). The Development of Metacognitive Knowledge in Children and Adolescents: Major Trends and Implications for Education. *Mind, Brain, and Education*, *2*(3), 114-121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schneider, W., & Lockl, K. (2008). Procedural metacognition in children: evidence for developmental trends. In *Handbook of metamemory and memory*. (pp. 391-409). Psychology Press.

Schneider, W., & Löffler, E. (2016). The development of metacognitive knowledge in children and adolescents. In J. Dunosky & S. K. Tauber (Eds.), *Oxford handbook of metamemory*. Oxford University Press.

Schulz, K. P., Fan, J., Magidina, O., Marks, D. J., Hahn, B., & Halperin, J. M. (2007). Does the emotional go/no-go task really measure behavioral inhibition? Convergence with measures on a non-emotional analog. *Archives of Clinical Neuropsychology*, *22*(2), 151-160. https://doi.org/10.1016/j.acn.2006.12.001

Schwaighofer, M., Fischer, F., & Buhner, M. (2015). Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators [Article]. *Educational Psychologist*, *50*(2), 138-166. https://doi.org/10.1080/00461520.2015.1036274

Scionti, N., Cavallero, M., Zogmaister, C., & Marzocchi, G. M. (2020). Is Cognitive Training Effective for Improving Executive Functions in Preschoolers? A Systematic Review and Meta-Analysis. *Frontiers in Psychology*, *10*(2812). https://doi.org/10.3389/fpsyg.2019.02812

Smid, C. R., Karbach, J., & Steinbeis, N. (2020). Toward a Science of Effective Cognitive Training. *Current Directions in Psychological Science*, *29*(6), 531-537. https://doi.org/10.1177/0963721420951599

Smortchkova, J., & Shea, N. (2020). Metacognitive Development and Conceptual Change in Children. *Review of Philosophy and Psychology*, *11*(4), 745-763. https://doi.org/10.1007/s13164-020-00477-7

Strobach, T., & Karbach, J. (2021). *Cognitive Training: An Overview of Features and Applications* (2nd ed.). Springer. https://doi.org/https://doi.org/10.1007/978-3-030-39292-5

Swanson, H. L. (2014). Does cognitive strategy training on word problems compensate for working memory capacity in children with math difficulties? *Journal of Educational Psychology*, *106*(3), 831-848. https://doi.org/10.1037/a0035838

Swanson, H. L. (2015). Cognitive strategy interventions improve word problem solving and working memory in children with math disabilities. *Frontiers in Psychology*, *6*(1099). https://doi.org/10.3389/fpsyg.2015.01099

Takacs, Z. K., & Kassai, R. (2019). The efficacy of different interventions to foster children's executive function skills: A series of meta-analyses. *Psychological Bulletin*, *145*(7), 653-697. https://doi.org/10.1037/bul0000195

Traut, H. J., Guild, R. M., & Munakata, Y. (2021). Why Does Cognitive Training Yield Inconsistent Benefits? A Meta-Analysis of Individual Differences in Baseline Cognitive Abilities and Training Outcomes. *Frontiers in Psychology*, *12*(1697). https://doi.org/10.3389/fpsyg.2021.662139

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/https://doi.org/10.1016/j.dr.2012.07.001

Wechsler, D. (2004). *The Wechsler intelligence scale for children—fourth edition*. Pearson Assessment.

Zelazo, P. D., Carlson, S. M., & Kesek, A. (2008). Development of executive function in childhood. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (Vol. 2, pp. 553-574). MIT Press.

Supplementary Material

Supplementary material S1: more detailed description of training programs

To allow for a systematic comparison of the three EF training groups and the active control group, all four training programs were very similar in terms of design, structure, and intensity. Specifically, each of the four programs consisted of eight training sessions, which lasted for approximately 20-25 minutes each. Each training session was programmed so that it had ten possible difficulty levels, with each level increasing in task difficulty (see detailed description of the respective training tasks below). For each training session, children worked on ten training blocks (independently of the difficulty level of each block). Each training block consisted of a fixed number of trials of the training task. In order to maximize potential training benefits, the difficulty of each block was individually adapted to children's performance: At the end of each block, children would either increase by one difficulty level for the next block (if their performance was above a certain cutoff value), decrease by one difficulty level for the next block (if their performance was below a certain cutoff value), or perform the next block on the same difficulty level (if their performance was between those two cutoffs; specific cutoffs of each task are detailed below). To further encourage children, keep them engaged, and maintain their motivation throughout the program, the difficulty level of the first training block of each training sessions increased across sessions: children started at the first difficulty level on the first training session, but automatically increased the starting level every two sessions (e.g., they started at the second difficulty level on the third and fourth training sessions, etc.). For the same reasons, game-like aspects were incorporated in all trainings (e.g., using cartoon characters as stimuli, using vivid colors, providing feedback with happy/sad smiley faces and information on how well children performed).

Updating training

As for the pre- and posttest tasks, the updating training tasks consisted of a spatial 2-back paradigm, for which children had to indicate for each trial whether a cartoon character was displayed in the same location as two trials before by pushing the green button (right arrow key) for hits and the red button (left arrow key) for non-hits. The first two difficulty levels were similar to the pre-posttest task, so that the cartoon character appeared in one of six boxes in a 3x2 grid. In levels three and four, the grid was replaced by a video-game like background map, with six circles outlining the locations where the cartoon character could appear. Levels five and six had a different background map and one more circle (i.e., seven total locations); levels seven and eight again had a different background map and an additional circle (i.e., eight total locations); levels nine and then had yet another background map and one more circle (i.e., nine total locations). Thus, difficulty was increased by adding more complex stimuli (backgrounds) and more spatial locations. Further, the duration of the fixation cross presented before each trial was of 950ms on level one and decreased by 50ms on each difficulty level. Similarly, the stimulus presentation was 3000ms on the first level, (followed by a 1000ms blank response window, if children did not already answer during the stimulus presentation), which decreased by 250ms on each level (the duration of the blank response window was kept constant). Difficulty was individually adapted to children's performance, so that they could advance to the next level if accuracy on all trials of one block was above .85, decrease by one difficulty level if their accuracy was below .60, and repeat the same difficulty level if their accuracy was between two cutoffs. Each training session consisted of 170 trials (50 hit trials), for a total of 1360 trials across the entire training (400 hit trials).

Switching training

As for the pre- and posttest tasks, the switching training tasks consisted of a task A/task B switching paradigm. For each trial, children were presented cartoon-like images of either one or two car(s) or plane(s), and their task was to indicate whether the image depicted a car or a plane (independently of the number of objects; task A), or one or two objects (independently of the type of object; task B) by pushing the left and right arrow keys. The paradigm switched between tasks on every third trial (AABBAA…) and an icon above the central stimulus indicate whether the number or the vehicle task had to be performed. To increase switching difficulty across levels, the stimulus presentation was 3000ms on the first level, (followed by a 2000ms blank response window, if children did not already answer during the stimulus presentation), which decreased by 250ms on each level (and the duration of the blank response window increased by 250ms). Similarly, the duration of the fixation cross presented before each trial was 950ms on level one and decreased by 50ms on each difficulty level. Difficulty was individually adapted to children's performance, so that they could advance to the next level if their accuracy on all trials of one block was above .80, decrease by one difficulty level if accuracy was below .60, and repeat the same difficulty level if their accuracy was between two cutoffs. Each training session consisted of 410 trials (200 switching trials), for a total of 3280 trials across the entire training (1600 switching trials).

Inhibition training

As for the pre- and posttest tasks, the inhibition training tasks consisted of a Go-NoGo response inhibition paradigm. For each trial, children were instructed to push the spacebar as fast as possible as soon as a cartoon-like image appeared on the screen (Go-trials) except for when the image corresponded to a specific category (NoGo-trials). To keep children engaged but also to increase difficulty throughout the training program, the images changed throughout the sessions and detecting the NoGo-images became more difficult (e.g., houses among images of objects in an early session versus different types of tools among other household utilities in a later session). Further, to increase difficulty across levels within each session, the duration of the stimulus presentation was 1000ms on the first level, (followed by a 3000ms blank response window, if children did not already answer during the stimulus presentation), which decreased by 50ms on each level (and the duration of the blank response window decreased by 100ms). Similarly, the duration of the fixation cross presented before each trial was between 1300ms and 1700ms on level one and the range decreased by 100ms on each difficulty level. Difficulty was individually adapted to children's performance, so that they could advance to the next level if their accuracy on all trials of one block was above .80, decrease by one difficulty level if accuracy was below .60, and repeat the same difficulty level if their accuracy was between two cutoffs. Each training session consisted of 320 trials (80 NoGo trials), for a total of 2560 trials across the entire training (640 NoGo trials).

Active control

For the active control training, children had to categorize images (similar to the categorization tasks of the switching paradigm, but without having to switch between different tasks). Specifically, each training session consisted of ten blocks for which participants had to indicate whether stimuli belonged to one of two categories (e.g., cars versus planes) by pushing the left/right arrow keys. Difficulty was individually adapted to children's performance, so that they could advance to the next level if their accuracy on all trials of one block was above .80, decrease by one difficulty level if accuracy was below .60, and repeat the same difficulty level if their accuracy was between two cutoffs. Task difficulty was adapted as in the other training paradigm (i.e., faster presentation of fixation cross and stimulus with increasing difficulty levels). To further engage and motivate children, categories changed between sessions. Overall, task structure and adaptive difficulty therefore was very similar to the other trainings, but performing the task did not significantly tapping into EF. Each training session consisted of 410 trials, for a total of 3280 trials across the entire training.

Supplementary material S2: Parameter estimates of Model 3

Regression weights, standard error of estimation, and p-values for predictors of latent changes in EFs in Model 3.

	Predicting change in updating				Predicting change in switching				Predicting change in inhibition			
Predictors	*b*	β	S.E.	*p*	*b*	β	S.E.	*p*	*b*	β	S.E.	*p*
Updating pretest	**-.52**	**-.46**	.07	< .001	-	-	-	-	-	-	-	-
Switching pretest	-	-	-	-	**-.73**	**-.70**	.05	< .001	-	-	-	-
Inhibition pretest	-	-	-	-	-	-	-	-	**-.66**	**-.55**	.07	<.001
Age	**.03**	**.13**	.01	.03	**-.04**	**-.17**	.01	< .001	-.01	-.06	.01	.33
Gender	.01	.02	.03	.73	-.02	-.03	.03	.54	.01	.03	.02	.65
Updating group	**.19**	**.34**	.05	< .001	.05	.06	.05	.35	-.01	-.03	.03	.64
Switching group	-.01	-.01	.05	.92	-.08	-.10	.05	.10	.01	.01	.03	.93
Inhibition group	-.02	-.09	.05	.20	.04	.04	.05	.49	-.01	-.01	.03	.95
Active control group	-.04	-.06	.05	.40	-.01	-.02	.05	.78	.03	.07	.03	.31
R^2	.30				.47				.33			

Note. b = raw regression weights. β = standardized regression weights. S.E. = standard error of estimation. Updating pretest = proportion of correctly detected hits minus the proportion of false alarms on non-hit trials; switching pre-test = switching cost in seconds (= mean reaction time on shift trials minus mean reaction time on non-shift trials on trials with correct responses only); inhibition scores = inhibition accuracy (proportion of correctly inhibited NoGo trials); age = children's age in years; gender = coded *0* for girls and *1* for boys. To code for group, four dummy variables were computed, with the passive control group as a reference. R^2 = Squared multiple correlation. Significant estimates in **bold**.