Received: 25 February 2019 | Revised: 30 September 2019 | Accepted: 4 October 2019

DOI: 10.1002/icd.2163

WILEY

RESEARCH ARTICLE

# The relation between children's aerobic fitness and executive functions: A systematic review

Hilde Van Waelvelde[1] | Karsten Vanden Wyngaert[1] | Tineke Mariën[1] | Dieter Baeyens[2] | Patrick Calders[1]

[1]Faculty of Medicine and Health Science, Department of Rehabilitation Sciences, Gent University, Gent, Belgium

[2]Faculty of Psychology and Educational Sciences Leuven, KU Leuven, Leuven, Belgium

**Correspondence**
Hilde Van Waelvelde, Faculty of Medicine and Health Science, Department of Rehabilitation Sciences, Gent University, Gent, Belgium.
Email: hilde.vanwaelvelde@ugent.be

## Abstract

A beneficial effect of physical activity on cognitive functioning is supposed, although to a certain extent, literature remains inconsistent. Furthermore, the mediating effect of aerobic fitness on this association remains unclear, especially in children. This review presents data from 26 articles on the relation between aerobic fitness and executive functions (i.e., inhibition, working memory, and cognitive flexibility) in prepubertal children (6–12 years). The qualitative synthesis was complicated by the use of diverse outcome measures to evaluate executive functions and the inconsistent inclusion of confounders. In general, cross-sectional studies reported enhanced executive functioning in children with higher aerobic fitness levels. Only three intervention studies with inconclusive results were retrieved. Whereas a beneficial effect of aerobic fitness on executive functioning in children is often supposed, the available research does not allow to draw a causal conclusion. Good quality randomized-controlled trials are necessary.

**Highlights:**

- This study reviewed the supposed mediating effect of aerobic fitness on the positive association between physical activity and executive functioning.
- Three intervention studies and 23 cross-sectional studies confirmed a positive correlation between aerobic fitness and certain measures of executive functioning.

 © 2019 John Wiley & Sons, Ltd. 

- A causal conclusion about the beneficial effect of aerobic fitness on executive functioning cannot be drawn from this review.

**KEYWORDS**

aerobic fitness, children, cognition, executive function, physical activity

## 1 | INTRODUCTION

The available evidence supports the overall conclusion that physical activity (PA) provides fundamental health benefits for children and youth (World Health Organization, 2019). The supposed association between a fit body and a fit brain in children only enhances the relevance of children's PA programmes. However, although a positive association between PA and cognitive performance in children is considered to be established, the available literature remains inconsistent (Best, 2010; Bidzan-Bluma & Lipowska, 2018; Diamond & Ling, 2018; Donnelly et al., 2016; Vandenbroucke, Seghers, Verschueren, Wijtzes, & Baeyens, 2016). In particular, the underlying mechanisms are still debated. There are several pathways by which PA may facilitate cognitive function in children: (a) the cognitive demands inherent in the structure of goal-directed PA, (b) the cognitive engagement required to execute complex movements, and (c) the physiological changes in the brain induced by aerobic exercise. To disentangle this association between body and brain, a clear framework with well-defined concepts is mandatory.

Physical fitness (PF), physical exercise, and PA are sometimes inappropriately used as interchangeable concepts. Regarding the World Health Organization (2019), PA is defined as any bodily movement produced by skeletal muscles that requires energy expenditure. Physical exercise is a subcategory of PA that is planned, structured, repetitive, and aims to improve or maintain one or more components of PF. Often a battery of tests is used to assess components such as cardiovascular fitness, muscular strength, and endurance, sometimes body composition or degree of fatness, flexibility, agility, coordination, balance, and reaction time. Cardiorespiratory or aerobic fitness (AF) represents an intermediate variable between PA behaviours and health outcomes that reflects the capacity of numerous bodily organs, such as the heart, lungs, and muscles, to support energy production during PA and exercise. The maximal oxygen consumption ($VO_2$ max) attained during a graded maximal exercise test to voluntary exhaustion has long since been considered as the single best indicator of cardiorespiratory fitness (Lang et al., 2018).

Several reviews focused already on the effect of PA intervention on cognition (Donnelly et al., 2016; Fedewa & Ahn, 2011; Keeley & Fox, 2009; Khan & Hillman, 2014; Vazou, Pesce, Lakes, & Smiley-Oyen, 2016). AF has been considered as the main link between PA and cognitive functioning in several studies (e.g., Khan & Hillman, 2014). It is hypothesized that a positive effect of PA on cognitive functions is partly caused by physiological changes in the body such as increased levels of brain-derived neurotrophic factor that facilitates learning and maintains cognitive functions by improving synaptic plasticity and acting as a neuroprotective agent, increased brain circulation and improved neuroelectric functionality (Hillman, Buck, Themanson, Pontifex, & Castelli, 2009). However, the relation between PA, PF, and more specific AF, cognition and brain function is particularly complex in children due to the developing brain and body (Dencker, Bugge, Hermansen, & Andersen, 2010). When bore in mind the inconsistency of children's activity levels, increasing PA may not necessarily result in enhanced AF (Fedewa & Ahn, 2011). It has been argued that motor proficiency is a more important mediator compared with AF. Haapala (2013) reviewed the literature to differentiate between AF and motor skills in relation to cognition and suggests that compared with less fit children, highly fit children have larger subcortical brain structures, more efficient brain activation, and a better neuroelectric efficiency during cognitive tasks, superior inhibitory control, working memory, attention, and better academic performance. Similarly, children with good compared with poor motor skills show better inhibitory control, attention capacity, and academic performance. Thus, the evidence suggests that both AF and motor skills play an important role in cognitive development during childhood. A cognitive stimulation hypothesis has been suggested by Schmidt, Jager, Egger, Roebers, and Conzelmann (2015). It assumes that coordinative demanding PA, learning new skills or sports, affects cognition in primary school children, but not aerobic exercise (Schmidt et al., 2015). To disentangle the mediating roles of AF and motor skills, this review focuses exclusively on the link between AF and cognitive functioning.

On the other hand, the inconsistency in the use of different measures, terms, and definitions for cognition such as academic achievement, learning, memory, and executive functions (EF) contributes to the difficult relation between AF and cognition. EF are a set of well-defined, basic processes necessary for the cognitive control of behaviour and supposed to be the underlying skills for mental and physical health; success in school and in life; and cognitive, social, and psychological development. EF are organized along the following three distinguishable yet interrelated component functions: (a) inhibition, (b) working memory, and (c) cognitive flexibility (Diamond, 2013). Inhibition includes response control and interference control encompassing selective attention and cognitive inhibition. Working memory is referred to as the ability to abstain information in mind and mentally work with it. This information can be distinguished by content: verbal working memory and visual–spatial working memory. Third, cognitive flexibility is defined as the ability to "think outside the box" and being able to adjust to situational demands or priorities. It can be subdivided in fluency and set-shifting tasks. Fluency tasks require abilities to be flexible in order to see something from different perspectives. Set-shifting paradigms rely on the ability to quickly switch between tasks or assignments (Davidson, Amso, Anderson, & Diamond, 2006; Diamond, 2013; Friedman & Miyake, 2016). EF has been seen as regulated by the prefrontal regions of the frontal lobes. However, putative EF measures do not load on a single EF factor, and different frontal and non-frontal brain regions are related to the different domains of EF (Nowrangi, Lyketsos, Rao, & Munro, 2014). Therefore, the relation between AF and EF can be dependent of the domain of EF and even the specific task and outcome measure.

This systematic review focuses specifically on the relation between AF and the different domains of EF. Definitions and subdomains of EF were based on the framework described by Diamond (2013). Age range is limited to pre-adolescent children of 6–12 years old because EF are maturing from birth to adult age, and the relation between AF and EF varies depending on the age of the children (Davidson et al., 2006). Specifically, the onset of puberty entails important hormonal and cognitive changes that differentiate adolescents from prepubescent children (McGivern, Andersen, Byrd, Mutter, & Reilly, 2002). Intervention studies are ultimately required to establish a relation of causality. Cross-sectional studies only allow to demonstrate "a" relationship, although in correlational studies causality can be more strongly suggested by controlling other potential mediators. Therefore, we aimed to make an inventory of the investigated confounding factors in the association between AF and EF.

This review does not consider the acute effect of aerobic exercise on EF and the effect of AF on specific brain structures or functions.

Previous reviews already compiled the literature on the association between AF and the broader concept of cognition (Haapala, 2013) and between PA and cognition (Keeley & Fox, 2009). However, by narrowing the focus in this review to the association between AF and EF, this review will be able to offer important additional insights into the relationship between cognition and PA.

## 2 | METHODS

This systematic review respects the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009).

## 2.1 | Information sources and search strategy

The searches were indexed until November 8, 2018, on electronic databases PubMed, Web of Science, Embase, and SPORTDiscus. Table 1 presents the construct of the search strategies. Finally, reference lists of reviews and selected studies were hand-searched to make this search as comprehensive as possible.

## 2.2 | Study selection

In this systematic review, 1,253 articles were screened. Articles were eligible if they objectively assessed AF and EF in typically developing children between 6 and 12 years old. Reviews, meta-analyses, case report forms, and letters to the editor were excluded. Literature was searched and screened in duplicate by the first two authors independently. Initially, all search results were screened on title and abstract, followed by a screening on full text when considered eligible. The inclusion and exclusion criteria for this final selection were delineated as reported hereafter to the following: (a) Only intervention studies including a cardiovascular exercise programme without high motor demands or evaluating the role of AF as mediating factor were included. (b) Acute bouts of aerobic exercise were excluded. (c) Studies adopting EF tasks requiring more as one reported subcomponent of EF and studies analysing derivative measures of an EF task were excluded. This last criterion was implemented based on the irrelevancy of comparing these derivative measures with other study results. Disagreements regarding eligibility for inclusion were resolved via development of consensus among both reviewers and the last author. The selection process, depicted in Figure 1, resulted in 26 eligible articles.

## 2.3 | Quality assessment

The Downs and Black checklist (1998) was used to evaluate methodological quality of interventional and non-randomized comparative trials (Aubut, Marshall, Bayley, & Teasell, 2013). This 27-item checklist reports on reporting bias (10 items), external validity (three items), overall bias (seven items), confounders (six items), and power (one item).

**TABLE 1** Search strategy

| Database | Entry terms | Number of articles |
|---|---|---|
| PubMed | ((“Physical fitness”[All Fields] OR “Cardiorespiratory fitness” [All Fields] OR “aerobic fitness” [All Fields]) AND ((“Executive”[All Fields] AND “Function”[All Fields]) OR “executive function”[All Fields] OR “executive functions”[All Fields] OR “executive functioning”[All Fields] OR “executive control”[All Fields] OR “executive controls”[All Fields] OR “inhibition (psychology)”[MeSH] OR (“inhibition”[All Fields] AND “psychology”[All Fields]) OR “cognitive flexibility”[All Fields] OR “Working memory”[All Fields])) | 316 |
| Web of Science | TS = (((“Physical fitness” OR “Cardiorespiratory fitness” OR “aerobic fitness”) AND ((“Executive” AND “Function”) OR “executive function” OR “executive functions” OR “executive functioning” OR “executive control” OR “executive controls” OR “inhibition (psychology)” OR (“inhibition” AND “psychology”) OR “cognitive flexibility” OR “Working memory”))) | 301 |
| SPORTDiscus | ((“Physical fitness” OR “Cardiorespiratory fitness” OR “aerobic fitness”) AND ((“Executive” AND “Function”) OR “executive function” OR “executive functions” OR “executive functioning” OR “executive control” OR “executive controls” OR “inhibition (psychology)” OR (“inhibition” AND “psychology”) OR “cognitive flexibility” OR “Working memory”)) | 125 |
| Embase | : (“fitness”/exp OR “fitness”) AND (“executive function”/exp OR “executive function” OR “executive functioning”/exp OR “executive functioning” OR “executive control”/exp OR “executive control” OR “inhibition (psychology)”/exp OR “inhibition (psychology)” OR “cognitive flexibility”/exp OR “cognitive flexibility”) | 511 |

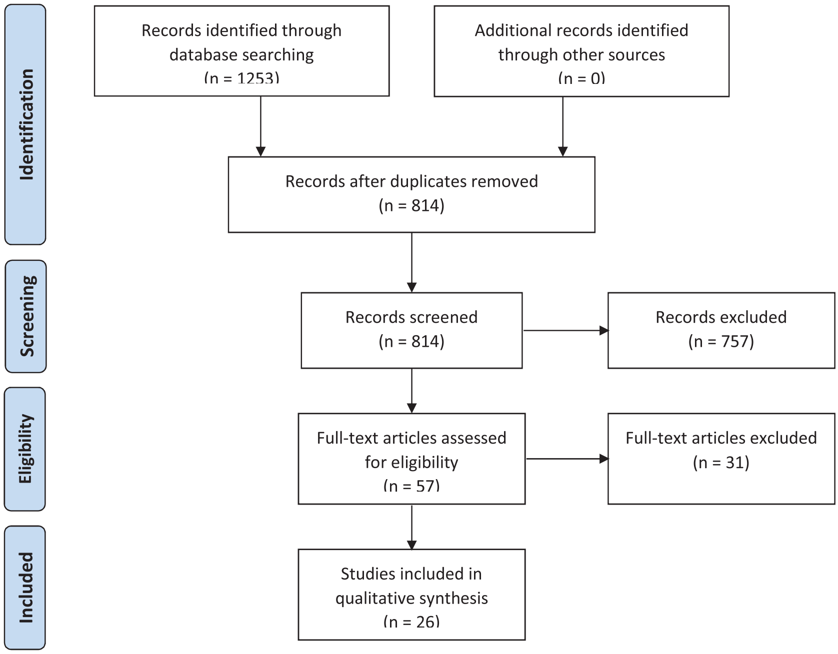


**FIGURE 1** Flowchart of the study selection

### 2.4 | Data items and collection

From each included study, the following data were extracted: (a) demographic variables of study participants, (b) type of AF measurements, (c) type of EF measurements, (d) type of confounding factors considered, and (e) results (Tables 2–4).

## 3 | RESULTS

Study design, participants, and the measures of AF and EF in the different studies are summarized in this section. Study results are reported per EF component (Tables 2–4). The qualitative analysis of the studies is reported at the end of Section 3 (Table 5).

### 3.1 | Study design

Tables 3–5 present the study characteristics. Eleven cross-sectional studies compared high-fit to low-fit participants based on normative AF data (Berchicci et al., 2015; Chaddock et al., 2010; Chaddock, Hillman, et al., 2012; Chaddock, Erickson, et al., 2012; Chaddock, Hillman, Buck, & Cohen, 2011; Hillman et al., 2009; Hillman, Castelli, & Buck, 2005; Pontifex et al., 2011; Pontifex, Scudder, Drollette, & Hillman, 2012; Voss et al., 2011; Wu et al., 2011). Children with an AF level below the 30th percentile were considered as low fit and above the 70th percentile as high fit. Eleven studies used correlational and regression analysis (Aadland et al., 2017; Buck, Hillman, & Castelli, 2008; Chojnacki et al., 2018; Drollette et al., 2016; Geertsen et al., 2016; Kao, Westfall, Parks, Pontifex, & Hillman, 2017; Moore et al., 2013; Pontifex et al., 2014; Scudder et al., 2014;

**TABLE 2** Studies examining the relationship between aerobic fitness and inhibition

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Aadland et al., 2017<br><br>Correlational study | 697 TD C (340/357)<br>10.2 years ± 0.3<br>est. $VO_2$ peak: 52.3 ± 7.5 (equals HF C) | Andersen test (m) | SCWT—number of words | Age<br>Pubertal status<br>SES<br>Birth weight<br>Body fat | *Higher levels of AF were associated with higher number of words in boys but not in girls.*<br>Correlation coefficients between AF and SCWT<br>Girls: $r_p$ = .15*<br>Boys: $r_p$ = .24*<br>Multiple regression analysis<br>Girls: stand. $\beta$ = .03, 95% CI [−0.11, 0.16]<br>Boys: stand. $\beta$ = .17, 95% CI [0.02, 0.31]* |
| Berchicci et al., 2015<br><br>Comparative study | HF C<br>21<br>10.0 years ± 0.6<br>$VO_2$ peak > Pc70<br><br>LF C<br>20<br>10.0 years ± 0.6<br>$VO_2$ peak < Pc 30 | $VO_2$ peak<br>(ml kg$^{-1}$ min$^{-1}$) | Modified Eriksen flanker task<br>Conditions:<br>Comp-congr<br>Comp-incongr<br>Incomp-congr<br>Incomp-incongr<br>RT and RA | | *HF C showed increased RA compared with LF C*<br>Repeated measures ANOVA to compare HF and LF C<br>mRT: HF C < LF C, $\eta^2$ = 0.06<br>mRA: HF C > LF C, $\eta^2$ = 0.21** |
| Buck et al., 2008<br><br>Correlational study | 74 TD C (41/33)<br>9.3 years ± 1.4<br>Laps: 23.2 ± 12.2<br>TD C | 20-m shuttle run (PACER, laps) | Paper and pencil version SCWT—number of words | Age<br>IQ<br>BMI (gender and SES were not correlated with AF or SCWT) | *Higher levels of AF were associated with higher number of words in all SCWT conditions*<br>Correlation coefficients<br>Word condition: $r_p$ = .36*<br>Colour condition: $r_p$ = .37*<br>Colour-word condition: $r_p$ = .37*<br>Stroop interference: $r_p$ = .37*<br><br>Hierarchical regression analyses<br>Word condition: stand. $\beta$ = .34***<br>Colour condition: stand. $\beta$ = .25*<br>Colour-word condition: stand. $\beta$ = .30*<br>Stroop interference: model was NS |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Castelli et al., 2011<br><br>Intervention study (pre-post uncontrolled design) | 59 TD C (33/26)<br>8.79 ± 0.54<br>$VO_2$ peak: 38.1 ± 7.0 | $VO_2$ peak ($ml\ kg^{-1}\ min^{-1}$), modified Balke protocol on motor-driven treadmill<br>Mean time above THZ during intervention | SCWT<br>Number of words<br>Posttest number subtracted from pretest number | Age | $VO_2$ peak and SWCT were significantly** enhanced from pretest to posttest<br>*Mean time above THZ was positively associated with SCWT colour-word condition*<br>Correlation between mean time in THZ and SCWT (colour-word condition): NS<br>Correlation between mean time above THZ and SCWT (colour-word condition): −.28 ($p$ = .02) |
| Chaddock et al., 2010<br><br>Comparative study | HF C ($VO_2$ peak > Pc 70)<br>25 (14/11)<br>10.0 years ± 0.6<br>$VO_2$ peak 52.5 ± 4.8<br><br>LF C ($VO_2$ peak < Pc 30)<br>18 (7/11)<br>10.1 years ± 0.5<br>$VO_2$ peak: 36.6 ± 4.4 | $VO_2$ peak ($ml\ kg^{-1}\ min^{-1}$), modified Balke protocol on motor-driven treadmill | Modified Eriksen flanker task<br>Conditions:<br>Congruent<br>Incongruent<br>RT, RA, and % interference =[(incongr-congr)/congr]*100 | Age, IQ, SES, and ADHD rating were not significant different between groups. | *No significant difference between groups in RT and RA in both conditions; HF C showed less percentage interference compared with LF C*<br>Independent $t$ test<br>Congruent<br>RA: NS, mRT: NS<br>Incongruent<br>RA: NS, mRT: NS<br>% interference RT: HF C < LF C* |
| Chaddock, Erickson, et al., 2012<br>Comparative study | HF C ($VO_2$ peak > Pc 70)<br>14 (9/5)<br>9.7 years ± 0.6<br>$VO_2$ peak:53.1 ± 4.5<br><br>LF C ($VO_2$ peak < Pc 30)<br>18 (7/11)<br>10.1 years ± 0.5<br>$VO_2$ peak: 36.6 ± 4.4 | $VO_2$ peak ($ml\ kg^{-1}\ min^{-1}$), modified Balke protocol on motor-driven treadmill | Modified Eriksen flanker task<br>Conditions:<br>Congruent<br>Incongruent<br>RT, RA, and comparison between early and late task blocks | Age, gender, IQ, and SES were not significant different between groups. | *HF C maintained RA better across task blocks compared with LF C in the incongruent condition*<br><br>Multivariate repeated measures ANOVAs<br>Congruent: ↑RT and ↓RA across blocks without fitness group interaction<br>Incongruent: ↓RT across blocks without fitness group interaction; HF C maintained RA across blocks, LF C ↓ RA across blocks ⇒ significant interaction between LF and HF C* |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Chaddock, Hillman, et al., 2012<br>Comparative study | Follow-up of the modified flanker task after 1 year<br>HF C ($VO_2$ peak > Pc 70)<br>14 (7/7)<br>10.0 years ± 0.6<br>$VO_2$ peak: 52.1 ± 4.4<br><br>LF C ($VO_2$ peak < Pc 30)<br>18 (8/10)<br>10.1 years ± 0.5<br>$VO_2$ peak: 35.8 ± 3.8 | $VO_2$ peak[a] (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task<br>Conditions:<br>Comp-congruent<br>Comp-incongruent<br>Incomp-congruent<br>Incomp-incongruent<br>RT and RA | Age, pubertal status, IQ, SES, and ADHD rating were not significant different between groups. | Multivariate repeated measures ANOVAs<br>RT: four conditions: initial LF = HFC; follow-up: HFC < LFC**<br>RA: main effect of fitness group LF C < HF C*; interaction between fitness group and condition (LF C more ↓RA in incompatible conditions compared with compatible conditions*) |
| Chojnacki et al., 2018<br><br>Correlational study | 233 TD C (100/133)<br>8.67 years ± 0.54<br>42.13 ± 7.2 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol<br><br>$VO_2FF$ | Modified Eriksen flanker task<br>Conditions:<br>Congr<br>Incongr<br>RT, RA, SDRT, CVRT | Age, gender pubertal status, IQ, SES, and %fat | *Higher levels of AF were associated with smaller intraindividual variability in inhibition tasks which required greater amounts of cognitive control.*<br><br>Correlation coefficients with FF $VO_2$ peak<br>Congr: RT, RA, and SDRT: NS<br>CVRT: $r_p$ = –0.13*<br>Incongr: RT:NS<br>RA: $r_p$ = .15*<br>SDRT: $r_p$ = –.18**<br>CVRT: $r_p$ = –.19*<br><br>Hierarchical linear regression analysis<br>Congruent:<br>RT, RA, SDRT, CVRT: $VO_2FF$ $\beta$ = NS<br>Incongruent:<br>RT, RA, SDRT: $VO_2FF$ $\beta$ = NS<br>CVRT: $VO_2FF$ $\beta$ = –.14* |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Hillman et al., 2005<br><br>Comparative study | HF C ($VO_2$ peak > Pc 70)<br>12 (7/5)<br>9.3 years ± 1.2<br>30.8 ± 13.1<br><br>LF C<br>12 (6/6)<br>9.8 years ± 0.6<br>12.6 ± 5.3 | 20-m shuttle run (PACER, laps) | Visual odd ball paradigm | Age, education level, IQ, SES, height and weight were not significant different between groups | *HF C responded more accurately and faster compared with LF C*<br><br>Mixed model ANOVA<br>RA: HF C > LF C*<br>RT: HF C < LF C** |
| Hillman et al., 2009<br><br>Comparative study | HF C ($VO_2$ peak < Pc 30)<br>19 (10/9)<br>9.3 years ± 0.9<br>31.2 ± 8.9<br><br>LF C<br>19 (10/9)<br>9.5 years ± 1.012.6 ± 3.7 | 20-m shuttle run[b] (PACER, laps) | Modified Eriksen flanker task<br>Conditions:<br>Congruent<br>Incongruent<br>RT and RA | Age, SES, IQ, and ADHD rating were not significant different between groups. BMI was not correlated with EF or AF parameters. | *HF C responded more accurately compared with LF C, regardless of task condition.*<br><br>Repeated measures MANOVA<br>Congruent + incongruent<br>RA: HF C > LF C*, $\eta^2 = 0.12$<br>mRT: NS |
| Moore et al., 2013<br><br>Correlational study | 93 TD C (54/39)<br>8.8 years ± 0.6<br>38.2 ± 7.3 (equals LF C) | $VO_2$ peak ($ml\ kg^{-1}\ min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task<br>Conditions:<br>Comp-congruent<br>Comp-incongruent<br>Incomp-congruent<br>Incomp-incongruent<br>RT and RA | Age<br>Gender<br>IQ<br>SES | *Higher levels of AF were associated with shorter RT in tasks which required greater amounts of cognitive control.*<br><br>Correlation coefficients:<br>Comp-congruent: mRT: NS; mRA: NS<br>Comp-incongruent: mRT:NS: mRA: NS<br>Incomp-congruent: mRT: $r_p = -.33^{**}$; mRA: NS<br>Incomp-incongruent: mRT: $r_p = -.34^{**}$; mRA: NS<br><br>Multiple hierarchical regression analyses |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| | | | | | Comp-congruent: mRT:NS; mRA: NS<br>Comp-incongruent: mRT:NS; mRA: NS<br>Incomp-congruent: mRT:stand. $\beta = -.26^*$; mRA: NS<br>Incomp-incongruent: mRT: stand. $\beta = -.26^*$; mRA: NS |
| Pindus et al., 2016<br><br>Correlational study | 74 TD C (40/34)<br>8.6 years ± 0.6)<br>43.4 ± 8 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task<br>Conditions:<br>Congruent<br>Incongruent<br>RT, RA, and interference (difference between congruent and incongruent condition) | Age<br>Gender<br>IQ<br>SES<br>Pubertal stage<br>ADHD ratings<br>Birth weight | *Higher levels of AF were only associated with lower accuracy interference. No additional associations between AF and inhibition were found in regression models predicting inhibitory control from daily moderate to vigorous physical activity.*<br><br>Partial correlation coefficients: NS, except for accuracy interference: $r = -.25^*$<br><br>Multiple hierarchical regression analyses<br>Incongruent mRT: NS; mRA: NS<br>mRT interference: NS<br>RA interference: NS |
| Pontifex et al., 2011<br><br>Comparative study | HF C ($VO_2$ peak > Pc 70)<br>24 (14/10)<br>10.0 years ± 0.6<br>52.6 ± 4.2<br><br>LFC ($VO_2$ peak < Pc 30)<br>24 (11/13)<br>10.0 years ± 0.6<br>35.7 ± 5.3 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task<br>Conditions:<br>Comp-congruent<br>Comp-incongruent<br>Incomp-congruent<br>Incomp-incongruent<br>RT and RA | Age<br>Pubertal status<br>IQ<br>SES<br>ADHD rating | *HF C responded more accurately compared with LF C.*<br><br>Multivariate repeated measures ANOVA<br>Four conditions together<br>RT: NS<br>RA: HF C > LF C**, $\eta^2 = 0.21$ |
| Pontifex et al., 2012 | HF C ($VO_2$ peak > Pc 70)<br>28 (17/11) | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), | Modifief Eriksen flanker task | Age<br>Gender | *HF C responded more accurately compared with LF C.* |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Comparative study | 10.0 years ± 0.6<br>52.8 ± 4.2<br><br>LF C ($VO_2$ peak < Pc 30)<br>34 (17/17)<br>10.1 years ± 0.6<br>35.7 ± 5.2 | motor-driven treadmill, Balke protocol | Conditions:<br>Congruent<br>Incongruent<br>RT and RA | IQ<br>SES<br>Pubertal status<br>ADHD rating | Multivariate repeated measures ANOVA<br>Congruent + incongruent condition<br>RT: NS<br>RA: HF C > LF C**, $\eta^2$ = 0.11 |
| Pontifex et al., 2014<br><br>Correlational study | 204 TD C (108/96)<br>8.8 years ± 0.6<br>38.2 ± 7.1 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task<br>Conditions:<br>Congruent<br>Incongruent<br>RT and RA | Age<br>Gender<br>IQ<br>SES<br>Pubertal status<br>Adiposity<br>Race | *Higher levels of AF were associated with greater RA in both congruent and incongruent trails.*<br><br>Correlation coefficients<br>Congruent: RA: $r_p$ = .25**<br>Incongruent: RA $r_p$ = .23**<br><br>Hierarchical linear regression analyses<br>Congruent: RA: stand. $\beta$ = .21**<br>Incongruent: RA stand. $\beta$ = .20** |
| Schmidt et al., 2015<br><br>Intervention study | 181 TD C (82/99) divided into three groups for an intervention study (team games, aerobic exercise, and control condition)<br>11.35 years ± .60<br>-? | Multistage 20-m shuttle run test (est. $VO_2$ peak, ml $kg^{-1}$ $min^{-1}$) | Modified Eriksen flanker task<br>Conditions:<br>Congruent<br>Incongruent<br>Interference cost RT | Age<br>Gender<br>Children with ADH<br>Physical activity level<br>Pubertal status<br>SES<br>BMI<br>Academic achievement | *RT decreased in the three groups, but the change did not differ significantly between the three groups based on a multilevel analysis (mixed models)*<br>*AF increased significantly* more in the two intervention groups, compared with the control group; the increase in the two intervention groups did not differ significantly from each other.*<br>*Change pretest to posttest* est. $VO_2$ peak, ml $kg^{-1}$ $min^{-1}$<br>Team games: 2.24 (4.59) |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| | | | | | Aerobic exercise: 1.191 (3.65)<br>Control condition: –07. (6.32) |
| Scudder et al., 2014<br><br>Correlational study | 397 TD C(172/225)<br>7.6 years ± 0.6<br>17.1 ± 8.9 | 20-m shuttle run (PACER, laps) | Modified Eriksen flanker task<br>Conditions:<br>Comp-congruent<br>Comp-incongruent<br>Incomp-congruent<br>Incomp-incongruent<br>RT and RA | Age<br>Grade<br>Gender<br>BMI<br>SES | *Higher levels of AF were associated with greater RA and faster RT, regardless of task condition (except for RA in the incompatible incongruent condition).*<br><br>Hierarchical linear regression analyses<br>Comp-congruent: RT: stand. $\beta$ = –.20**; RA: stand. $\beta$ = .12*<br>Comp-incongruent: RT stand. $\beta$ = –.18**; RA: stand. $\beta$ = .11*<br>Incomp-congruent: RT: stand. $\beta$ = –.12**; RA: stand. $\beta$ = .13*<br>Incomp-incongruent: RT: stand. $\beta$ = –.11**; RA: NS |
| Scudder et al., 2016<br><br>Correlational study | Follow-up study TD C<br>290 (115/175)<br>Time 1<br>8.1 years ± 0.04<br>Laps: 16.7 ± 0.5<br>Time 2<br>10.6 ± 0.04<br>Laps: 25.1 ± 0.7 | 20-m shuttle run (PACER, laps) | Modified Eriksen flanker task<br>Conditions:<br>Comp-congruent<br>Comp-incongruent<br>Incomp-congruent<br>Incomp-incongruent<br>RT, RA, and interference cost | Age<br>Grade<br>Gender<br>BMI<br>SES | *Higher levels of AF were associated with faster RT greater RA and greater interference control in some of the conditions, not consequent over time.*<br><br>*Only significant effects were reported.*<br>Hierarchical regression analyses<br>Time 1<br>Comp-congruent: RT: stand. $\beta$ = –.21<br>Comp-incongruent: RT stand. $\beta$ = –.20<br>Incomp-congruent: RT: stand. $\beta$ = –.19<br>Incomp-incongruent: RT: stand. $\beta$ = –.17<br><br>Time 2<br>Comp-incongruent: RA: stand. $\beta$ = .20 |

(Continues)

**TABLE 2** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| | | | | | comp RA Interference: stand. $\beta$ = –.16<br><br>ΔAF<br>Comp-incongruent: RA: stand. $\beta$ = .13<br>Comp- RA interference cost: stand. $\beta$ = .11 |
| Voss et al., 2011<br><br>Comparative study | HFC ($VO_2$ peak > Pc 70)<br>18 (10/8)<br>9.8 years ± 0.6<br>52.8 ± 5.2<br><br>LFC ($VO_2$ peak < Pc 30)<br>18 (7/11)<br>9.9 years ± 0.6<br>36.9 ± 4.4 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task during fMRI scanning<br>Conditions:<br>Congruent<br>Incongruent<br>Neutral<br>RT, RA, RA interference cost | Age<br>Gender<br>Pubertal status<br>IQ | *HF C showed only greater RA in the incongruent condition and less interference cost compared with LF C.*<br><br>Repeated measures ANOVA<br>Congruent: RT: NS; RA: NS<br>Incongruent: RT: NS; RA: HF C > LF C*<br>Interference cost RA: HF C > LF C* |
| Wu et al., 2011<br><br>Comparative study | HFC ($VO_2$ peak > Pc 70)<br>24 (14/10)<br>10.0 years ± 0.6<br>52.6 ± 4.3<br><br>LF C ($VO_2$ peak < Pc 30)<br>24 (13/11)<br>10.1 years ± 0.6<br>35.7 ± 5.3 | $VO_2$ peak [a] (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Modified Eriksen flanker task<br>Conditions:<br>Comp-congruent<br>Comp-incongruent<br>Incomp-congruent<br>Incomp-incongruent<br>mRT, SD RT, CV RT, RA | Age<br>Gender<br>IQ<br>Pubertal status<br>ADHD rating | *HF C showed less variable RT and higher RA than LF C.*<br><br>Multivariate repeated measures ANOVA<br>Four conditions together<br>mRT: NS<br>SD RT: HF C > LF C*, $\eta^2$ = 0.09<br>SD RT: HF C > LF C**, $\eta^2$ = 0.21<br>RA: HF C > LF C***, $\eta^2$ = 0.21 |

Abbreviations: AF, aerobic fitness; ADHD, attention deficit hyperactivy disorder; BMI, body mass index; comp, compatible; CVRT, intraindividual coefficient of variation of response time; est, estimated; HF C, high-fit children; incomp, incompatible; IQ, intelligence quotient; LF C, low-fit children; mRA, mean response accuracy; mRT, mean response time; NS, no significant fitness based difference; PACER, progressive aerobic cardiovascular endurance run; Pc, percentile; RA, response accuracy; RT, response time; SES, socio-economic status; troop Colour-Word Test (SCWT); TD C, typical developing children; THZ, target heart zone, maximum oxygen uptake; $VO_2FF$, $VO_2$ peak adjusted for fat free mass; ΔAF, change in AF between time 1 and 2.

*$p$ < .05. **$p$ < .01. ***$p$ < .001.

**TABLE 3** Studies examining the relationship between aerobic fitness and working memory

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Aadland et al., 2017<br>Correlational study | 697 TD C (340/357)-10.2 years ± 0.3<br>est. $VO_2$ peak: 52.3 ± 7.5 (equals HF C) | Andersen test (m) | Digit span backward number of sequences | Age<br>Pubertal status<br>SES<br>Body fat<br>Birth weight | *Higher fitness levels were associated with higher number of sequences in boys but not in girls.*<br>Correlations between AF and digit span backward<br>Girls: $r_p$ = .16*<br>Boys: $r_p$ = .15*<br>Multiple regression analysis<br>Girls: stand. $\beta$ = .01, 95% CI [−0.13, 0.16]<br>Boys: stand. $\beta$ = .16, 95% CI [0.01, 0.32]* |
| Chaddock et al., 2011<br>Comparative study | HF C ($VO_2$ max > Pc 70)<br>22 (12/10)<br>9.9 years ± 0.5<br>$VO_2$ peak:48.7 ± 3.6<br>LF C ($VO_2$ peak < Pc 30)<br>24 (11/13)<br>9.9 years ± 0.6<br>$VO_2$ peak: 35.2 ± 4.6 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Memory task in nonrelation encoding condition versus relational encoding condition<br>RA | Age<br>Pubertal status<br>IQ<br>ADHD | *HF C showed higher RA in relational encoding tasks than LF C, no significant difference was found for nonrelational encoding.*<br>Univariate ANCOVA<br>Relational encoding: $\eta^2$ = 0.11*<br>Nonrelational encoding: $\eta^2$ = 0.001 |
| Drollette et al., 2016 (Study 1)<br>Correlational study | 97 TD C(55/42)<br>9.4 years ± 0.1<br>$VO_2$ peak: 39.4 ± 1.0 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | OSPAN<br>RA | Age<br>Grade<br>IQ<br>Gender<br>Gender × fitness | *Higher AF levels were associated with greater OSPAN RA for boys but not girls.*<br>*Hierarchical linear regression analyses*<br>Girls: stand. $\beta$ = −.18, 95% CI [−0.017, 0.006]<br>Boys: stand. $\beta$ = .26, 95% CI [0.001, 0.013]* |
| Drollette et al., 2016 (Study 2)<br>Correlational study | 95 TD C (50/45)<br>8.8 years ± 0.2<br>$VO_2$ peak: 39.5 ± 0.6 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Serial N-back task (0-, 1-, and 2-back)<br>$d'$ | Age<br>Grade<br>Gender<br>Gender × fitness<br>IQ | *Higher AF levels were associated with greater 2-back $d'$ accuracy for boys but not girls, only in 2-back condition*<br>Hierarchical linear regression analyses2-back condition, $d'$ accuracy<br>Boys: stand. $\beta$ = .28, 95% CI [0.001, 0.06]*<br>Girls: no significant model |

(Continues)

**TABLE 3** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Drollette et al., 2016 (Study 3)<br><br>Correlational study | 83 TD C (46/37)<br>7.8–9.9 years (8.9 years ± 0.1)<br>$VO_2$ peak: 34.2 ± 1.16 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Sternberg task (1, 3, or 5 letters)—*d′* | Age<br>Grade<br>Gender<br>Gender × fitness<br>IQ | *Higher AF levels were associated with greater Sternberg task d′ for boys but not girls, only in 5 letter condition.*<br>Hierarchical linear regression analyses5 letter condition, *d′*<br>Boys: stand. $\beta$ = .29, 95% CI [0.004, 0.05]*<br>Girls: stand. $\beta$ = −.00, 95% CI [−0.04, 0.03] |
| Geertsen et al., 2016<br><br>Correlational study | 423 TD C (214/209)<br>9.3 years ± 0.4<br>Yo-Yo test (m): 713 ± 421 | Yo-Yo intermittent recovery level 1 children's test (m) | Spatial WMe test CANTAB—number of errors | Aage<br>Grade<br>Gender<br>Pubertal status<br>Municipality | *Higher AF levels were associated with fewer errors in spatial WMe test CANTAB.*<br>Linear mixed effect model<br>Estimate slope coefficient = −.30 (*SE* = 0.10)* |
| Kao et al., 2017<br><br>Correlational study | 79 TD C (44/35)<br>10.1 years ± 0.6<br>$VO_2$ peak: 44.3 ± 6.7 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), motor-driven treadmill, Balke protocol | Serial N-back task, 1- and 2-back condition—*d′* | Age<br>Grade<br>Gender<br>IQ<br>SES<br>Muscular fitness | *Higher AF levels were associated with greater d′ in serial 2-back task.*<br>Correlation between AF and *d′* serial 2-back task<br>$r_s$ = .26*<br>Hierarchical linear regression analyses<br>Serial 2-back task: stand. $\beta$ = .26* (after entering muscular fitness, AF is no longer a significant factor) |
| Koutsandreou et al., 2016<br><br>Correlational study | 71 TD C (32/39) divided into three groups for an intervention study (cardiovascular exercise, motor exercise, control)<br>9.35 years ± 0.6<br>SR level: 5.6 ± 1.8 | 20-m shuttle run test (number of laps) | Letter digit span RA | Age<br>Height<br>Weight<br>Pubertal status<br>AF<br>Motor fitness | *Children from the "motor exercise" group improved WMe more than children from cardiovascular exercise group, which improved more than control group.*<br>Repeated measures ANOVA<br>Significant interaction effect group-time on AF: only cardiovascular exercise group improved AF**<br>Significant interaction effect group-time on WMe: control group: NS |

(Continues)

**TABLE 3** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| | | | | | Cardiovascular group improved Wme***<br>Motor exercise group improved Wme*** |
| Pindus et al., 2016<br><br>Correlational study | 74 TD C (40/34)<br>8.6 years ± 0.6<br>$VO_2$ peak: 43.4 ± 8 | $VO_2$ peak ($ml\ kg^{-1}\ min^{-1}$), motor-driven treadmill, Balke protocol | OSPAN<br>Four different scores | Age<br>Gender<br>IQ<br>ADHD ratings<br>Birth weight | *No significant association between AF and OSPAN scores*<br>Partial correlation between AF and OSPAN: NS<br>*Multiple hierarchical regression analyses*<br>Factor AF: NS |
| Schmidt et al., 2015<br><br>Intervention study | 181 TD C (82/99) divided into three groups for an intervention study (team games, aerobic exercise and control condition)<br>11.35 years ± .60<br>-? | Multistage 20-m shuttle run test (est. $VO_2$ peak, $ml\ kg^{-1}\ min^{-1}$) | Nonspatial N-back task, 2-back condition<br>Number of correct answers | Age<br>Gender<br>Children with ADH<br>Physical activity level<br>Pubertal status<br>SES<br>BMI<br>Academic achievement | *RA increased in the three groups, but the change did not differ significantly between the three groups based on a multilevel analysis (mixed models).*<br>*AF increased significantly* more in the two intervention groups, compared with the control group; the increase in the two intervention groups did not differ significantly from each other.*<br>*Change pretest to posttest* est. $VO_2$ peak, $ml\ kg^{-1}\ min^{-1}$<br>Team games: 2.24 (4.59)<br>Aerobic exercise: 1.191 (3.65)<br>Control condition: –07. (6.32) |
| Scudder et al., 2014<br><br>Correlational study | 397 TD C (172/225)<br>7.6 years ± 0.6<br>Laps: 17.1 ± 8.9 | 20-m shuttle run (PACER, laps) | Spatial N-back task<br>0, 1-, and 2-back condition *d′* | Grade<br>Gender<br>SES<br>BMI | *Higher AF levels were associated with greater d′ in 1- and 2-back condition*<br>Hierarchical linear regression analyses<br>*d′* 1-back condition, AF: stand. $\beta$ = .17**<br>*d′* 2-back condition, AF: stand. $\beta$ = .19** |
| Scudder et al., 2016 | Follow-up study TD C<br>290 (115/175)<br>Time 1 | 20-m shuttle run (PACER, laps) | Spatial N-back task<br>0-, 1-, and 2-back condition *d′* | Grade<br>Gender<br>SES | *Higher AF levels were associated with greater d′ in 1- and 2-back condition*<br>Hierarchical regression analyses |

(Continues)

**TABLE 3** (Continued)

| Authors<br>Study design | Participants<br>Sample size (male/female)<br>Age: Average ± SD<br>Aerobic fitness: average ± SD | **Aerobic fitness measurement** | **Executive function measurement** | **Considered confounding variables** | **Results** |
|---|---|---|---|---|---|
| Correlational study | 8.1 years ± 0.04<br>Laps: 16.7 ± 0.5<br>Time 2<br>10.6 ± 0.04<br>Laps: 25.1 ± 0.7 | | | BMI | Time 1<br>*d′* 1-back condition, AF: stand. $\beta$ = .20**<br>*d′* 2-back condition, AF: stand. $\beta$ = .24**<br><br>Time 2<br>*d′* 2-back condition, AF: stand. $\beta$ = .27**<br><br>ΔAF<br>*d′* 2-back condition, AF: stand. $\beta$ = .20* |

Abbreviations: AF, aerobic fitness; BMI, body mass index; *d′* = z (adjusted hit rate)- z (adjusted FA rate); est, estimated; HF C, high-fit children; IQ, intelligence quotient; LF C, low-fit children; mRA, mean response accuracy; mRT, mean response time; NS, no significant fitness based difference; PACER, progressive aerobic cardiovascular endurance run; Pc, percentile; RA, response accuracy; RT, response time; SES, socio-economic status; TD C, typical developing children; $VO_2$ peak, maximum oxygen uptake; WMe, working memory.
**p* < .05. ***p* < .01. ****p* < .001.

**TABLE 4** Studies examining the relationship between aerobic fitness and cognitive flexibility

| Study | Participants<br>Sample size (male/female)<br>Age: average ± SD<br>Aerobic fitness: average ± SD | Aerobic fitness measurement | Executive function measurement | Considered confounding variables | Results |
|---|---|---|---|---|---|
| Aadland et al., 2017<br><br>Correlational Study | 697 TD C (340/357)<br>10.2 years ± 0.3<br>est. $VO_2$ peak: 52.3 ± 7.5 (equals HF C) | Andersen test (m) | TMT—part A + B - time required to complete<br>VF—number of animals | Age<br>Pubertal status<br>SES<br>Birth weight<br>Body fat | *AF was positively associated with TMT and VF in boys but not in girls.*<br>Correlations between AF and TMT<br>Girls: $r_p$ = –.12*<br>Boys: $r_p$ = –.23*<br>Correlations between AF and VF<br>Girls: $r_p$ = –.01<br>Boys: $r_p$ = .13*<br>Multiple regression analysis TMT<br>Girls: stand. $\beta$ = –.06, 95% CI [–0.20, 0.08]<br>Boys: stand. $\beta$ = –.17, 95% CI [–0.32, –0.02]*<br>Multiple regression analysis VF<br>Girls: stand. $\beta$ = –.12, 95% CI [–0.26, 0.02]<br>Boys: stand. $\beta$ = .03, 95% CI [–0.12, 0.18] |
| Castelli et al., 2011<br><br>Intervention study (pre-post uncontrolled design) | 59 TD C (33/26)<br>8.79 ± 0.54<br>$VO_2$ peak: 38.1 ± 7.0 | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), modified Balke protocol on motor-driven treadmill<br>Mean time above THZ during intervention | TMT part A + B<br>Time required to complete<br>Posttest time subtracted from pretest time | Age | *Time in THZ was positively associated with TMT part B.*<br>Correlations between mean time above THZ and Trails A: NS<br>Correlations between mean time above THZ and Trails A: .33 ($p$ < .01) |
| Pontifex et al., 2014<br><br>Correlational Study | 204 TD C (108/96)<br>8.8 years ± 0.6<br>38.2 ± 7.1 (p19.7 ± 21.8) | $VO_2$ peak (ml $kg^{-1}$ $min^{-1}$), modified Balke protocol on motor-driven treadmill | Colour shape switch task—RA<br>Homogeneous condition<br>Heterogeneous condition | Age<br>Gender<br>IQ<br>Pubertal status<br>SES<br>Adiposity<br>Race | *AF was positively associated with greater RA in the heterogeneous condition of the switch task.*<br>Correlations between AF and RA in homogeneous condition: $r_p$ = .11<br>Correlations between AF and RA in heterogeneous condition: $r_p$ = .22** |

(Continues)

**TABLE 4** (Continued)

| Study | Participants<br>Sample size (male/female)<br>Age: average ± SD<br>Aerobic fitness: average ± SD | Aerobic fitness measurement | Executive function measurement | Considered confounding variables | Results |
|---|---|---|---|---|---|
| | | | | | Hierarchical regression analyses in heterogeneous condition: stand. $\beta$ = .20** |
| Schmidt et al., 2015<br><br>Intervention study | 181 TD C (82/99) divided into three groups for an intervention study (team games, aerobic exercise and control condition)<br>11.35y ± .60<br>-?<br><br>Pretest to posttest the change in est. $VO_2$ peak was 2.24 ( ±4.59) for the "team games" group; 1.91 (±3.56) for the "aerobic exercise" group and –.07 (±6.32) for the control group | Multistage 20-m shuttle run test (est. $VO_2$ peak, ml $kg^{-1}$ $min^{-1}$) | Additional block in the Eriksen flanker task with congruent and incongruent trials and an additional rule cued by the colour of the trials.<br>Difference in RT between the mixed and the standard block. | Age<br>Gender<br>Children with ADHD<br>Physical activity level<br>Pubertal status<br>SES<br>BMI<br>Academic achievement | *RT decreased in the three groups, but the improvement in the group "team games" was significantly stronger compared with the "aerobic exercise group"* and the control group* based on a multilevel analysis* (*mixed models*)<br>*AF increased significantly* more in the two intervention groups, compared with the control group; the increase in the two intervention groups did not differ significantly from each other.*<br>*Change pretest to posttest* est. $VO_2$ peak, ml $kg^{-1}$ $min^{-1}$<br>Team games: 2.24 (4.59)<br>Aerobic exercise: 1.191 (3.65)<br>Control condition: –07. (6.32) |
| van der Niet et al., 2014<br><br>Correlational Study | 263 TD C (145/118)<br>9.5 years ± 1.2<br>-? | 20-m shuttle run (EUROFIT, laps) | Trail Making Test (Parts A and B)—RT | Age<br>Gender<br>SES | AF was positively associated with TMT<br>*Structural equation modelling*<br>RT (Part B—Part A) ↓ with AF ↑* |

Abbreviations: AF, aerobic fitness; ADHD, attention deficit hyperactivity disorder; BMI, body mass index; est, estimated; HF C, high-fit children; IQ, intelligence quotient; RA, response accuracy; RT, response time; SES, socio-economic status; TD C, typical developing children; THZ, target heart zone; TMT, Trail Making Test; VF, verbal fluency; $VO_2$ peak, maximum oxygen uptake; ↑, increased; ↓, decreased.
*$p$ < .05. **$p$ < .01.

**TABLE 5** Scores on the "Downs and Black" checklist

| Studies | Reporting | External validity | Internal validity bias | Internal validity confounding | Power | Total Score | Score (%) |
|---|---|---|---|---|---|---|---|
| Aadland et al., 2017 | 8/8 | 2/2 | 3/3 | 1/1 | / | 14/14 | 100 |
| Berchicci et al., 2015 | 5/8 | 0/2 | 3/3 | 3/3 | / | 11/16 | 69 |
| Buck et al., 2007 | 8/8 | 0/2 | 3/3 | 1/1 | / | 12/14 | 86 |
| Castelli et al., 2011 | 7/9 | 0/2 | 0/1 | 3/3 | 0/1 | 10/16 | 63 |
| Chaddock et al., 2010 | 7/9 | 0/2 | 0/1 | 3/3 | / | 13/16 | 81 |
| Chaddock et al., 2011 | 8/8 | 0/2 | 3/3 | 3/3 | / | 14/16 | 88 |
| Chaddock, Erickson, et al., 2012 | 7/8 | 0/2 | 3/3 | 3/3 | / | 13/16 | 81 |
| Chaddock, Hillman, et al., 2012 | 8/8 | 0/2 | 3/3 | 3/3 | / | 14/16 | 88 |
| Chojnacki et al., 2018 | 7/8 | 0/2 | 3/3 | 1/1 | / | 13/16 | 81 |
| Drollette et al., 2016a (Study 1) | 8/8 | 0/2 | 3/3 | 1/1 | / | 12/14 | 86 |
| Geertsen et al., 2016 | 7/8 | 1/2 | 3/3 | 1/1 | / | 12/14 | 86 |
| Hillman et al., 2005 | 7/8 | 0/2 | 3/3 | 3/3 | / | 13/16 | 81 |
| Hillman et al., 2009 | 7/8 | 0/2 | 3/3 | 3/3 | / | 13/16 | 81 |
| Kao et al., 2017 | 7/8 | 0/2 | 3/3 | 1/1 | / | 11/14 | 79 |
| Koutsandreou et al., 2016 | 7/10 | 0/2 | 6/7 | 4/5 | 0/1 | 17/25 | 68 |
| Moore et al., 2013 | 8/8 | 0/2 | 3/3 | 3/3 | / | 14/16 | 88 |
| Pindus et al., 2016 | 8/8 | 0/2 | 3/3 | 1/1 | / | 12/14 | 86 |
| Pontifex et al., 2011 | 8/8 | 0/2 | 3/3 | 3/3 | / | 14/16 | 88 |
| Pontifex et al., 2012 | 7/8 | 0/2 | 3/3 | 3/3 | / | 13/16 | 81 |
| Pontifex et al., 2014 | 6/8 | 0/2 | 3/3 | 1/1 | / | 10/14 | 71 |
| Schmidt et al., 2015 | 9/10 | 0/2 | 6/7 | 5/5 | 0/1 | 20/25 | 80 |
| Scudder et al., 2014 | 8/8 | 0/2 | 3/3 | 1/1 | / | 12/14 | 86 |
| Scudder et al., 2016 | 7/8 | 0/2 | 3/3 | 1/1 | / | 11/14 | 79 |
| van der Niet et al., 2014 | 4/8 | 0/2 | 3/3 | 1/1 | / | 8/14 | 57 |
| Voss et al., 2011 | 4/8 | 0/2 | 3/3 | 1/1 | / | 8/14 | 57 |
| Wu et al., 2011 | 8/8 | 0/2 | 3/3 | 3/3 | / | 14/16 | 88 |

Scudder et al., 2016; van der Niet, Hartman, Smith, & Visscher, 2014). Only three studies were interventional studies (Castelli, Hillman, Hirsch, Hirsch, & Drollette, 2011; Koutsandreou, Wegner, Niemann, & Budde, 2016; Schmidt et al., 2015).

## 3.2 | Participants

Participants mean age varied between 7.6 ± 0.6 (Scudder et al., 2014) and 11.35 ± 0.6 years (Schmidt et al., 2015). Sample size varied between 24 (Hillman et al., 2005) and 697 children (Aadland et al., 2017).

## 3.3 | Measures of AF

AF was evaluated using different methods: Sixteen studies assessed the $VO_2$ max during a maximal incremental treadmill test (Berchicci et al., 2015; Castelli et al., 2011; Chaddock et al., 2010; Chaddock et al., 2011; Chaddock, Erickson, et al., 2012; Chaddock, Hillman, et al., 2012; Chojnacki et al., 2018; Drollette et al., 2016; Kao et al., 2017; Moore et al., 2013; Pindus et al., 2016; Pontifex et al., 2011; Pontifex et al., 2012; Pontifex et al., 2014; Voss et al., 2011; Wu et al., 2011), eight studies used a 20-m shuttle run test (Buck et al., 2008; Hillman et al., 2005; Hillman et al., 2009; Scudder et al., 2014; Scudder et al., 2016; van der Niet et al., 2014), one study implemented the Andersen task (Aadland et al., 2017), and another one made use of the Yoyo intermittent recovery level 1 (Geertsen et al., 2016). The study of Castelli et al. (2011) used the time above the target heart zone as a measure of intensity of the cardiovascular effort during intervention.

## 3.4 | Measures of EF

There was a large variety between studies in the tasks and protocols used to measure EF. Inhibition was measured with three different tests. Seventeen studies used a modified Eriksen flanker task (Berchicci et al., 2015; Chaddock et al., 2010; Chaddock, Erickson, et al., 2012; Chaddock, Hillman, et al., 2012; Chojnacki et al., 2018; Hillman et al., 2009; Moore et al., 2013; Pindus et al., 2016; Pontifex et al., 2011; Pontifex et al., 2012; Pontifex et al., 2014; Schmidt et al., 2015; Scudder et al., 2014; Scudder et al., 2016; Voss et al., 2011; Wu et al., 2011). In this test, a directional response is assigned to a central target stimulus. The target is flanked by nontarget stimuli that correspond either to the same directional response as the target (congruent flankers), to the opposite response (incongruent flankers), or to neither (neutral flankers). Various forms of the task are used to measure information processing and selective attention. Of these 16 studies, seven added an incompatible condition to increase cognitive control by letting participants answer in the direction opposite to the central flanker (Berchicci et al., 2015; Chaddock, Hillman, et al., 2012; Moore et al., 2013; Pontifex et al., 2011; Scudder et al., 2014; Scudder et al., 2016; Voss et al., 2011; Wu et al., 2011). Three studies used the Stroop Colour-Word Test to measure cognitive inhibition (Aadland et al., 2017; Buck et al., 2008; Castelli et al., 2011). In this test, subjects are required to read three different tables as fast as possible. Two of them represent congruent conditions. In the third incongruent condition, colour-words are printed in an inconsistent colour. One study assessed response control with a visual odd-ball paradigm (Hillman et al., 2005). This paradigm has typically been used in combination with measurement of event-related brain potentials to evaluate ongoing neuroelectric activity, as was the case in this study. Participants are required to react quickly to an infrequently presented target stimulus, whereas no response is required for the frequent nontarget stimulus.

Working memory was administered through eight different task protocols. Two spatial working memory tasks were used. The spatial N-back task presents a stream of visuo-spatial stimuli to the participant, which they must monitor and respond to when they believe the current stimulus is the same as the stimulus presented "n" items ago (Aadland et al., 2017; Scudder et al., 2014; Scudder et al., 2016). The spatial working memory task from the CANTAB (Geertsen et al., 2016) requires retention and manipulation of visuospatial information. Verbal working memory was examined by four different tests: a digit span backward task (WISC IV; Aadland et al., 2017), a serial N-back task (Drollette et al., 2016; Kao et al., 2017), an operation word span task (Drollette et al., 2016; Pindus et al., 2016), and a letter digit span (Koutsandreou et al., 2016). A modified version of a relational memory task was used by Chaddock et al. (2011).

Drollette et al. (2016) used a Sternberg task, placing greater demand on storage and maintenance operations and as such matching more a short-term memory task. The key variable in the task is the list length manipulation of the memory set of items, which are encoded during the first phase of the trial and then retained over a short (seconds long) retention interval, in order to respond appropriately to a final probe item. Probe judgments indicate whether the item is positive (was a member of the current trial memory set) or negative (not a member of the memory set).

Cognitive flexibility was assessed by four different tasks. Three studies used the non-verbal trail-making task, requiring a subject to connect a sequence of consecutive targets, in a similar manner to a child's connect-the-dots puzzle (Aadland et al., 2017; Castelli et al., 2011; van der Niet et al., 2014). There are two parts to the test: In the first, the targets are all numbers and the test taker needs to connect them in sequential order; in the second part, the subject alternates between numbers and letters. Three tasks were only used in one study: the verbal fluency task, requiring the participants to name as many animals as possible in 60 s (Aadland et al., 2017), an additional condition of the Eriksen flanker task (Schmidt et al., 2015), and a colour-shape switch task evaluating the task switch cost (in the shape task, you need to respond to circles and shapes while ignoring the colour; in the colour task, you need to respond to the colour while ignoring the shape; Pontifex et al., 2014).

## 3.5 | Study results

Results are summarized per investigated EF in Tables 3–5. Only significant results are reported in the text below. Effect sizes of the studies comparing a group of high fit with low-fit children were reported in six out of 11 studies (Berchicci et al., 2015; Chaddock et al., 2011; Hillman et al., 2009; Pontifex et al., 2011; Pontifex et al., 2012; Wu et al., 2011) with a maximal reported eta-square coefficients of 0.21 in the studies from Wu et al. (2011) and Berchicci et al. (2015). Both studies investigated the association between the Eriksen flanker task and the $VO_2$ max, although Berchicci et al. (2015) related the behavioural results also to electro-encephalography. It seems that the behavioural data are mainly from the same group of participants in both studies. The mean response accuracy in the Eriksen flanker task was in both studies higher in the group of high-fit children with the same effect size of $\eta^2$ = 0.21. The correlational studies reported coefficients varying between nonsignificant and .37 in the study of Buck et al. (2008), a correlation between the Stroop Colour-Word Test and the 20-m shuttle run test. In the studies performing an hierarchical regression analysis, the standardized β coefficient varied between nonsignificant and .30 in the same study of Buck et al. (2008). Effect sizes were not available from the intervention studies.

## 3.6 | Inhibition

### 3.6.1 | Response control

Higher fit children responded more quickly and accurately compared with their lower fit peers during the visual-odd ball task (Hillman et al., 2005).

### 3.6.2 | Interference control

Greater levels of AF were associated with better task performance at the Stroop Colour-Word Test, regardless of the required amount of cognitive inhibition and after adjustment for age, intelligence quotient (IQ), and body mass index (BMI; Buck et al., 2008). A more recent and more powerful study partially confirmed this positive association, albeit only in boys and not girls (Aadland et al., 2017). In the intervention study of Castelli et al. (2011), only the increase in the number of words read in the complex colour-word condition of the Stroop Colour-Word Test after intervention was significantly related to the mean time above the target heart zone during the intervention. None of the Stroop Colour-Word Test conditions were significantly related to the mean time in, or below the target heart zone.

Sixteen studies used a modified Eriksen flanker task, although different conditions and different outcome measures were used. Nine studies compared a group of high-fit children with a group of lower fit peers (Berchicci et al., 2015; Chaddock et al., 2010; Chaddock, Erickson, et al., 2012; Chaddock, Hillman, et al., 2012; Hillman et al., 2009; Moore et al., 2013; Pontifex et al., 2012; Voss et al., 2011; Wu et al., 2011). These studies reported consistently more accurate responses for higher fit children compared with lower fit peers regardless of the required amount of cognitive control. Only one study by Chaddock, Hillman, et al. (2012) did not establish this response accuracy difference between groups during the compatible trails but did also find them during the incompatible trails, which required a higher amount of cognitive control. Additionally, higher fit children showed less response accuracy interference cost (Voss et al., 2011) and less percent response time interference (smaller difference between congruent and incongruent condition) than lower fit children (Chaddock et al., 2010). No significant interactions could be established between response time and AF. Six studies, adopting an Eriksen flanker task, were correlational studies (Buck et al., 2008; Chojnacki et al., 2018; Moore et al., 2013; Pindus et al., 2016; Pontifex et al., 2014; Scudder et al., 2014; Scudder et al., 2016). Only two of the six studies reported an increasing accuracy with increasing levels of AF (Pontifex et al., 2014; Scudder et al., 2014). Additionally, one study reported quicker and less variable response time with increasing levels of AF (Scudder et al., 2014), and one study reported less intraindividual variability with increasing levels of AF in the incongruent condition (Scudder et al., 2014). The only intervention study that investigated the modified Eriksen flanker task could not show a better performance in two intervention groups, compared with a control group although both intervention groups showed increased AF compared with the control group (Schmidt et al., 2015). The reader should note that this study used the difference in response time between congruent and incongruent trials, which is an uncommon choice to use this measure as the unique outcome measure for the flanker task.

## 3.7 | Working memory

Only one comparative study regarding working memory was included in the present review. They reported an increased response accuracy for relational encoding for higher fit children compared with lower fit peers. No significant fitness based differences were reported for nonrelational encoding (Chaddock et al., 2011). Eight out of the nine included correlational studies confirmed these results. Increasing levels of AF were positively related to more accurate responses (Aadland et al., 2017; Drollette et al., 2016; Geertsen et al., 2016; Kao et al., 2017; Scudder et al., 2014; Scudder et al., 2016). This association was stronger when working memory demands were increased during the 2-back conditions, regardless of the task that was used (Scudder et al., 2014). The current findings provide supporting evidence regarding the selectivity hypothesis, which states that there is a stronger association in tasks that require a greater amount of executive control. However, an important side note has to be made as studies that separated the results of boys and girls could only reproduce this positive relation in boys and not girls (Aadland et al., 2017; Drollette et al., 2016). In contrast with the study of Drollette et al. (2016), Pindus et al. (2016), adopting the same EF task, could not find a significant association between AF and the OSPAN task and could not confirm the better performance of boys compared with girls on the OSPAN task. However, although boys did not outperform girls significantly based on an independent *t* test, a linear regression analysis revealed a main effect of sex in the study of Drollette et al. (2016). This finding illustrates the difficulty to control possible confounders in cross-sectional studies.

Two intervention studies evaluated working memory and differentiated between a programme offering mainly cardiovascular exercises and a PA programme encompassing also team games and/or motor skill training. The study by Schmidt et al. (2015) used a nonspatial N-back task to compare in a short 6-week programme an intervention group focusing on aerobic exercises, a group focusing on team games, and a control group. Both intervention groups showed a similar increase in AF, compared with the stable level of AF in the control group. No difference in working memory could be shown between the three groups (Schmidt et al., 2015). Koutsandreou, Wegner, Niemann, and Budde, (2016) differentiated within a 10-week PA programme between a cardiovascular exercise group and a motor exercise group and compared them with a control group. Both intervention groups benefited from their PA programme regarding working memory performance (measured by a letter digit span task), opposed to the control group. However, the increase in response accuracy was significantly larger for children in the motor exercise group compared with the cardiovascular exercise group.

## 3.8 | Cognitive flexibility

Three included studies that researched cognitive flexibility were correlational studies (Aadland et al., 2017; Pontifex et al., 2014; van der Niet et al., 2014). Results were similar to the findings in other EF domains. Shorter response time during the Trail Making Test (van der Niet et al., 2014) and increased accuracy on the heterogeneous condition of the colour-shape switch task (Pontifex et al., 2014) were associated with AF. In the latter study, this relation was not present in the homogeneous condition, which suggests that the relation is only notable in tasks requiring a higher amount of cognitive flexibility (Pontifex et al., 2014). Aadland et al. (2017) reported increasing task performance in the verbal fluency task with increasing PF. In the intervention study of Castelli et al. (2011), only the decrease in time in the complex condition B of the Trail Making Test after intervention was significantly related to the mean time above the target heart zone during the intervention. None of the conditions of the Trail Making Test were significantly related to the mean time in, or below the target heart zone.

## 3.9 | Possible confounders

All reviewed studies described some variables that could possibly interact as confounding factors, except Berchicci et al. (2015). Nevertheless, the implementation of these factors in the analyses was diverse. In the comparative studies, a lack of significant difference between the high-fit and low-fit groups for a specific factor was used to control that possible confounder. In the correlational studies, a lack of significant correlation between the possible confounder and EF was mostly used to disregard the factor in further analysis. However, not all studies controlled for all possible confounders.

AF was reported to be positively related to age (Drollette et al., 2016; Scudder et al., 2014; Scudder et al., 2016), IQ (Drollette et al., 2016), and socio-economic status (SES; Drollette et al., 2016; Pontifex et al., 2014; Scudder et al., 2014; Scudder et al., 2016) and negatively related to BMI or % adiposity (Buck et al., 2008; Pontifex et al., 2014; Scudder et al., 2014; Scudder et al., 2016). Boys were fitter compared with girls (Aadland et al., 2017; Drollette et al., 2016; Pindus et al., 2016; Pontifex et al., 2014; Scudder et al., 2014; Scudder et al., 2016; van der Niet et al., 2014). About half of the studies controlled the pubertal status of the children (Aadland et al., 2017; Castelli et al., 2011; Chaddock et al., 2011; Chaddock, Erickson, et al., 2012; Chaddock, Hillman, et al., 2012; Chojnacki et al., 2018; Koutsandreou et al., 2016; Pontifex et al., 2011; Pontifex et al., 2012; Pontifex et al., 2014; Schmidt et al., 2015; Voss et al., 2011; Wu et al., 2011). Pontifex et al. (2014) identified a positive relation between AF and the Tanner scale.

The relation between EF tasks and possible confounders has been reported less frequently. If a significant relation was identified, this was only in part of the tasks and/or outcome measures. In that way, EF were positively related to age (Buck et al., 2008; Castelli et al., 2011; Chojnacki et al., 2018; Kao et al., 2017; Pontifex et al., 2014) and SES (Chojnacki et al., 2018; Kao et al., 2017; Pontifex et al., 2014) and negatively related to BMI (Chojnacki et al., 2018; Pontifex et al., 2014). Some differences were found between boys and girls regarding EF but not exclusively in one direction (Aadland et al., 2017; Chojnacki et al., 2018; Drollette et al., 2016; Kao et al., 2017; Pindus et al., 2016; Pontifex et al., 2014).

### 3.10 | Quality assessment

Table 5 reports the quality assessment. Nineteen out of 26 studies noted good quality, with a score of at least 80%. However, because of convenience sampling, all studies scored 0/2 on external validity, with the exception of one (Aadland et al., 2017). Sampling bias could be present in 18 studies that recruited children from a similar region (East Central Illinois, USA), in a similar way (Berchicci et al., 2015; Buck et al., 2008; Castelli et al., 2011; Chaddock et al., 2010; Chaddock et al., 2011; Chaddock, Erickson, et al., 2012; Chaddock, Hillman, et al., 2012; Chojnacki et al., 2018; Drollette et al., 2016; Hillman et al., 2005; Hillman et al., 2009; Kao et al., 2017; Moore et al., 2013; Pindus et al., 2016; Voss et al., 2011; Wu et al., 2011). Hence, overlapping data may be present between some articles. The most frequently reported problem in the studies with lower quality scores was insufficient description of patient characteristics and insufficient assessment of possible confounders, such as SES, pubertal stage, and IQ. Out of the three interventional studies, merely the study of Schmidt et al. (2015) was of good quality.

## 4 | DISCUSSION

Present, several PA programmes for children focus on AF as, for example, the “Daily Mile” programme for schools. These programmes suppose that the increase in AF explains how PA impacts cognition. Based on this review, the level of evidence for the effect of enhanced AF on EF in typically developing children is low. Therefore, further exploration of the effect of programmes focusing on AF compared with more cognitive engaging PA programmes are necessary. Only three intervention studies fulfilled the selection criteria for this review. In two of these studies, aerobic exercise was compared with cognitive more engaging PA. The study of Schmidt et al. (2015) investigated only a short intervention programme and could not reveal effects on EF, except for cognitive flexibility, and this only in the group exercising with high cognitive engagement. Hence, from this study, AF enhancement does not seem to be responsible for increased cognitive flexibility following a PA programme. In contrast, the study of Kousandréou et al. (2015) revealed a positive effect of a 10-week programme on working memory in both training groups, opposed to a control group, and reported a larger effect size for the motor training group compared with the cardiovascular exercise group. Hence, they suggested that a mediating role of enhanced AF was present, although the cognitive stimulation during the motor exercise programme seemed to have an additional effect. The study of Castelli et al. (2011) evaluated a mixed PA programme, not allowing to identify the mediating role of AF. But they tried to relate the gain in EF to the time spend in the target heart zone during intervention to isolate the effect of the cardiovascular effort. Only in the complex EF test conditions and only in the “mean time above the target heart zone” a significant correlation was found. We can conclude that unless the available evidence for the positive effect of PA programmes on different aspects of cognition, the mediating role of enhanced AF in this effect remains unclear.

On the other hand, although results were not always conclusive and effect sizes or measures of association were only small to moderate, the majority of the selected cross-sectional studies confirmed some positive relations between AF and the different EF. Different test protocols and outcome measures to evaluate AF and EF make it difficult to compare studies. Most studies used AF tests, which were validated to evaluate or estimate $VO_2$ max, although it cannot be excluded that the variety in AF test protocols had possibly an impact on the study findings. Without doubt, the variety in paradigms, protocols, and outcome measures to evaluate EF interferes when comparing studies to each other. Even in the 15 studies using the Eriksen flanker test paradigm, protocols lacked uniformity. Methodological and reporting quality of the included cross-sectional studies was moderate to high in the majority of studies. Selection bias, insufficient control of confounders, and absence of an independent assessment of outcome were the most common methodological flaws. Most studies included only a small sample size. Although the current results supports the evidence for an overall positive relation between EF and AF in children, the diversity in the results reinforce "the unity and diversity" hypothesis and the need to differentiate between different executive functioning domains (Friedman & Miyake, 2016).

## 4.1 | Inhibition

There is converging evidence that AF is positively related to response accuracy in the Eriksen flanker task. When analysing results of the congruent and incongruent trials separately, it appeared that tasks requiring a higher level of inhibition were more sensitive to changes related to fitness levels. Furthermore, results regarding response time in the Eriksen flanker test were more inconclusive. Interference cost scores in the Eriksen flanker task, for both response accuracy and response time, were most effective in revealing significant fitness based differences between groups. Although both studies evaluating inhibition by mean of the Stroop Colour-Word test confirmed a positive relation with AF, Aadland et al. (2017) suggest that this relation is gender specific and that it is only valid for boys. The findings that increased AF are associated with less behavioural interference to misleading and irrelevant cues and are in agreement with results found in adults (Bherer, Erickson, & Liu-Ambrose, 2013).

## 4.2 | Working memory

Positive effects between AF and working memory were established in five included studies, with three different tasks. Task accuracy showed again to be the most responsive parameter. Similar to the inhibition results, the relation showed to be stronger in tasks with increased working memory demands. Noteworthy, both studies, which researched the results of boys and girls separately, revealed important differences regarding the relation between working memory and AF (Aadland et al., 2017; Drollette et al., 2016), indicating that an increase in levels of fitness improved working memory performance in boys but not in girls. Nevertheless, future research is needed to clarify this gender difference and makes thorough conclusions.

## 4.3 | Cognitive flexibility

Although the findings on cognitive flexibility are in line with the studies evaluating inhibition and working memory, they should be interpreted with more caution as the results are merely based on five studies adopting four different paradigms. Higher fit children seemed to have a superior ability to flexible allocate cognitive control processes and alter their strategies to effectively meet task demands (Aadland et al., 2017; Pontifex et al., 2012; Schmidt et al., 2015; van der Niet et al., 2014). In this domain of EF, Aadland et al. (2017) identified a gender difference, and the relation between AF and cognitive flexibility could only be identified in boys as well.

## 4.4 | Confounders

Several demographical factors were considered as confounders in the relation between AF and EF. Again, the diversity in the different studies regarding these confounders makes it difficult to make conclusions. From the studies of Aadland et al. (2017) and Drollette et al. (2016), it became clear that gender is a possible confounder. Age was in most studies controlled by the small age range of the studies, but likely, the relation between AF and EF is age dependent, and pubertal status is a mediating factor. BMI, SES, and IQ seemed to be possible determinants as well. A limitation to all studies, although important to compare studies, was the lack of reporting the sequence of the different assessments and the time between EF and AF assessments. This is necessary to adequately compare the results regarding to possible fatigue after the AF measurement or to take the acute bout of exercise into the equation (Ludyga, Gerber, Brand, Holsboer-Trachsler, & Puhse, 2016).

### 4.5 | Strengths, limitations, and recommendations for future research

One the one hand, the limited focus on the relation between AF and EF in this review allows to disentangle at least one component of the complex relation between PA and cognition. On the other hand, this limited focus leads to other important aspects not being addressed: the possible effect of other fitness capacities as, for example, muscular fitness that might also be related to EF and the acute effect of physical exertion. And not considering studies investigating the possible neural underlying mechanisms can also be seen as a limitation to this review.

Good quality randomized clinical trials comparing PA programmes, with and without cognitive engagement, are the only way to unravel the mediating role of AF. Lack of these studies made it impossible to come to strong conclusions from this review. We can only address the necessity and importance to conduct more of these studies. Additionally, PA programmes should be of sufficient length and intensity to allow enhancement of AF. Furthermore, longitudinal studies are needed to explore the long-term effect of childhood AF on later executive functioning. Based on the reported gender differences, it is recommended to control future studies for gender, SES, BMI, IQ, age, and pubertal status. And at last, as EF has been supposed to be crucial in learning, not only to typical developing children but also to children with attention deficit hyper activity disorder (Craig et al., 2016), developmental coordination disorder (Leonard, Bernardi, Hill, & Henry, 2015), and autism spectrum disorder (Craig et al., 2016), more studies with these groups of children are also necessary. Not at least because the relationship between PA, EF, and learning is probably even more complex in children with neurodevelopmental disorders.

## 5 | CONCLUSION

The importance of regular PA to the developing body and mind cannot be overestimated. Results suggest that AF is positively related to inhibition, working memory, and probably also to cognitive flexibility, although not all results were conclusive and effect sizes limited. Good quality randomized controlled trials that unravel the association between AF and the different domains of executive functioning are missing. Nevertheless, they are essential to determine whether the focus should be on cognitive demanding PA, teaching adaptive goal-directed problem-solving skills or if the AF component is the most important factor and the heart rate zone needs to be monitored to exercise at the physiological most efficient level.


### ACKNOWLEDGEMENTS

The authors would like to thank the children, their parents, and the WAVO school for their enthusiastic participation and Ellen Devos, Loreen Van den Bulcke, Dieter Van de Walle, and Nathalie Van Gelder for assistance with data collection.


## CONFLICT OF INTERESTS

The authors report no conflicts of interest.


### ORCID

*Hilde Van Waelvelde* https://orcid.org/0000-0002-7330-7615

## REFERENCES

Aadland, K. N., Moe, V. F., Aadland, E., Anderssen, S. A., Resaland, G. K., & Ommundsen, Y. (2017). Relationships between physical activity, sedentary time, aerobic fitness, motor skills and executive function and academic performance in children. *Mental Health and Physical Activity*, *12*, 10–18. https://doi.org/10.1016/j.mhpa.2017.01.001

Aubut, J. A., Marshall, S., Bayley, M., & Teasell, R. W. (2013). A comparison of the PEDro and Downs and Black quality assessment tools using the acquired brain injury intervention literature. *NeuroRehabilitation*, *32*(1), 95–102. https://doi.org/10.3233/nre-130826

Berchicci, M., Pontifex, M. B., Drollette, E. S., Pesce, C., Hillman, C. H., & di Russo, F. (2015). From cognitive motor preparation to visual processing: The benefits of childhood fitness to brain health. *Neuroscience*, *298*, 211–219. https://doi.org/10.1016/j.neuroscience.2015.04.028

Best, J. R. (2010). Effects of physical activity on children's executive function: Contributions of experimental research on aerobic exercise. *Developmental Review*, *30*(4), 331–551. https://doi.org/10.1016/j.dr.2010.08.001

Bherer, L., Erickson, K. I., & Liu-Ambrose, T. (2013). A review of the effects of physical activity and exercise on cognitive and brain functions in older adults. *Journal of Aging Research*, *2013*, 657508. https://doi.org/10.1155/2013/657508

Bidzan-Bluma, I., & Lipowska, M. (2018). Physical activity and cognitive functioning of children: A systematic review. *International Journal of Environmental Research and Public Health*, *15*(4). https://doi.org/10.3390/ijerph15040800

Buck, S. M., Hillman, C. H., & Castelli, D. M. (2008). The relation of aerobic fitness to Stroop task performance in preadolescent children. *Medicine and Science in Sports and Exercise*, *40*(1), 166–172. https://doi.org/10.1249/mss.0b013e318159b035

Castelli, D. M., Hillman, C. H., Hirsch, J., Hirsch, A., & Drollette, E. (2011). FIT kids: Time in target heart zone and cognitive performance. *Preventive Medicine*, *52*(Suppl 1), S55–S59. https://doi.org/10.1016/j.ypmed.2011.01.019

Chaddock, L., Erickson, K. I., Prakash, R. S., VanPatter, M., Voss, M. W., Pontifex, M. B., … Kramer, A. F. (2010). Basal ganglia volume is associated with aerobic fitness in preadolescent children. *Developmental Neuroscience*, *32*(3), 249–256. https://doi.org/10.1159/000316648

Chaddock, L., Erickson, K. I., Prakash, R. S., Voss, M. W., VanPatter, M., Pontifex, M. B., … Kramer, A. F. (2012). A functional MRI investigation of the association between childhood aerobic fitness and neurocognitive control. *Biological Psychology*, *89*(1), 260–268. https://doi.org/10.1016/j.biopsycho.2011.10.017

Chaddock, L., Hillman, C. H., Buck, S. M., & Cohen, N. J. (2011). Aerobic fitness and executive control of relational memory in preadolescent children. *Medicine and Science in Sports and Exercise*, *43*(2), 344–349.

Chaddock, L., Hillman, C. H., Pontifex, M. B., Johnson, C. R., Raine, L. B., & Kramer, A. F. (2012). Childhood aerobic fitness predicts cognitive performance one year later. *Journal of Sports Sciences*, *30*(5), 421–430.

Chojnacki, M. R., Raine, L. B., Drollette, E. S., Scudder, M. R., Kramer, A. F., Hillman, C. H., & Khan, N. A. (2018). The negative influence of adiposity extends to intraindividual variability in cognitive control among preadolescent children. *Obesity*, *26*(2), 405–411. https://doi.org/10.1002/oby.22053

Craig, F., Margari, F., Legrottaglie, A. R., Palumbi, R., de Giambattista, C., & Margari, L. (2016). A review of executive function deficits in autism spectrum disorder and attention-deficit/hyperactivity disorder. *Neuropsychiatric Disease and Treatment*, *12*, 1191–1202. https://doi.org/10.2147/ndt.s104620

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

Dencker, M., Bugge, A., Hermansen, B., & Andersen, L. B. (2010). Objectively measured daily physical activity related to aerobic fitness in young children. *Journal of Sports Sciences*, *28*(2), 139–145. https://doi.org/10.1080/02640410903460726

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

Diamond, A., & Ling, D. S. (2018). Aerobic-exercise and resistance-training interventions have been among the least effective ways to improve executive functions of any method tried thus far. *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2018.05.001

Donnelly, J. E., Hillman, C. H., Castelli, D., Etnier, J. L., Lee, S., Tomporowski, P., … Szabo-Reed, A. N. (2016). Physical activity, fitness, cognitive function, and academic achievement in children: A systematic review. *Medicine and Science in Sports and Exercise*, *48*(6), 1197–1222. https://doi.org/10.1249/mss.0000000000000901

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, *52*(6), 377–384.

Drollette, E. S., Scudder, M. R., Raine, L. B., Moore, R. D., Pontifex, M. B., Erickson, K. I., & Hillman, C. H. (2016). The sexual dimorphic association of cardiorespiratory fitness to working memory in children. *Developmental Science*, *19*(1), 90–108. https://doi.org/10.1111/desc.12291

Fedewa, A. L., & Ahn, S. (2011). The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: A meta-analysis. *Research Quarterly for Exercise and Sport*, *82*(3), 521–535. https://doi.org/10.1080/02701367.2011.10599785

Friedman, N. P., & Miyake, A. (2016). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*. https://doi.org/10.1016/j.cortex.2016.04.023

Geertsen, S. S., Thomas, R., Larsen, M. N., Dahn, I. M., Andersen, J. N., Krause-Jensen, M., … Lundbye-Jensen, J. (2016). Motor skills and exercise capacity are associated with objective measures of cognitive functions and academic performance in preadolescent children. *PLoS ONE*, *11*(8).

Haapala, E. A. (2013). Cardiorespiratory fitness and motor skills in relation to cognition and academic performance in children—A review. *Journal of Human Kinetics*, *36*, 55–68. https://doi.org/10.2478/hukin-2013-0006

Hillman, C. H., Buck, S. M., Themanson, J. R., Pontifex, M. B., & Castelli, D. M. (2009). Aerobic fitness and cognitive development: Event-related brain potential and task performance indices of executive control in preadolescent children. *Developmental Psychology*, *45*(1), 114–129. https://doi.org/10.1037/a0014437

Hillman, C. H., Castelli, D. M., & Buck, S. M. (2005). Aerobic fitness and neurocognitive function in healthy preadolescent children. *Medicine and Science in Sports and Exercise*, *37*(11), 1967–1974. https://doi.org/10.1249/01.mss.0000176680.79702.ce

Kao, S. C., Westfall, D. R., Parks, A. C., Pontifex, M. B., & Hillman, C. H. (2017). Muscular and aerobic fitness, working memory, and academic achievement in children. *Medicine and Science in Sports and Exercise*, *49*(3), 500–508. https://doi.org/10.1249/MSS.0000000000001132

Keeley, T. J., & Fox, K. R. (2009). The impact of physical activity and fitness on academic achievement and cognitive performance in children. *International Review of Sport and Exercise Psychology*, *2*(2), 6.

Khan, N. A., & Hillman, C. H. (2014). The relation of childhood physical activity and aerobic fitness to brain function and cognition: A review. *Pediatric Exercise Science*, *26*(2), 138–146. https://doi.org/10.1123/pes.2013-0125

Koutsandreou, F., Wegner, M., Niemann, C., & Budde, H. (2016). Effects of motor versus cardiovascular exercise training on children's working memory. *Medicine and Science in Sports and Exercise*, *48*(6), 1144–1152. https://doi.org/10.1249/mss.0000000000000869

Lang, J. J., Wolfe Phillips, E., Orpana, H. M., Tremblay, M. S., Ross, R., Ortega, F. B., … Tomkinson, G. R. (2018). Field-based measurement of cardiorespiratory fitness to evaluate physical activity interventions. *Bulletin of the World Health Organization*, *96*(11), 794–796. https://doi.org/10.2471/blt.18.213728

Leonard, H. C., Bernardi, M., Hill, E. L., & Henry, L. A. (2015). Executive functioning, motor difficulties, and developmental coordination disorder. *Developmental Neuropsychology*, *40*(4), 201–215. https://doi.org/10.1080/87565641.2014.997933

Ludyga, S., Gerber, M., Brand, S., Holsboer-Trachsler, E., & Puhse, U. (2016). Acute effects of moderate aerobic exercise on specific aspects of executive function in different age and fitness groups: A meta-analysis. *Psychophysiology*, *53*(11), 1611–1626. https://doi.org/10.1111/psyp.12736

McGivern, R. F., Andersen, J., Byrd, D., Mutter, K. L., & Reilly, J. (2002). Cognitive efficiency on a match to sample task decreases at the onset of puberty in children. *Brain and Cognition*, *50*(1), 73–89. https://doi.org/10.1016/s0278-2626(02)00012-x

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264–269, w264. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Moore, R. D., Wu, C. T., Pontifex, M. B., O'Leary, K. C., Scudder, M. R., Raine, L. B., … Hillman, C. H. (2013). Aerobic fitness and intra-individual variability of neurocognition in preadolescent children. *Brain and Cognition*, *82*(1), 43–57. https://doi.org/10.1016/j.bandc.2013.02.006

Nowrangi, M. A., Lyketsos, C., Rao, V., & Munro, C. A. (2014). Systematic review of neuroimaging correlates of executive functioning: Converging evidence from different clinical populations. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *26*(2), 114–125. https://doi.org/10.1176/appi.neuropsych.12070176

Pindus, D. M., Drollette, E. S., Scudder, M. R., Khan, N. A., Raine, L. B., Sherar, L. B., … Hillman, C. H. (2016). Moderate-to-vigorous physical activity, indices of cognitive control, and academic achievement in preadolescents. *Journal of Pediatrics*, *173*, 136–142. https://doi.org/10.1016/j.jpeds.2016.02.045

Pontifex, M. B., Kamijo, K., Scudder, M. R., Raine, L. B., Khan, N. A., Hemrick, B., … Hillman, C. H. (2014). The differential association of adiposity and fitness with cognitive control in preadolescent children. *Monographs of the Society for Research in Child Development*, *79*(4), 72–92. https://doi.org/10.1111/mono.12131

Pontifex, M. B., Raine, L. B., Johnson, C. R., Chaddock, L., Voss, M. W., Cohen, N. J., … Hillman, C. H. (2011). Cardiorespiratory fitness and the flexible modulation of cognitive control in preadolescent children. *Journal of Cognitive Neuroscience*, *23*(6), 1332–1345. https://doi.org/10.1162/jocn.2010.21528

Pontifex, M. B., Scudder, M. R., Drollette, E. S., & Hillman, C. H. (2012). Fit and vigilant: The relationship between poorer aerobic fitness and failures in sustained attention during preadolescence. *Neuropsychology*, *26*(4), 407–413. https://doi.org/10.1037/a0028795

Schmidt, M., Jager, K., Egger, F., Roebers, C. M., & Conzelmann, A. (2015). Cognitively engaging chronic physical activity, but not aerobic exercise, affects executive functions in primary school children: A group-randomized controlled trial. *Journal of Sport & Exercise Psychology*, *37*(6), 575–591. https://doi.org/10.1123/jsep.2015-0069

Scudder, M. R., Drollette, E. S., Szabo-Reed, A. N., Lambourne, K., Fenton, C. I., Donnelly, J. E., & Hillman, C. H. (2016). Tracking the relationship between children's aerobic fitness and cognitive control. *Health Psychology*, *35*(9), 967–978. https://doi.org/10.1037/hea0000343

Scudder, M. R., Lambourne, K., Drollette, E. S., Herrmann, S. D., Washburn, R. A., Donnelly, J. E., & Hillman, C. H. (2014). Aerobic capacity and cognitive control in elementary school-age children. *Medicine and Science in Sports and Exercise*, *46*(5), 1025–1035. https://doi.org/10.1249/mss.0000000000000199

van der Niet, A. G., Hartman, E., Smith, J., & Visscher, C. (2014). Modeling relationships between physical fitness, executive functioning, and academic achievement in primary school children. *Psychology of Sport and Exercise*, *15*(4), 319–325. https://doi.org/10.1016/j.psychsport.2014.02.010

Vandenbroucke, L., Seghers, J., Verschueren, K., Wijtzes, A. I., & Baeyens, D. (2016). Longitudinal associations between objectively measured physical activity and development of executive functioning across the transition to first grade. *Journal of Physical Activity & Health*, *13*(8), 895–902. https://doi.org/10.1123/jpah.2015-0708

Vazou, S., Pesce, C., Lakes, K., & Smiley-Oyen, A. (2016). More than one road leads to Rome: A narrative review and meta-analysis of physical activity intervention effects on cognition in youth. *International Journal of Sport and Exercise Psychology*, 1–26. https://doi.org/10.1080/1612197X.2016.1223423

Voss, M. W., Chaddock, L., Kim, J. S., Vanpatter, M., Pontifex, M. B., Raine, L. B., … Kramer, A. F. (2011). Aerobic fitness is associated with greater efficiency of the network underlying cognitive control in preadolescent children. *Neuroscience*, *199*, 166–176. https://doi.org/10.1016/j.neuroscience.2011.10.009

World Health Organization. Global strategy on diet, physical activity and health. Retrieved from: on https://www.who.int/dietphysicalactivity/pa/en/ 29/8/2019.

Wu, C. T., Pontifex, M. B., Raine, L. B., Chaddock, L., Voss, M. W., Kramer, A. F., & Hillman, C. H. (2011). Aerobic fitness and response variability in preadolescent children performing a cognitive control task. *Neuropsychology*, *25*(3), 333–341. https://doi.org/10.1037/a0022167

**How to cite this article:** Van Waelvelde H, Vanden Wyngaert K, Mariën T, Baeyens D, Calders P. The relation between children's aerobic fitness and executive functions: A systematic review. *Inf Child Dev*. 2020;29:e2163. https://doi.org/10.1002/icd.2163