Exercising attention within the classroom

LIAM HILL | JUSTIN H G WILLIAMS | LORNA AUCOTT | JUNE MILNE | JENNY THOMSON | JESSIE GREIG | VAL MUNRO | MARK MON-WILLIAMS

1 Child Health, College of Medicine and Life Sciences, University of Aberdeen, UK. 2 Curriculum Support, Aberdeen City Council, Aberdeen, UK. 3 Institute of Psychological Sciences, University of Leeds, UK.

Correspondence to Liam Hill at Child Health, University of Aberdeen, Royal Aberdeen Children’s Hospital, Westburn Road, Aberdeen AB25 2ZD, UK. E-mail: liam_hill@abdn.ac.uk

This article is commented on by Hill on page 888 of this issue.

AIM To investigate whether increased physical exercise during the school day influenced subsequent cognitive performance in the classroom.

METHOD A randomized, crossover-design trial (two weeks in duration) was conducted in six mainstream primary schools (1224 children aged 8–11y). No data on sex was available. Children received a teacher-directed, classroom-based programme of physical exercise, delivered approximately 30 minutes after lunch for 15 minutes during one week and no exercise programme during the other (order counterbalanced across participants). At the end of each school day, they completed one of five psychometric tests (paced serial addition, size ordering, listening span, digit-span backwards, and digit-symbol encoding), so that each test was delivered once after exercise and once after no exercise.

RESULTS General linear modelling analysis demonstrated a significant interaction between intervention and counterbalance group (p<0.001), showing that exercise benefitted cognitive performance. Post-hoc analysis revealed that benefits occurred in participants who received the exercise intervention in the second but not the first week of the experiment and were also moderated by type of test and age group.

INTERPRETATION Physical exercise benefits cognitive performance within the classroom. The degree of benefit depends on the context of testing and participants’ characteristics. This has implications for the role that is attributed to physical exercise within the school curriculum.

There have long been anecdotal reports of neuropsychological benefits arising from regular physical exercise, and these now receive support from a substantial body of empirical evidence. Recent reviews suggest that physical exercise can be useful in the treatment of depression, can be used to maintain cognitive vitality in old age, and can even reduce stereotypic and disruptive behaviours in children with autism, learning disorders, or behavioural disorders.

In children, associations between greater aerobic fitness and higher cognitive functioning have been reported. Also, there is evidence of aerobic fitness level moderating aspects of academic performance and children exhibiting increased ‘on-task’ classroom behaviour after brief exercise interventions. Increases in the amount of curriculum time devoted to physical activity have also been reported to result in either no change or an improvement in students’ academic achievement, suggesting that increased exercise capably counteracts the negative impact of reduced academic instruction time, perhaps through improving cognition.

Given these findings, recent initiatives to encourage increased childhood physical activity may have important implications for mental as well as physical health. In adults executive functions such as working memory, goal-oriented processing, and inhibition are particularly improved by exercise interventions. If this also applies in children, one might expect that the largest effects of exercise would be seen in individuals who have poor executive function, such as those with attention-deficit–hyperactivity disorder. The past 15 to 20 years have seen rapid increases in the diagnosis of this disorder, and this raises the question whether changes in educational practices, such as decreased physical activity, might have contributed to this problem. Aside from diagnostic issues, evidence that poor attentional skills in the classroom have adverse affects on academic achievement calls for greater investigation into strategies that may remedy this problem.

Nevertheless, robust evidence of the effect of exercise on cognition in the classroom is sparse. Most studies are observational, and few experimental investigations in children have been reported. Improvements after exercise have been reported in three studies of performance on speeded arithmetic tasks before and after physical activity, but these studies all lacked adequate control conditions for practice. Improvements in concentration after physical exercise, perhaps through improving cognition in the classroom.
Further studies,\textsuperscript{17–19} but again these findings may be better explained by practice effects\textsuperscript{19} or uncontrolled between-group variance.\textsuperscript{17,18}

We sought to address the lack of research in this area by conducting a pilot study of the effects of physical exercise on psychometric tests, which would bear ecological validity in relationship to classroom performance. To address previous methodological issues we used a crossover, counterbalanced, within-participant study design that controlled for all factors other than age and classroom physical exercise.

**METHOD**

**Participants**

Six primary schools in Aberdeen, northeast Scotland, were invited to participate in the study by Aberdeen City Council’s Curriculum Support officers. Selected schools had two classes of equivalent size and average ability at each level. Only children from the senior section of the schools (Scottish primary classes 4, 5, 6, and 7; average ages 8y, 9y, 10y, and 11y respectively), who are accustomed to written tests, were included, resulting in a sample size of 1224 children. Consent was provided by participants and their parents or guardians. To blind participants to the experimental aims, consents were obtained from them separately for the exercise intervention and psychometric tests, and full explanation was given only at debriefing. Parent or guardian consent was fully informed from the outset. Ethical approval for the study was obtained from the local ethics committee of the University of Aberdeen.

**Apparatus**

The classroom exercise programme (CEP; Summary of Exercise Intervention available online only) was developed by the Curriculum Support team for Physical Education, Health and Wellbeing in Aberdeen City Council. Students completed all exercises while standing behind their desks. The regimen incorporated stretching and aerobic physical exercises (e.g. running on the spot, hopping sequences to music). Lasting 10 to 15 minutes, it was intended to be moderately intensive for the average student. Teachers reported that participants were beginning to perspire and were slightly out of breath on its completion.

To administer psychometric tests en masse in the classroom, common age-appropriate cognitive tests were modified so that test instructions and stimuli were given from an audio compact disk on a portable player, with children making their responses while seated at their desks by writing in answer booklets. Testing was carried out in the classroom at the end of the school day. Tests were selected to be definable as mental tracking tasks, ‘requiring subjects to track two or more stimuli or associated ideas simultaneously, alternatively or sequentially on double or multiple tracking tests involving dividing and or shifting attention’,\textsuperscript{20} and thus they placed a high demand on participants’ attention and executive functions.

The following tests were used: paced serial addition, size ordering, listening span, digit-span backwards, and digit-symbol encoding (transcripts of audio files of test components are available online only). In paced serial addition,\textsuperscript{21} children heard a sequence of 21 single-digit numbers and were required to add pairs together (i.e. adding the first number to the second, second to the third, and so on, scoring 1 point for each correct addition); the first set had a gap of 3 seconds between digits, the second set only 2 seconds, and the third set only 1 second. In the size-ordering test,\textsuperscript{22} children listened to a jumbled sequence of different-sized objects and were required to recall them in size order (2 points awarded for each object recalled in correct order). In the sentence-span test,\textsuperscript{23} participants heard nine trivia statements of approximately equal sentence length and were asked to score the statements as true or false; they also needed to recall the last word in each sentence at the end of the test (2 points scored for every correct recall; the trivia responses served only as a distracter). In the digit-span backwards test,\textsuperscript{24} children recalled a sequence of verbally presented numbers (from 1–9) in reverse order, scoring two points for each correct trial. For digit-symbol encoding,\textsuperscript{24} children had 2 minutes to write the corresponding abstract symbols (e.g. \(\ddagger, \pi\) under a sequence of numbers, according to a legend (1 point scored for each correctly correlated symbol).

For each test two versions of equivalent difficulty were created, \(x\) and \(\beta\), to reduce practice effects from retesting. For example, the \(x\) and \(\beta\) of the digit-span backwards exercise used different sets of randomized number sequences. Participants completed the \(x\) version on week 1 of the experiment and the \(\beta\) version on week 2.

**Procedure**

The experiment had a counterbalanced design. Within each school, the two classes of similar size at each primary level were randomly designated to either group A or group B by a blinded member of staff. Given that schools did not stream these parallel classes (e.g. for sex or ability), this form of randomization made any between-group differences in age, sex, or IQ very unlikely. Group A received the CEP in week 1 and no CEP in week 2. Group B had this order reversed.

The CEP was conducted approximately 30 minutes after lunch. On the non-CEP week teachers followed their normal curriculum plan, allowing a practical comparison between the CEP and existing educational practice. On both weeks, each participant completed one mental tracking test each day for 5 days, at the end of the school day. Therefore, participants received the CEP approximately 1 hour before testing. Tests were delivered in a fixed order (e.g. on Monday of both weeks participants completed the paced serial addition task). See Figure 1 for a schematic of the design.

Teachers were instructed not to inform their students of the experimental hypothesis and had only an invigilating role in test administration. Given the time between intervention and
testing, the frequency with which children underwent other school assessments similar to the test battery, and the natural week-to-week variability of the curriculum, it was likely that participants were sufficiently naive to the experimental aims.

With version \(a\) of the tests administered in week 1 and version \(b\) in week 2, their delivery was counterbalanced across conditions resulting in equal distributions of test versions between CEP and control conditions.

**Analysis**

An independent scorer, naive to the experimental aims, marked the booklets. Response to any particular test was excluded from further analysis if it had not been validly completed by the participant on both weeks of the experiment. Non-completion was primarily due to absence from school or a miscomprehension of the pre-recorded instructions leading to a non-response or an invalid response.

The data from the individual tests were converted into percentages of the maximum possible score on each test and then averaged to give the child’s overall performance (OP) score. If a participant had not completed a given test on both weeks this test was excluded when calculating their OP score.

To investigate whether the CEP influenced test performance, a repeated-measure mixed-design model using general linear modelling was conducted specifying intervention (CEP versus no CEP) as a within-participant factor and primary level (an approximation of age) and group (A or B) as between-participant factors. This was done for each test and for the OP score. Post-hoc analyses of age main effects were conducted using the least significant difference method. Statistical advice was that adjustment for multiple comparisons was overcautious.\(^{25}\) Post-hoc analyses of the differential effects of intervention with age were conducted using Wilcoxon signed-rank tests. To investigate group by intervention interactions (without being confounded by practice effects), between-group differences in weekly performance were analysed using Mann–Whitney tests.

**RESULTS**

In total, 1224 students completed at least one component of the test battery, and 1074 participants (88%) completed three or more of the tests on both weeks. No data on sex were recorded. The numbers of participants who successfully completed each component and their distribution by counterbalance group and age are shown in Table I.

On the general linear modelling repeated-measure analysis of the OP score, significant main effects were found for intervention \((F_{(1,1216)}=6.851; p=0.009; \eta^2=0.006)\) and age \((F_{(3,1216)}=70.514; p<0.001; \eta^2=0.148)\), but not for group \((p=0.437)\). A significant two-way interaction between intervention and group was also observed \((F_{(3,1215)}=291.111; p<0.001; \eta^2=0.193)\). All other two- and three-way interactions were non-significant \((p>0.05)\).

Further analysis of the OP score indicated a group by intervention interaction. Examination of weekly differences between groups revealed no significant difference in week 1 performance between groups (mean [SD] OP scores: group A 51.97 [17.68], group B 51.77 [15.81]; \(p=0.984\)), whereas group B performed significantly better than group A in week 2 (group A 58.20 [18.03], group B 60.19 [19.38]; \(p=0.045\)). This interaction is displayed in Figure 2.

**Table I**: Students successfully completing each test

<table>
<thead>
<tr>
<th>Test</th>
<th>Total, n</th>
<th>Group A</th>
<th>Group B</th>
<th>Class 4</th>
<th>Class 5</th>
<th>Class 6</th>
<th>Class 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paced serial addition</td>
<td>310</td>
<td>53</td>
<td>47</td>
<td>16</td>
<td>15</td>
<td>30</td>
<td>39</td>
</tr>
<tr>
<td>Size ordering</td>
<td>1039</td>
<td>54</td>
<td>46</td>
<td>26</td>
<td>30</td>
<td>26</td>
<td>18</td>
</tr>
<tr>
<td>Listening span</td>
<td>1025</td>
<td>50</td>
<td>50</td>
<td>27</td>
<td>29</td>
<td>24</td>
<td>21</td>
</tr>
<tr>
<td>Digit-span backwards</td>
<td>839</td>
<td>50</td>
<td>50</td>
<td>27</td>
<td>29</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Digit-symbol encoding</td>
<td>794</td>
<td>54</td>
<td>47</td>
<td>30</td>
<td>27</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>At least one test</td>
<td>1224</td>
<td>50</td>
<td>50</td>
<td>26</td>
<td>28</td>
<td>24</td>
<td>22</td>
</tr>
</tbody>
</table>
Analysis of age effects for the OP score indicated significant increases in mean performance from primary class 4 to class 5 (mean increase 9%; 95% confidence interval [CI] 6–12%, \( p = 0.001 \)) and from class 5 to class 6 (7%; 95% CI 3–10%; \( p = 0.001 \)), and a non-significant decrease from class 6 to class 7 (–1%; 95% CI –4% to 3%; \( p = 0.561 \)). More conservative analysis of OP score, including only participants who had completed three or more of the psychometric tests, did not significantly alter the results.

Results of the individual tests are summarized in Table II. General linear modelling analysis of the paced serial addition, listening span, and digit-span backwards tests showed similar patterns to those already reported for the OP score. For all three tests, a significant main effect of age and a significant interaction between intervention and group were found. Post-hoc analysis showed no significant performance differences between groups for all three tests on week 1 or for paced serial addition on week 2. However, group B’s performance was significantly better than that of group A on week 2 for both the listening span (mean [SD] score: group A 65% [25.6], group B 69% [28.1]; \( p = 0.001 \)) and digit-span backwards (group A 59% [33.0], group B 62% [35.3]; \( p = 0.03 \)).

For size ordering a significant interaction was reported between age and intervention; the CEP had a significant positive effect for students in class 7 (average age 11y; \( p = 0.015 \)), a negative effect for students in class 4 (average age 8y; \( p = 0.016 \)), and no significant effect for students in classes 5 and 6 (average ages 9y and 10y respectively).

For digit-symbol encoding we found a three-way interaction between all factors. On week 1 those who had the CEP (group A) showed significantly higher performance than those who did not (group B) if they were in primary class 4 (mean [SD] scores: group A 36% [9.1], group B 34% [9.0]; \( p = 0.024 \)) or class 6 (group A 55% [25.0], group B 42% [9.3]; \( p = 0.04 \)). In week 2, those in primary class 6 who had the CEP (group B) performed worse than those who did not (group A 69% [22.7], group B 49% [11.6]; \( p = 0.007 \)). All other pairwise comparisons were non-significant.

**DISCUSSION**

We found that children’s performance on a set of attentionally demanding cognitive tests improved as a result of a classroom-based exercise programme. This result is consistent with previous findings of cognitive benefits from physical exercise in older people and academic benefits in children.

Because of the counterbalanced design, this pattern of results cannot be explained simply by differential difficulty levels of the \( x \) and \( \beta \) versions of the test battery or as a result of practice effects, as benefits occurred over and above practice effects (see Figure 2). Group differences can also be excluded as an explanation, because post-hoc analysis failed to show group B outperforming group A on both weeks, which would be expected if group B had higher cognitive ability. Furthermore, our method of randomizing classes within schools ensured that both arms were of approximately equal demography. This, along with the large size of the sample, makes it highly unlikely that any significant group differences in cognitive ability existed between groups or that group differences in ability could account for the results. Therefore the increased cognitive performance on week 2 can be explained only as a result of the exercise programme.

The main effect of exercise was mediated by the counterbalance group. Post-hoc analysis predominantly found significant exercise benefits only for those receiving the intervention on the second week of the experiment (group B). This was the case for overall performance and all individual tests apart from size ordering (where the main effect of exercise was mediated by age), and for students in primary classes 4 to 6 on the visual-coding task.

Theoretical models propose that exercise influences cognition acutely by affecting arousal. If this is the case, we would expect that exercise would exert greatest benefits when the children were least aroused, such as at the end of the day, when they are most tired, or when the tests are already familiar (i.e. on week 2). Our findings are consistent with this. The exercise intervention had benefits only once the novelty of the testing situation had worn off by the second week. The digit-symbol encoding was the only exception to this; however, it is a highly routine test and was administered on a Friday at the end of the first week. In two instances the effect of exercise was moderated by age; for size ordering, positive effects were observed only in children in primary class 7. Again, it would follow that reduced task difficulty with age could render a task more routine and therefore open to the arousing effects of exercise. Future research is needed to specifically explore whether the effects of exercise on executive functions differ as a function of task difficulty or patterns of attentional demand.
We took a novel approach in our study by using en-masse test administration techniques. This made research with a very large sample feasible at a low cost in the ecologically valid environment of the classroom. A disadvantage was that testing could not be tailored to meet an individual’s needs. For example, comparatively few participants completed the paced serial addition task. Teachers and markers indicated that this was because many participants failed to comprehend the recorded instructions.

A final criticism that might be levelled at this study is that the intervention simply provided the children with a break from cognitive effort during the school afternoon, giving them more energy to complete the tasks during testing. This seems unlikely given that the physical exercise contained significant cognitive demands (e.g. carefully following instructions from teaching staff). Also, the afternoon curricular activities typical of a primary school would not normally be construed as highly cognitively tiring (e.g. art, music, drama). It could be argued that some element of the intervention other than physical activity was responsible for the increase in cognitive performance (e.g. the novelty). However, normal curricular practice is highly varied, and it is unlikely that participants viewed their activities on the intervention week as an exceptionally novel experience. Nevertheless, future studies might include a control activity such as gentle stretching. We did not include such a condition for two reasons: the study was already at the limits of intrusion into the prescribed curriculum, and the schools assured us that the teachers were using a variety of non-exercise activities to keep the children on task; thus, best current educational practice served as the control condition. We would argue strongly that this control condition has great practical significance, illustrating that an increase in physical activity within the school curriculum could boost students’ cognitive performance compared with the best current educational practice.

This research represents the largest and most methodologically thorough investigation to date into the effect of exercise on childhood cognition, and it provides compelling evidence that physical activity between lessons is a valuable component of the school curriculum, for academic as well as physical development.

ACKNOWLEDGEMENTS
This research was supported by a grant from Action Medical Research and funding from Scottish Enterprise and Aberdeen City Council. We are extremely grateful to all of the teachers, parents, guardians, and especially children who made the research possible. We also thank Jennifer Wild and Bipin Bhakta for helpful discussions.

ONLINE MATERIAL/SUPPORTING INFORMATION
Additional material and supporting information may be found in the online version of this article.

Table II: Influence of exercise, age, and group on individual test performance

<table>
<thead>
<tr>
<th>Test</th>
<th>Main effects</th>
<th>Interaction</th>
<th>Interaction-age</th>
<th>Interaction-group</th>
<th>Interaction·age</th>
<th>Interaction·group</th>
<th>Interaction·age·group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paced serial addition</td>
<td>0.970</td>
<td>&gt;0.001</td>
<td>0.198</td>
<td>0.341</td>
<td>0.021</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>Size ordering</td>
<td>0.855</td>
<td>&lt;0.001</td>
<td>0.024</td>
<td>0.002</td>
<td>0.002</td>
<td>0.013</td>
<td>0.013</td>
</tr>
<tr>
<td>Listening span</td>
<td>0.003</td>
<td>&lt;0.001</td>
<td>0.009</td>
<td>0.004</td>
<td>0.004</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>Digit-span backwards</td>
<td>0.342</td>
<td>&lt;0.001</td>
<td>0.029</td>
<td>0.030</td>
<td>0.030</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>Digit-symbol encoding</td>
<td>0.872</td>
<td>&lt;0.001</td>
<td>0.025</td>
<td>0.025</td>
<td>0.025</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>

*The effect of the Classroom Exercise Programme versus no exercise on test performance within participants.

*The effect of the Classroom Exercise Programme versus no exercise on test performance between participants.

g² shows effect size.
REFERENCES