The Stroop revisited: a meta-analysis of interference control in AD/HD

Rosa van Mourik, Jaap Oosterlaan, and Joseph A. Sergeant

Department of Clinical Neuropsychology, Vrije Universiteit, Amsterdam, The Netherlands

Background: An inhibition deficit, including poor interference control, has been implicated as one of the core deficits in AD/HD. Interference control is clinically measured by the Stroop Colour-Word Task. The aim of this meta-analysis was to investigate the strength of an interference deficit in AD/HD as measured by the Stroop Colour-Word Task and to assess the role of moderating variables that could explain the results. These moderating variables included: methods of calculating the interference score, comorbid reading and psychiatric disorders, AD/HD-subtypes, gender, age, intellectual functioning, medication, and sample size. **Methods:** Seventeen independent studies were located including 1395 children, adolescents, and young adults, in the age range of 6–27 years. A meta-analysis was conducted to assess the effect sizes for the scores on the word and the colour card as well as the interference score. **Results:** Children with AD/HD performed more poorly on all three dependent variables. The effect sizes for word reading ($d = .49$) and colour naming ($d = .58$) were larger and more homogeneous than the effect size for the interference score ($d = .35$). The method used to calculate the interference score strongly influenced the findings for this measure. When interference control was calculated as the difference between the score on the colour card minus the score on the colour-word card, no differences were found between AD/HD groups and normal control groups. **Discussion:** The Stroop Colour-Word Task, in standard form, does not provide strong evidence for a deficit in interference control in AD/HD. However, the Stroop Colour-Word Task may not be a valid measure of interference control in AD/HD and alternative methodologies may be needed to test this aspect of the inhibitory deficit model in AD/HD. **Keywords:** AD/HD, Stroop Colour-Word Task, interference control, rapid naming, meta-analysis. **Abbreviations:** AD/HD: Attention Deficit/Hyperactivity Disorder; C: Colour; CD: Conduct Disorder; CW: Colour-Word; ODD: Oppositional Defiant Disorder; W: Word.

Numerous authors have highlighted the role of executive dysfunction in Attention Deficit/Hyperactivity Disorder (AD/HD) (Pennington & Ozonoff, 1996; Sergeant, Geurts, & Oosterlaan, 2002). A key process in executive functioning is response inhibition (Barkley, 1997). Barkley (1997) distinguished three interrelated processes believed to constitute response inhibition: (1) inhibiting a prepotent response, (2) stopping an ongoing response, and (3) interference control. The Stroop Colour-Word Task (Stroop, 1935, see for review MacLeod, 1991) is widely used as a measure of interference control in studies with AD/HD groups and is recommended as part of a psychological test battery in clinical settings (Doyle, Biederman, Seidman, Weber, & Faraone, 2000). Given both the clinical and research interest in the Stroop Colour-Word Task with respect to AD/HD, we report here a quantitative meta-analysis of studies that compare children with AD/HD and normal controls on the Stroop Colour-Word Task, as opposed to a head-count (Sergeant et al., 2002).

The standard Stroop Colour-Word Task (Golden, 1978) consists of three conditions, represented by three different cards. There are different versions, but in the AD/HD literature, the 'Golden' version is the most widely used. On the first card, the 'word' card, the speed of word reading is measured: the subject is required to read rows containing four different colour words (red, green, yellow and blue) printed in black ink and presented in random order. On the second card, the colour card, the subject has to name the colours of either rows of four Xs or blocks that are printed in the colours red, green, yellow, and blue. This condition measures colour-naming speed. On the third card, the colour-word card, the subject is required to name colours of the colour-content mismatching colour-words; for example, the colour-word red may be presented in blue ink. On the third card, the distracter is the colour meaning of the word. Interference occurs when the to-be-named colour differs from the to-be-ignored word (incongruent). This causes response conflict (Posner & DiGirolamo, 1998). On all three cards, the subject is timed for 45 seconds and the number of correct responses is counted. All stimuli are presented in a random order and the subject is required to name them as quickly as possible. A lower score on the colour-word card, in the presence of normal scores on the word and colour card, reflects the interference effect (Golden, 1978).

Generally, there are two different theoretical models to explain the interference effect in the Stroop Colour-Word Task: sequential models and parallel models. In sequential explanations, processing in one stage must be completed (or almost completed) before the next stage can begin. Interference is supposed to occur only at the response stage. Sequential theories have not been very

successful in explaining all the effects found with different task manipulations (MacLeod, 1991). Parallel theories emphasise the speed of processing and the automaticity of word reading and colour naming. Cohen, Dunbar, and McClelland (1990) have developed a parallel model that states that the two features of the stimuli on the colour-word card: word (meaning) and colour, are processed simultaneously. The relative automaticity of the two dimensions (word processing and colour processing) determines the direction and the degree of interdimensional interference (MacLeod & MacDonald, 2000). In this account, the two processes run in parallel through activation moving along pathways of different strength in the system. The degree of automaticity is a function of the strength of each pathway. The difference between this parallel model and previous parallel models is that interference can also occur during the processing of the stimuli and not simply at the end of a 'horse race'.

Neuro-imaging studies show that the brain areas that are active, while subjects perform Stroop-like tasks, include the anterior cingulate cortex (Adleman et al., 2002; Bush, Luu, & Posner, 2000; Carter, Mintun, & Cohen, 1995b; Pardo, Pardo, Janer, & Raichle, 1990; Peterson et al., 1999, 2002), a region of the frontal cortex associated with the frontal executive networks (Posner & DiGirolamo, 1998). Other regions that consistently show differential increases in activation for the incongruent condition compared to a neutral control condition are: the frontal polar cortex (Bench et al., 1993; Carter et al., 1995b), the lateral prefrontal cortex (Peterson et al., 1999; Zysset, Müller, Lohmann, & von Cramon, 2001), the inferior frontal regions (Adleman et al., 2002; Chung Leung, Skudlarski, Gatenby, Peterson, & Gore, 2000; Peterson et al., 1999), and the inferior parietal lobule (Adleman et al., 2002; Carter et al., 1995b; Peterson et al., 1999). Note that neuro-imaging studies differ with respect to the neutral control condition (for example, coloured Xs, coloured neutral words or congruent colour-words) and that the tasks used were paced rather than self-paced, as conducted in the clinical Stroop card procedure. The neural networks, which are activated when subjects perform Stroop-like tasks, are also considered to be implicated in AD/HD. Especially the frontal cortex has been hypothesised to play a major role in AD/HD pathology (see for review Barkley, Grodzinsky, & DuPaul, 1992). Therefore, it might seem surprising that the search for a deficit in interference control in AD/HD, as measured by Stroop-type tasks, has yielded conflicting findings (Nigg, 2001).

These conflicting findings can be explained in at least three possible ways. First, rapid naming deficiencies have also been observed in children with AD/HD (Tannock, Martinussen, & Frijters, 2000). Thus, a lower score on the CW-card may also be due to slower rapid naming instead of poorer interference control. Not all studies that reported deficits in interference control in AD/HD controlled for reading ability (the word condition) or naming speed (the colour condition). Second, estimates of children with AD/HD who also have a comorbid reading disorder range from 25% to 40% (Dykman & Ackerman, 1991; Semrud-Clikeman et al., 1992). If a child cannot read well, it is probably easier to ignore the word meaning on the colour-word card. This could lead to relatively faster responses on the colour-word card in children with AD/HD who are comorbid for reading disorder as compared to children with AD/HD without a reading disorder. However, this is not always the case. Children with a reading disability actually show more interference than normal controls in some studies (Everatt, Warne, Miles, & Thomson, 1997; Helland & Asbjornsen, 2000). This suggests that a deficit in interference control might not be specific to AD/HD. Third, an alternative explanation for the conflicting results might be that children with AD/HD often have other comorbid disorders such as a disruptive disorder or an anxiety disorder (Angold, Costello, & Erkanli, 1999). Results may be confounded by the high comorbidity of AD/HD with other psychiatric disorders. Inhibition deficits have also been found in comorbid disruptive disorders (Oosterlaan, Logan, & Sergeant, 1998), whereas anxiety disorders have been associated with an increased ability to inhibit (Oosterlaan, 2001). Thus, the presence of rapid naming difficulties, comorbid reading, or psychiatric disorders might have affected the interference scores found in the various studies.

Since the interference score is also determined by reading and rapid naming ability, the first goal of this meta-analysis is to test if children with AD/HD have lower word or colour scores, indicating rapid naming, and/or reading problems. A second goal is to determine the strength of an interference deficit in AD/HD. Third, we will examine the influence of eight possible moderating factors. These moderating variables include: methods of calculating the interference score, comorbid reading and psychiatric disorders, AD/HD-subtypes, gender, age, intellectual functioning, the use of medication, and sample size. To assess if a deficit in interference control is specific to AD/HD, we will compare AD/HD groups with reading disorder and psychiatric disorder groups. Furthermore, the AD/HD inattentive subtype is compared with the AD/HD combined subtype. This issue is theoretically important because of the discussion on the validity of the distinction between these subtypes. AD/HD inattentive subtype and AD/HD hyperactive/impulsive subtype have been characterised as distinct, unrelated disorders (Milich, Balentine, & Lynam, 2001). Barkley (1997) explicitly states that his behavioural inhibition model of AD/HD refers only to the AD/HD combined and hyperactive/impulsive subtypes but not to the AD/HD inattentive subtype. On average, children with AD/HD have a lower IQ than their normal peers

(Barkley, 1997). We wish to test if differences on the Stroop Colour-Word Task might be partly attributable to differences in IQ. The possible moderating effect of sample size is assessed to ascertain if the meta-analytic results are influenced by a publication bias.

Method

Description of the studies

This meta-analysis covers 17 studies published between 1990 and 2002. Table 1 summarises the main features of these studies. In column 1, the authors are listed. Column 2 shows the subject groups and the number of subjects in each of these groups. Column 3 provides the mean age and the age range for each of the groups. Information on the IQ of the children is summarised in column 4. We describe in column 5 how the different studies deal with various possible moderators, including reading and psychiatric disorders, AD/HD-subtypes, gender, age, IQ, and medication. Column 6 summarises the main results. In column 7 appear some remarks on the study. The studies were located in Pubmed, PsycInfo, Science Direct, Web of Science, and Picarta. We combined search terms related to the Stroop Colour-Word Task (such as Stroop, interference, executive) with search terms related to AD/HD (such as AD/HD, hyperactive, attention). The reference lists of published articles were used to locate other relevant studies.

To be included in the meta-analysis, studies had to meet the following criteria: (1) the studies contained at least one AD/HD group and a comparison group of normal control children, (2) studies had to use the standard Stroop Colour-Word Task, and (3) for the interference score: studies were required to use one of two methods (described in the following section) to calculate the interference score. Where studies did not report the interference score (Schmitz et al., 2002; Perugini, Harvey, Lovejoy, Sandstrom, & Webb, 2000) or another interference score was used than the two proposed in this meta-analysis (Willcutt et al., 2001), attempts were made to locate the primary author to provide the group means and the standard deviations of the group means in order to allow computation of the C–CW interference score. This meta-analysis reports on the results of 15 studies for the colour and word score; furthermore, meta-analytic results for the interference score pertain to 14 studies. With a single exception (Reeve & Schandler, 2001), all studies in this meta-analysis used DSM criteria (DSM-III-R; American Psychiatric Association, 1987; DSM-V; APA, 1994) to establish a diagnosis of ADHD. More specifically, for the studies included in the meta-analysis, 291 children were diagnosed as ADHD using DSM-III-R criteria and 306 children using DSM-IV criteria.

Studies were excluded if the same subject data were also (partly) published in another study. Therefore, we excluded the following studies: Barkley et al. (1992); Seidman et al. (1995); Seidman, Biederman, Faraone, Weber, and Oullette (1997); and the study by Grodzinsky and Barkley (1999). We excluded studies that were published before 1990 (Cohen, Weiss, & Minde, 1972;

Boucugnani & Jones, 1989; Gorenstein, Mammato, & Sandy, 1989) for one or both of the following reasons. First, these studies did not use DSM-IIIR or DSM-IV criteria. Second, some studies did not report the findings for the interference score and we were unable to locate the primary author of older studies in order to obtain the data that allows computation of the interference score.

Computerised versions of the Stroop Colour-Word Task are not comparable with the standard version in terms of control condition, response mode, and Stroop stimuli. The studies by Carter, Krener, Chaderjian, Northcutt, and Wolfe (1995a), Miller, Kavcic, and Leslie (1996) and Gaultney, Kipp, Weinstein, and McNeill (1999) used a computerised Stroop. These studies were excluded from the present meta-analysis.

Dependent variables

This meta-analysis focused on the following three dependent variables derived from the Stroop Colour-Word Task:

1. The number of words named correctly in 45 seconds on the word card: a rough indication of reading ability and rapid naming.
2. The number of colours named correctly in 45 seconds on the colour card: an indication of speed of colour processing and rapid naming.
3. The interference score: the measure of interference control in the Stroop Colour-Word Task. This measure quantifies how much slower colour naming becomes, when word reading interferes with colour naming. Two widely used methods of calculating the interference score are available. The first method controls only for colour naming, the second for both colour naming and word reading.

In the first method (the classical method), the score derived from the colour-word card was subtracted from the score on the colour card (Hammes, 1971). In the second method (Golden, 1978), correction for colour naming and word reading was achieved as follows. First, a colour-word (CW) score was predicted. This predicted score was subtracted from the uncorrected raw CW score. The predicted CW score can be calculated either by using a regression formula or by a theoretical formula (Golden, 1978). The regression formula is based on a mean score corrected for the subjects' age and education. The theoretical model suggests that the time to read one colour-word is actually the time to read one word followed by the time to name one colour. The following formula (Golden, 1978) can be deduced from the theoretical model:

Interference score
$$= \text{CW score} - [(\text{W score} \cdot \text{C score}) \div (\text{W score} + \text{C score})]$$

In which the W score is the score on the word card, the C score represents the score on the colour card, and the CW score is the score on the colour-word card.

The interference score is positive when a subject is able to inhibit word reading, and negative when word reading actively interferes with the colour naming process. The two methods used to predict a CW score yield highly comparable results ($r = .96$), suggesting that

Table 1 Summary of articles on AD/HD and the Stroop Colour-Word Task

Study	Participants	Age	IQ	Confounding Variables	Results	Remarks
1. Golden & Golden, 2002	43 LD (=RD) 43 NC 43 PD* 43 AD/HD (AD/HD- subtypes: 24 AD/HD-C; 14 AD/HD-H; 5 AD/HD-I)	10.0 9.9 9.4 9.9 Range: 6–15		Not Confounding: – Excluded: comorbidity: groups mutually exclusive – Matched on: All groups are matched with the LD group on: age, gender education, ethnicity – Statistically not different: age, education Not Reported: medication, IQ, AD/HD- subtype	Interference: AD/HD = PD = NC > LD Colour naming: AD/HD = PD = NC > LD Word reading: AD/HD = PD = NC > LD Colour-Word: NC > AD/HD = PD = LD	PD = 6 CD, 11 ODD; 1 Mood Disorder NOS; 5 Anxiety Disorder; 18 Adjustment Disorder; 2 PTSD
2. Nigg et al., 2002	69 AD/HD-C 35 AD/HD-I 51 NC	9.6 9.9 9.7 Range: 7–12	101.5 104.9 109.4	Not Confounding: – Excluded: Autism, Tourette, Depression, and Bipolar Disorder – Matched on: age, gender, recruitment source – Tested as a covariate: IQ, RD, ODD, CD, medication – Statistically not different: AD/HD-subtype, gender	Interference: AD/HD = NC Colour naming: NC > AD/HD Word reading: NC > AD/HD Colour-Word: NC > AD/HD AD/HD-I = AD/HD-C	More data than presented in the article are used. In calculating the overall effect size, only AD/HD-C and NC are included.
3. Rucklidge & Tannock, 2002	35 AD/HD 12 RD 24 AD/HD + RD 37 NC	15.2 15.1 14.9 15.0 Range: 13–16	102.2 99.9 101.1 111.0	Not Confounding: – Excluded: medication – Matched on: age – Statistically not different: gender – Tested as a covariate: ODD/CD, PD, IQ, SES Confounding: Statistically different: comorbid RD Not Reported: – AD/HD-subtypes	Interference: AD/HD = RD = AD/HD + RD Colour naming: NC > AD/HD, AD/HD + RD Word-reading: NC > RD, AD/HD + RD and AD/HD > AD/HD + RD Colour-Word: NC, AD/HD, RD > AD/HD + RD and NC > AD/HD	
4. Scheres et al., in press	18 AD/HD 20 NC (AD/HD-subtypes: 8 AD/HD-I, 10 AD/HD-C)	9.3 9.9 Range: 8–12	99.6 104.8	Not Confounding: – Excluded: girls, medication, PD – Statistically not different: AD/HD-subtype Possible Confounding: – ODD/CD: symptoms correlated high with AD/HD symptoms Confounding: – Tested as a covariate: age, IQ Not Reported: – RD	Interference: AD/HD > NC AD/HD-I = AD/HD-C	A selection of subjects from the original study were analysed because not all children performed the Stroop.

Table 1 Continued

Study	Participants	Age	IQ	Confounding Variables	Results	Remarks
5. Schmitz et al., 2002	10 AD/HD-H 10AD/HD-I 10 AD/HD-C 60 NC	14.4 14.1 14.1 13.8 Range: 12–16	91.3 87.8 85.8 92.9	Not Confounding – Excluded: medication – Statistically not different: gender, IQ Confounding: –Tested as a covariate: education, age, SES –Statistically different: AD/HD-subtype Not Reported: – ODD/CD, PD, RD	Word reading: AD/HD = NC Colour-Word: NC, AD/HD-H, AD/HD-C > AD/HD-I	In calculating the effect size for word reading, only the AD/HD-C and NC are included.
6. Reeve & Schandler, 2001	10 AD/HD 10 NC	15.3 15.2 Range: 12–17	98.9 100.6	Not Confounding –Excluded: Medication, LD, ODD, CD – Matched on: age, gender Not Reported: – PD, AD/HD-subtype	Interference: AD/HD > NC Colour naming: NC > AD/HD Word-reading: NC = AD/HD Colour-Word: NC > AD/HD	
7. Seidman, et al., 2001	21 AD/HD + 2 LD* 32 AD/HD + AD 16 AD/HD + RD 79 AD/HD 127 NC	14.3 13.8 15.2 15.1 15.1 Range: 6–17	96.1 105.4 102.2 113.3 118.0	Not Confounding: – Excluded: girls – Tested as a covariate: IQ, age, PD, ODD/CD, SES – Statistically not different: medication Confounding: – Statistically different: RD and AD Not Reported AD/HD-subtype,	Interference: AD/HD + RD > NC Colour naming: NC > AD/HD + LD; AD/HD > AD/HD + LD > AD/HD + 2LD Word reading: NC > AD/HD > AD/HD + LD; AD/HD + AD > AD/HD + 2 LD Colour-Word: NC > AD/HD > AD/HD + LD, AD/HD + 2 LD	2 LD = combined reading and arithmetic disorder
8. Spalletta et al., 2001	8 AD/HD 8 NC (7 AD/HD-C, 1 AD/HD-I)	9.4 9.0 Range: 6–14		Not Confounding – Excluded: PD, CD, medication – Matched on: age, gender, education. Not Reported: IQ, ODD, RD, AD/HD-subtype	Colour naming: NC > AD/HD Colour-Word: NC > AD/HD	PET-study
9. Willcutt et al., 2001	121 NC 93 RD 52 AD/HD 48 RD + AD/HD	10.7 10.4 10.8 10.6 Range: 8–16	113.3 100.1 101.1 99.2	Not Confounding – Excluded: Medication – Tested as a covariate: IQ, Reading ability, gender, ODD/CD Confounding: – Statistically different: age (older AD/HD more impaired), RD Not Reported: AD/HD-subtype	Interference: NC = AD/HD = RD = AD/HD + RD Colour naming: NC = AD/HD > AH/HD + RD Word reading: NC = AD/HD > RD, AD/HD + RD Colour-Word: NC = AD/HD = RD > AD/HD + RD	

Table 1 Continued

Study	Participants	Age	IQ	Confounding Variables	Results	Remarks
10. Perugini, et al., 2000	21 AD/HD 22 NC	9.6 9.2 Range: 6–12	110.1 114.2	Not Confounding: – Excluded: girls, medication – Statistically not different: age, IQ, ODD/CD Not Reported: – PD, RD, AD/HD-subtype	Colour-Word: AD/HD = NC	
11. Seidman, et al., 2000	40 AD/HD 116 NC (sibling of (AD/HD-child 118 NC (no sibling with AD/HD)	15.5 15.5 15.0 Range: 6–27	108.4 111.5 113.2	Not Confounding: – Tested as a covariate: gender, PD, ODD/CD, LD, IQ, SES – Statistically not different: medication Confounding: – {LD, PD, SES} only on word reading Not Reported: – Age, AD/HD-subtype.	Interference: AD/HD = NC Colour naming: NC > AD/HD Word reading: NC > AD/HD* Colour-Word: NC > AD/HD (No differences between the NC - groups)	*If controlled for PD, LD, SES: AD/HD = NC on Word reading
12. Semrud-Clikeman et al., 2000	10 AD/HD 11 NC	12.9 15.1 Range: 8–18	120.5 125.4	Not Confounding: – Excluded: Girls, PD, LD – Matched on: age and IQ Possible Confounding: – Medication Not Reported: – ODD/CD, AD/HD–subtype	Colour naming: NC > AD/HD Word reading: NC > AD/HD Colour-Word: NC > AD/HD	MRI study
13. Houghton et al., 1999	32 AD/HD-I 62 AD/HD-C 28 NC	10.5 9.9 10.2 Range: 6–12	V: 107 P: 119 V:100 P:111 V:107 P:116	Not Confounding: – Excluded: PD, LD, medication – Tested as a covariate: gender, age Confounding: – Statistically different: AD/HD-subtype	Interference: NC = AD/HD-I = AD/HD-H Colour naming: NC > AD/HD-C Word reading: NC > AD/HD-C, AD/HD-I Colour-Word: NC > AD/HD-C	In calculating the overall Effect size, only AD/HD-C and NC are included.
14. Seidman et al., 1997	43 AD/HD 36 NC	11.4 11.9 Range: 6–17	106.0 112.1	Not Confounding: – Excluded: boys – Matched on: age, SES, grade – Tested as a covariate: family history, PD, LD Possible Confounding: – Medication ·Not Reported: AD/HD-subtype, IQ	Interference: NC = AD/HD Colour naming: NC = AD/HD Word reading: NC = AD/HD Colour-Word: NC = AD/HD	

Table 1 Continued

Study	Participants	Age	IQ	Confounding Variables	Results	Remarks
15. MacLeod & Prior, 1996	12 NC 12 AD/HD 10 CD 12 PD*	15.5 14.5 15.8 15.6Range: 12–18	112.5 107.2 103.6 106.4	Not Confounding: – Excluded: PD and CD: groups mutually exclusive, medication. – Tested as a covariate: IQ and age Not Reported: gender, RD, AD/HD-subtype	Interference: AD/HD, CD, PD > NC	*PD = 3 anorexia, 4 schizophrenia, 3 depression, 2 school refusals
16. Grodzinsky & Diamond, 1992	66 AD/HD 34 young 32 old 64 NC 30 young 34 old	7.6 10.2 7.5 10.4 Range: 6–11	112.9 110.1 112.9 115.5	Not Confounding: – Excluded: girls, LD, PD medication, AD/HD-I – Statistically not different: age, IQ, motor performance Possible Confounding: – Tested as a covariate: SES Not Reported: – ODD/CD	Colour naming: NC > AD/HD Word reading: NC > AD/HD Colour-Word: NC > AD/HD	
17. Lufi et al., 1990	29 AD/HD 21 PD* 20 NC	12.7 13.2 12.7 Range: 9–16		Not Confounding: – Excluded: girls, medication Not Reported: – RD, PD, ODD/CD, AD/HD-subtype, IQ	Interference: AD/HD > PD Colour naming: NC = AD/HD = PD Word reading: AD/HD > PD Colour-Word: NC = AD/HD = PD	*PD = various conduct and anxiety disorders

Note: Interference >= more interference. Other measures >= a better score, thus a faster response time.

Index of abbreviations:

AD = Arithmetic Disorder

AD/HD = Attention Deficit/ Hyperactivity Disorder

AD/HD-I = predominantly inattentive subtype

AD/HD-H = predominantly hyperactive subtype

AD/HD-C = combined subtype

CD = Conduct Disorder

IQ = Intelligence Quotient

LD = Learning disorders (reading disorder or artithmetic disorder)

NC = Normal Controls

ODD = Oppositional Defiant Disorder

PD = Psychiatric Disorders (other than AD/HD, ODD and CD)

RD = Reading Disorder

SES = Social Economic Status.

these two methods are interchangeable. Both methods will be referred to as the 'Golden' method, since Golden (1978) proposed both methods.

We compared the 'Golden' method (irrespective of how the CW score was predicted) with the classical method (C score – CW score, Hammes, 1971) as far as was possible with the available data. If no interference score was reported, the classical interference score was calculated using the raw mean data. We estimated the group standard deviation (*SD*) with the following formula:

$$SD(\text{C score} - \text{CW score})$$
$$= \sqrt{\{SD(\text{C score})^2 + SD(\text{CW score})^2}$$
$$- 2[SD(\text{C score}) \cdot SD(\text{CW score}) \cdot r(\text{C score} - \text{CW score})]\}$$

In this formula, *r*(C score – CW score) represents the average correlation between the number of correct responses on the colour card and the number of correct responses on the colour-word card. The *r*(C score – CW score) was set at .7 (C. J. Golden, personal communication, 25 February 2003).

Statistical analyses

Analyses were conducted using a computer program developed by Borenstein and Rothstein (1999). The effect sizes (in terms of Cohen's *d*) for the word, colour and interference score were calculated for each study separately. An overall effect size was computed by weighting all the effect sizes with the sample size of the study. Following Cohen's guidelines, effect sizes of .20, .50, and .80 were used as thresholds to define small, medium, and large effects, respectively (Cohen, 1988). To test if the variability in effect sizes exceeded that expected from sampling error alone, a test of heterogeneity was conducted (Borenstein & Rothstein, 1999).

Since it is unreasonable to assume that all of the heterogeneity in the effect sizes of the studies can be explained, the possibility of 'residual heterogeneity' must be acknowledged in the statistical analysis. The appropriate analysis is, therefore, a 'random effects' rather than a 'fixed effects' meta-regression model (Thompson & Higgins, 2002). In the 'fixed effects' model, it is assumed that all studies are derived from a common population, and the only source of variation between the studies is random error. With a sufficiently large sample, this error will approach zero and the estimates of the effect sizes reflect together the true combined effect size. In the 'random effects' model, it is assumed that the effect sizes may differ because the subject characteristics vary from one study to another. When an attempt is made to combine data, two sources of variance need to be dealt with: random error and variance that reflects real differences between the populations from which subjects are sampled. A fixed effects analysis estimates the assumed common effect, whereas a random effects analysis estimates the mean of a distribution of effects across studies. If residual heterogeneity exists, a random effects analysis appropriately yields wider confidence intervals for the combined effect size than a fixed effects analysis (Thompson & Higgins, 2002).

The random combined effect sizes were calculated for the word, colour, and interference score of studies (1) comparing AD/HD groups without comorbid reading disorder and AD/HD groups with comorbid reading disorder, (2) comparing AD/HD groups and reading disorder groups, (3) comparing AD/HD groups and psychiatric control groups and, (4) comparing AD/HD primarily inattentive subtype groups and AD/HD combined subtype groups. Note that for the third comparison, it was not possible to calculate the effect sizes for the word and colour score because there were too few studies to conduct this comparison. Age, IQ, and sample size were correlated with the effect sizes for the word, colour, and interference score. The correlations for age and IQ were weighted for the relative number of subjects in the study (Stevens, 1996). The two methods for calculating the interference score were compared using a Wilcoxon rank order test. Alpha was set at .05 for all analyses.

Results

The results of the three dependent measures of interest: the word score, the colour score, and the interference score are summarised in Tables 2 to 4.

The combined random effect size for the word condition (Table 2) was .49 and significant (*p* < .001). This effect is close to Cohen's standard for a medium effect size. The effect sizes for the word condition were heterogeneous (*p* = .003), indicating that there were large variations in the magnitude of the difference between children with AD/HD and normal controls. Two effect sizes were close to zero (Golden & Golden, 2002; Schmitz et al., 2002) and 14 effect sizes were positive.

The combined random effect size for the colour condition, .58 (see Table 3), is significant (*p* < .001) and corresponds to a medium effect size. Again, effect sizes were heterogeneous (*p* = .003). All the effect sizes for the colour condition were positive, which means that only the magnitude of the effect varied between studies. This indicates that in all studies, normal controls performed better than children with AD/HD in the colour naming condition.

The combined random effect size of the variable of primary interest, the interference score, was .35 (see Table 4) and significant (*p* = .004). This is considered a small effect size. The effect sizes for the interference scores were heterogeneous (*p* < .001). One effect size for the interference score was negative (Perugini et al., 2000). In eight studies, the effect sizes were around zero (Golden & Golden, 2002; Houghton et al., 1999; Nigg, Blaskey, Huang-Pollock, & Rappley, 2002; Rucklidge & Tannock, 2002; Seidman et al., 1997; Seidman, Biederman, Monuteaux, Weber, & Faraone, 2000; Seidman, Biederman, Monuteaux, Doyle, & Faraone, 2001; Willcutt et al., 2001) and in five studies the effect sizes were positive (Lufi et al., 1990; MacLeod & Prior, 1996; Reeve & Schandler, 2001; Scheres et al., in press; Spalletta et al., 2001). This indicates that studies report inconsistent results for the difference

between children with AD/HD and their normal peers with regard to the interference score.

Moderating variables

Methods for calculating the interference score. The overall effect size for the interference score (.35) was calculated with the interference score as reported by the authors. The effect size for the C – CW interference score was .26, not significant and heterogeneous ($p = .04$) while the effect size for the Golden interference score was .40 and significant ($p = .01$) but heterogeneous ($p < .001$).

If both the raw mean data and the Golden score were available, a C – CW score was computed and compared with the Golden score. This was done for seven studies (Houghton et al., 1999; Lufi et al., 1990; Nigg et al., 2002; Reeve & Schandler, 2001; Scheres et al., in press; Seidman et al., 2000, 2001). The random combined effect size for the C – CW interference score was −.003, not significant and homogeneous. The effect size for the Golden interference score, was .29, significant ($p = .03$) but heterogeneous ($p = .02$). No significant difference was found between these two methods ($Z = −1.69, p = .09$), but this result suggests that there is a trend for the Golden score being larger than the C – CW score.

Table 1 indicates how the studies deal with the moderating variables described in the following section.

Reading disorder and psychiatric disorders. The meta-analytic results of the group comparisons between: (1) AD/HD – AD/HD and comorbid reading disorder, (2) AD/HD– reading disorder, and (3) AD/HD and psychiatric disorders are presented in Table 5.

As can be seen in Table 5, children with AD/HD perform better at word reading and colour naming than children with AD/HD and a comorbid reading disorder but there is no significant difference in the interference score. Compared with children with a reading disorder, children with AD/HD perform better at word reading, there is a trend for better performance at colour naming but they have a reduced resistance to interference. There is no significant difference between children with AD/HD and children with various psychiatric disorders in interference control.

AD/HD-Subtypes. Three studies (Houghton et al., 1999; Nigg et al., 2002; Scheres et al., in press) compared children with AD/HD inattentive subtype (AD/HD-I) and AD/HD combined subtype (AD/HD-C) and found no differences. Meta-analytic results, however, reveal a small, but significant and homogeneous combined random effect size of −.35 ($p = .02$) for the interference score. This effect size indicates that children with AD/HD-I have less

resistance to interference than children with AD/HD-C. The effect sizes for the time-to-read words and the time-to-name colours were not significantly different between the subtypes (combined random effect size: −.14, *ns*, and .21, *ns*, respectively) and homogeneous.

Gender. Research has failed to find a substantial difference in the Stroop Colour-Word Task dependent measures between men and women at any age (MacLeod, 1991), although women may be somewhat faster, especially in naming colours. In this meta-analysis, the proportion of boys and girls was approximately equal across AD/HD groups and the normal control groups. Hence, there is no reason to suspect an influence of gender on the dependent variables.

Age. The interference effect begins early in the school years, rising to its highest level around grades 2 to 3 as reading skills develop (Schiller, 1966). Cognitive control is still developing after grades 2 and 3 with an accompanying improvement in interference control. No developmental changes have been reported until approximately 60 years, at which age interference control begins to decrease (Comalli, Wapner, & Werner, 1962).

No significant correlations were found between the effect sizes for each of the dependent variables and mean age. Thus, it seems that the differences on the Stroop Colour-Word Task between children with AD/HD and their normal peers remain the same across the age range studied here.

Intellectual functioning. No significant correlations were found between the difference between AD/HD groups and normal control groups in IQ scores and the effect sizes for the word: $r(14) = −.31$, *ns*, colour: $r(13) = −.20$, *ns*, and interference score: $r(12) = .11$, *ns*.

Medication. Methylphenidate (MPH) is the most common pharmacological treatment for children with AD/HD (Greenhill, Halperin, & Abikoff, 1999; MTA Cooperative Group, 1999). Recently, it has been shown that MPH improves colour naming and word reading, but that it has no effect on response interference (Bedard, Ickowicz, & Tannock, 2002). There were too few studies using the Stroop Colour-Word Task to analyse the effects of medication in this meta-analysis.

Sample size. There was a strong negative correlation between sample size and the effect sizes for the dependent variables in this meta-analysis for the colour: $r(16) = −.68$, $p < .01$, word: $r(16) = −.42$, *ns* and interference score: $r(14) = −.60$, $p = .02$. This means that studies with larger samples report small effect sizes, while studies with small samples report large effect sizes. These correlations may reflect the

Table 2 Comparison of the effect sizes for the word score in AD/HD studies

Study	Effect Size	p	Sample size		Schematic representation of effect sizes with 95% confidence interval				
			AD/HD	Controls	−4.0	−2.0	0.0	2.0	4.00
Golden & Golden, 2002	−.06	.80	43	43					
Grodzinsky & Diamond, 1992	.85	.00	32	34					
Grodzinsky & Diamond, 1992	.93	.00	34	30					
Hougton et al., 1999	.57	.01	62	28					
Lufi et al., 1990	.71	.02	29	20					
Nigg et al., 2002	.53	.00	69	51					
Perugini et al., 2000	.48	.13	22	21					
Reeve & Schandler, 2001	.44	.34	10	10					
Rucklidge & Tannock, 2002	.58	.02	35	37					
Scheres et al., in press	.87	.01	18	20					
Schmitz et al., 2002	.10	.76	10	60					
Seidman et al., 1997	.35	.12	43	36					
Seidman et al., 2000	.38	.04	40	118					
Seidman et al., 2001	.22	.13	79	127					
Semrud-Clikeman et al., 2000	1.53	.00	10	11					
Willcut et al., 2001	.28	.09	52	121					
Combined(16)	**.49**	**.00**	**588**	**767**					

Note: Positive effect sizes indicate better performance for normal controls as compared to children with AD/HD. Two different age groups are included for the Grodzinsky & Diamond study (1992).The random combined effect size is reported in the last row.

Table 3 Comparison of the effect sizes for the colour score in AD/HD studies

Study	Effect Size	p	Sample size		Schematic representation of effect sizes with 95% confidence interval				
			AD/HD	Controls	−4.0	−2.0	0.0	2.0	4.00
Golden & Golden, 2002	.35	.11	43	43					
Grodzinsky & Diamond, 1992	.93	.00	32	34					
Grodzinsky & Diamond, 1992	.58	.03	34	30					
Hougton et al., 1999	.69	.00	62	28					
Lufi et al., 1990	.70	.02	29	20					
Nigg et al., 2002	.41	.03	69	51					
Perugini et al., 2000	.29	.35	21	22					
Reeve & Schandler, 2001	1.21	.01	10	10					
Rucklidge & Tannock, 2002	.93	.00	35	37					
Scheres et al., in press	.82	.02	18	20					
Seidman et al., 1997	.32	.17	43	36					
Seidman et al., 2000	.40	.03	40	118					
Seidman et al., 2001	.17	.24	79	127					
Semrud-Clikeman et al., 2000	1.95	.00	10	11					
Spalleta et al., 2001	1.09	.05	8	8					
Willcut et al., 2001	.20	.23	52	121					
Combined(16)	**.58**	**.00**	**585**	**716**					

Note: Positive effect sizes indicate better performance for normal controls as compared to children with AD/HD. Two different age groups are included for the Grodzinsky & Diamond study (1992).The random combined effect size is reported in the last row.

difficulty of publishing studies including small samples and reporting no group differences.

Discussion

Impairments in interference control have been implicated as one of the core deficits in AD/HD (Barkley, 1997). The Stroop Colour-Word Task has been frequently used to demonstrate this deficit and as an aid in clinical diagnosis. Seventeen independ-

ent studies, encompassing large groups of children, were analysed to determine the degree of this deficit in interference control in children with AD/HD compared with normal controls. The role of the following moderator variables was assessed: comorbid reading and psychiatric disorders, AD/HD-subtypes, gender, age, IQ and sample size. The results reported here indicate that a deficit in interference control, as measured with the Stroop Colour-Word Task, is either absent or very small in children with AD/HD and depends heavily on the method of

Table 4 Comparison of the effect sizes for the interference score in AD/HD studies

Study	Interference score	Effect Size	p	AD/HD	Controls	Schematic representation of effect sizes with 95% confidence interval
Golden & Golden, 2002	C–CW	.15	.50	43	43	
Perugini et al., 2000	C–CW	−.29	.35	21	22	
Scheres et al., in press	C–CW	.70	.04	18	20	
Spalleta et al., 2001	C–CW	1.30	.02	8	8	
Willcut et al., 2001	C–CW	.07	.65	52	121	
Combined C–CW (5)		**.26**	**.18**	**142**	**214**	
Hougton et al., 1999	Golden	.11	.64	62	28	
Lufi et al., 1990	Golden	.69	.02	29	20	
MacLeod & Prior, 1996	Golden	2.00	.00	12	12	
Nigg et al., 2002	Golden	.16	.40	69	51	
Reeve & Schandler, 2001	Golden	1.45	.00	10	10	
Rucklidge & Tannock, 2002	Golden	.10	.66	35	37	
Seidman et al., 1997	Golden	.15	.50	43	36	
Seidman et al., 2000	Golden	.13	.48	40	118	
Seidman et al., 2001	Golden	−.03	.81	79	127	
Combined Golden (9)		**.40**	**.01**	**379**	**439**	
Combined (14)		**.35**	**.00**	**521**	**653**	

Note: Positive scores indicate better performance for normal controls as compared to children with AD/HD. Separate combined effect sizes are shown in the 6th row for the C–CW score, in the 16th row for the 'Golden' score and in the last row for both methods combined.

calculation. Children with AD/HD had lower word reading and colour naming scores than normal controls. Comorbid reading disorder was found to have a negative impact on colour naming and word reading, but there was no consistent effect on the interference score. Compared with children with a reading disorder, children with AD/HD had a better word and colour score, but a lower interference score. There was no significant difference between children with AD/HD and children with various psychiatric disorders on the interference score. A small difference was found in interference control between the AD/HD-subtypes: children with predominantly inattentive subtype had poorer control over interference than the children with AD/HD-combined subtype. No effects of gender, age, and IQ were noted, but the correlations between the effect sizes and sample size suggest a publication bias.

Study limitations

The negative correlation between sample size and the effect sizes for the colour and interference scores

may be an indication of a publication bias. Small studies with significant results will probably be published more often than small studies with no significant results (see for a review Rosenthal, 1979).

Some children in the AD/HD-inattentive subtype group may be just one hyperactivity symptom below the threshold for the AD/HD-combined subtype or may be formerly children with AD/HD-combined subtype, who have outgrown one or two symptoms of hyperactivity/impulsivity over time. The distinction between AD/HD-inattentive subtype and AD/HD-combined subtype may be confounded by problems of contamination of the inattentive subtype with subthreshold combined subtype cases.

The comparisons between children with AD/HD and children with various psychiatric disorders, between children with AD/HD with and without a comorbid reading disorder, and between children with AD/HD and children with a reading disorder are based on a limited number of studies (three studies). Thus, the results pertaining to these group comparisons should be interpreted with caution.

Table 5 Random combined effect sizes for the word, colour and interference score in studies with AD/HD groups, reading disorder groups and psychiatric disorder groups

Group comparison	Sample size	Studies	Word		Colour		Interference	
			d	p	d	p	d	p
AD/HD – AD/HD + RD	254	3	.96[a]	.00	.54[a]	.00	−.26[b]	.42
AD/HD – RD	278	3	.64[a]	.00	.29[a]	.06	−.32[a]	.02
AD/HD – PD	160	3	–	–	–	–	−.36[b]	.28

Note: Dashes indicate that the effect size was not calculated. Positive effect sizes indicate better performance for the AD/HD group as compared to the other groups. *d* = random combined effect size, RD = reading disorder, PD = various psychiatric disorders. [a]Homogeneous effect. [b]Heterogeneous effect.

No significant correlations were found between, on the one hand, age and IQ, and, on the other, the effect sizes for the word, colour, and interference scores. However, these correlations probably underestimate the associations that would be found if this analysis were conducted using data at a subject level. Furthermore, one study (Scheres et al., in press) found that covarying for age (and IQ) reduced the differences between children with AD/HD and their normal peers.

What is the best method to calculate interference control?

This meta-analysis shows that the method of calculating interference is crucial to the interpretation of the results. When interference is calculated by subtracting the CW score from the C score, there is no difference in interference control between children with AD/HD and normal controls. Thus, because children with AD/HD are slower on both cards (C card and CW card) compared with normal controls, there is no difference in the interference score. The Golden method is better in differentiating children with AD/HD from normal controls than the classical C–CW score. It should be borne in mind that the interference score proposed by Golden (1978) is based on a comparison of an estimation of a CW score and the real CW score. This estimation is based on the assumption that the time to read one colour-word is actually the time to read one word followed by the time to name one colour. This assumption corresponds with older, sequential explanations of the Stroop effect: that processing in one stage must be completed (or almost completed) before processing in the next stage may begin. Neural imaging research on the Stroop Colour-Word Task supports the notion that Stroop stimuli are processed in parallel in a network of brain areas (Atkinson, Drysdale, & Fulham, 2003; Ukai et al., 2002; West & Alain, 1999). Therefore, the theoretical model on which the formula is founded does not stand on strong ground. For this reason, we suggest that the traditional method of calculating the interference score may be a more 'pure measure' of interference.

Do children with AD/HD have a reduced resistance to interference?

This meta-analysis suggests that there is little support for a deficit in interference control in AD/HD, as measured by the Stroop Colour-Word Task. The fact that no deficit in interference control was observed using the traditional method to calculate interference, and that children with AD/HD-inattentive subtype may have less resistance to interference than children with AD/HD-combined subtype, does not support the inhibition deficit hypothesis (Barkley, 1997; Pennington & Ozonoff, 1996), which pertains to the AD/HD combined subtype in particular.

Results of other studies, using a different design to measure interference control, are mixed. Scheres et al. (in press) and Jonkman et al. (1999) measured interference control with a Flanker Task and found an interference effect on errors. Cornoldi et al. (2001) found that children with AD/HD had difficulties in controlling interference related to working memory. When a computerised version of the Stroop was used, Carter et al. (1995a) found a difference in reaction time between children with AD/HD and normal controls, while Gaultney et al. (1999) did not find such an effect.

An interesting finding emerges from interference studies with fMRI in which AD/HD groups and normal control groups are compared on brain activation during a 'counting Stroop' (Bush et al., 1999) and a 'go-nogo' task (Durston et al., 2003). Activation patterns indicated that the normal adults activated the anterior cingulate cortex; specifically the cognitive division (Bush et al., 1999) and normal children activated fronto-striatal regions (Durston et al., 2003). In contrast, adults with AD/HD failed to activate the anterior cingulate cortex, and children with AD/HD failed to activate fronto-striatal regions. In both studies, the AD/HD groups relied on a more diffuse network of regions, although in the study by Bush et al. (1999) no performance differences were observed between the control group and the AD/HD group. Bush and colleagues interpreted these finding as demonstrating that adults with AD/HD may compensate for impairments by recruiting a different and less responsive pathway. Based on only the card version of the Stroop Colour-Word Task, one cannot conclude that children with AD/HD have no deficit in interference control. This is because results from other interference tasks and imaging research indicate that AD/HD is related to problems in interference control. The fact that this is not shown by the card version of the Stroop Colour-Word Task may indicate that this is not a *generalised* deficit but may be context dependent.

Rapid naming

Interference scores need to be controlled for at least colour naming. If this is not done, differences on the CW card may also reflect differences in rapid naming. Deficiencies on the W, C, and CW cards have been related to abnormalities in brain structure (Semrud-Clikeman et al., 2000). Semrud-Clikeman et al. (2000) demonstrated that poorer performance on all three cards of the Stroop Colour-Word Task was significantly related to reversed asymmetry of the caudate. Thus, a slower retrieval of colour-names and a slower reading speed may be an indication of abnormalities in brain structure in AD/HD. Therefore, it is important to assess these deficits in AD/HD. Slow processing speed is frequently reported in children with AD/HD compared to normal controls (e.g., Mason, Humphreys, & Kent, 2003; Sergeant,

Oosterlaan, & van der Meere, 1999). This general slowing has been interpreted as reflecting a 'non-optimal activation state' (see, for review, Sergeant & Van der Meere, 1990, 1991; Sergeant et al., 1999; Van de Meere, 1996). Other evidence that children with AD/HD may be less able than their normal peers to maintain the state required for optimal task performance can be derived from the work of Leth-Steensen, King Elbaz, and Douglas (2000). Their results confirmed that the mean slower reaction times of boys with AD/HD were not due to a generalised slowing of all responses but was due to a greater proportion of abnormally slower responses, as shown earlier by Sergeant (1988). Children with AD/HD may be less able than their normal peers to maintain a stable reaction time over trials. This result can explain the slower naming and reading speeds and is consistent with the hypothesis that AD/HD involves a non-optimal activation state. Unfortunately, the present data does not allow this theoretical explanation to be tested. Future studies should address this issue.

Clinical practice and future research

Based on this meta-analysis, we cannot recommend the Stroop Colour-Word Task in its standard form for use in clinical practice in AD/HD. Another reason to advise against the use of the Stroop Colour-Word Task in clinical practice is its low negative predictive power: a normal score can be obtained despite the fact that the child has AD/HD (Doyle et al., 2000; Grodzinsky & Barkley, 1999). The predictive validity can be improved when used in combination with other executive tests (Perugini et al., 2000). Therefore, if the Stroop Colour-Word Task is used in clinical practice, it should always be used in combination with other executive function tests.

The interference score cannot differentiate between children with AD/HD and children with various other psychiatric disorders. The interference score can differentiate between children with AD/HD and children with a reading disorder. This difference probably reflects the fact that reading is less automatised in children with a reading disability. Word reading will thus interfere less with colour naming on the CW-card in children with a reading disorder.

A better alternative for research and clinical use may be a 'trial-by-trial' computerised version of the Stroop Colour-Word Task. Perlstein, Carter, Barch, and Baird (1999) showed that a trial-by-trial version of the Stroop Colour-Word Task showed greater sensitivity to attentional pathology. A second advantage is that a computer allows response times and the response variability to be measured with high accuracy. A computerised Stroop Colour-Word Task and variations on this task have already been used in various studies (Bush et al., 1999, Gaultney et al., 1999, Carter et al., 1995a, Miller, Kavic, & Leslie, 1996).

Conclusion

The results obtained with the Stroop Colour-Word Task do not provide strong evidence for a core deficit in interference control in AD/HD. This result argues against current theoretical models, which emphasise inhibitory control deficits in AD/HD (Barkley, 1997; Pennington & Ozonoff, 1996). Studies using other measures of interference control, however, do provide evidence in favour of the interference control deficit hypothesis, which suggests that there might be a subtler and contextually dependent interference deficit in AD/HD. Interestingly, in this meta-analysis, rapid naming deficiencies are more pronounced in AD/HD than a deficit in interference control. Should we reject the Stroop Colour-Word Task in its standard form if we want to investigate interference control in children with AD/HD? Our conclusion is affirmative to this question. The Stroop Colour-Word Task is not a golden standard to demonstrate an interference deficit in AD/HD.

Acknowledgements

Correspondence to

R. van Mourik, Department of Clinical Neuropsychology, Vrije Universiteit, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands; Email: R.van.mourik@psy.vu.nl

References

Adleman, N.E., Menon, V., Blasey, C.M., White, C.D., Warsofsky, I.S., Glover, G.H., & Reiss, A.L. (2002). A developmental fMRI study of the Stroop Color-Word Task. *Neuroimage, 16,* 61–75.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd edn, rev.). Washington, DC: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th edn.). Washington, DC: Author.

Angold, A.E., Costello, J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40,* 57–87.

Atkinson, C.M., Drysdale, K.A., & Fulham, W.R. (2003). Event-related potentials to Stroop and reverse Stroop

stimuli. *International Journal of Psychophysiology*, *47*, 1–21.

Barkley, R.A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin, 121*, 65–94.

Barkley, R.A., Grodzinsky, G., & DuPaul, G.J. (1992). Frontal-lobe functions in attention-deficit disorder with and without hyperactivity: A review and research report. *Journal of Abnormal Child Psychology, 20*, 163–188.

Bedard, A.C., Ickowicz, A., & Tannock, R. (2002). Methylphenidate improves Stroop naming speed, but not response interference, in children with attention deficit hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology, 12*, 301–309.

Bench, C.J., Frith, C.D., Grasby, P.M., Friston, K.J., Paulesu, E., Frackowiak, R.S., & Dolan, R.J. (1993). Investigations of the functional anatomy of attention using the Stroop test. *Neuropsychologica, 31*, 907–922.

Borenstein, M., & Rothstein, H. (1999). *Comprehensive meta-analysis, a computer program for research synthesis*. New Jersey: Biostat Inc.

Boucugnani, L.L., & Jones, R. W. (1989). Behaviors analogous to frontal lobe dysfunction in children with attention deficit hyperactivity disorder. *Archives of Clinical Neuropscyhology, 4*, 161–173.

Bush, G., Frazier, J.A., Rauch, S.L., Seidman, L.J., Whalen, P.J., Jenike, M.A., Rosen, B.R., & Biederman, J. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. *Biological Psychiatry, 45*, 1542–1552.

Bush, G., Luu, P., & Posner, M.I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4*, 215–222.

Carter, C.S., Krener, P., Chaderjian, M., Northcutt, C., & Wolfe, V. (1995a). Abnormal processing of irrelevant information in attention-deficit hyperactivity disorder. *Psychiatry Research, 56*, 59–70.

Carter, C.S., Mintun, M., & Cohen, J.D. (1995b). Interference and facilitation effects during selective attention: An H2 15O PET study of Stroop Task performance. *Neuroimage, 2*, 264–272.

Chung Leung, H., Skudlarski, P., Gatenby, J.C., Peterson, B.S., & Gore, J.C. (2000). An event-related functional MRI study of the Stroop Color Word interference task. *Cerebral Cortex, 10*, 552–560.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn). Hillsdale, NJ: Erlbaum.

Cohen, J.D., Dunbar, K., & MacClelland, J.L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97*, 332–361.

Cohen, N.J., Weiss, G., & Minde, K. (1972). Cognitive styles in adolescents previously diagnosed as hyperactive. *Journal of Child Psychology and Psychiatry, 13*, 203–209.

Comalli, P.E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology, 100*, 46–53.

Cornoldi, C., Marzocchi, G.M., Belotti, M., Garoli, M.G., De Meo, T., & Braga, C. (2001). Working memory interference control deficit in children referred by teachers for ADHD symptoms. *Child Neuropsychology, 7*, 230–240.

Doyle, A.E., Biederman, J., Seidman, L.J., Weber W., & Faraone, S.V. (2000). Diagnostic efficiency of neuropsychological test scores for discriminating boys with and without attention deficit-hyperactivity disorder. *Journal of Consulting and Clinical Psychology, 68*, 477–488.

Durston, S., Tottenham, N.T., Thomas, K.M., Davidson, M.C., Eigsi, I.M., Yang, Y., Uluǧ A.M., & Casey, B.J. (2003). Differential patterns of striatal activation in young children with and without ADHD. *Biological Psychiatry, 53*, 871–878.

Dykman, R. & Ackerman, P.T. (1991). ADD and specific reading disability: Separate but often overlapping disorders. *Journal of Learning Disabilities, 28*, 272–290.

Everatt, J., Warne, J., Miles, T.R., & Thomson, M.E. (1997). The incidence of Stroop interference in dyslexia. *Dyslexia, 3*, 222–228.

Gaultney, J.F., Kipp, K., Weinstein, J., & McNeill, J. (1999). Inhibition and mental effort in attention deficit hyperactivity disorder. *Journal of Developmental and Physical Disabilities, 11*, 105–114.

Golden, C.J. (1978). *The Stroop color and word test*. Chicago, IL: Stoelting Company.

*Golden, Z.L., & Golden, C.J. (2002). Patterns of performance on the Stroop color and word test in children with learning, attentional, and psychiatric disabilities. *Psychology in the Schools, 39*, 489–495.

Greenhill, L.L., Halperin, J.M., & Abikoff, H. (1999). Stimulant medications. *Journal of the American Academy for Child and Adolescent Psychiatry, 38*, 503–512.

*Grodzinsky, G.M., & Diamond, R. (1992). Frontal-lobe functioning in boys with attention-deficit hyperactivity disorder. *Developmental Neuropsychology, 8*, 427–445.

Grodzinsky, G.M., & Barkley, R.A. (1999). Predictive power of frontal lobe tests in the diagnosis of attention deficit hyperactivity disorder. *The Clinical Neuropsychologist, 13*, 12–21.

Gorenstein, E.E., Mammato, C.A., & Sandy, J.M. (1989). Performance of inattentive-overactive children on selected measures of prefrontal-type function. *Journal of Clinical Psychology, 45*, 619–632.

Hammes, J.G.W. (1971). *De Stroop Kleur-Woord Test*. Handleiding. Lisse: Swets and Zeitlinger.

Helland, T., & Asbjornsen, A. (2000). Executive functions in dyslexia. *Child Neuropsychology, 6*, 37–48.

*Houghton, S., Douglas, G., West, J., Whiting, K., Wall. M., Langsford, S., Powell, L., & Carroll, A. (1999). Differential patterns of executive function in children with attention-deficit hyperactivity disorder according to gender and subtype. *Journal of Child Neurology, 14*, 801–805.

Jonkman, L.M., Kemner, C., Verbaten, M.N., Van Engeland, H., Kenemans, J.L., Camfferman, G., Buitelaar, J.K., & Koelega, H.S. (1999). Perceptual

---

* References marked with an asterisk indicate studies included in the meta-analysis.

and response interference in children with attention-deficit hyperactivity disorder and the effects of methylphenidate. *Psychophysiology, 3*, 419–429.

Leth-Steensen, C., King Elbaz, Z., & Douglas, V.I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica, 104*, 167–190.

Lufi, D., Cohen, A., & Parish-Plass, J. (1990). Identifying attention deficit hyperactive disorder with the WISC-R and the Stroop color and word test. *Psychology in the Schools, 27*, 28–34.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163–203.

MacLeod, C.M., & MacDonald, P.A. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*, 383–391.

*MacLeod, D., & Prior, M. (1996). Attention deficits in adolescents with ADHD and other clinical groups. *Child Neuropsychology, 2*, 1–10.

Mason, D.J., Humphreys, G.W., & Kent, L.S. (2003). Exploring selective attention in ADHD: Visual search through space and time. *Journal of Child Psychology and Psychiatry, 44*, 1–20.

Milich, R., Balentine, A.C., & Lynam, D.R. (2001). ADHD combined type and ADHD predominantly inattentive type are distinct and unrelated disorders. *Clinical Psychology: Science and Practice, 8*, 463–488.

Miller, D.C., Kavcic, V., & Leslie, J.E. (1996). ERP changes induced by methylphenidate in boys with attention deficit hyperactivity disorder. *Journal of Attention Disorders, 1*, 95–113.

MTA Cooperative Group. (1999). A 14-month randomized clinical trial of treatment strategies for Attention-Deficit/Hyperactivity Disorder. *Archives of General Psychiatry, 56*, 1073–1085.

Nigg, J.T. (2001). Is ADHD a disinhibitory disorder? *Psychological Bulletin, 127,* 571–598.

*Nigg, J.T., Blaskey, L.G., Huang-Pollock, M.A., & Rappley, M.D. (2002). Neuropsychological executive functions and DSM-IV ADHD subtypes. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 59–66.

Oosterlaan, J. (2001). Behavioral inhibition and the development of childhood anxiety disorders. In W.K. Silverman & P.D.A. Treffers (Eds.), *Anxiety disorders in children and adolescents: Research, assessment and intervention* (pp. 45–71). Cambridge: University Press.

Oosterlaan, J., Logan, G., & Sergeant, J.A. (1998). Response inhibition in AD/HD, CD, comorbid AD/HD+CD, anxious, and control children: A meta-analysis of studies with the stop task. *Journal of Child Psychology and Psychiatry, 39*, 411–425.

Pardo, J.V., Pardo, P.J., Janer, K.W., & Raichle, M.E. (1990). The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. *Neurobiology, 87*, 256–259.

Pennington, B.F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry, 37*, 51–87.

Perlstein, W.M., Carter, C.S., Barch, D.M., & Barid, J.W. (1998). The Stroop task and attention deficits in schizophrenia: A critical evaluation of card and single-trial Stroop methodologies. *Neuropsychology, 12*, 414–425.

*Perugini, E.M., Harvey, E.A., Lovejoy, D.W., Sandstrom, K., & Webb, A.H. (2000). The predictive power of combined neuropsychological measures for attention-deficit/hyperactivity disorder in children. *Child Neuropsychology, 6*, 101–114.

Peterson, B.S., Skudlarski, P., Gatenby, J.C., Zhang, H., Anderson, A.W., & Gore, J.C. (1999). An fMRI study of Stroop word-color interference: Evidence for cingulate subregions subserving multiple distributed attentional systems. *Biological Psychiatry, 45*, 1237–1258.

Peterson, B.S., Kane, M.J., Alexander, G.M., Lacadie, C., Skudlarski, P., Leung, H.C., May, J., & Gore, J.C. (2002). An event-related functional MRI study comparing interference effects in the Simon and Stroop tasks. *Cognitive Brain Research, 13*, 427–440.

Posner, M.I., & DiGirolamo, G.J. (1998). Executive attention: Conflict, target detection, and cognitive control. In R. Parasuraman (Ed.), *The attentive brain* (pp. 401–424). Cambridge: MIT Press.

*Reeve, W.V., & Schandler, S.L. (2001). Frontal lobe functioning in adolescents with attention deficit hyperactivity disorder. *Adolescence, 36*, 749–765.

Rosenthal, R. (1979). 'The file drawer problem' and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

*Rucklidge, J.J. & Tannock, R. (2002). Neuropsychological profiles of adolescents with ADHD: Effects of reading difficulties and gender. *Journal of Child Psychology and Psychiatry, 43*, 988–1003.

*Scheres, A., Oosterlaan, J., Geurts, H.M., Morein-Zamir, S., Meiran, N., Schut, H., Vlasveld, L., & Sergeant, J.A. (in press). Executive functioning in ADHD: Primarily and inhibition deficit? *Archives of Clinical Neuropsychology*.

Schiller, P.H., (1966). Developmental study of color-word interference. *Journal of Experimental Psychology, 72*, 105–108.

*Schmitz, M., Cadore, L., Paczko, M., Kipper, L., Chaves, M., Rohde, L.A., Moura, C., & Knijnik, M. (2002). Neuropsychological performance in DSM-IV ADHD subtypes: An exploratory study with untreated adolescents. *Canadian Journal of Psychiatry, 47*, 863–869.

Seidman, L.J., Biederman, J., Faraone, S.V., Milberger, S., Norman, D., Seiverd, K., Benedict, K., Guite, J., Mick, E., & Kiely, K. (1995). Effects of family history and comorbidity on the neuropsychological performance of children with ADHD: Preliminary findings. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*, 1015–1024.

*Seidman, L.J., Biederman, J., Faraone, S.V., Weber, W., Mennin, D., & Jones, J. (1997). A pilot study of neuropsychological function in girls with ADHD. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 366–373.

Seidman, L.J., Biederman, J., Faraone, S.V., Weber, W., & Ouellette, C. (1997). Toward defining a neuropsychology of attention deficit-hyperactivity disorder: Performance of children and adolescents from a large clinically referred sample. *Journal of Consulting and Clinical Psychology, 65,* 150–160.

*Seidman, L.J., Biederman, J., Monuteaux, M.C., Weber, W., & Faraone, S.V. (2000). Neuropsychological functioning in nonreferred siblings of children with attention deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *109*, 252–265.

*Seidman, L.J., Biederman, J., Monuteaux, M.C., Doyle, A.E, & Faraone, S.V. (2001). Learning disabilities and executive dysfunction in boys with attention-deficit/hyperactivity disorder. *Neuropsychology*, *15*, 544–556.

Semrud-Clikeman, M., Biederman, J., Sprich-Buckminster, S., Lehman, B.K., Faraone, S.V., & Norman, D. (1992). Comorbidity between ADDH and learning disability: A review and report in a clinically referred sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 439–448.

*Semrud-Clikeman, M., Steingard, R.J., Filipek, P., Biederman, J., Bekken, K., & Renshaw, P.F. (2000). Using MRI to examine brain–behavior relationships in males with attention deficit disorder with hyperactivity. *Journal of the American Academy of Child and Adolescent Psychiatry*, *39*, 477–484.

Sergeant, J. (1988). From DSM-III attentional deficit disorder to functional defects. In L.M. Bloomingdale & J. Sergeant (Eds.), *Attention deficit disorder, criteria, cognition, intervention: A book supplement to the Journal of Child Psychology and Psychiatry, number 5* (pp. 183–198). Oxford: Pergamon Press.

Sergeant, J.A., & Van der Meere, J.J. (1990). Converging approaches on localizing the hyperactivity deficit. In B.B. Lahey & A.E. Kazdin (Eds.), *Advances in clinical child psychology, 13* (pp). 207–245. New York: Plenum Press.

Sergeant, J.A. & Van der Meere, J.J. (1991). Ritalin effects and information processing in hyperactivity. In L.L. Greenhill & B.B. Osman (Eds.), *Ritalin: Theory and patient management* (pp. 1–13). New York: Mary Ann Liebert.

Sergeant, J.A., Oosterlaan, J., & Van der Meere, J.J. (1999). Information processing and energetic factors in attention-deficit/hyperactivity disorder. In H.C. Quay & A. Hogan (Eds.), *Handbook of disruptive behavior disorders* (pp. 75–104). New York: Plenum Press.

Sergeant, J.A., Geurts, H., & Oosterlaan, J. (2002). How specific is a deficit of executive functioning for Attention-Deficit/Hyperactivity Disorder? *Behavioural Brain Research*, *130,* 3–28.

*Spalletta, G., Pasini, A., Pau, F., Guido, G., Menghini, L., & Caltagirone, C. (2001). Prefrontal blood flow dysregulation in drug naive ADHD children without structural abnormalities. *Journal of Neural Transmission*, *108*, 1203–1216.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd edn). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–662.

Tannock, R., Martinussen, R., & Frijters, J. (2000). Naming speed performance and stimulant effects indicate effortful, semantic processing deficits in attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology*, *28*, 237–252.

Thompson, S.G. & Higgins, J.P.T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21,* 1559–1573.

Ukai, S., Shinosaki, K., Ishii, R., Ogawa, A., Mizuno-Matsumoto, Y., Inouye, T., Hirabuki, N., Yoshimine, T., Robinson S.E., & Takeda, M. (2002). Parallel distributed processing neuroimaging in the Stroop task using spatially filtered magnetoencephalography analysis. *Neuroscience Letters*, *334*, 9–12.

Van der Meere, J.J. (1996). The role of attention. In S.T. Sandberg (Ed.), *Monographs on child and adolescent psychiatry: Hyperactivity disorders of childhood* (pp. 109–146). New York: Cambridge University Press.

West, R., & Alain, C. (1999). Event-related neural activity associated with the Stroop task. *Cognitive Brain Research*, *8*, 157–164.

*Willcutt, E.G., Pennington, B.F., Boada, R., Ogline, J.S., Tunick, R.A., Chabildas, N.A., & Olson, R.K. (2001). A comparison of the cognitive deficits in reading disability and attention-deficit/hyperactivity disorder. *Journal of Abnormal Psychology*, *110*, 157–172.

Zysset, S., Müller, K., Lohmann, G., & von Cramon, D.Y. (2001). Color-word matching Stroop task: Separating interference and response conflict. *Neuroimage*, *13*, 29–36.