Working memory and cognitive flexibility-training for children with an autism spectrum disorder: a randomized controlled trial

Marieke de Vries,1,2 Pier J.M. Prins,1 Ben A. Schmand,1,3 and Hilde M. Geurts1,2,4

1University of Amsterdam, Amsterdam; 2Dutch Autism and ADHD Research Center, Amsterdam; 3Academic Medical Center Amsterdam, Amsterdam; 4Dr. Leo Kannerhuis, Amsterdam, the Netherlands

Background: People with autism spectrum disorders (ASDs) experience executive function (EF) deficits. There is an urgent need for effective interventions, but in spite of the increasing research focus on computerized cognitive training, this has not been studied in ASD. Hence, we investigated two EF training conditions in children with ASD.

Methods: In a randomized controlled trial, children with ASD (n = 121, 8–12 years, IQ > 80) were randomly assigned to an adaptive working memory (WM) training, an adaptive cognitive flexibility-training, or a non-adaptive control training (mock-training). Braingame Brian, a computerized EF-training with game-elements, was used. Outcome measures (pretraining, post-training, and 6-week-follow-up) were near-transfer to trained EFs, far-transfer to other EFs (sustained attention and inhibition), and parent’s ratings of daily life EFs, social behavior, attention deficit hyperactivity disorder (ADHD)-behavior, and quality of life. Results: Attrition-rate was 26%. Children in all conditions who completed the training improved in WM, cognitive flexibility, attention, and on parent’s ratings, but not in inhibition. There were no significant differential intervention effects, although children in the WM condition showed a trend toward improvement on near-transfer WM and ADHD-behavior, and children in the cognitive flexibility condition showed a trend toward improvement on near-transfer flexibility. Conclusion: Although children in the WM condition tended to improve more in WM and ADHD-behavior, the lack of differential improvement on most outcome measures, the absence of a clear effect of the adaptive training compared to the mock-training, and the high attrition rate suggest that the training in its present form is probably not suitable for children with ASD.

Keywords: Executive functioning, working memory, flexibility, cognitive training, autism.

Introduction

Executive functions (EFs) are cognitive functions needed for goal directed behavior that is adaptive to the environment. EFs are important in many aspects of life, such as school success, and well-developed EFs can predict lifelong achievement (Diamond, 2013). A growing body of evidence suggests EFs can be trained (e.g., Melby-Lervåg & Hulme, 2013; Zinke, Einert, Pfennig, & Kliegel, 2012), although generalization to other functions and to daily life (Holmes, Gathercole, & Dunning, 2009), and long term effects (Roughan & Hadwin, 2011) are not well established. EF training has been studied in typically developing (TD) people (e.g., Alloway, Bibile, & Lau, 2013) and clinical groups, such as children with Attention Deficit Hyperactivity Disorder (ADHD: e.g., Klingberg et al., 2005). One of the mainstream cognitive theories on autism spectrum disorders (ASD) postulates that EF deficits are the underlying problem (Russell, 1997). Moreover, EF deficits in ASD are related to adaptive behavior (Gilotty, Kenworthy, Sirian, Black, & Wagner, 2002), which in turn is associated with quality of life (Chiang & Wineman, 2014). However, only a few studies focused on EF interventions in ASD (Baltruschat et al., 2011; Fisher & Happé, 2005; Kenworthy et al., 2014). ASDs are characterized by difficulties in social interaction and communication, and restricted and repetitive behavior and interests (American Psychiatric Association, 2000). These difficulties interfere with daily life, and may partly be caused by EF deficits (Damasio & Maurer, 1978; Geurts, Vries, & van den Bergh, 2014; Russell, 1997). Hence, an EF intervention may be fruitful for ASD.

Some intervention studies in ASD included EF as an outcome measure. Although a Theory of Mind (ToM)-training did not influence EF (Fisher & Happé, 2005), positive reinforcement (Baltruschat et al., 2011) and a social competence intervention (Stichter et al., 2010) did show positive effects on EF, although studied without a control group. A Social Skills training also improved EF, but to a lesser extent than a direct EF intervention (Kenworthy et al., 2014). The effects of direct EF interventions in ASD have hardly been studied, and findings are mixed; both positive (Kenworthy et al., 2014) and null effects (Fisher & Happé, 2005) have been reported. Children with ASD do show development of EFs (Pellicano, 2010a), which implies some plasticity. In sum, there are tentative indications that targeting EFs in interventions for ASD is promising, and that the best approach is to target EFs directly.

Working memory (WM)-training has been studied quite extensively in both clinical and TD groups (e.g.,...
Melby-Lervåg & Hulme, 2013), in particular, the computerized WM-training Cogmed (Chacko et al., 2013; Shipstead, Hicks, & Engle, 2012). Findings are inconsistent. Both near-transfer (improvement of the trained function) and far-transfer (improvement of other functions) have been reported, even when compared to an active control group, who performed a nonadaptive version of the training task (e.g., Holmes et al., 2009; Klingberg et al., 2005; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009). However, recent reviews and meta-analyses are critical. Short term near-transfer to verbal and visual-spatial WM is reported, but long term and far-transfer (to attention, ADHD-behavior, and reasoning) is small and non-significant (Melby-Lervåg & Hulme, 2013). Hence, the effects are not (yet) evident, also due to lack of replication and caveats in the studies (Shipstead, Redick, & Engle, 2012). Clearly, more methodologically sound WM-training research is needed. Moreover, there is a striking absence of studies on WM-training in ASD.

Although flexibility deficits in ASD are not consistently reported in the research setting, inflexible behavior in daily life is evident (Geurts, Corbett, & Solomon, 2009; Kenworthy, Yerys, Anthony, & Wallace, 2008; Leung & Zakzanis, 2014). Therefore, a flexibility-training might be particularly fruitful. The few studies that focused on flexibility-training reported improved flexibility, interference control, and verbal WM in children with ADHD (Kray, Karbach, Haenig, & Freitag, 2011) and TD people (Karbach & Kray, 2009). Hence, flexibility seems trainable.

In spite of evident EF deficits in ASD and positive results of WM and flexibility-training in related disorders, studies on EF training in ASD are lacking. Therefore, we studied the effect of a WM and a flexibility-training compared to an active control condition (mock-training) in a large randomized controlled trial of children (8–12 years) with ASD. Repetition of monotonous tasks in EF training reduces motivation and increases attrition-rates, and adding game-elements enhances motivation, and may improve the training effect (Prins, Dovis, Ponsioen, Ten Brink, & Van der Oord, 2011). We, therefore, used an EF training with game-elements: Braingame Brian (Prins et al., 2013), that showed promising pilot-results in ADHD (Oord, Ponsioen, Geurts, Brink, & Prins, 2014). We expected near-transfer to WM and flexibility respectively, and far-transfer to other EF tasks (inhibition, sustained attention), and to parent’s ratings of daily life EFs and behavior (Figure S1).

Methods

Participants

Families were approached through mental health care institutions and internet advertisements. Of the initially 166 signed up families, 34 did not participate because of the required time investment (see Figure 1). The remaining 132 children were screened for (a) a prior independent clinical ASD diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV: American Psychiatric Association, 2000), diagnosed by a multidisciplinary team specialized in ASD, (b) age 8–12 years, (c) IQ above 80, and (d) absence of a seizure disorder. First, the diagnosis was verified if the score on the Social Responsiveness Scale parent report (SRS: Constantino et al., 2003; Roeyers, Thys, Druart, De Schryver, & Schittekatte, 2011) was above the specified Dutch clinical ASD cut-off of 57 (raw score), if Autism Diagnostic Interview Schedule-Revised (ADI-R: De Jonge & de Bildt, 2007; Lord, Rutter, & Couteur, 1994) cut-offs were reached on two out of the three domains, and if ASD was recognized before the age of 36 months (Gray, Tonge, & Sweeney, 2008). Second, IQ was

Figure 1 Flow chart of inclusion for RCT (conforming to CONSORT)
estimated with two subtests of the Dutch version of the Wechsler Intelligence Scale for Children (WISC-III: Kort et al., 2002; Sattler, 2001). Eleven children were excluded after screening (2 ADI-R, 6 IQ, 3 personal circumstances). One hundred and twenty-one children were randomly (double blind) assigned to the three intervention-conditions. Six children dropped out after pretraining. Finally, 115 children performed the training (40 WM, 37 flexibility, 38 mock-training). Twenty-five children did not complete 25 sessions (9 WM, 10 flexibility, and 6 mock-training; similar in the three conditions $[\chi^2(2) 1.3, p = .52]$). Forty children used psychotropic medication (17 abstained during test-appointments, 23 continued). Of the 90 children who completed the training, 24 children used medication (10 abstained, 14 continued). Years since diagnosis was 2.1 ($SD = 1.7$, $N = 84$, as six parents did not indicate the date of diagnosis).

**Tasks and questionnaires**

Measures include screening and four steps of outcome measures: (a) near-transfer to WM and flexibility tasks resembling the training tasks, (b) near-transfer to WM and flexibility tasks differing from the training tasks, (c) far-transfer to inhibition and sustained attention tasks, and (d) far-transfer to daily life (see Figure 2).

**Inclusion/screening.** IQ (Kort et al., 2002). The WISC-III subtests Vocabulary and Block Design, used to estimate IQ, correlate highly with Full Scale IQ, and have good reliability (Legerstee, van der Reijden-Lakeman, Lechner-van der Noort, & Ferdinand, 2004; Sattler, 2001).

**Diagnosis:** The SRS (Constantino et al., 2003; Dutch version: Roeyers et al., 2011) is a reliable and valid questionnaire measuring autism characteristics (Bolte, Poustka, & Constantino, 2008). The ADI-R (Lord et al., 1994; Rutter, Lord, & LeCouteur, 2003) is an extensive semi-structured interview including three domains of autism symptoms (social, communication, and repetitive behavior). Reaching specified cut-offs on these domains, and clear indications of abnormalities in development before 36 months, suggests a DSM-IV (American Psychiatric Association, 2000) or ICD-10 (World Health Organisation, 1992) diagnosis of autism (Gray et al., 2008; Rutter et al., 2003).

**WM and flexibility tasks resembling the training tasks.** WM: The Corsi block tapping task (Corsi-BTT: Corsi, 1972) measures visual-spatial WM (WM-training near-transfer, flexibility-training far-transfer). In the currently used computerized version, nine blocks on a computer screen light up (500 ms each) in a certain sequence, which has to be repeated with mouse-clicks. The initial sequence length is three blocks, and after every four trials, the sequence length increases with one, up to a maximum of eight blocks. The task ends after four incorrectly repeated sequences of the same length. The outcome measure is the longest correctly repeated sequence length.

**Flexibility:** The Gender-emotion switch-task (de Vries & Geurts, 2012) is an adaptation of the classical switch-task (Rogers & Monsell, 1995; flexibility-training near-transfer, WM-training far-transfer). Pictures of easy to recognize happy or angry male and female faces (Lundqvist, Flykt, & Öhman, 1998) are displayed on the computer screen, and have to be sorted, alternatingly, on gender or emotion by pressing designated keys on the keyboard (de Vries & Geurts, 2012). The sorting rule changes randomly after two, three, or four trials (30% switch-trials). Children with ASD are quite capable to recognize these basic emotions (Geurts, Luman, & van Meel, 2008), hence emotion recognition most likely does not influence performance. Outcome measures are error rates (ER) and reaction time (RT) switch costs (switch-trial minus repeat-trial scores).

**WM and flexibility tasks different from the training tasks.** WM: In the N-back task (Casey et al., 1995; Smith & Jonides, 1999) participants have to remember pictures that are presented on a computer screen and indicate if the current picture matches the picture N trials before. Three levels with increasing WM-load were included: 0-back (baseline, indicate if the picture is a dog or not), 1-back, and 2-back (de Vries & Geurts, 2014). The outcome measures were the ER and RT on the 2-back, and 1-back task (corrected by subtraction for ER and RT on the 0-back task).

**Flexibility:** In the Number-gnome switch-task, an adaptation of the number-switch-task (Cepeda, Cepeda, & Kramer, 2000) pictures of one, two or three gnomes (Geirnaert, 2010) with a speech bubble with the number ‘1’, ‘2’, or ‘3’ is displayed. Participants have to report the number of gnomes, or the number in the speech bubble, by pressing the 1, 2, or 3 on the keyboard. The task design is identical to the gender-emotion switch-task. Outcome measures were ER, and RT switch costs.

**Far-transfer to inhibition and sustained attention tasks.** Inhibition: In the current adaptation of the classical Stop-task (Logan, 1994) a picture of a yellow dog facing

![Figure 2](image-url)
left or right appears on the computer screen. Participants have to press a corresponding left or right key on the keyboard as fast as possible. On 30% of the trials, the dog turns red (stop-signal), and the response has to be inhibited (de Vries & Geurts, 2014). The outcome measure was the stop-signal reaction time (SSRT; Logan & Cowan, 1984).

**Sustained attention:** In the Sustained attention response task (SART; Robertson, Manly, Andrade, Baddeley, & Viend, 1997) numbers 1–9 appear on the computer screen (250 ms) followed by a cross (1000 ms or until response). After each number, except ‘3’, when the cross appears, the spacebar has to be pressed. The outcome measures were ER and RT.

**Far-transfer to daily life. EF:** The Behavior Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000; Dutch version: Smidts & Huizinga, 2009; 75 items, 3-point Likert-scale) has high to very high internal consistency and test-retest stability (Huizinga & Smidts, 2010). The subdomains Shift and WM, and the total scale were the outcome measures.

**Social behavior:** The Children’s Social Behavior Questionnaire (CSBQ, Dutch version: Hartman, Luteijn, Moorlag, De Bildt, & Minderaa, 2007; 49 items, 3-point Likert-scale) has good internal consistency and test-retest stability (Huizinga & Smidts, 2006). The total scale was the outcome measure.

**ADHD-behavior:** The Dutch parent version of the Disruptive Behavior Disorders Rating Scale (Oosterlaan, Scheres, Antrop, Roeyers, & Sergeant, 2000; DBDRS: Pelham, Gnagy, Greenslade, & Milich, 1992; 42 items, 4-point Likert-scale) has good psychometric properties (Oosterlaan et al., 2000). The subscales inattention, and hyperactivity/impulsivity combined (DBDRS-ADHD) were used as the outcome measure.

**Quality of life:** The Pediatric Quality of Life Inventory (PedsQL: Bastiaansen, Koot, Borgers, Varni, & Verhulst, 2004; Varni, Seid, & Kurtin, 2001; 23 items, 5-point Likert-scale) has satisfactory reliability and validity (Varni et al., 2001). The total scale was used as the outcome measure.

**Intervention**
Each child performed one of three intervention-conditions of ‘Braingame Brian’ (Prins et al., 2013): a WM-training, a flexibility-training, or a mock-training (Appendix S1 and Figure S3). In each intervention-condition, all EF tasks were performed but whether the level was adaptive differed per intervention-condition. The WM-training included five adaptive WM-training tasks with increasing difficulty, and the other tasks remained at a low level. The flexibility-training included an adaptive flexibility task, and the other tasks remained at a low, nonadaptive level. In the flexibility-training and mock-training only the first most basic task of the five WM-training tasks was performed.

**Procedure**
First, written informed consent, SRS, and screening questionnaires were obtained. Next, training and test-appointments (screening, pretraining, post-training [after 6 weeks], and follow-up [FU, after 6 more weeks]) were scheduled. During screening, the ADI-R was administered to the parents, and the WISC-III subtests to the children. Parents, children, and assessors were kept blind to intervention-condition until after follow-up. At pretraining, post-training, and FU (at the University of Amsterdam) all outcome tasks were administered to the children, and parents filled out the questionnaires. Pretraining a laptop with Braingame Brian installed was handed over and children and parents were instructed. Children and parents were told that there were three intervention-conditions, but not how these conditions were designed. All received similar basic training instructions and were told that tasks could change over time, but that this was not always noticeable. During the six training-weeks, parents were contacted weekly about the child’s progress. In these weekly telephone calls, parents were asked about the current progress in the training, motivation of the child, and possible reward systems used. After the training, parents were asked about the amount of parental support their child received during the training. The number of completed sessions was retrieved from the automatically saved log-files of the training. The test-appointments lasted 50–150 min (as additional, for the current study irrelevant, tasks and questionnaires were administered). Children received a small gift for participating. Parents received a report with the ADI-R results. Travel expenses were largely covered.

**Data analysis**
The main (per-protocol) analyses included children who finished 25 sessions. If applicable, inverse efficiency scores (IES = mean RT/(1-ER; Bruyer & Brysbaert, 2011) were used as an outcome measure, to take possible speed-accuracy trade-off into account. IES are reliable when RT and ER are correlated, and ER <10% (Bruyer & Brysbaert, 2011), which was satisfied for the SART.

First, t-tests and chi-square tests were conducted on screening and pretraining measures to test if children who finished the training differed from the drop-outs, and a MANOVA was conducted to test if children in the three intervention-conditions differed. Second, two repeated measures ANOVAs were conducted to test whether children in the WM, and flexibility intervention-condition improved on the within-training tasks (WM: mean final reached sequence length on the five training tasks, flexibility: final level reached). Third, training effects were examined with four (to increase power, missing data-points 1.47%) repeated measures MANOVA’s with intervention (WM, flexibility, and mock-training) as between subject variable, and time (pretraining, post-training, and FU) as within subject variable. Outcome measures were (a) Corsi-BTT and n-back task (near-transfer WM), (b) gender-emotion switch-task and number-gnome switch-task (near-transfer flexibility), (c) SART, and stop task, and (d) questionnaires (far-transfer). Additional intention-to-treat analyses (last observation carried forward) were conducted: the four MANOVA’s on intervention effects were repeated including all participants, with number of completed sessions as covariate. Finally, we explored in children who performed all sessions if within-training-improvement (mean of five WM-training tasks, and reached level of flexibility-training task) was correlated with near-transfer training-outcome (Corsi-BTT and gender-emotion task improvement between pretraining and post-training).

Bonferroni corrections for multiple testing were applied (screening and pretraining: 0.05/25 = 0.002, training effects: 0.05/4 = 0.0125). Analyses were conducted both with and without children who used medication during test-appointments. Due to missing data, numbers differed slightly for different measures (see Table 1).

**Results**

**Screening and pretraining**
The drop-outs of the intervention used medication slightly more often [χ²(1) = 6.5, p = .04], and showed
Table 1 Mean (standard deviations) and MANOVAs for near-transfer and far-transfer measures

<table>
<thead>
<tr>
<th>Group Comparison</th>
<th>Time</th>
<th>Time*Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre-training</td>
<td>Post-training</td>
</tr>
</tbody>
</table>

**Group Comparisona**

<table>
<thead>
<tr>
<th>Measure</th>
<th>WM-training</th>
<th>n = 31</th>
<th>Flexibility-training</th>
<th>n = 26</th>
<th>Mock-training</th>
<th>n = 29</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corsi-BTT N-backbc</td>
<td>5.2 (0.7)</td>
<td>6.3 (1.0)</td>
<td>6.2 (0.8)</td>
<td>5.5 (0.9)</td>
<td>6.0 (1.0)</td>
<td>5.8 (0.8)</td>
</tr>
<tr>
<td>ER1-backb</td>
<td>9.0 (12.1)</td>
<td>1.8 (5.3)</td>
<td>0.8 (6.2)</td>
<td>4.3 (7.9)</td>
<td>1.9 (6.8)</td>
<td>1.0 (4.8)</td>
</tr>
<tr>
<td>ER2-back</td>
<td>18.1 (12.6)</td>
<td>12.2 (8.5)</td>
<td>11.5 (8.1)</td>
<td>16.3 (7.7)</td>
<td>12.0 (8.8)</td>
<td>7.8 (6.0)</td>
</tr>
<tr>
<td>RT1-back</td>
<td>67.6 (41.4)</td>
<td>43.8 (54.6)</td>
<td>47.5 (44.4)</td>
<td>65.2 (39.5)</td>
<td>51.1 (41.9)</td>
<td>38.0 (32.9)</td>
</tr>
<tr>
<td>RT2-back</td>
<td>77.3 (60.5)</td>
<td>71.3 (64.9)</td>
<td>72.0 (39.8)</td>
<td>79.0 (61.6)</td>
<td>89.1 (51.4)</td>
<td>59.4 (62.2)</td>
</tr>
</tbody>
</table>

Switch costs on the gender-emotion switch-task

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time</th>
<th>Time*Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>109.3 (73.0)</td>
<td>96.7 (58.2)</td>
</tr>
<tr>
<td>ER</td>
<td>5.2 (8.0)</td>
<td>8.8 (6.9)</td>
</tr>
</tbody>
</table>

Switch costs on the number-gnome switch-task

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time</th>
<th>Time*Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>64.0 (79.3)</td>
<td>56.7 (58.0)</td>
</tr>
<tr>
<td>ER</td>
<td>6.8 (6.4)</td>
<td>7.2 (7.1)</td>
</tr>
</tbody>
</table>

**Switch costs on the gender-emotion switch-task**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time</th>
<th>Time*Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>233.3 (61.3)</td>
<td>216.8 (61.6)</td>
</tr>
<tr>
<td>ER</td>
<td>269.5 (79.6)</td>
<td>242.7 (66.9)</td>
</tr>
</tbody>
</table>

**Switch costs on the number-gnome switch-task**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time</th>
<th>Time*Intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRIEF-shift</td>
<td>18.4 (3.0)</td>
<td>17.3 (3.1)</td>
</tr>
<tr>
<td>BRIEF-WM</td>
<td>24.3 (5.3)</td>
<td>23.3 (5.0)</td>
</tr>
<tr>
<td>BRIEF-total</td>
<td>164.6 (22.2)</td>
<td>157.6 (22.1)</td>
</tr>
<tr>
<td>CSBQ</td>
<td>47.8 (14.1)</td>
<td>44.2 (15.3)</td>
</tr>
<tr>
<td>DBDRS-ADHD</td>
<td>28.6 (10.8)</td>
<td>27.0 (10.8)</td>
</tr>
<tr>
<td>PedsQI</td>
<td>34.2 (12.0)</td>
<td>29.8 (12.8)</td>
</tr>
</tbody>
</table>

BRIEF, Behavior Rating Inventory of Executive Function; Corsi-BTT, Corsi block tapping task; CSBQ, Children Social Behavior Questionnaire; DBDRS-ADHD, inattention and hyperactivity/impulsivity subscales Disruptive Behavior Disorders Rating Scale; ER, % error; FU, 6-week-follow-up; PedsQI, Pediatric Quality of Life Inventory; RT, reaction time ms; SART-IES, sustained attention response task, inverse efficiency score; WM, working memory.

a df were corrected (Greenhouse-Geisser) if assumption of sphericity was violated (Mauchly’s test).

b Corrected for 0-back.

c Analyses without outliers did not change the main findings.

*p < .05; **p < .1; ***p < .001.

© 2014 Association for Child and Adolescent Mental Health.
slightly less flexible behavior according to parents ($T_{1[118]} = 2.3, p = .02, r = .21, \text{NS after Bonferroni correction, Table S1}$). The RTs on the 1-back task were slightly higher in the mock-training condition ($F_{[2, 87]} = 3.4, p = .04, \text{NS after Bonferroni correction, see Table S2}$). There were no further differences between the three intervention-conditions on screening or pretraining variables.

**Within training**

Children in the WM and flexibility intervention-conditions improved significantly in sequence lengths in the WM-training tasks, and level in the flexibility-training task ($p's < .001$, see Figure 3, and Table S3). Hence, the adaptive function in the training was effective.

**Training effects**

Near-transfer. WM: In all three intervention-conditions there was a significant effect of time on the Corsi-BTT ($p < .001$, see Table 1). Post hoc analysis revealed a significant increase between pretraining, and post-training scores ($p < .001$), but not between post-training and FU scores ($p > .05$). There was a time x intervention interaction trend ($p = .04$, NS after Bonferroni correction). Children in the WM-training condition tended to improve more than children in the flexibility and mock-training conditions.

In all intervention-conditions there was a significant effect of time on ER and RT on the 1-back task, and ER on the 2-back task (see Table 1). Post hoc analysis revealed only a significant decrease in IES between pretraining and post-training ($p < .01$). There was no significant time x intervention interaction. Hence, sustained attention improved in all intervention-conditions, but did not improve more in the adaptive intervention-conditions than in the mock-training condition.

Far-transfer. Sustained attention and inhibition: There was a significant effect for time on the SART-IES. Post hoc analysis revealed only a significant decrease in IES between pretraining and post-training ($p < .01$). There was no significant time x intervention interaction. Hence, sustained attention improved in all intervention-conditions, but did not improve more in the adaptive intervention-conditions than in the mock-training condition.

There was no significant time effect, nor time x intervention interaction in SSRT on the stop task (see Table 1). Inhibition did not improve.

Far-transfer to daily life. There was a significant effect for time on all questionnaires (see Table 1). Post hoc analysis revealed a significant difference between pretraining and post-training ($p's < .01$, except for BRIEF-WM: $p = .03$), but not between post-training and FU ($p > .01$). There was a time x intervention interaction trend on DBDRS-ADHD ($p = .05$, NS after Bonferroni correction).

**Figure 3** Mean and standard errors of (A) the sequence length on the five training tasks in the WM-training and (B) level of cognitive flexibility-training.
Hence, EFs, social behavior, and quality of life improved equally in all three intervention-conditions, but ADHD-behavior improved slightly more in the WM intervention-condition.

Re-running the MANOVAs on training effect without children who used medication during testing did not change the main findings, except that the time x intervention interaction trend on DBDRS-ADHD was not significant anymore, probably due to a lack of power.

**Intention-to-treat analysis**

Intention-to-treat analyses [WM \( N = 41 \), flexibility \( N = 40 \), mock-training \( N = 40 \), \( M(\text{SD}) \) performed sessions: WM: 21 (7), flexibility: 20 (9), mock: 22 (7)] revealed that overall children did not improve in Corsi-BTT performance, but there was a marginal time x number of sessions interaction (\( F_{[3, 228]} = 4.3, p = .02, \eta^2_p = .04, \text{NS after Bonferroni correction} \)), and a significant time x intervention interaction (\( F_{[4, 228]} = 4.4, p = .002, \eta^2_p = .07 \)). Corsi-BTT performance was positively related to the number of finished sessions, and improved the most in the WM intervention-condition between pre- and post-training (\( F_{[2, 119]} = 5.7, p = .004, \eta^2_p = .09 \)), but not between post-training and FU (\( p > .01 \)). Children in the WM intervention-condition improved more in Corsi-BTT performance than children in the flexibility and mock-training, and remained at a higher level at FU.

There was a time x intervention interaction trend on the switch costs in ER in the gender-emotion switch-task (\( F_{[6, 234]} = 2.9, p = .02, \eta^2_p = .05, \text{NS after Bonferroni correction} \)). Switch costs in ER improved more in the flexibility intervention-condition than in the WM and mock-training condition between pre- and post-training (\( F_{[2, 117]} = 4.6, p = .01, \eta^2_p = .07 \)), but not between post-training and FU (\( p > .05 \)).

On the BRIEF Shift, there was a marginal time x number of sessions interaction (\( F_{[2, 230]} = 3.2, p = .04, \eta^2_p = .03, \text{NS after Bonferroni correction} \)). Children who performed more sessions improved slightly more on the BRIEF-shift scale.

On the DBDRS-ADHD, the time effect was not significant (\( p > .05 \)), but there was a time x intervention interaction trend (\( F_{[3.7, 212.2]} = 2.6, p = .04, \eta^2_p = .04, \text{NS after Bonferroni correction} \)). Post hoc analyses revealed that between pretraining and posttraining, there was neither time nor a time x intervention interaction (\( p > .05 \)), and between posttraining and FU there was no significant time effect (\( F_{[1, 116]} = 0.1, p = .80, \eta^2_p = .00 \), but a significant time x intervention interaction (\( F_{[2, 117]} = 5.3, p = .006, \eta^2_p = .08 \)). Overall, children in the WM intervention-condition decreased more in their parent rated ADHD-behavior than children in the other intervention-conditions, and after the post-training this improvement continued only in the WM intervention-condition.

In line with per-protocol analysis, there were no time, nor interaction effects on the number-gnome switch-task and the stop task. In contrast to per-protocol analyses, there were no significant time nor interaction effects on the n-back task, SART, BRIEF-WM, BRIEF-total, CSBQ, and PedsQL.

**Relation between within training improvement and training effect**

The improvement in the training tasks was not significantly correlated to the primary outcome measures (WM \( r = -.09 \) to .16, flexibility \( r = -.33 \) to .06). Although children did improve within the training, this improvement was not significantly related with performance on similar tasks (Corsi-BTT, and gender-emotion switch-task; see Table S4).

**Discussion**

Compared to a mock-training, a WM-training and flexibility-training did not induce near-transfer nor far-transfer in children with ASD in this randomized controlled trial. All children improved in WM, flexibility, attention, and parent rated EF, social behavior, ADHD-behavior, and quality of life, but the adaptive-intervention-conditions did not result in a larger improvement than the mock-training. Overall, differently measured flexibility and inhibition did not improve. The WM-training did tend to induce more improvement in WM (near-transfer) and ADHD-behavior (far-transfer) than the other intervention-conditions. The medium effect sizes and the significant findings in intention-to-treat analyses suggest a trend, whereas the non-significant treatment x time interaction in the per-protocol analyses might have resulted from a lack of power. Moreover, intention-to-treat analyses suggested a trend toward near-transfer of the flexibility-training, and a trend that children who performed more sessions, improved more in daily life flexibility.

The trend of near-transfer of the WM-training is consistent with studies in children with ADHD (e.g., Chacko et al., 2013). However, the lack of generalization to the broader WM construct and other EFs makes the trend in parent reported improvement in ADHD-behavior difficult to interpret. Although WM is related to ADHD symptoms (Oord et al., 2012; Rapport et al., 2009), training induced improvements in WM have not unequivocally shown improvement in ADHD symptoms (Melby-Lervåg & Hulme, 2013; Rapport, Orban, Kofler, & Friedman, 2013). Moreover, WM improvement was found in a recent study on Braingame Brian in children with ADHD (S. Dovis, S. Oord, R.W. Wiers & P.J.M. Prins, unpublished data), but far-transfer to ADHD-behavior was not found. Hence, the currently found trend of improved WM and ADHD-behavior in children with ASD should be further examined in future research.

The trend of near-transfer of the flexibility-training in the intention-to-treat analysis is consistent with previous studies reporting positive effects of this type.
of training (Karbach & Kray, 2009; Kray et al., 2011). However, this small effect was not revealed in the main analyses. As children with poor flexibility seem to benefit more from flexibility-training (Karbach & Kray, 2009), and as not all children with ASD show flexibility deficits (Pellicano, 2010b; de Vries & Geurts, 2012), there might have been too little room for improvement. Moreover, improvement may relapse when discontinuing a task-switching training (Karbach & Kray, 2009). Hence, including FU might reveal a less strong effect as compared to a pretraining/post-training design. In sum, although the current findings suggest that our computerized flexibility-training is ineffective for ASD, there is limited evidence of a near-transfer effect.

In line with previous research (Thorell et al., 2009), we found no improvement in inhibition. Recent research suggests that a combined EF intervention might be more successful to improve inhibition (S. Dovis, S. Oord, R.W. Wiers & P.J.M. Prins, unpublished data). The equal improvement in all intervention-conditions indicated that both the WM-training and the flexibility-training did not induce far-transfer to attention, daily EFs, social behavior, and quality of life. The overall improvement might be due to test-retest-, practice-, and expectancy effects, or regression to the mean (Melby-Lervåg & Hulme, 2013). However, nonspecific training aspects, such as the needed sustained attention to perform low level EF tasks over 25 sessions (Rueda, Checa, & Còmbita, 2012); and interacting with characters in the virtual world (Best, 2014) might have induced improvement in all intervention-conditions. Moreover, parents likely give their child extra attention during the training by jointly visiting the test-sessions, and motivating and supporting their child to complete the training. This treatment related parental involvement may have ameliorated behavioral problems (Forehand, Jones, & Parent, 2013).

The finding that the WM-training showed more effect than the flexibility and mock-training may be due to the fact that this training included five different WM-training tasks, while the other conditions only included one WM-training task, and the flexibility-training only included one adaptive flexibility-training task. Performing various tasks may be more demanding, but also more motivating, and importantly, more effective (Karbach & Kray, 2009). The WM-training includes remembering, manipulating, and updating of information. Targeting a broader range of functions, even though all part of WM, may contribute to generalization of the training effect (Buitenweg, Murre, & Ridderinkhof, 2012; Karbach & Kray, 2009). In contrast, the solely increasing speed in the flexibility-training may be more difficult to generalize to daily life. Moreover, the gap between cognitive flexibility in the research setting and in daily life may be too large. A cognitive behavioral approach may be more appropriate to improve cognitive flexibility (Kenworthy et al., 2014), while a computerized WM-training may be promising to improve WM. Moreover, it may be fruitful to train more EFs simultaneously, while training a single EF is perhaps too limited to result in far-transfer.

Attrition was relatively high (26%), compared to previous studies (5%-10%; Kenworthy et al., 2014; Melby-Lervåg & Hulme, 2013; Stichter et al., 2010). Children who dropped out used medication more often, indicating that these children possibly experienced more severe problems. There were no differences on other measures, but uncontrolled factors might be relevant to attrition. We unfortunately did not quantify parental support, or children’s attitude toward the training-game, such as motivation to succeed on the training tasks, or to complete the training. These factors may have contributed to success or failure to complete the training, or benefit from the training. The weekly phone calls did indicate more enthusiasm about the game world of the training than about the training tasks, and integrating the training tasks more thoroughly in the game world might make the training more appealing and rewarding to the children.

One might argue that the currently used mock-training also trained EFs to a certain extent. However, the EF demand in this mock-training was very low, and undoubtedly much smaller than in the other training conditions. Completely ruling out EFs may be difficult, if not impossible, in virtually any task, and including a completely different task is hard to compare to a computer training. Thus, we consider our mock-training a proper control condition, but a challenge for future research may be to design an active, though non-EF applying control condition.

The current study has some limitations. Firstly, some children used psychotropic medication. However, medication-use was similar in all conditions, and analyses without these children gave similar results. Secondly, we did not use teacher ratings. However, teacher ratings would probably not have shown larger effects (e.g., Klingberg et al., 2005). Thirdly, although on a group level EF deficits are present in children with ASD, not every individual within this group experiences EF deficits (de Vries & Geurts, 2014). The training may be effective only for children with ASD who show EF deficits. It will be an important next step to identify specific subgroups that may benefit from this training.

In sum, the current study indicates that WM-training shows only marginal effects on WM and ADHD-behavior, and effects of the flexibility-training are even less clear, though we found limited evidence for near-transfer. The training in its present form is not a feasible treatment for all children with ASD. However, given the major individual differences in ASD (Pellicano, 2010b; de Vries & Geurts, 2012,
2014), pursuing whether the training is useful for specific subgroups is a worthwhile enterprise.

Supporting information
Additional Supporting Information may be found in the online version of this article:
Appendix S1. Description of Braingame Brian.
Table S1. Mean (M) and SD of screening and pretraining measures for children who did, and did not finish the training.
Table S2. Mean (M) and SD of screening and pretraining measures for children in the three training groups.
Table S3. Within training: Mean (M), and SD of the number of sequences at the end of each training task in the working memory training, and reached level in the cognitive flexibility-training.
Table S4. Correlation between within training improvement, and training effect.
Appendix S2. Expected and actual pattern of findings and visual representation of WM tasks.
Figure S1. Expected pattern of findings.
Figure S2. Pattern of findings.
Figure S3. Visual representation of the WM-training and cognitive flexibility tasks.

Key points
- Children with autism spectrum disorders (ASD) show executive functioning (EF) deficits.
- EF training seems promising in other populations.
- A working memory (WM) and cognitive flexibility-training may be promising for children with ASD.
- WM-training shows marginal effects on WM and ADHD-behavior in children with ASD, and marginal effects of flexibility-training on flexibility are indicated.
- The training in its current form is no feasible treatment for children with ASD.

Note
1. RCI’s (Jacobson & Truax, 1991) based on test-retest reliability (BRIEF: Smidts & Huizinga, 2009; CSBQ: Hartman et al., 2007; DBD-ADHD: Oosterlaan et al., 2008), and Crohnbachs z (PedsQL: Varni, Burwinkle, Seid, & Skarr, 2003) indicated that between pretraining and post-training relatively few children improved significantly (one sided 95% confidence interval; BRIEF 12.4%, CSBQ 14.6%, DBD-ADHD 7.9%, PedsQL 28.1%).

References
Bryuer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? Psychologica Belgica, 51, 5–15.


disorder rating scale for children from 6 to 16 years old]. Amsterdam: Harcourt Assessment BV.

Accepted for publication: 11 August 2014
Published online: 26 September 2014

© 2014 Association for Child and Adolescent Mental Health.