# Introduction

School-based physical activity (PA) interventions are often rapidly disseminated because of their harmonious fit into the school day and perceived contributions to children's daily PA, even with limited evidence of their educational impact (Fairhurst & Hotham, 2017; Forman et al., 2008). The appeal of under-investigated interventions seems appropriate considering they may, quite obviously, address children's need for more daily PA (Chief Medical Officers., 2011; Department of Health and Social Care, 2018; Tremblay et al., 2016). One novel intervention, The Daily Mile™ (TDM), has recently been (i) implemented across thousands of primary schools both nationally and internationally (The Daily Mile Foundation, 2019), (ii) rolled out nationwide in Scotland (Scottish Government., 2018) and (iii) recommended for all UK primary schools in the new Childhood Obesity Plan: Part II (Department of Health and Social Care, 2018). Nevertheless, evidence of TDM's effectiveness is limited.

To date, only one published study has assessed the long-term effectiveness of TDM, concluding positive changes on PA levels and body fat percentage (Chesham et al., 2018). Closer inspection of this study reveals the low-quality methodological design. As highlighted by Daly-Smith et al. (2019), the study used different exposure durations for the control and intervention groups, resulting in unequal dose-response conditions and creating inequitable opportunities for benefits. The study's final sample size was small, meaning the findings cannot be generalised, and no attempt was made to confirm the treatment dose of PA during TDM, reducing confidence in the outcomes (Daly-Smith et al., 2019). Consequently, recommending TDM based on this one weak study seems premature. More rigorous research is needed to confidently establish both the short- and long-term effectiveness of TDM on health, cognitive and academic outcomes for children.

TDM seems to be driven by a powerful social media presence, claiming a host of mostly unsubstantiated beneficial acute effects, including ‘greater focus and concentration levels in the classroom’ (The Daily Mile Foundation, 2019). To date, there is no high-quality evidence to base these anecdotal claims. Despite this, TDM has become a favoured approach for schools and teachers providing an active 15-minute break away from the classroom. While acute bouts of PA – similar to TDM – seem to promote some, but inconsistently, small positive improvements in executive functions (EF; top-down mental processes essential in everyday life) and academic performance (AP), these findings arise from laboratory-based studies (Drollette et al., 2012; Hillman et al., 2009b). In response, it is imperative to assess these claims in practical settings as they are being used to continue to push the wide-spread adoption of TDM.

Recent systematic reviews summarising the effect of acute PA on EFs and/or AP are inconclusive at best (Chang et al., 2012; Daly-Smith, 2018; Daly-Smith et al., 2018; Donnelly et al., 2016); the dominance of low-medium quality studies undermines confidence in their outcomes (Daly-Smith et al., 2018). Common problems with studies assessing the acute effects of PA on EF in schools derive from not: (i) reporting the process used to randomise participants, (ii) including a familiarisation protocol and (iii) confirming the treatment fidelity by objectively assessing PA levels at the individual level (Daly-Smith et al., 2018). Naturally, a body of empirical evidence using high-quality translational study designs is warranted to assess PA interventions such as TDM.

Similar to other acute PA interventions, for TDM to affect AP, a corresponding effect would be required on the underpinning cognitive processes, namely EF’s (Tomporowski et al., 2015). EFs refer to a set of top-down mental processes allowing for goal-directed, purposeful and controlled behaviour (Diamond, 2013). EFs have three core dimensions (Diamond, 2013; Miyake et al., 2000). *Inhibitory control* refers to one’s ability to control their attention, behaviours and thoughts (Macdonald et al., 2014). *Working memory* refers to the ability to hold information in one's mind and mentally work with it (Diamond, 2013). *Cognitive flexibility* refers to the ability to change perspective spatially and interpersonally (Diamond, 2013). Broadly, for children, positive development of EFs has been associated with better classroom behaviour, grades and academic achievement (Howie & Pate, 2012; Tomporowski et al., 2015). More specifically, an abundance of research suggests EFs – specifically inhibitory control and working memory – underly the 'domain-specific competencies' of mathematical achievements, including factual knowledge, procedural skills and conceptual understanding (Cragg et al., 2017; Geary et al., 2012). However, to the best of our knowledge, only laboratory-based, high-quality studies have assessed acute effectiveness of quantitative PA (e.g., running, Pesce, 2012) on mathematical performance (Hillman et al., 2009b; Pontifex et al., 2013).

There remains a limited understanding of the acute effects of run-walk initiatives within the school setting. Studies that have explored the acute effects of PA interventions on EF and AP provide inconsistent conclusions. Interventions on classroom exercise breaks identified 10 minutes of moderate-to-vigorous PA (MVPA) were sufficient for EF improvements (Howie et al., 2014; Janssen et al., 2014). In contrast, a meta-analysis (Chang et al., 2012) suggests 20 minutes or more of quantitative PA is required to enhance cognition. To date, there are no finite conclusions on the optimal duration of PA needed to elicit enhancements on EF and AP. A similar pattern emerges across children's characteristics (Hillman et al., 2009a; Tomporowski et al., 2008; Vazou & Smiley-Oyen, 2014). For example, previous studies identified differing impacts of PA on EFs were moderated by children's current fitness level and academic achievements (Chang et al., 2012; Jager et al., 2015). Further investigations are warranted to increase the evidence base on run-walk programmes such as TDM, as well as consider potential moderating variables and intervention components that may affect the impact of PA on EFs and AP.

It is also imperative to understand the individual dose of MVPA during TDM (Daly-Smith et al., 2018). Unlike a laboratory-based study design, using tightly controlled environments, a real-world setting creates variability in physical engagement levels during an intervention. Due to the unsubstantiated claims of TDM's benefits (The Daily Mile Foundation, 2019), it is equally important to explore the effects - if any - of EFs and AP within a classroom environment instead of a laboratory. If a specific dose of MVPA is required to evoke cognitive benefits, treating children as a homogenous group is not appropriate as there may not be universal benefits.

In the first instance, it is essential to prioritise the immediate impact of TDM before progressing to longer-term investigations and broader implications. Understanding the acute effects may help inform teachers to strategically use TDM – depending on the findings - to boost performance in the subsequent lessons. Assessing the acute effect may also enable teachers to feel more confident in supporting the initial adoption of a PA intervention without fear of a reduction in immediate performance in the classroom. Using high-quality study designs in a translational setting are needed to confidently determine the efficacy and effectiveness to justify widespread, evidence-based, endorsement and adoption (Flay et al., 2005). Findings on such outcomes could influence the future promotion of TDM and support future implementation and tailoring of the intervention.

Therefore, this study has three objectives. First, ascertain the amount of MVPA from one session of TDM. Second, any acute effects of TDM on children's EF and maths fluency versus usual academic lessons. Third, identify how individual levels of PA engagement through TDM influences EF and maths fluency scores versus usual academic lessons.

# Methodology

## Design

A randomised control trial was conducted using a between-subjects design. Children were randomly assigned to one of two conditions: (i) TDM and (ii) control – continuing with the classroom-based academic lesson. EFs and maths fluency were assessed before (pre-test) and immediately after the intervention (post-test). Conditions were conducted at the same time, completing the tests together, as a class, taking place in the children's classroom. Children's PA levels were assessed with accelerometers to evaluate the treatment fidelity of TDM.

## Subjects

A total sample of 422 primary school-aged children gave assent, and their parents/guardians gave consent to participate in the study, with a final sample of 303 children providing a complete profile (see Figure 1, CONSORT flow diagram). Ethical approval was obtained by Leeds Beckett University Ethics Committee (Ref:38381), and consent was obtained from every school's headteacher. An *a priori* power analysis (Faul et al., 2007) was used following Benzing et al. (2016) similar and comparable study design (1 –beta error probability=0.80; alpha error probability=0.05; effect size $f$=0.15; conditions=2, outcome measurements=8; and correlation between the repeated measures $r$=0.75) requiring a minimal sample size of 274. Any child with special educational needs (learning difficulties) was excluded from the final sample size due to differing abilities and impact of cognition (Pontifex et al., 2013). Table 1 provides an overview of children's characteristics, stratified by condition, which did not statistically differ.

## Procedure and Randomisation

Data collection occurred between September 2017 to February 2018. Each class was visited on two occasions, separated by seven days, following afternoon registration (start time *M*=1:17 pm). On day one, children (i) were familiarised with the EF battery of tests and maths fluency test; (ii) had their height, weight, waist and hip circumference measurements taken, and (iii) completed a multistage fitness test (Institute of Medicine, 2012; Léger & Lambert, 1982). Following the familiarisation day, to prevent selection bias, the lead researcher used a 4-block randomisation process to assign children equally to both conditions (Suresh, 2011). Children were stratified by gender and ranked on their familiarisation EF battery and maths fluency scores. The 4-block randomisation process has six options: (1) A-A-B-B, (2) A-B-A-B, (3) A-B-B-A, (4) B-B-A-A, (5) B-A-B-A and (6) B-A-A-B, where 'A' denoted TDM condition and 'B' the control condition. Using an online randomiser tool (https://www.randomizer.org), numbers from one to six were randomly generated, determining which of the six patterns and in what order, would be used to allocate the children to either TDM or control condition.

On the testing day (day seven), children were fitted with accelerometers and completed the pre-EF and maths fluency tests in the same classroom (duration M=35.00±6.95 minutes). Children then completed the condition they were randomised to, either completing TDM in the playground (intervention) or remaining in the classroom for a typical classroom-based lesson (active control). After the children completing TDM returned to the classroom, a five-minute break was administered for children in both conditions for standardisation purposes, before the post EF and maths fluency tests commenced (duration M=33.60±7.46minutes). Following the tests, children removed the accelerometers.

**Experimental Conditions**

For each class, TDM and control condition had matched duration times (M=14.63±2.31). TDM was led by teachers, with two ways of administration: (1) some classes completed TDM precisely as instructed by TDM foundation (a 15-minute break seat-to-seat, focused on time, not distance, M=15.0±0.0minutes, The Daily Mile Foundation, 2019) and (2) some classes were asked to complete TDM under usual provision for that class without any changes (M=14.07±3.01minutes). All testing for TDM condition took place in mild weather conditions when outdoors. One class completed TDM indoors following usual provision due to adverse weather. On two occasions, data collection was postponed due to adverse weather (e.g., snow). The control consisted of academic classroom-based lessons, excluding mathematics to eliminate any new learning that may impact the maths fluency test.

## Manipulation Check Variables

Physical activity was measured using GT9X and GT3X+ accelerometers, using the same engineering structure, matching accuracy, and repeatability (Actigraph, Pensacola, FL). Accelerometers were worn on the right hip, using 100hertz and 15-second epoch lengths. Evenson cut-points (Evenson et al., 2008) were used to provide arbitrary thresholds for sedentary time, light PA (LPA) and MVPA. Children wore the accelerometers all afternoon, and segmented data analyses were conducted to look specifically at the conditions.

## Dependent Variables

### Executive Functions.

Four paper-based tests were used to assess inhibitory control and working memory, adapted from previously valid and reliable tests used with children (Howie et al., 2014; Jager et al., 2014, 2015; Szűcs et al., 2014). Tests were adapted to allow a class of children to complete the tests concurrently as well as providing greater ecological validity. Children were asked to complete as many questions as fast and accurately as possible in a set amount of time. The dependent variables for each test were assessed using the total number correct responses, the total number of errors made (omission and commission) and the total score (correct minus errors). EF tests were completed in the same order: Trail Making Task, Digit Recall, Flanker and Animal Stroop. Test versions were randomised and counterbalanced, essential in minimising order effects.

The Trail Making Task (TMT) is a valid and feasible test (Howie et al., 2014; Lezak et al., 2004) assessing visual search, selective attention, psychomotor speed, working memory and cognitive flexibility (Schiff et al., 2016), providing an overall measure of EF. TMT-A (45 seconds) involved drawing a line between numbers in numerical order (e.g., 1-2-3). The first page consisted of five numbers, one additional number on each subsequent question. TMT-B (45 seconds) involved drawing a line from numbers to letters in numerical and alphabetical order (e.g., 1-A-2-B-3-C). Page one started with four numbers, and four letters, with the next chronological number and letter, added onto the following question. For standardisation, different versions were matched by mirroring each page of the task ensuring numbers and letters were the same distance apart.

The Digit Recall is a validated measure of working memory (Gathercole et al., 2003; Howie et al., 2014) and involved children listening to a pre-recorded standardised sequence of single-digit numbers and writing the sequence down in chronological order within five-seconds. The sequences started with three digits and progressively got harder, finishing with seven-digit sequences. The pre-recorded sequence was played to the whole class to complete the task independently concurrently. Sequence difficulty was matched between versions based on how many numbers the children recorded and how many numbers the children had to re-order to record the numbers chronologically.

The Flanker task has been extensively used to assess inhibitory Control (Eriksen & Eriksen, 1974; Jager et al., 2014). The Flanker task typically is computer-based; modifications were made to allow a class of children to complete the task individually, at the same time. The Flanker consists of three trial blocks, (i) Standard Flanker (ii) Progression

Block and (iii) Mixed Flanker (45 seconds each), all following the consistent rule: circling an arrow matching the direction of the target stimuli (fish). The Standard Flanker consisted of five black fish, and the middle fish was the target stimuli. The Progression Block consisted of five white fish; four outside fish are the target stimuli. The Mixed Flanker combines the previous two blocks. All three blocks consisted of equal weighting on congruent (no interference, fish all facing the same way) and incongruent questions (interference fish facing different directions), which were randomly ordered across all versions.

The Animal Stroop task has been previously validated in assessing inhibitory control (Bryce et al., 2010; Szűcs et al., 2014). The Animal Stroop test has been predominantly conducted in verbal or electronic format. For this study, modifications were made to allow a class to complete the test individually and therefore completed on paper. Children had 30 seconds to complete as many trials as possible. Each trial consisted of placing a cross in a box to correctly identify which animal was larger in real life from a choice of two images for each question. Both versions had equal weighting on congruent (bigger animal in real life and bigger on the page) and incongruent trials (bigger animal in real life but smaller on the page).

**Maths fluency.**

Maths fluency was assessed using a specifically designed assessment ‘Maths Addition and Subtraction, Speed and Accuracy Test’ (MASSAT). The test was based on the UK National Curriculum (Department of Education., 2013). Construct validity was assessed at 100% by three practising maths teachers. Four-day test re-test reliability (r=0.85) was assessed using a sample of 26 children in year five (age 9-10). The MASSAT test consisted of five sets of 15 questions. Each page consisted of a set of questions with three rows (addition, subtraction, and inverse operations) of five questions. Each subsequent page of questions increased in difficulty, defined by the largest number in the question. Using this approach, the five levels of difficulty were; <10, 11-20, 21-50, 51-100, 101-1000. Children were instructed to complete each question, working from left to right, one row at a time. If a child could not answer a specific question, they were required to place an ‘X’ in the answer box before moving to the next question. When scoring, any blank or ‘X’ boxes were classed as an error. The MASSAT provided correct answers, errors, and a total score (correct minus errors).

**Anthropometric and Fitness Variables**

Children’s height, weight, waist, and hip circumferences were taken per the alpha fitness testing battery (Espana-Romero et al., 2010). BMI and BMI z scores were calculated and classified against British 1990 growth reference (UK90) distribution (Cole et al., 1995), in line with UK NCMP. Cardiorespiratory fitness was assessed by the multistage fitness test, a reliable and feasible protocol (Institute of Medicine, 2012; Léger & Lambert, 1982) for children of the same age (Fairclough et al., 2013; Ridgers et al., 2018; Taylor et al., 2017). The multistage fitness test is a 20-metre shuttle run that took place in the school sports hall/playground and was individually assessed using the number of completed shuttles (Ridgers et al., 2018; Taylor et al., 2017).

**Statistical Analyses**

Analyses were conducted using R (v.3.5.0) in R-Studio (v.1.1.447). Independent t-tests examined differences in baseline characteristics between the two conditions. The first objective and estimate the effect of TDM (binary variable: TDM versus control condition) on MVPA (continuous variable: minutes), LPA (continuous variable: minutes) and sedentary time (continuous variable: minutes). A series of two-level regression models were conducted (Muthen & Muthen, 1998), allowing for the nesting of children (level 1) within classes (level 2). The second objective and estimate the effect of TDM on the progression of maths fluency (continuous variable: number of correct responses, number of errors and the total score (correct responses minus errors)) and EFs (continuous variable: number of correct responses, number of errors and total score) over time. A series of three-level regression models were conducted, allowing for the nesting of measurement occasions (level 1), within students (level 2), within classes (level 3). Separate models were run for each of the outcome variables. All multi-level models included random intercepts for classes and children. Models on EFs and maths fluency over time additionally included random slopes for time. School-level was not included in the model due to the significant correlation with the class level (one class per school). The third objective was to determine whether the effects of the TDM on the progression of maths fluency and EF over time were sensitive to the duration of MVPA children engaged in. To answer this question, children in the TDM group were only included in analyses if they achieved 10+ minutes of MVPA. These children were then contrasted with the children in the Control condition. Effect sizes were calculated as the ratio of the beta coefficient to the within-group pooled standard deviation (Feingold, 2013). Statistical significance was set at $p<0.05$.

# Results

## Physical Activity

Children in TDM condition engaged in significantly greater levels of MVPA versus control ($b$=10.25, SE=0.14, 95% CI=9.97, 10.53, $p\leq0.001$, $d$=4.92), achieving 10.67±2.74 minutes of MVPA during TDM compared to the control (0.44±0.95 minutes). Conversely, sedentary time was also significantly lower during TDM versus control (0.72±0.93 vs. 9.91±3.5 minutes, $b$=-9.23, SE=0.19, 95% CI=-9.60, -8.85, $p\leq0.001$, $d$=3.61). Children in TDM also spent significantly less time in LPA versus control (3.45±2.03 vs. 4.32±2.79 minutes, $b$=-0.87, SE=0.18, 95% CI=-1.22, -0.52, $p\leq0.001$, $d$=0.36). Figure 2 explores treatment fidelity of TDM, revealing large variability; while the most active child spent the duration of TDM engaging in MVPA, yet the least active child spent only 33% of TDM engaging in MVPA, spending the remaining time engaging in LPA.

**Maths Fluency**

As shown in Table 2, the total score for maths fluency (in the MASSAT) revealed a significant interaction between time and condition, with a small effect size ($b$=2.12, SE=0.98, 95% CI=0.21, 4.04, $p$=0.031, $d$=0.25). However, the effect of TDM condition pre-to-post, was not significantly different ($b$=1.03, SE=0.69, 95% CI=-0.32, 2.39, $p$=0.136, $d$=0.25). Similarly, there were no significant differences over time in the control ($b$=-1.10, SE=0.70, 95% CI=-2.48, 0.28, $p$=0.117, $d$=0.25). A significant time by condition interaction was observed for total errors ($b$=1.15, SE=0.54, 95% CI=-2.21, -0.09, $p$=0.034, $d$=0.24). On closer inspection, there were no significant differences over time in TDM ($b$=-0.63, SE=0.36, 95% CI=-1.34, 0.08, $p$=0.08) or the control ($b$=0.61, SE=0.44, 95% CI=-0.26, 1.47, $p$=0.168). No interactions were found for the correct responses ($p$=0.263, $d$=0.13). Looking at the effect of the maths fluency test in children who achieved at least 10 minutes of MVPA in TDM condition revealed a similar pattern (Table 2, model 2); an interaction was found between time and condition ($b$=2.32, SE=1.04, 95% CI=0.28, 4.35, $p$=0.026, $d$=0.29). Similarly, the effect of TDM condition pre-to-post, was not significantly different ($b$=1.23, SE=0.77, 95% CI=-0.29, 2.74, $p$=0.113, $d$=0.27), but the confidence intervals are closer together and the negative crossover of zero was smaller.

**Executive Function**

No significant interactions were found in any of the EFs total scores ($p>0.05$, Table 3, model 1). Both conditions demonstrated a small non-significant increase pre-to-post in every test score (Figure 3). Similarly, no significant interactions were found in any of the EFs totals scores when only including children in TDM that had achieved the 10 minutes of MVPA ($p>0.05$, Table 3, model 2).

# Discussion

The present study is the first to assess the accumulation of MVPA from one session of TDM and the acute effects on EFs and maths fluency. Multi-level modelling revealed that the 15-minute daily mile averaged 10 minutes more MVPA and 9 minutes less sedentary time versus a classroom-based lesson. MVPA accumulation within TDM varied greatly from five minutes by the least active child to 15 minutes by the most active child. For maths fluency scores – a measure of AP - a significant interaction was revealed between conditions and time, in favour of TDM condition. While TDM condition scores increased post-intervention, and the control condition scores decreased post-intervention, neither condition secured significance, suggesting the interaction was due to a magnitude of change in the two conditions. TDM revealed no significant interactions in EFs – irrespective of the individual dose of MVPA, as demonstrated through the 10-minute threshold analysis. Concluding, there was no impact on the amount of MVPA accumulated within TDM on the underlying processes behind mathematical ability (Cragg & Gilmore, 2014; Cragg et al., 2017).

**Physical Activity Levels During The Daily Mile**

The first objective was to ascertain the amount of MVPA that arises from one session of TDM, identifying the contribution towards the in-school daily MVPA recommendations of 30 minutes(Department of Health and Social Care, 2018). During one session of TDM, children accumulated significantly more MVPA (71.3% of TDM session; 35.6% of the in-school daily PA guidelines) and less sedentary time versus a typical classroom-based lesson. Conversely, children accumulated significantly more LPA during the control group versus TDM, which may be explained by low movement or fidgeting in a classroom lesson (Bailey et al., 2012; Nettlefold et al., 2010). Assessing PA at the individual level revealed a highly varied intervention effect. The least active child accumulated 16.7% of their 30-minute in-school target, and the most active child achieved 50% of in-school target. Suggesting a varied appetite to engage in MVPA and supports the need for TDM to be part of a whole-school multifaceted PA programme. Also, these findings support the need to respond to individual responsiveness from different intervention types (Morris et al., 2019; Norris et al., 2015) and unpick the characteristics that influence individual-level engagement.

These findings provide an important insight into the PA potential during one session of TDM. However, previous studies have shown that children may compensate their PA levels later in the day (Goodman et al., 2011) or over the following days (Ridgers et al., 2014). Meaning daily or weekly MVPA levels may not differ from the implementation of TDM. Assessing PA levels across the day or week may combat this. Future studies that start to explore daily PA levels may favour exploring the potential long-term effects from TDM. MVPA levels over time may increase (i.e., children get fitter and can run for longer) or decrease (i.e. children may get bored with a lack of cognitive challenges). It is crucial that studies aiming to understand the long-term effects use high-quality designs that conform to quality assessment tools (e.g., Downs & Black, 1998).

**Acute effect of The Daily Mile on Maths Fluency**

Statistical outcomes suggest a time by condition interaction in maths fluency total score; favourable improvements were revealed over time in TDM condition. These changes may be explained by a reduction in the number of errors made, due to a similar interaction, favouring the intervention. However, the diagnostics within the maths fluency errors model was not acceptable and therefore, should be interpreted with caution. The magnitude of this effect agrees with similar intervention studies that assessed time-on-task behaviour following a physically active lesson (Grieco et al., 2009; Grieco et al., 2016). While the two conditions (intervention and control) demonstrated changes over time, going in different directions, these changes in isolation were not great enough to detect noteworthy, significant changes. They should be interpreted cautiously until replicated.

To date, the present study is the first to assess the acute impact of TDM on maths fluency performance. The findings corroborate previous laboratory-based studies assessing AP that revealed some significant improvements in arithmetic (Pontifex et al., 2013) and reading (Hillman et al., 2009b; Pontifex et al., 2013). Field-based studies that have assessed maths performance – predominantly focused on cognitively-engaging PA - have revealed inconsistencies (Graham et al., 2014; Howie et al., 2014). One study identified significant improvements in 10- and 20- minute bouts of PA but not in 5-minutes, suggested a potential threshold for improvements (Howie et al., 2014). However, this study used different measures of assessing maths performance and did not objectively evaluate treatment fidelity as accurately as the present study.

The learning-focused control condition may offer another explanation for the opposing effect found overtime in maths fluency between conditions. The increase in maths fluency following TDM could be attributed to either the bout of PA or because children in the intervention took a break from academic learning. However, the present study is not the first to compare a PA intervention to learning-focused control (Grieco et al., 2016; Mahar, 2011; Schmidt et al., 2016). Previous studies have assessed cognitively-engaging or physically active learning interventions compared to classroom lessons. Future studies may benefit from a third condition such as a sedentary break away from the classroom. Controlling the classroom environment is one of the biggest challenges for field-based research due to variability in a myriad of factors (e.g., teaching styles, school cultures, implementation of TDM). While this should be acknowledged, the importance of testing initiatives being widely implemented within real-world settings warrants considerable investigation. Additionally, comparing PA interventions to classroom lessons replicates 'normal' school behaviour in its truest form, justifying this approach.

The 10-minute MVPA threshold analyses revealed a significant interaction between time and condition, with marginally positive improvements. Significance was underpinned by an improvement in TDM and a decrease in the control. Assessing the changes over time in isolation revealed no significant differences. While these findings show some threshold effects, other influential participant characteristics may have been overlooked. Regardless of the 10-minute threshold analysis, this level of PA variability limits confidence in confirming the effect of TDM on maths fluency performance.

It is essential to understand the variable response to school-based translational work. Researchers' should continue to objectively assess PA at the individual level and continue to explore subgroup analysis to ensure subgroup benefits are not masked by aggregate findings (Morris et al., 2019). Additionally, future research is warranted to replicate similar field-based run-walk studies to assess the impact on other measures of AP important to the school system.

**Impact of The Daily Mile on Executive Functions**

There were no significant improvements in EFs in either TDM or control condition. Given the previous laboratory- and field-based studies using quantitative exercise found inconsistent effects (Chen et al., 2014; Drollette et al., 2012; Hillman et al., 2009b; Jager et al., 2015), these outcomes were not surprising. Higher-quality studies have consistently found no effects on working memory (Benzing et al., 2016; Egger et al., 2018; Jager et al., 2014, 2015). For inhibitory control, mixed outcomes have been established; laboratory-based studies demonstrated improved accuracy (Drollette et al., 2012; Hillman et al., 2009b; Pontifex et al., 2015) while higher-quality field-based studies have shown shorter response times post-exercise (Chen et al., 2014). In agreement with the current study, previous studies have found no effects on response time or accuracy (Egger et al., 2018; Jager et al., 2015).

The 10-minute MVPA threshold analyses revealed no significant interactions in any of the EF outcomes. A previous classroom-based study using exercise breaks suggested a minimum of 10-minutes MVPA may be required for positive EF effects (Howie et al., 2014). Our findings align with the meta-analysis by Chang et al. (2012) who found no positive impact from quantitative exercise bouts ranging from 11-20 minutes. Therefore, the volume (duration x intensity) of TDM may not be sufficient to cause the release of catecholamines, which positively influences cognition (McMorris, 2018). While it may be challenging for schools to increase the duration of TDM, as suggested in previous research, they could combine quantitative PA with cognitively engaging tasks (Daly-Smith et al., 2018) to facilitate positive effects on the intervention on EF (Best, 2010; Pesce, 2012).

**Understanding the Differences in Maths Fluency and Executive Functions**

The findings in the maths fluency and EF tests are not harmonious. One explanation for why a significant interaction was found in the maths fluency test but not in any EFs may be due to the small effects not being detectable by individual tests of EF. A contextualised composite of multiple EF processes – such as the maths fluency test – relying on numerous processes that may uncover a detectable change (Cragg & Gilmore, 2014; Cragg et al., 2017). Tests such as the MASSAT provide familiar territory for the children and provide high ecological validity unlike the tests used to assess the subdomains within EFs. Completing the EF tests might have subconsciously altered the learning environment (e.g., children getting overexcited or distracted, resulting in reduced concentration levels). A learning effect may also have occurred; recommending an extended familiarisation testing protocol. Another explanation may be the order of tests. While previous studies have randomised and counterbalanced test order (Jager et al., 2015; van den Berg et al., 2018), we could not deploy this approach. In support, previous studies found no issues administering one order of tests (Chen et al., 2014; Cooper et al., 2018). Although several of these limitations could be revised, future studies may strengthen their designs by placing more focus on contextually relevant and ecologically validated tests.

**Strengths and Limitations**

The present study is the first to explore the immediate impacts of TDM on children's PA levels, EFs and AP. The robust methodology encompassed (i) a statistically powered sample size, (ii) randomisation at the individual-level and (iii) an objective assessment of treatment fidelity for PA in each participating child, providing new insights into TDM's impact. Moreover, the use of multi-level modelling helps overcome any statistical dependence found in nested data (Raudenbush & Bryk, 2001). That said, the study is not without limitations; first, no attempt was made to blind the children and researchers. Blinding is an on-going challenge for school-based research designs. Second, the EF tests demonstrated medium test-retest reliability scores (r=0.56, 0.65), resulting in large unexplained variability. However, the scores were similar to a previous, field-based study administrating tests to a cohort of children at once (r=0.56, 0.63, Howie et al., 2014). While standardised, reliable and robust EF tests are easy to administer in laboratory-based studies, this methodological rigour cannot be transferred into school-based study designs as it undermines the purpose of translational research. Consequently, classroom-based studies may best include standardised AP measures with greater relevance to each school.

**Conclusion and Future Directions**

Based on unsubstantiated claims, TDM has been implemented in over 8,800 schools worldwide (The Daily Mile Foundation, 2019). The current study uses a robust and rigorous study design that offers exemplary translational evidence of the impact of an acute bout of

TDM-based PA on children's EFs and maths fluency. The only existing published study on TDM (Chesham et al., 2018) explored long-term outcomes, meaning that the present study offers original high-quality evidence establishing the acute effects on PA levels, EFs and AP. While TDM secured higher levels of MVPA and overall improvements in maths fluency, no significant differences were found in EFs. Unlike previous studies that evaluated similar bouts of PA (Chen et al., 2014; Drollette et al., 2012; Hillman et al., 2009b; Tomporowski et al., 2008), these findings deserve future investigation based on similar, improved protocols. Such work will overcome the current challenges of assessing EFs in a school-based setting and move towards assessing multiple measures of AP. Caution is also warranted regarding the PA findings; while they provide a snapshot of the PA potential of TDM, they do not give answers to PA levels overtime or the varied implementation approaches across different schools. Further, impacts on children's enjoyment levels and attitudes towards TDM and PA uptake within the school context are yet to be published.

# References

Bailey, D. P., Fairclough, S. J., Savory, L. A., Denton, S. J., Pang, D., Deane, C. S., & Kerr, C. J. (2012). Accelerometry-assessed sedentary behaviour and physical activity levels during the segmented school day in 10–14-year-old children: the HAPPY study. *European Journal of Pediatrics, 171,* 1805–1813. doi:10.1007/s00431-012-1827-0

Benzing, V., Heinks, T., Eggenberger, N., & Schmidt, M. (2016). Acute Cognitively Engaging Exergame-Based Physical Activity Enhances Executive Functions in Adolescents. *PloS One, 11*(12), e0167501. doi:10.1371/journal.pone.0167501

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.

Bryce, D., Szűcs, D., Soltész, F., & Whitebread, D. (2010). The development of inhibitory control- An averaged and single-trial Lateralized Readiness Potential study. *NeuroImage, 57,* 671-685. doi:10.1016/j.neuroimage.2010.12.006

Chang, Y. K., Labban, J. D., Gapin, J. I., & Etnier, J. L. (2012). The effects of acute exercise on cognitive performance: a meta-analysis. *Brain Research, 1453,* 87-101. doi:10.1016/j.brainres.2012.02.068

Chen, A.-G., Yan, J., Yin, H.-C., Pan, C.-Y., & Chang, Y.-K. (2014). Effects of acute aerobic exercise on multiple aspects of executive function in preadolescent children. *Psychology of Sport and Exercise, 15*(6), 627-636. doi:10.1016/j.psychsport.2014.06.004

Chesham, R. A., Booth, J. N., Sweeney, E. L., Ryde, G. C., Gorely, T., Brooks, N. E., & Moran, C. N. (2018). The Daily Mile makes primary school children more active, less sedentary and improves their fitness and body composition: a quasi-experimental pilot study. *BMC Medicine, 16*(1). doi:10.1186/s12916-018-1049-z

Chief Medical Officers. (2011). *Start Active, Stay Active: A Report on Physical Activity for Health from the Four Home Countries Chief Medical Officers.* London: Chief Medical Officers of England Scotland Wales and Northern Ireland.

Cole, T. J., Freeman, J. V., & Preece, M. A. (1995). Body mass index reference curves for the UK, 1990. *Archives of Disease in Childhood., 73*, 25-29.

Cooper, S. B., Dring, K. J., Morris, J. G., Sunderland, C., Bandelow, S., & Nevill, M. E. (2018). High intensity intermittent games-based activity and adolescents' cognition: moderating effect of physical fitness. *BMC Public Health, 18*(1), 603. doi:10.1186/s12889-018-5514-6

Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3*(2), 63-68. doi:10.1016/j.tine.2013.12.001

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition, 162*, 12-26. doi:10.1016/j.cognition.2017.01.014

Daly-Smith, A., Morris, J. L., Hobbs, M., & McKenna, J. (2019). Commentary on a recent article on the effects of the 'Daily Mile' on physical activity, fitness and body composition: addressing key limitations. *BMC Medicine, 17*, 96.

Daly-Smith, A. J. (2018). A review of school-based studies on the effect of acute physical activity on cognitive function in children and young people. In R. Meeusen, S. Schaefer, P. D. Tomporowski, & R. Bailey (Eds.), *Physical Activity and Educational Achievement: Insights from Exercise Neuroscience*. London and New York: Routledge.

Daly-Smith, A. J., Zwolinsky, S., McKenna, J., Tomporowski, P. D., Defeyter, M. A., & Manley, A. (2018). Systematic review of acute physically active learning and classroom movement breaks on children's physical activity, cognition, academic performance and classroom behaviour: understanding critical design features. *BMJ Open Sport Exerc Med, 4*(1), e000341. doi:10.1136/bmjsem-2018-000341

Department of Education. (2013). *The National Curriculum in England Key Stage 1 and 2 Framework Document*. (Reference: DFE-00178-2013). UK.

Department of Health and Social Care. (2018). *Childhood Obesity: A Plan to Action. Chapter 2*.

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

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. doi: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., Shishido, T., Pontifex, M. B., & Hillman, C. H. (2012). Maintenance of Cognitive Control during and after Walking in Preadolescent Children. *Medicine & Science in Sports & Exercise, 44*(10), 2017-2024.

Egger, F., Conzelmann, A., & Schmidt, M. (2018). The effect of acute cognitively engaging physical activity breaks on children's executive functions: Too much of a good thing? *Psychology of Sport and Exercise, 36*, 178-186. doi:10.1016/j.psychsport.2018.02.014

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143–149. doi:doi: 10.3758/BF03203267

Espana-Romero, V., Artero, E. G., Jimenez-Pavon, D., Cuenca-Garcia, M., Ortega, F. B., Castro-Pinero, J., . . . Ruiz, J. R. (2010). Assessing health-related fitness tests in the school setting: reliability, feasibility and safety; the ALPHA Study. *International Journal of Sports Medicine, 31*(7), 490-497. doi:10.1055/s-0030-1251990

Evenson, K. R., Catellier, D. J., Gill, K., Ondrak, K. S., & McMurray, R. G. (2008). Calibration of two objective measures of physical activity for children. *Journal Sports Science, 26*(14), 1557-1565. doi:10.1080/02640410802334196

Fairclough, S. J., Hackett, A. F., Davies, I. G., Gobbi, R., Mackintosh, K. A., Warburton, G. L., . . . Boddy, L. M. (2013). Promoting healthy weight in primary school children through physical activity and nutrition education: A pragmatic evaluation of the CHANGE! Randomised intervention study. *BMC Public Health, 2*(13), 626. doi:doi: 10.1186/1471-2458-13-626.

Fairhurst, A., & Hotham, S. (2017). Going further than the ‘Daily Mile’. *Perspectives in Public Health, 137*(2), 83-84. doi:10.1177/1757913916685639

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modelling of group differences. *Review of General Psychology, 17*(1), 111-121. doi:10.1037/a0030048

Flay, B. R., Biglan, A., Boruch, R. F., Gonzalez Castro, F., Gottfredson, D., Kellam, S., . . . Ji, P. (2005). Standards of Evidence: Criteria for Efficacy, Effectiveness and Dissemination. *Prevention Science, 6*(3), 151-175.

Forman, S. G., Olin, S. S., Hoagwood, K. E., Crowe, M., & Saka, N. (2008). Evidence-Based Interventions in Schools: Developers' Views of Implementation Barriers and Facilitators. *School Mental Health, 1*(1), 26-36. doi:10.1007/s12310-008-9002-5

Gathercole, S. E., Brown, L., & Pickering, S. J. (2003). Working memory assessments at school entry as longitudinal predictors of National Curriculum attainment leveld. *Educational and Child Psychology, 20*, 109-222.

Geary, D. C., Hoard, M. K., & Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. *Journal of Experimental Child Psychology, 113*(1), 49–65.

Goodman, A., Mackett, R. L., & Paskins, J. (2011). Activity compensation and activity synergy in British 8-13 year olds. *Preventive Medicine, 53*, 293–298. doi:doi:10.1016/j.ypmed.2011.07.019

Graham, D. J., Lucas-Thompson, R. G., & O'Donnell, M. B. (2014). Jump In! An Investigation of School Physical Activity Climate, and a Pilot Study Assessing the Acceptability and Feasibility of a Novel Tool to Increase Activity during Learning. *Frontiers in Public Health, 2*(58). doi:10.3389/fpubh.2014.00058

Grieco, L. A., Jowers, E. M., & Bartholomew, J. B. (2009). Physically active academic lessons and time on task: the moderating effect of body mass index. *Medicine and Science in Sports and Exercise, 41*(10), 1921-1926.

Grieco, L. A., Jowers, E. M., Errisuriz, V. L., & Bartholomew, J. B. (2016). Physically active vs. sedentary academic lessons: A dose response study for elementary student time on task. *Preventive Medicine, 89*, 98-103. doi:10.1016/j.ypmed.2016.05.021

Hillman, C. H., Buck, S. M., Themanson, J. R., Pontifex, M. B., & Castelli, D. M. (2009a). 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. doi:10.1037/a0014437

Hillman, C. H., Pontifex, M. B., Raine, L. B., Castelli, D. M., Hall, E. E., & Kramer, A. F. (2009b). The effect of acute treadmill walking on cognitive control and academic achievement in preadolescent children. *Neuroscience, 159*(3), 1044-1054. doi:10.1016/j.neuroscience.2009.01.057

Howie, E. K., Beets, M. W., & Pate, R. R. (2014). Acute classroom exercise breaks improve on-task behavior in 4th and 5th grade students: A dose-response. *Mental Health and Physical Activity, 7*(2), 65-71. doi:10.1016/j.mhpa.2014.05.002

Howie, E. K., & Pate, R. R. (2012). Physical activity and academic achievement in children: A historical perspective. *Journal of Sport and Health Science, 1*(3), 160-169. doi:10.1016/j.jshs.2012.09.003

Institute of Medicine. (2012). *Fitness Measures and Health Outcomes in Youth*. Washington, DC: National Academies Press.

Jager, K., Schmidt, M., Conzelmann, A., & Roebers, C. M. (2014). Cognitive and physiological effects of an acute physical activity intervention in elementary school children. *Frontiers in Psychology, 18*(5), 1473. doi:10.3389/fpsyg.2014.01473

Jager, K., Schmidt, M., Conzelmann, A., & Roebers, C. M. (2015). The effects of qualitatively different acute physical activity interventions in real-world settings on executive functions in preadolescent children. *Mental Health and Physical Activity, 9*, 1-9. doi:10.1016/j.mhpa.2015.05.002

Janssen, M., Chinapaw, M. J. M., Rauh, S. P., Toussaint, H. M., van Mechelen, W., & Verhagen, E. (2014). A short physical activity break from cognitive tasks increases selective attention in primary school children aged 10-11. *Mental Health and Physical Activity, 7*(3), 129-134. doi:10.1016/j.mhpa.2014.07.001

Léger, L. A., & Lambert, J. (1982). A maximal multistage 20-m shuttle run rest to predict VO2 max. *European Journal of Applied Physiology and Occupational Physiology, 49*(1), 1-12.

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological Assessment.* (4th Ed. ed.). New York: Oxford University Press.

Macdonald, J. A., Beauchamp, M., Crigan, J. A., & Anderson, P. J. (2014). Age related differences in inhibitory control in the: early school years. *Child Neuropsychology, 20*, 509-526.

Mahar, M. T. (2011). Impact of short bouts of physical activity on attention-to-task in elementary school children. *Preventive Medicine*, *52 Suppl 1*, S60-S64. doi:10.1016/j.ypmed.2011.01.026

McMorris, T. (2018). The Development of the Acute Exercise--Catecholamines-Cognition Interaction Theory: Implications for Learning and Memory. In R. Meeusen, Schaefer, S., Tomporowski, P. & Bailey, R. (Ed.), *Physical Activity and Educational Achievement.* (pp. 89–110.). New York: Routledge.

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

Morris, J. L., Daly-Smith, A., Defeyter, M. A., McKenna, J., Zwolinsky, S., Lloyd, S., . . . Graham, P. L. (2019). A Pedometer-Based Physically Active Learning Intervention: The Importance of Using Preintervention Physical Activity Categories to Assess Effectiveness. *Pediatric Exercise Science*, 1-7. doi:10.1123/pes.2018-0128

Muthen, L. K., & Muthen, B. O. (1998). *Mplus: Statistical Analysis with Latent Variables User's Guide.* **Los Angeles, CA.**

Nettlefold, L., McKay, H. A., Warburton, D. E. R., McGuire, K. A., Bredin, S. S. D., & Naylor, P. J. (2010). The challenge of low physical activity during the school day: at recess, lunch and in physical education. *British Journal of Sports Medicine, 45*, 813-819. doi:10.1136/bjsm.2009.068072

Norris, E., Shelton, N., Dunsmuir, S., Duke-Williams, O., & Stamatakis, E. (2015). Physically active lessons as physical activity and educational interventions: a systematic review of methods and results. *Preventive Medicine*(72), 116–125. doi:10.1016/j.ypmed.2014. 12.027

Pesce, C. (2012). Shifting the Focus From Quantitative to Qualitative Exercise Characteristics in Exercise and Cognition Research. *Journal of Sport & Exercise Psychology, 34*, 766-786.

Pontifex, M. B., Parks, A. C., Henning, D. A., & Kamijo, K. (2015). Single bouts of exercise selectively sustain attentional processes. *Psychophysiology, 52*(5), 618-625. doi:10.1111/psyp.12395

Pontifex, M. B., Saliba, B. J., Raine, L. B., Picchietti, D. L., & Hillman, C. H. (2013). Exercise improves behavioral, neurocognitive, and scholastic performance in children with attention-deficit/hyperactivity disorder. *Journal of Pediatrics, 162*(3), 543-551. doi:10.1016/j.jpeds.2012.08.036

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical Linear Models: Applications and Data Analysis Methods.* (2nd edition ed.). Inc, Thousand Oaks.: SAGE Publications.

Ridgers, N. D., Barnett, L. M., Lubans, D. R., Timperio, A., Cerin, E., & Salmon, J. (2018). Potential moderators of day-to-day variability in children's physical activity patterns. *Journal of Sports Science, 36*(6), 637-644. doi:10.1080/02640414.2017.1328126

Ridgers, N. D., Timperio, A., Cerin, E., & Salmon, J. O. (2014). Compensation of Physical Activity and Sedentary Time in Primary School Children. *Medicine & Science in Sports & Exercise, 46*, 1564–1569.

Schiff, S., Amodio, P., Testa, G., Nardi, M., Montagnese, S., Caregaro, L., . . . Sellitto, M. (2016). Impulsivity toward food reward is related to BMI: Evidence from intertemporal choice in obese and normal-weight individuals. *Brain and Cognition, 110,* 112-119. doi:10.1016/j.bandc.2015.10.001

Schmidt, M., Benzing, V., & Kamer, M. (2016). Classroom-Based Physical Activity Breaks and Children's Attention: Cognitive Engagement Works! *Frontiers in Psychology, 7*, 1474-1474.

Scottish Government. (2018). Scotland: a Daily Mile Nation. Retrieved from https://beta.gov.scot/news/scotland-a-daily-mile-nation/

Suresh, K. P. (2011). An overview of randomization techniques: An unbiased assessment of outcome in clinical research. *Journal of Human Reproductive Sciences, 4*(1), 8-11.

Szűcs, D., Devine, A., Soltesz, F., Nobes, A., & Gabriel, F. (2014). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex, 49*(10), 2674-2688. doi:http://dx.doi.org/10.1016/j.cortex.2013.06.007

Taylor, S. L., Curry, W. B., Knowles, Z. R., Noonan, R. J., McGrane, B., & Fairclough, S. J. (2017). Predictors of Segmented School Day Physical Activity and Sedentary Time in Children from a Northwest England Low-Income Community. *International Journal of Environmental Research and Public Health, 14*(5), pii: E534. doi:10.3390/ijerph14050534

The Daily Mile Foundation. (2019). The Daily Mile. Retrieved from https://thedailymile.co.uk

Tomporowski, P. D., Davis, C. L., & Lambourn, K. (2008). Task Switching in Overweight Children: effects of acute exercise and age. *Journal of Sport and Exercise Psychology, 36*(5), 474-485.

Tomporowski, P. D., McCullick, B., Pendleton, D. M., & Pesce, C. (2015). Exercise and children's cognition: The role of exercise characteristics and a place for metacognition. *Journal of Sport and Health Science, 4*(1), 47-55. doi:10.1016/j.jshs.2014.09.003

Tremblay, M. S., Barnes, J. D., González, S. A., Katzmarzyk, P. T., Onywera, V. O., Reilly, J. J., & Tomkinson, G. R. (2016). Global Matrix 2.0: Report Card Grades on the Physical Activity of Children and Youth Comparing 38 Countries. *Journal of Physical Activity and Health, 11*((11 Suppl 2)), S343–366. doi:http://dx.doi.org/10.1123/jpah.2016-0594

van den Berg, V., Saliasi, E., Jolles, J., de Groot, R. H. M., Chinapaw, M. J. M., & Singh, A. S. (2018). Exercise of Varying Durations: No Acute Effects on Cognitive Performance in Adolescents. *Frontiers in Neuroscience, 12*. doi:10.3389/fnins.2018.00672

Vazou, S., & Smiley-Oyen, A. (2014). Moving and academic learning are not antagonists: acute effects on executive function and enjoyment. *Journal of Sport and Exercise Psychology, 36*(5), 474-485. doi:10.1123/jsep.2014-0035

# Tables

Table 1

*Characteristics of Participants Stratified by Condition*

| | **TDM (n=158)** *M (SD)* | **Control (n=145)** *M (SD)* | ***p*** |
|---|---|---|---|
| Age (year) | 8.99 (0.5) | 8.99 (0.5) | 0.98 |
| Girls (%) | 56 | 57 | |
| Year Group (%) | | | |
| Year 3 | 1 (1) | 3 (4) | |
| Year 4 | 71 (112) | 70 (101) | |
| Year 5 | 28 (45) | 27 (40) | |
| Offset Maturity (y) | -3.08 (0.68) | -3.06 (0.64) | 0.85 |
| Height (cm) | 133.57 (6.58) | 133.94 (6.11) | 0.61 |
| Weight (kg) | 31.98 (8.14) | 32.61 (7.48) | 0.49 |
| BMI (kg/m$^2$) | 17.72 (3.21) | 18.07 (3.39) | 0.36 |
| BMI z-score | 0.45 (1.23) | 0.6 (1.18) | 0.27 |
| Weight Category (%) | | | |
| Normal Weight | 75 | 71 | |
| Overweight | 15 | 17 | |
| Obese | 10 | 12 | |
| Waist Circumference (cm) | 61.23 (9.82) | 62.15 (8.3) | 0.38 |
| Hip Circumference (cm) | 67.15 (8.68) | 67.32 (8.33) | 0.86 |
| WHR | 0.91 (0.08) | 0.92 (0.06) | 0.12 |
| Pupil Premium (%) | 20 | 25 | |
| Academic Level (%) | | | |
| Low | 22 | 19 | |
| Mid | 56 | 56 | |
| High | 22 | 25 | |
| CRF (Number of Shuttles) * | 28.64 (15.41) | 26.15 (14.01) | 0.16 |

*Note.* CRF: Cardiorespiratory fitness; Pupil premium: Percentage of children receiving free school meals, as a measure of socioeconomic status; *Numbers of children differed here: TDM=149, Control=129.

Table 2

*Regression Models for Maths Fluency Performance using Multilevel Modelling*

| | **Model One (*n*=303)** | | | | **Model Two (*n*=272)** | | | |
|---|---|---|---|---|---|---|---|---|
| | *b* (SE) | 95% CI | p | *d* | b (SE) | 95% CI | p | *d* |
| **MASSAT Total Score** | | | | | | | | |
| Intercept | 24.24 (1.61) | 21.08, 27.41 | ≤0.0001 | | 24.29 (1.56) | 21.22, 27.35 | ≤0.0001 | |
| Time | -1.04 (0.76) | -2.55, 0.45 | 0.172 | | -1.10 (0.73) | -2.52, 0.34 | 0.136 | |
| Condition | -0.97 (1.61) | -4.13, 2.20 | 0.550 | | -1.00 (1.70) | -4.34, 2.34 | 0.556 | |
| Time: Condition | 2.12 (0.98) | 0.21, 4.04 | 0.031 | 0.25 | 2.32 (1.04) | 0.28, 4.35 | 0.026 | 0.29 |
| **MASSAT Correct Answers** | | | | | | | | |
| Intercept | 29.36 (1.17) | 27.06, 31. 76 | ≤0.0001 | | 29.35 (1.19) | 27.02, 31.69 | ≤0.0001 | |
| Time | -0.49 (0.61) | -1.68, 0.70 | 0.418 | | -0.50 (0.61) | -1.69, 0.69 | 0.413 | |
| Condition | -0.65 (1.23) | -3.06, 1.76 | 0.595 | | -0.63 (1.30) | -3.17, 1.91 | 0.627 | |
| Time: Condition | 0.80 (0.71) | -0.60, 2.20 | 0.263 | 0.13 | 0.94 (0.76) | -0.55, 2.43 | 0.218 | 0.16 |
| **MASSAT Total Errors** | | | | | | | | |
| Intercept | 5.00 (0.59) | 3.85, 6.15 | ≤0.0001 | | 4.94 (0.53) | 3.90, 5.99 | ≤0.0001 | |
| Time | 0.53 (0.40) | -0.27, 1. 23 | 0.194 | | 0.56 (0.40) | -0.22, 1.33 | 0.159 | |
| Condition | 0.31 (0.67) | -0.99, 1.62 | 0.638 | | 0.41 (0.71) | -0.97, 1.80 | 0.559 | |
| Time: Condition | -1.15 (0.54) | -2.21, -0.09 | 0.034 | 0.24 | -1.16 (0.58) | -2.30, -0.03 | 0.045 | 0.26 |

*Note.* b: unstandardised beta coefficient; CI: confidence intervals; d: Cohen's d for effect size; SE: standard error.

Table 3

*Regression Models for Executive Function Tests using Multilevel Modelling*

| | **Model 1 (n=303)** | | | | **Model 2 (n=272)** | | | |
|---|---|---|---|---|---|---|---|---|
| | b (SE) | 95% CI | p | d | b (SE) | 95% CI | p | d |
| **TMT-A TS** | | | | | | | | |
| Intercept | 26.18 (1.04) | 24.13, 28.22 | ≤0.0001 | | 26.19 (1.05) | 24.13, 28.24 | ≤0.001 | |
| Time | 3.36 (0.72) | 1.95, 4.76 | ≤0.001 | | 3.37 (0.72) | 1.96, 4.78 | ≤0.001 | |
| Condition | 0.84 (0.91) | -0.94, 2.62 | 0.356 | | 1.12 (0.95) | -0.75, 2.98 | 0.242 | |
| Time: Condition | -0.42 (0.84) | -2.07, 1.23 | 0.620 | 0.06 | -0.32 (0.87) | -2.04, 1.39 | 0.712 | 0.05 |
| **TMT-B TS** | | | | | | | | |
| Intercept | 21.93 (0.91) | 20.15, 23.71 | ≤0.0001 | | 21.94 (0.93) | 20.11, 23.77 | ≤0.0001 | |
| Time | 2.72 (0.68) | 1.38, 4.05 | 0.001 | | 2.73 (0.67) | 1.41, 4.04 | 0.0001 | |
| Condition | -0.23 (0.98) | -2.16, 1.70 | 0.813 | | 0.00 (1.01) | -2.00, 2.00 | 0.999 | |
| Time: Condition | -1.20 (0.94) | -3.03, 0.64 | 0.202 | 0.15 | -1.37 (0.98) | -3.28, 0.55 | 0.163 | 0.18 |
| **Digit Recall TS** | | | | | | | | |
| Intercept | 19.78 (1.65) | 16.54, 21.08 | ≤0.0001 | | 19.79 (1.66) | 16.54, 23.04 | ≤0.0001 | |
| Time | 2.82 (1.64) | -0.41, 6.04 | 0.088 | | 2.80 (1.56) | -0.26, 5.86 | 0.074 | |
| Condition | 4.35 (1.84) | 0.74, 7.96 | 0.019 | | 4.47 (1.96) | 0.62, 8.31 | 0.024 | |
| Time: Condition | -1.34 (1.61) | -4.49, 1.81 | 0.405 | 0.09 | -1.41 (1.74) | -4.83, 2.00 | 0.417 | 0.10 |
| **Standard Flanker TS** | | | | | | | | |
| Intercept | 12.66 (0.37) | 11.93, 13.38 | ≤0.0001 | | 12.65 (0.38) | 11.91, 13.38 | ≤0.0001 | |
| Time | 2.23 (0.50) | 1.26, 3.21 | ≤0.0001 | | 2.23 (0.51) | 1.22, 3.24 | ≤0.0001 | |
| Condition | 0.33 (0.43) | -0.52, 1.17 | 0.448 | | 0.22 (0.44) | -0.65, 1.08 | 0.621 | |
| Time: Condition | -0.21 (0.40) | -0.98, 0.56 | 0.593 | 0.06 | -0.06 (0.42) | -0.89, 0.76 | 0.882 | 0.02 |
| **Mixed Flanker TS** | | | | | | | | |
| Intercept | 10.75 (0.39) | 9.98, 11.52 | ≤0.0001 | | 10.74 (0.39) | 9.97, 11.51 | ≤0.0001 | |
| Time | 0.84 (0.40) | 0.06, 1.62 | 0.036 | | 0.85 (0.39) | 0.08, 1.62 | 0.032 | |
| Condition | 0.29 (0.45) | -0.60, 1.17 | 0.527 | | 0.23 (0.48) | -0.70, 1.17 | 0.622 | |
| Time: Condition | 0.16 (0.43) | -0.69, 1.01 | 0.710 | 0.04 | 0.15 (0.46) | -0.75, 1.06 | 0.741 | 0.04 |
| **Animal Stroop TS** | | | | | | | | |
| Intercept | 10.30 (0.40) | 9.52, 11.08 | ≤0.0001 | | 10.31 (0.43) | 9.48, 11.15 | ≤0.0001 | |
| Time | 1.01 (0.23) | 0.55, 1.47 | ≤0.0001 | | 1.02 (0.21) | 0.60, 1.44 | ≤0.0001 | |
| Condition | 0.05 (0.35) | -0.64, 0.73 | 0.894 | | 0.19 (0.36) | -0.52, 0.89 | 0.605 | |
| Time: Condition | 0.12 (0.32) | -0.51, 0.74 | 0.713 | 0.04 | 0.06 (0.31) | -0.56, 0.67 | 0.853 | 0.03 |

*Note.* b: unstandardised beta coefficient; CI: confidence intervals; d: Cohen's d for effect size; SE: standard error; TS: total score.

# Figures

*All figures are 2-column fitting images. All figures are attached separately.*

**Figure 1:** CONSORT Flow Diagram.

**Figure 2:** Individual Variability of Physical Activity Levels for Children Completing The Daily Mile.

**Figure 3:** Pre and Post Scores for the MASSAT and Executive Function Tests Stratified by TDM and Control Conditions.

*Note*. A: MASSAT correct responses; B: MASSAT total errors made; C: MASSAT total score; D: Trail Making Task - A total score; E: Trail Making Task - B total score; F: Digit Recall total score; G: Standard Flanker total score; H: Mixed Flanker total score; I: Animal Stroop total score.