Pilot Physical Activity Intervention Reduces Severity of ADHD Symptoms in Young Children

Alan L. Smith¹, Betsy Hoza², Kate Linnea², Julia D. McQuade², Meghan Tomb², Aaron J. Vaughn², Erin K. Shoulberg², and Holly Hook¹

Abstract

Objective: Physical activity associates with mental health and neurocognitive function, showing potential for addressing ADHD symptoms. As a preliminary assessment of this potential, the authors piloted a before-school physical activity intervention for young children. Method: Seventeen children (Grades K-3) exhibiting four or more hyperactivity/impulsivity symptoms on the Disruptive Behavior Disorders Rating Scale (Pelham, 2002) completed about 26 min of continuous moderate-to-vigorous physical activity daily over eight school weeks. The authors administered cognitive, motor, social, and behavioral functioning measures at pre- and postprogram, assessed response inhibition weekly, and coded negative behaviors daily. Results: Several measures showed significant or marginally significant change over time (effect size = 0.35-0.96) with additional measures showing meaningful effect size values (≥ 0.20). Response inhibition effects were most consistent. Most participants (64% to 71%) exhibited overall improvement according to postprogram parent, teacher, and program staff ratings. Conclusion: Physical activity shows promise for addressing ADHD symptoms in young children. (J. of Att. Dis. 2013; 17(1) 70-82)

Keywords

ADHD treatment, attention, exercise, school-based intervention

ADHD is a chronic health condition that is considered the most common neurobehavioral disorder experienced by children (American Academy of Pediatrics, 2000). Present in 3% to 7% of the school-aged population, ADHD is expressed in inattentive, hyperactive/impulsive, and combined forms (American Psychiatric Association, 2000). The key features of inattention and/or hyperactivity/impulsivity are generally present by the age of 7 and foster impairments across at least two settings, most often home and school. Such impairments include academic and social difficulties, family distress and dysfunction, and challenges in various extracurricular settings such as sports (Hoza, Owens, & Pelham, 1999). Difficulties often persist and expand (e.g., driving accidents, illicit drug use, school dropout) over the adolescent and adult years (Mannuzza & Klein, 2000), making chronic management of the disorder essential. In light of the chronic nature of ADHD and the substantial economic impact of ADHD in children and adolescents, estimated to be US$36 to US$52.4 billion annually (Pelham, Foster, & Robb, 2007), it is recognized as a major public health issue (National Institutes of Health, 2000).

The primary treatment approaches for ADHD are medication management, behavioral treatment, or a combination of the two. The National Institute of Mental Health Multimodal Treatment Study of Children with ADHD (MTA; MTA Cooperative Group, 1999a, 1999b; Richters et al., 1995) significantly extended the database on long-term treatment and serves as a benchmark for research quality (Schachar et al., 2002). The MTA findings suggest that a rigorously monitored medication management plan offers meaningful benefits to school-aged children, whether alone or in combination with behavioral treatment. Medication and combined treatments show an advantage over behavioral and community care for ADHD outcomes. However, when extending assessment beyond ADHD symptoms to other functional outcomes (e.g., academic performance, aggression, internalizing, social skills/relationships), combined approaches may be preferred (MTA Cooperative

¹Purdue University, West Lafayette, IN, USA
²University of Vermont, Burlington, USA

Corresponding Author:
Alan L. Smith, Department of Health and Kinesiology, 800 West Stadium Avenue, Purdue University, West Lafayette, IN 47907-2046, USA
Email: alsmith7@purdue.edu
Group, 1999a). Consistent with this conclusion, secondary analyses involving a composite outcome variable demonstrated a small benefit of combined treatment over medication management and moderate to moderately large benefit over behavioral treatment and community care (Conners et al., 2001). Anchoring these outcomes to clinical relevance, additional secondary analyses indicated that 68% of those receiving combined treatment achieved parent and teacher average symptom ratings of “just a little” or below, considered a threshold for successful treatment, followed by medication management (56%), behavioral treatment (34%), and community care (25%; J. M. Swanson et al., 2001).

Despite these documented benefits, MTA follow-up assessments show reduction of effects by 10 months post-treatment (MTA Cooperative Group, 2004) and no detectable treatment group differences 22 months after completion of study treatment (Jensen et al., 2007). Moreover, there are a variety of challenges associated with ADHD treatment. As evident in the MTA findings, there is substantial discrepancy between what is offered to children in community care settings and what yields greatest success. Demands associated with most successful treatment such as frequent visits to the physician and/or therapist, taking medicine one or more times daily, and chronic behavioral management are onerous. Moreover, parents may be hesitant to pursue stimulant medication management, especially in young children, because of tolerability (T. Wigal et al., 2006), documented side effects such as trouble in sleeping and reduced growth rate (J. Swanson et al., 2006; T. Wigal et al., 2006), and extant debate on cardiovascular risk (Biederman, Spencer, Wilens, Prince, & Faraone, 2006; Nissen, 2006). Thus, additional intervention options are needed that are amenable to long-term maintenance and more easily incorporated into the lifestyles of children. In this pilot work, we explore the potential for physical activity to serve as such an option.

Physical activity is any movement produced by skeletal muscles and resulting in energy expenditure (Caspersen, Powell, & Christenson, 1985). For children this can come in the form of active transport (e.g., walking or biking to school), informal play, sports, physical education, and deliberately structured activity to promote fitness components (i.e., exercise). Though the bulk of work examining physical activity and health has focused on adults, emerging evidence shows an association between physical activity and a host of physical and mental health markers in young people (Stensel, Gorely, & Biddle, 2008). The indicators of mental health that are most extensively examined in the youth physical activity literature are self-concept, depression, and anxiety. The extant literature suggests that physical activity involvement positively associates with various components of self-concept and negatively associates with anxiety and depression (for reviews, see Calfas & Taylor, 1994; Strong et al., 2005). Specifically examining children aged 6 to 14 years with ADHD, a recent study by Kiluk, Weden, and Culotta (2009) showed greater sport involvement to be associated with lower anxious-depressed, internalizing problems, and affective T-scores on the Child Behavior Checklist (Achenbach, 1991). Such associations were not observed for a comparison sample of children with learning disabilities, suggesting that youth at risk for mood problems may particularly benefit from physical activity. Moreover, for both boys and girls with ADHD, those engaging in three or more sports showed significantly lower anxious-depressed T-scores than those engaging in two or fewer sports. Thus, consistent involvement in physical activity may be an important matter when considering physical activity as a strategy for addressing ADHD symptoms.

Beyond the link to mental health described above, physical activity also shows associations with neurocognitive function. Early interest in this connection was largely focused on later adulthood, where physically active lifestyles and aerobic physical activity interventions have been linked to the preservation and enhancement of cognitive function, particularly executive control processes (Hall, Smith, & Keele, 2001; Kramer et al., 1999). Extant animal and human work suggests that such physical activity affects the brain in ways that would be expected to impact executive function. For example, animal studies suggest that, in rodents, aerobic physical activity benefits learning and memory, plasticity in the hippocampal formation, neurogenesis, and neurotrophin expression (see Cotman, Berchtold, & Christie, 2007; Olson, Eadie, Ernst, & Christie, 2006). Also, recent human studies suggest that aerobic physical activity enhances frontal brain function (Colcombe et al., 2004), affects frontal brain structure (Colcombe et al., 2006), increases serum levels of brain-derived neurotrophic factor (BDNF; Ferris, Williams, & Shen, 2007; Tang, Chu, Hui, Helmeste, & Law, 2008), and promotes hippocampal neurogenesis (Pereira et al., 2007). Such findings are relevant to ADHD in that executive control deficits, in particular behavioral inhibition challenges, are core impairments associated with ADHD (Barkley, 1997).

Importantly, ADHD symptoms extend beyond executive function deficits. Indeed, it appears that multiple brain regions and neurochemical pathways are implicated in the disorder (see Nigg, 2006). Compared with typically developing individuals, those with ADHD have been shown to have less brain volume not only in frontal regions associated with executive function but also in the cerebellum and other cortical regions (Halperin & Healey, 2011; Seidman, Valera, & Makris, 2005). Moreover, despite children with hyperactive/impulsive and combined-type ADHD ostensibly being quite active as a function of hyperactivity, children with ADHD are at risk for movement skill deficits and motor problems as well as poor levels of physical fitness (Diamond, 2000; Harvey & Reid, 2003; Nigg, 2006). Patterns of chronic physical activity engagement of children with ADHD have not been examined, though the
documented movement skill and fitness deficits suggest that hyperactivity in itself does not yield movement and fitness outcomes consistent with typically developing children. Thus, exposing children with ADHD to consistent, deliberately structured physical activity may help accommodate for observed motor and fitness deficits associated with the disorder as well as other cognitive, behavioral, and social deficits. In light of its broad ranging effects on the brain, physical activity may be well suited for addressing a disorder such as ADHD, which is characterized by imprecise etiology and broad neurocognitive effects. 

More recently, attention has been paid to the potential link of physical activity with cognitive, behavioral, and academic performance in children (see Best, 2010; Centers for Disease Control and Prevention, 2010; Tomporowski, Davis, Miller, & Naglieri, 2008). Sibley and Etnier (2003) conducted a meta-analysis of initial efforts exploring the activity to cognitive function link, including 44 studies (125 effect sizes) in their analysis. The overall effect size was 0.32, and moderator analyses showed that middle school-aged children (effect size = 0.48 ± 0.27) followed by young elementary children (effect size = 0.40 ± 0.26) showed the largest effect sizes. The authors speculated that findings for young elementary-aged children may be explained by the importance of movement to young children’s cognitive development. This is an interesting consideration in light of the above-described movement challenges experienced by many children with ADHD and recent animal work addressing physical activity and neurogenesis. Kim et al. (2004) demonstrated age-related differences in the effect of treadmill exercise on cell proliferation within the dentate gyrus of rats. Specifically, they found treadmill exercise to increase cell proliferation of 4-, 8-, and 62-week-old rats, with the most active proliferation in the youngest rats and less active proliferation with increasing age. This work suggests that employing physical activity as an intervention to address ADHD may be especially beneficial during young childhood.

The association of physical activity with positive mental health and neurocognitive function in children, the relatively broad effects of physical activity, and the plasticity of the young brain point to physical activity as a potentially valuable intervention for children with ADHD symptoms. Aside from descriptive work on physical activity and cognitive function (Gapin & Etnier, 2010) and research examining the impact of acute physical activity bouts (Tantillo, Kesick, Hynd, & Dishman, 2002; S. B. Wigal et al., 2003) on children with ADHD, we located no published investigations of an extended aerobic physical activity protocol for the management of ADHD symptoms. An extended protocol presumably would be necessary for effective ADHD management and to afford any benefits to brain development. The present study was designed as a preliminary effort to address the viability of such a protocol. Specifically, the purpose of the present study was to pilot a before-school physical activity intervention for reducing ADHD symptoms in young children. We hypothesized that completing physical activity before each school day over a 9-week time frame would foster adaptive change in cognitive, behavioral, motor, and social symptoms associated with ADHD.

Method

Participants

Eligible participants (N = 17) were identified using parent and teacher ratings on the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV) version of the Disruptive Behavior Disorders (DBD) Rating Scale (Pelham, 2002; Pelham, Gnagy, Greenslade, & Milich, 1992). The DBD is comprised of the DSM-IV symptoms of ADHD, oppositional defiant disorder (ODD), and conduct disorder (CD); however, CD items were not used given the ages of participants. Kindergarten through third graders in a diverse, low-SES Vermont community (54% of students on free or reduced lunch) with parental consent were screened. We sought children at risk for ADHD and set four or more symptoms of hyperactivity/impulsivity as a requirement for eligibility. The mean number of symptoms endorsed was 7.1 for hyperactivity/impulsivity (SD = 1.9), 5.6 for inattention (SD = 2.7), and 4.9 for ODD (SD = 2.6). Participants were 5.2 to 8.7 years of age at entry (M = 6.7, SD = 1.0). Three children dropped out in the first 2 weeks and 14 completed the program; data from these 14 children, all medication naïve, are summarized in the “Results” section. Six boys and 8 girls comprised the completers. In the final sample, 12 participants were White and the 2 remaining participants were Black.

Procedure

The physical activity program was designed to maximize participant involvement in continuous moderate-to-vigorous physical activity (MVPA) within an engaging, before-school setting. Each day the program was for 30 min, with participants completing four stations within small groups (about five children) that each lasted 6 min. Stations were designed to foster sustained MVPA within the context of games and activities that required participants to employ a variety of motor skills (e.g., moving objects to different locations in the activity area, various forms of locomotion—skipping, running, hopping, crab walking). Staff members were trained to limit the provision of verbal instructions with the children being sedentary. Instead, staff members modeled desired behaviors and provided instruction while the children were moving. One staff member supervised each station. In addition to these stations, one large-group
absence of prior studies indicating what measures may be
necessary. Given the preliminary nature of this pilot work, and the
feasibility of implementing a behavior management system, the
inclusion of such a system may be prudent. However, confounds
were not formally addressed, and no formal behavior management
system was in place and staff were instructed to handle negative
behavior as best they could; they were not permitted to use time out
or other behavioral strategies for negative behavior.

Program staff received approximately 20 hr of training and
practice in the implementation of the program prior to
program initiation to ensure reliable administration of the
intervention and assessment procedures. Most of the staff
who ran the physical activity program had previous experience
with youth with ADHD or teaching physical activity skills in camp settings, and all were undergraduate
or graduate students in psychology.

Preassessments listed in the “Pre–Post Program Measures” section were completed the week immediately before
the start of the program. Eight weeks of physical activity
were then completed over a 9-week period. Two weeks (9
school days) of the intervention were conducted, followed
by the participants’ 1.5-week spring break, a partial (2 days)
week of intervention, and then five full intervention weeks.
In addition to the pre–post measures, various weekly and
daily ecological/behavioral assessments were completed
(described below). Postassessments were completed the
week immediately following the program; teachers, par-
ents, and staff also provided global improvement ratings
(described below) at the end of the program.

Children were provided school bus transportation to the
program. Participants were not compensated monetarily,
but attendance was tracked daily and children were given a
star on a chart for each day they attended. Children with
good attendance received good attendance certificates peri-
odically throughout the program and an attendance award
(a small trophy) on the last day of the program. Rates of
attendance were high, with 86% (12 of 14) of children
meeting or exceeding 75% attendance. A problem arose
partway through the program with children sitting down to
rest as soon as they got tired or out-of-breath; hence, a sys-
tem was initiated whereby children who “kept moving” the
totality of at least three of the activity stations earned a
small trinket from a grab bag that day. This was contingent
only on movement, not any other behavior(s). To avoid
confounds, no formal behavior management system was in
place and staff were instructed to handle negative behavior
as best they could; they were not permitted to use time out
or other behavioral strategies for negative behavior.

Measures

Given the preliminary nature of this pilot work, and the
absence of prior studies indicating what measures may be
sensitive to physical activity effects on ADHD symptoms,
a broad range of measures was employed. Direct measures
of motor, cognitive, social, and behavioral capacities were
used, as well as behavioral and improvement ratings by
multiple informants. We report results of all measures
described below, emphasizing effect sizes rather than sig-
nificance of comparisons, in the hope that these preliminary
data are useful to future investigators exploring physical
activity effects across a variety of outcome domains.

Pre–Post Program Measures

Bruininks–Oseretsky Test of Motor Proficiency, 2nd edition
(BOT-2). The BOT-2 (Bruininks & Bruininks, 2005) is a
widely used, reliable, valid, and age-normed assessment
of children’s motor proficiency, including both gross-motor
and fine-motor tasks. The short form of the BOT-2 was
administered in the present study. It consists of the follow-
ing fine-motor tasks: drawing lines through crooked paths
on a sheet of paper, folding paper on lines, copying a square,
copying a star, and transferring pennies one at a time from
one hand to the other (and then placing in a box) for 15 s,
and the following gross-motor tasks: synchronized jumping
in place, synchronized tapping of fingers and feet, walking
-forward on a line, standing on one leg on a small beam for
10 s (eyes open), hopping on one foot for 15 s, dropping
and catching a tennis ball after one bounce using both hands,
dribbling a tennis ball (alternating hands), knee push-ups
for 30 s, and sit ups for 30 s. The short form of the BOT-2
produces a single score standardized to a mean of 50
and standard deviation of 10, on which higher scores index
greater motor proficiency. The norm sample for the BOT-2
consisted of randomly selected individuals from ages 4
through 21 years, stratified within age groups by sex, race/
ethnicity, socioeconomic status, and disability status
(including ADHD; Bruininks & Bruininks, 2005). For the
short form of the BOT-2 with knee push-ups, the 5- through
8-year-old children in the norm sample exhibited internal
consistency reliability values ranging from 0.75 to 0.86. Test–
retest reliability for a subgroup of 4- through 7-year-old
children was 0.86. Interrater reliability was 0.98 (see Bruin-
inks & Bruininks, 2005).

Motor timing task. Each child completed a motor timing
task (Zelaznik, Spencer, & Ivry, 2008) that purportedly
reflects cerebellar function. Specifically, participants
attempted to synchronize pressing the space bar on a com-
puter keyboard to a metronome set at a rate of 2 Hz (500 ms
period). Following 10 s (19 intervals) of synchronization,
participants continued to press without the metronome for
about 20 s, attempting to maintain the rate. Participants
completed 15 to 20 trials of this task. The time series of
inter-tap intervals from the continuation (i.e., no metronome)
portion of the task were detrended and the coefficient of
variation was calculated. The coefficient of variation is
defined as the standard deviation of the time series divided
by the average intertap interval, converted to a percentage. The best six trials (i.e., smallest coefficients) were averaged and used in the primary data analysis. Using best trials limits the degree to which timing variability stems from changing participant tapping strategies within and across trials (see Zelaznik, Spencer, & Ivry, 2002).

**Shape School.** The Shape School (Espy, 1997; Espy, Bull, Martin, & Stroup, 2006) assesses setting and response inhibition in young children using a storybook format that presents tasks of increasing difficulty. Condition A assesses cognitive-processing speed, where children are asked to name the colors of the depicted figures. The subsequent conditions assess executive function. Condition B requires children to name colors of “happy-faced figures” and inhibit naming “sad-faced figures”; Condition C requires naming the colors of figures without hats and naming the shapes of figures with hats; Condition D requires children to name the colors of happy-faced figures without hats, inhibit naming sad-faced figures without hats, name the shapes of happy-faced figures wearing hats, and inhibit naming sad-faced figures with hats. In the present study, the efficiency scores (number correct divided by time to completion) for Conditions A through D were analyzed. Espy et al.’s (2006) validation sample for the Shape School was 219 typically developing preschoolers, 54% girls, 85% White, and described as possessing varied vocabulary skills. Executive conditions of the Shape School (i.e., all conditions except A) had internal consistency reliability values exceeding 0.70 and showed significant correlations with other measures purported to assess executive function (see Espy et al., 2006). Moreover, recent work shows inattention/hyperactivity to correspond with lower inhibitory efficiency on the Shape School (Pritchard & Woodward, 2011).

**Mazes.** The Mazes subtest from either the Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI-R; Wechsler, 1989) or the Wechsler Intelligence Scale for Children–III (WISC-III; Wechsler, 1991) was administered depending on the child’s age. Mazes assesses “the child’s planning ability and perceptual-organizational ability—that is, the ability to plan and follow a visual pattern” (Sattler, 2001, p. 294). In the present study, age-normed scaled scores \((M = 10, SD = 3)\) were derived and analyzed with higher scores indexing more advanced planning ability. The standardization sample for the WPPSI-R was 1,700 children ranging in age from 3 years to 7 years 3 months; the standardization sample for the WISC-III was 2,200 children ranging in age from 6 to 16 years (Sattler, 2001). The mean internal consistency reliabilities for the WPPSI-R and WISC-III Mazes, respectively, are 0.77 and 0.70 (Sattler, 2001).

**Finger Windows.** The Finger Windows subtest from the Wide Range Assessment of Memory and Learning–second edition (WRAML-2; Sheslow & Adams, 2003) was employed as a measure of spatial working memory. Specifically, children are shown an 8 × 11 in. card that has nine holes cut in various locations. The examiner puts a pencil through the holes in various prespecified sequences of increasing difficulty and the child is required to correctly replicate each one using her or his index finger. Scores are derived based on the number of sequences correctly replicated and are converted to scaled scores \((M = 10, SD = 3)\) based on age norms. The standardization sample for the WRAML-2 was 1,200 individuals, ranging in age from 5 to 90 years (Sheslow & Adams, 2003). Higher scores indicate better spatial working memory. As reported in the test manual, the coefficient alpha reliabilities of the Finger Windows subtest ranged from 0.76 to 0.83 across the 5- to 8-year-11-month age range (Sheslow & Adams, 2003).

**Sentence Memory.** The Sentence Memory subtest, also from the WRAML-2 (Sheslow & Adams, 2003), was employed as a measure of verbal working memory. Specifically, children are read a series of sentences of increasing difficulty and asked to recall each one exactly as stated. Points are deducted for each error of omission, commission, or change of a word (e.g., tense, contracting two words). Scores are derived based on the degree of correct recall and are converted to scaled scores \((M = 10, SD = 3)\) based on age norms. Higher scores indicate better verbal working memory. As reported in the test manual, the coefficient alpha reliabilities of the Sentence Memory subtest ranged from 0.69 to 0.78 across the 5-to 8-year-11-month age span (Sheslow & Adams, 2003).

**Numbers Reversed.** The Numbers Reversed subtest from the Woodcock–Johnson III Tests of Cognitive Abilities (WJ-III; Woodcock, McGrew, & Mathur, 2001) was administered, which requires children to listen to sequences of number digits read by the examiner and then to reproduce them in reverse order. This task measures working memory skills that require mental manipulation and also has been shown to assess planning (e.g., Schofield & Ashman, 1986). Scores are derived based on the degree of correct performance and are converted to standard scores \((M = 100, SD = 15)\) based on age norms. The standardization sample for the WJ-III included 8,818 individuals selected to be representative of persons aged 2 to 90+ years from 100 different communities spanning the Northeast, Midwest, South, and West portions of the United States (McGrew & Woodcock, 2001). Higher scores indicate better verbal working memory. According to Appendix A of the technical manual reliability coefficients for the Numbers Reversed subtest ranged from .84 to .92 across the 5- to 8-year-old age range in the standardization sample.

**Weekly Measures**

**Simon Says.** The proportion of inhibition failures during the popular children’s game Simon Says was used as a weekly ecological measure of response inhibition. This task has been used previously in research (Strommen, 1973) to assess inhibition errors in young children. In a given
administration of the task, 30 commands were provided with 10 instances absent the preliminary statement “Simon Says.” These 10 instances were coded for inhibition failure, and were differently sequenced from week to week to prevent improvements stemming from memory of a command pattern. Timing of commands was constant across an administration of the task. Interrater reliability (agreements divided by the sum of agreements and disagreements) was assessed on detection of inhibition errors beginning in Week 2 and averaged 0.94 over the remainder of the program.

Red Light/Green Light. Similarly, the proportion of inhibition errors during a weekly Red Light/Green Light game was used as an ecological measure of response inhibition. Specifically, children were permitted to take one step following a “red light” call, and any additional steps or loss of balance involving touching the ground (with any body part except one’s feet) were coded as inhibition errors. In a given administration of the task, 27 commands were provided with 15 instances being a “red light” and coded by observers. The command sequence was changed from week to week and the length of “green light” portions varied, ranging from 1 to 4 s, within a session to prevent improvements stemming from memory of the command pattern. Interrater reliability was assessed on detection of inhibition errors beginning in Week 2 and averaged 0.95 over the remainder of the program.

Pittsburgh Modified Conners Teacher Rating Scale. Teacher ratings of behavior on the Pittsburgh Modified Conners Rating Scale (Pelham, 2002), which includes the widely used 10-item Abbreviated Conners (Goyette, Conners, & Ulrich, 1978) and Iowa Conners scales (Loney & Milich, 1982), as well as a Peer Interaction Scale developed by Pelham (2002), were collected on at least a weekly basis. For children with more frequent teacher ratings, items were averaged to produce a weekly item score. Teachers rated on a 4-point scale (ranging from not at all to very much) the extent to which each problem behavior was present. Scale scores were derived separately for the Abbreviated Conners, the inattention/overactivity and oppositional/defiant subscales of the Iowa Conners, as well as the Peer Interactions Scale, by averaging across raters. Similarly, the averaged frequency of observed behaviors per program station over the first half of the program were compared with averaged data over the second half of the program. In light of the relatively limited power to detect statistically significant change over time, effect size values were calculated and interpreted.

Daily Observations of Behavior. Trained behavioral observers recorded frequencies of the following negative behaviors nicely, interrupting, intentional aggression, unintentional aggression, and not following adult directions. Observational methods were derived specifically for this study to fit the ages of participants and the research goals but were modeled after established procedures (Pelham & Hoza, 1996). Observers received approximately 20 hr of training and practice with the observational procedures prior to commencing the program. Because of limited resources and the staffing demands required to administer this pilot program, only sporadic checks of the reliability of the observation system could be conducted. There was inconsistency of interrater reliability values (complete disagreement through complete agreement) across these spot checks. This in part is likely because the checks covered short periods of time and were infrequent, meaning a low number of overall behaviors would be observed (i.e., any rater disagreement would substantially impact reliability). Accordingly, results pertaining to behavioral observations should be interpreted with caution.

Post Program Only Improvement Ratings. Ratings of perceived degree of improvement were completed by teachers, parents, and program staff at post program following procedures used extensively in prior work by Pelham and colleagues (Hoza & Pelham, 1995; Pelham et al., 2000; Pelham & Hoza, 1996). Global ratings of overall improvement as well as ratings in specific areas (see Table 2) were assessed on a 7-point scale (1 = very much worse; 7 = very much improved). Consistent with prior work, children for whom raters did not perceive a problem in a given area were rated as “no problem” and not included in analyses for that domain. In addition, both teachers and program staff completed end-of-program ratings indexing the degree to which each child’s behavior with (a) peers and (b) adults was “like that of a normal child” (0 = very much like a normal child; 6 = not at all like a normal child). Finally, teachers and staff rated the degree to which they found interacting with the child pleasant by the end of the program (0 = very pleasant; 6 = very unpleasant). Parent and teacher ratings were each provided by a single informant. As there were multiple program staff providing ratings of each child, a single score for each child was derived by averaging across raters.

Data Analysis

Dependent t tests were employed to test change over the program on pre- and post program measures, weekly measures, and daily observations of behavior. For weekly measures, averaged data over the first half of the program were compared with averaged data over the second half of the program. Similarly, the averaged frequency of observed behaviors per program station over the first half of the program was compared with the averaged frequency over the second half of the program. In light of the relatively limited power to detect statistically significant change over time, effect size values were calculated and interpreted.
Table 1. Descriptive Statistics, Dependent t-Values, and Effect Size Values for Pre–Post and First Half–Second Half Comparisons

<table>
<thead>
<tr>
<th>Measure</th>
<th>Preprogram or first half</th>
<th>Postprogram or second half</th>
<th>n&lt;sup&gt;a&lt;/sup&gt;</th>
<th>t</th>
<th>p</th>
<th>Effect size&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-post program measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOT-2 short form&lt;sup&gt;c&lt;/sup&gt;</td>
<td>40.77 (3.35)</td>
<td>44.00 (5.97)</td>
<td>13</td>
<td>−3.27</td>
<td>&lt;.01</td>
<td>0.96</td>
</tr>
<tr>
<td>Continuation timing</td>
<td>15.68 (9.55)</td>
<td>14.21 (6.47)</td>
<td>12</td>
<td>0.72</td>
<td>.49</td>
<td>0.15</td>
</tr>
<tr>
<td>Shape School, Condition A</td>
<td>1.25 (0.29)</td>
<td>1.22 (0.43)</td>
<td>10</td>
<td>0.31</td>
<td>.76</td>
<td>−0.10</td>
</tr>
<tr>
<td>Shape School, Condition B</td>
<td>0.75 (0.23)</td>
<td>0.83 (0.25)</td>
<td>12</td>
<td>−3.06</td>
<td>&lt;.05</td>
<td>0.35</td>
</tr>
<tr>
<td>Shape School, Condition C</td>
<td>0.48 (0.19)</td>
<td>0.54 (0.18)</td>
<td>11</td>
<td>−1.78</td>
<td>.11</td>
<td>0.32</td>
</tr>
<tr>
<td>Shape School, Condition D</td>
<td>0.37 (0.11)</td>
<td>0.41 (0.16)</td>
<td>12</td>
<td>−1.42</td>
<td>.18</td>
<td>0.36</td>
</tr>
<tr>
<td>Mazes&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.43 (2.56)</td>
<td>9.00 (2.08)</td>
<td>14</td>
<td>−0.87</td>
<td>.40</td>
<td>0.22</td>
</tr>
<tr>
<td>Finger Windows&lt;sup&gt;c&lt;/sup&gt;</td>
<td>8.07 (2.84)</td>
<td>8.29 (3.02)</td>
<td>14</td>
<td>−0.26</td>
<td>.80</td>
<td>0.08</td>
</tr>
<tr>
<td>Sentence Memory&lt;sup&gt;c&lt;/sup&gt;</td>
<td>10.38 (3.28)</td>
<td>10.85 (3.08)</td>
<td>13</td>
<td>−0.79</td>
<td>.45</td>
<td>0.14</td>
</tr>
<tr>
<td>Numbers Reversed&lt;sup&gt;c&lt;/sup&gt;</td>
<td>99.50 (10.59)</td>
<td>104.08 (15.22)</td>
<td>12</td>
<td>−1.23</td>
<td>.25</td>
<td>0.43</td>
</tr>
<tr>
<td><strong>Weekly measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Simon Says</td>
<td>0.29 (0.22)</td>
<td>0.33 (0.29)</td>
<td>13</td>
<td>−0.51</td>
<td>.62</td>
<td>−0.18</td>
</tr>
<tr>
<td>Red Light/Green Light</td>
<td>0.33 (0.15)</td>
<td>0.24 (0.14)</td>
<td>14</td>
<td>2.42</td>
<td>&lt;.05</td>
<td>0.60</td>
</tr>
<tr>
<td>PMCTRS—Abbreviated Conners</td>
<td>0.94 (0.53)</td>
<td>0.62 (0.42)</td>
<td>14</td>
<td>2.18</td>
<td>&lt;.05</td>
<td>0.60</td>
</tr>
<tr>
<td>PMCTRS—Iowa I/O</td>
<td>1.20 (0.53)</td>
<td>0.83 (0.51)</td>
<td>14</td>
<td>2.23</td>
<td>&lt;.05</td>
<td>0.70</td>
</tr>
<tr>
<td>PMCTRS—Iowa O/D</td>
<td>0.67 (0.54)</td>
<td>0.45 (0.44)</td>
<td>14</td>
<td>2.66</td>
<td>&lt;.05</td>
<td>0.41</td>
</tr>
<tr>
<td>PMCTRS—Peer interactions</td>
<td>0.46 (0.43)</td>
<td>0.29 (0.31)</td>
<td>14</td>
<td>2.03</td>
<td>.06</td>
<td>0.40</td>
</tr>
<tr>
<td><strong>Daily observational measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not speaking nicely</td>
<td>0.15 (0.12)</td>
<td>0.17 (0.18)</td>
<td>14</td>
<td>−0.70</td>
<td>.50</td>
<td>−0.18</td>
</tr>
<tr>
<td>Interrupting</td>
<td>0.23 (0.14)</td>
<td>0.12 (0.09)</td>
<td>14</td>
<td>4.10</td>
<td>&lt;.01</td>
<td>0.78</td>
</tr>
<tr>
<td>Intentional aggression</td>
<td>0.27 (0.21)</td>
<td>0.30 (0.34)</td>
<td>14</td>
<td>−0.44</td>
<td>.67</td>
<td>−0.12</td>
</tr>
<tr>
<td>Unintentional aggression</td>
<td>0.26 (0.13)</td>
<td>0.21 (0.12)</td>
<td>14</td>
<td>1.96</td>
<td>.07</td>
<td>0.40</td>
</tr>
<tr>
<td>Not following adult directions</td>
<td>1.31 (0.86)</td>
<td>1.32 (0.75)</td>
<td>14</td>
<td>−0.27</td>
<td>.79</td>
<td>−0.02</td>
</tr>
</tbody>
</table>

Notes: PMCTRS = Pittsburgh Modified Conners Teacher Rating Scale; I/O = inattention/overactivity; O/D = oppositional/defiant.

<sup>a</sup>Reduced n for some analyses because of missing data, administrator/recording error, and/or participant failure to understand directions.

<sup>b</sup>Effect size calculated by dividing pre-post difference by preprogram standard deviation (positive value assigned to finding in expected/adaptive direction).

<sup>c</sup>Scaled/standard scores.

Effect size was calculated by dividing the pre-post program difference by the standard deviation at preprogram. A positive sign was assigned to values corresponding to change in an adaptive direction and a negative sign was assigned to values corresponding to change in a maladaptive direction. Values of 0.2, 0.5, and 0.8 were interpreted as small, medium, and large effect sizes, respectively (Cohen, 1988).

**Results**

Mean and standard deviation values for pre-post program measures, weekly measures, and daily observational measures are found in Table 1. Statistically significant pre-to-post (or first half-to-second half) program changes were observed for BOT-2, Shape School Condition B, Red Light/Green Light, Abbreviated Conners, Iowa Conners (both inattention/overactivity and oppositional/defiant subscales), and Interrupting scores (see Table 1). All significant changes were in adaptive directions. The associated effect sizes were of small or medium magnitude, with the exception of the BOT-2, which had a large effect size. The BOT-2 effect size would be expected in light of the nature of the program and provides support for the efficacy of the physical activity intervention. Six additional measures yielded positive effect sizes greater than 0.20, indicating improvement that did not reach statistical significance. Overall, the relatively larger positive effect size values tended to correspond with measures ostensibly assessing response inhibition/impulsiveness. There was no evidence of an adverse effect of the program, with all negative effect sizes being less than 0.20 in magnitude.

Table 2 shows the percentage of parents, teachers, and staff rating participants as “somewhat,” “much,” or “very much” improved by the program. Depending on rater and the focal domain, from 29% to 71% of participants were rated as exhibiting some degree of improvement. The greatest proportions of improvement were reported for the overall rating. Never did more than two participants receive parent ratings that indicated worsening on a given domain.
activity intervention was designed in a way that afforded from pre- to postprogram. This suggests that the physical meaningful effect size were in a maladaptive direction, suggesting that physical activity is unlikely to be harmful when participating essentially served as a second piece of feedback available from viewing coactive tasks. Many children could return to their previous body position before their behaviors were coded as inhibition failures. That is, we were able to reliably measure behavioral responses on this particular task, children could correct themselves before this manipulation timing task. Because this task is simplified and appears to involve a variety of brain regions (see Nigg, 2006), such a dynamic physical activity intervention seems appropriate.

Of note is the absence of intervention effect on the continuous timing task. Because this task is simplified and highly timing dependent, it is strongly linked to cerebellar function (Ivry, Keele, & Diener, 1988). The findings suggest that if any benefits of a physical activity intervention to motor timing and cerebellar function exist, they would be less marked than those for broader motor functioning and/or would require an extended intervention time frame to detect. Pursuit of long-term intervention investigations will offer opportunity to understand this finding better. In light of the imprecise etiology of ADHD, it is reasonable to expect that physical activity benefits on particular symptoms or brain systems would be neither uniform nor accrued at equal rates.

A broad range of tasks were employed to measure cognitive function of study participants, with measures assessing response inhibition, set shifting, planning, and working memory. Such executive functions were selected for examination because deficits in these functions represent core impairments associated with ADHD (Barkley, 1997) and emerging work indicates that fitness levels and physical activity may be linked to their preservation and enhancement (Etman & Chang, 2009; Hillman, Buck, Themanson, Pontifex, & Castelli, 2009). The strength of findings was mixed across tasks. For example, working memory tasks did not show meaningful effect size values (i.e., ≥ 0.20) except when involving additional cognitive load such as the Numbers Reversed task (effect size = 0.43). The latter task involves planning, which in children has been shown to benefit from exercise training (Davis et al., 2007). However, the Mazes task, which purportedly measures planning, yielded a small effect size in our study. The more consistent effects were found for the tasks requiring response inhibition. The effect size values for these tasks ranged from 0.35 to 0.60, with the exception of the Simon Says task. Though we were able to reliably measure behavioral responses on this particular task, children could correct themselves before their behaviors were coded as inhibition failures. That is, children could return to their previous body position before fully adopting the body position to inhibit. Many children appeared to use the feedback available from viewing coacting peers, which essentially served as a second piece of

### Table 2. Percentage of Participants Perceived as Improved by Adult Raters

<table>
<thead>
<tr>
<th>Measure</th>
<th>Parent</th>
<th>Teacher</th>
<th>Staff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name calling/teasing/aggression</td>
<td>33</td>
<td>33</td>
<td>58</td>
</tr>
<tr>
<td>Defiance/noncompliance</td>
<td>42</td>
<td>50</td>
<td>69</td>
</tr>
<tr>
<td>Self-esteem</td>
<td>67</td>
<td>54</td>
<td>71</td>
</tr>
<tr>
<td>Responsibility</td>
<td>58</td>
<td>31</td>
<td>29</td>
</tr>
<tr>
<td>Social skills</td>
<td>67</td>
<td>Not rated</td>
<td>57</td>
</tr>
<tr>
<td>Following rules</td>
<td>46</td>
<td>54</td>
<td>43</td>
</tr>
<tr>
<td>Overall</td>
<td>69</td>
<td>64</td>
<td>71</td>
</tr>
</tbody>
</table>

Note: Rater indicated “somewhat improved,” “much improved,” or “very much improved.”

For teacher ratings, on no occasion did more than one participant receive such a rating. Normalization ratings for behavior with peers (teacher rating: \( M = 2.00, SD = 1.52 \); staff rating: \( M = 1.77, SD = 0.90 \)) and with adults (teacher rating: \( M = 2.07, SD = 1.54 \); staff rating: \( M = 1.77, SD = 0.88 \)) corresponded to participants behaving more “like a normal child” than not. Finally, teachers (\( M = 2.07, SD = 1.64 \)) and staff (\( M = 1.56, SD = 0.96 \)) rated participants overall as relatively more pleasant than not to interact with by the program end.

### Discussion

The data obtained in the present study suggest that there is potential value in exploring regular physical activity as a management tool for ADHD. Specifically, our preliminary work suggests that sustained involvement in structured physical activity may offer benefits to motor, cognitive, social, and behavioral functioning in young people exhibiting ADHD symptoms. Almost half of the measures showed significant or marginally significant change over the program, with effect sizes of mostly small to medium magnitude. Measures of response inhibition showed the most consistently favorable findings, suggesting that physical activity may be especially helpful in addressing this core manifestation of ADHD. Moreover, key informants perceived some degree of improvement in overall functioning for about two thirds of the participants. No outcomes of meaningful effect size were in a maladaptive direction, suggesting that physical activity is unlikely to be harmful when employed as a management strategy for ADHD. In light of the overall pattern of findings, we believe that investing in controlled, larger-scale examinations of physical activity effects on the management of ADHD are warranted.

The assessment of overall motor proficiency (i.e., BOT-2; Bruininks & Bruininks, 2005) showed a large change from pre- to postprogram. This suggests that the physical activity intervention was designed in a way that afforded development of fitness parameters and motor control. The emphasis on sustained MVPA within a framework of games and activities that employ a variety of movement skills would be expected to yield such an outcome, and is recommended for future work with young children. Designing the intervention in this way offers the stimulus necessary to promote fitness and motor skill development while offering the variety and challenge necessary to sustain participant interest. Also, because ADHD is a disorder appearing to involve a variety of brain regions (see Nigg, 2006), such a dynamic physical activity intervention seems appropriate.

Of note is the absence of intervention effect on the continuation timing task. Because this task is simplified and highly timing dependent, it is strongly linked to cerebellar function (Ivry, Keele, & Diener, 1988). The findings suggest that if any benefits of a physical activity intervention to motor timing and cerebellar function exist, they would be less marked than those for broader motor functioning and/or would require an extended intervention time frame to detect. Pursuit of long-term intervention investigations will offer opportunity to understand this finding better. In light of the imprecise etiology of ADHD, it is reasonable to expect that physical activity benefits on particular symptoms or brain systems would be neither uniform nor accrued at equal rates.

A broad range of tasks were employed to measure cognitive function of study participants, with measures assessing response inhibition, set shifting, planning, and working memory. Such executive functions were selected for examination because deficits in these functions represent core impairments associated with ADHD (Barkley, 1997) and emerging work indicates that fitness levels and physical activity may be linked to their preservation and enhancement (Etman & Chang, 2009; Hillman, Buck, Themanson, Pontifex, & Castelli, 2009). The strength of findings was mixed across tasks. For example, working memory tasks did not show meaningful effect size values (i.e., ≥ 0.20) except when involving additional cognitive load such as the Numbers Reversed task (effect size = 0.43). The latter task involves planning, which in children has been shown to benefit from exercise training (Davis et al., 2007). However, the Mazes task, which purportedly measures planning, yielded a small effect size in our study. The more consistent effects were found for the tasks requiring response inhibition. The effect size values for these tasks ranged from 0.35 to 0.60, with the exception of the Simon Says task. Though we were able to reliably measure behavioral responses on this particular task, children could correct themselves before their behaviors were coded as inhibition failures. That is, children could return to their previous body position before fully adopting the body position to inhibit. Many children appeared to use the feedback available from viewing coacting peers, which essentially served as a second piece of
information (beyond the leader’s instruction) about the correct response. The children did not have sufficient time to make such corrections in the Red Light/Green Light task, which was fast paced and had maximally concise commands. Accordingly, we believe the latter task was a particularly sensitive ecological assessment of response inhibition. The overall findings lead us to conclude that physical activity may offer cognitive benefits to young children with ADHD symptoms, particularly with regard to executive tasks requiring inhibitory control.

Social and behavioral functioning was assessed by weekly teacher ratings and daily observations of social behaviors. The weekly teacher ratings showed meaningful effect size values (0.40 to 0.70), suggesting improved school-day social and behavioral functioning across the time frame of the before-school physical activity program. This finding is notable, suggesting that possible physical activity effects on social and behavioral symptoms can extend beyond the physical activity period and context. Such carryover is essential if physical activity is to be a viable option for ADHD intervention. Within the program setting itself, the observed behaviors of interrupting (effect size = 0.78) and unintentional aggression (effect size = 0.40) showed meaningful improvement, whereas the behaviors of not speaking nicely, intentional aggression, and not following adult directions failed to meaningfully change. Interrupting and unintentional aggressive acts are underpinned by impulsiveness, and therefore these behavioral findings align closely with the response inhibition findings. This noted, the reader is reminded that the reliability of our behavioral coding system is uncertain and therefore the daily behavior findings must be cautiously interpreted.

In line with the behavioral observations, fewer children were rated by adult informants as showing some degree of improvement over the program in name calling/teasing/aggression and following rules than in some other areas, although it is important to remember that only those children who had problems in these areas could be rated as improved. A comparable outcome was observed for responsibility. More encouraging findings were obtained for adult ratings of child self-esteem, social skills, and overall functioning. Defiance/noncompliance improvement rating values were between the values for these sets. Overall, we interpret the ratings to offer some support for the validity of our behavioral observation system and to suggest a generalized potential benefit of the before-school program to our participants. The finding for overall functioning is comparable with improvement percentages for children upon initial stimulant use (Spencer et al., 1996), suggesting that physical activity may hold promise as an intervention strategy for ADHD.

This work is clearly preliminary and must be considered in light of its limitations, including lack of a control condition, lack of a comparison group of typically developing peers, a relatively small sample, and the administration of weekly tasks (i.e., Simon Says, Red Light/Green Light) that in themselves could improve inhibitory control independent of the physical activity intervention. Important delimitations include the exclusive enrollment of young children and those exhibiting hyperactivity/impulsivity in the study. The present work cannot speak of possible physical activity benefits to older children or those with the inattentive-only form of ADHD. However, we believe the results stemming from this pilot before-school physical activity program are noteworthy because (a) all meaningful changes were in adaptive directions, (b) the lack of meaningful maladaptive outcomes suggests such an intervention program has low probability of doing harm to participants, (c) observable changes were detected despite a small sample size, and (d) none of the medication naïve participants initiated medication use over the span of the program, though families were free to pursue this option. Moreover, these encouraging program outcomes are observed within an intervention context yielding high participant adherence. Program attendance was strong, suggesting that reasonable adherence to a physical activity intervention protocol can be expected in a before-school setting over about a 2-month time frame. Future work to assess adherence rates over longer time spans is warranted, as chronic physical activity participation will be necessary to manage persistent ADHD symptoms.

Beyond an extended and controlled field investigation of physical activity intervention effects on ADHD symptoms that addresses the limitations noted above, future work would be strengthened by adopting a translational approach where basic research on mechanisms of ADHD and physical activity effects on behavior, as well as brain function and development, informs clinical research. Recent animal work by Hopkins, Sharma, Evans, and Bucci (2009), for example, showed exercise to yield adaptive effects on attention and social behavior in female mice possessing behavioral and neurobiological characteristics of ADHD compared with controls. The effects were not observed in male counterparts, suggesting that sex may be an important moderator of physical activity effects on ADHD. This finding should be carefully explored within human intervention trials. If physical activity is established as an effective intervention for ADHD, it will also be important to address possible complementary effects of physical activity and existing treatment strategies, dose–response relationships, costs of physical activity intervention, long-term adherence challenges, and developmental implications of early physical activity intervention among other issues. We believe that the present pilot findings justify the expenditure of time, effort, and resources necessary to pursue these interesting and clinically valuable research avenues.
Acknowledgments

The authors thank David Bucci and John Green for thoughtful feedback on an earlier draft of the manuscript, J. D. DeFreese for assistance with assessment scoring, and Howard Zelaznik for guidance on the motor timing methods and analysis. Finally, the authors thank the participating children, parents, teachers, and school administrators that contributed to the success of this project.

Authors’ Notes

Aaron J. Vaughn is now with the Division of Behavioral Medicine, Cincinnati Children’s Hospital, OH. Meghan Tomb is now with Columbia University Medical Center/New York Presbyterian Hospital. She completed her PhD and is now a post doc there. Hoza and Smith are equal contributors to this research.

Declaration of Conflicting Interests

The author(s) disclosed receipt of the following financial support to the research, authorship, and/or publication of this article: The Oseretsky test of motor proficiency, 728-734. doi:10.1249/mss.0b013e31802f04c7.


**Bios**

**Alan L. Smith**, PhD, is a professor of health and kinesiology at Purdue University. His research addresses social and motivational aspects of children’s sport involvement and physical activity behavior.

**Betsy Hoza**, PhD, is a professor of psychology at the University of Vermont. Her research addresses mechanisms of ADHD, as well as treatment outcomes, from a developmental psychopathology perspective.

**Kate Linnea**, BA, is a doctoral student in clinical psychology at the University of Vermont. Her research examines peer relationships in children with ADHD and the role of sport involvement and skill.

**Julia D. McQuade**, BA, is a doctoral student in clinical psychology at the University of Vermont. Her research addresses cognitive risk and protective factors of emotional and social adjustment for children with ADHD from a developmental psychopathology perspective.

**Meghan Tomb**, PhD, is a postdoctoral fellow at Columbia University Medical Center/New York Presbyterian Hospital. Her research interests include social and cognitive processes of children with ADHD and related disorders.

**Aaron J. Vaughn**, PhD, is an assistant professor at Cincinnati Children’s Hospital Medical Center. His research pertains to assessment and treatment of ADHD.

**Erin K. Shoulberg**, MA, is a doctoral student in developmental psychology at the University of Vermont. Her research examines how children’s and adolescents’ relationships influence their social and academic functioning.

**Holly Hook**, MS, is a physical education and health teacher in Indiana.