Neuropsychology
2007, Vol. 21, No. 2, 251–262

0894-4105/07/$12.00 DOI: 10.1037/0894-4105.21.2.251

Stroop Interference and Attention-Deficit/Hyperactivity Disorder: A Review and Meta-Analysis

Marieke M. Lansbergen and J. Leon Kenemans
Utrecht University

Herman van Engeland
University Medical Center Utrecht

Previous reviews and meta-analyses that addressed abnormal Stroop interference in attention-deficit/hyperactivity disorder (ADHD) yielded mixed results. The authors of the present study argue that the inconsistencies may reflect the problematic nature of 2 frequently used methods to quantify Stroop interference—the difference score and Golden's method (C. J. Golden, 1978). Golden's method correction for base-word reading is inadequate, and the difference score is sensitive to the nature of the outcome variable. The latter can be remedied with a ratio score. Contrasting previous meta-analyses, this meta-analysis covers all age groups and all Stroop test variants, and it excludes studies using the Golden quantification method. Mean effect sizes for interference in ADHD as quantified by difference scores relative to control scores were 0.24 across all studies but 1.11 for time-per-item studies; outcome variable was a significant moderator variable, reflecting the sensitivity of the difference score to this variable. Consistency analysis of ratio scores across 19 studies reveals more interference for the ADHD groups relative to the control groups. It is concluded that interference control is consistently compromised in individuals with ADHD.

Keywords: meta-analysis, Stroop interference, attention-deficit/hyperactivity disorder, Stroop color and word task

Supplemental materials: http://dx.doi.org/10.1037/0894-4105.21.2.251.supp

Attention-deficit/hyperactivity disorder (ADHD) is defined by inappropriate degrees of inattention, hyperactivity, and impulsivity. Barkley (1997) suggested that behavioral disinhibition is the primary deficit in ADHD (predominantly hyperactive-impulsive type and combined type) and distinguished three interrelated forms of behavioral inhibition. The first form is inhibition of an initial *prepotent response* to an event, that is, a response for which immediate reinforcement is available or with which reinforcement has been previously associated. The second form is *stopping an ongoing response* or response pattern. This causes a delay in the decision to respond or to continue responding. The third form is *interference control*, defined as protecting the period of delay and self-directed responses that occur within it from disruption by competing events and responses.

Continuous performance tests (CPTs) and stop-signal tasks can be used to assess both inhibition of prepotent responses and stopping of ongoing responses. Several reviews and meta-analyses (Corkum & Siegel, 1993; Lijffijt, Kenemans, Verbaten, & van Engeland, 2005; Losier, McGrath, & Klein, 1996; Oosterlaan, Logan, & Sergeant, 1998; Riccio, Reynolds, & Lowe, 2001) demonstrated that individuals with ADHD perform worse on a CPT and stop-signal task as compared with normal control participants, reflecting poor response inhibition in ADHD patients. The Stroop test (Stroop, 1935) can be used to quantify interference control. Inconsistent findings have been reported regarding deficient interference control in ADHD.

Stroop Test and ADHD

Three versions of the Stroop test have been used in ADHD research: the standard Stroop test, the Golden Stroop test (Golden, 1978), and computerized Stroop tests. In the standard Stroop test, participants are required to read 100 black color words (word card), to name the colors of 100 solid squares or XXXXs (color card), and to name the colors of 100 incongruent color words (color–word card) as fast as possible. Here, total time needed for each card is the outcome variable. In the Golden Stroop test, participants are required to name as many items as they can in 45 s for each card. The outcome variables are the number of items completed for the word card (W = raw word score), the color card (C = raw color score), and the color–word card (CW = raw color–word score), respectively. In computerized Stroop tests, congruent color words, neutral non-color words, as well as incongruent color words are presented one at a time, and reaction time for each item can be recorded.

In the color–word part of the Stroop test, the automatic response of word reading has to be suppressed and prevented from interfering with naming the color of the color word. *Stroop interfer-*

Marieke M. Lansbergen and J. Leon Kenemans, Departments of Experimental Psychology and Psychopharmacology, Utrecht University, Utrecht, The Netherlands; Herman van Engeland, Department of Child and Adolescent Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands.

This study was supported by the Dutch Organization for Scientific Research (NWO). We thank Maarten van der Smagt and Michiel Stapelkamp for the fruitful discussions on quantifying Stroop interference, and we thank Marijn Lijffijt for sharing his knowledge on meta-analysis. We also thank M. C. Goldberg, J. J. Rucklidge, and A. Scheres for providing us their data, which were necessary to conduct the analyses.

Correspondence concerning this article should be addressed to Marieke M. Lansbergen, Department of Experimental Psychology, Utrecht University, Heidelberglaan 2, 3584 CS Utrecht, The Netherlands. E-mail: M.Lansbergen@fss.uu.nl

ence, then, is the extent of delay in naming the color of an incongruent color word relative to naming the color of a congruent color word or of a neutral non-color word.

Recently, four meta-analyses have reviewed the literature on the Stroop test and ADHD. Whereas Homack and Riccio (2004) reported an effect size of 0.75 for Stroop interference, two meta-analyses reported relatively small effect sizes for this variable (Hervey, Epstein, & Curry, 2004, effect size = 0.15; Van Mourik, Oosterlaan, & Sergeant, 2005, effect size = 0.35). Frazier, Demaree, & Youngstrom (2004) found that the Stroop variable (not further specified) was significantly sensitive to ADHD (effect size = 0.56). It has been claimed that the meta-analyses cover the complete literature on Stroop and ADHD (Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005), but there are still some important unresolved issues. First, previous meta-analyses have thus far addressed specifically either children (Homack & Riccio, 2004; Van Mourik et al., 2005) or adults (Hervey et al., 2004) with ADHD. Second, previous meta-analyses excluded computerized Stroop tests (Hervey et al., 2004; Homack & Riccio, 2004; Van Mourik et al., 2005). Frazier et al. (2004) did not report on age effects, or on which Stroop test was used, or on which studies were actually used. Furthermore, and of most noted importance, previous meta-analyses included studies that used different and incorrect quantification methods for Stroop interference. In comparing patient and control groups on interference, it is of utmost importance that valid and sensitive quantifications of Stroop interference are used. In the review of Frazier et al. (2004), the authors mention "number of trials completed or time required to complete the task" (548) but not at all in a relative sense (e.g., CW − C), which is a minimum requirement for valid quantification. Hervey et al. (2004) included only studies that used Golden's interference score, and two meta-analyses (Homack & Riccio, 2004; Van Mourik et al., 2005) included studies that used the difference score or Golden's interference score. Here, we argue that two frequently used quantification methods, that is, difference score and Golden's interference score, are problematic and that a ratio score may be preferred.

Quantifying Stroop Interference

The standard method to assess Stroop interference is to calculate the *difference score* (I_D), that is, the difference between the color (C) and color–word (CW) scores, with $I_D = C - CW$ when the number of items has been scored and $I_D = CW - C$ for time per item. A lower difference score means less interference from incongruent words when naming the colors in the color–word condition. Golden (1978) proposed an alternative method to quantify interference in the Stroop test. The first step is to calculate a predicted color–word score (P_{CW}) on the basis of the word (W) and color (C) scores. The predicted numbers of items named in 45 s in the color–word condition (P_{CW}) is calculated as

$$P_{CW} = 45/\{[(45 \times W) + (45 \times C)]/(W \times C)\} = (W \times C)/(W + C). \quad (1)$$

Given P_{CW}, Golden's interference score (I_G) is calculated by subtracting this score from the actual color–word score ($I_G = CW - P_{CW}$). Golden and Freshwater (2002) argued that the time to name a color–word item is equal to the time needed to suppress the reading of a word plus the time to identify a color. In case of impaired interference control, reading the word in the color–word condition will actively interfere with naming the color and switching from one to the other will go slowly, resulting in a smaller color–word score relative to the predicted score and thus a negative Golden's interference score.

Validly assessing interference in the Stroop test depends on the adequate use of an appropriate control condition. Performance in the color–word condition can be assumed to depend both on speed of naming colors and on speed of word reading per se. A general slowing of speed of word reading and color naming can bias the interference estimate. For instance, given a constant ability to suppress irrelevant response tendencies, individuals who are relatively slow in color naming have more problems with the ability to suppress the word reading, resulting in an overestimation of the normal "true" interference level. This is correctly counteracted by applying I_G or I_D. Regarding word reading, individuals who are relatively slow readers have less problems with the ability to suppress word reading, resulting in an underestimation of the normal true interference level (given a constant ability to suppress irrelevant response tendencies). As illustrated in Equation 1, the way in which I_G corrects for base-word reading has exactly the same effect as the correction for base-color naming, indicating a further underestimation of the normal true interference level. Thus, although I_G and I_D adequately correct for base-color naming, the way in which I_G corrects for base-word reading is not valid.

So, I_G suffers from computational problems. These problems do not hold for I_D, but we note a perhaps more serious problem that might prompt researchers to refrain from using I_D: its sensitivity to transformations of the raw data. Table 1 shows hypothetical data from a Golden Stroop test as acquired in a pathological sample and in healthy control participants, both as the number of items in 45 s in the color and color–word tasks and after transformation into time for each item (e.g., $45/50 \times 1000 = 900$ ms). Difference scores were calculated by subtracting numbers of items in the color–word condition from those in the color condition ($I_{DN} = C - CW$) and by a reversed subtraction for time per item ($I_{DT} = CW - C$). As can be seen in Table 1, I_{DN} may not differ between the two hypothetical groups (the difference between the color and color–word condition is 25 items in 45 s in both groups), whereas differences in $I_{DT,}$ which was based on exactly the same data (i.e., CW − C after transformation into time for each item), are sizable (900 ms per item in patients vs. 150 ms per item in control participants).

A rather simple alternative for the difference score is the ratio estimated, or ratio score (I_R), with $I_R = CW/C$ when the number

Table 1
Fictive Data of Golden Stroop Test

Score	Number of items in 45 s		Time for each item (in ms)	
	Patient	Control	Patient	Control
Color score (C)	50	100	900	450
Color–word score (CW)	25	75	1,800	600
Difference score (I_D)	25	25	900	150
Ratio score (I_R)	0.50	0.75	0.50	0.75

of items has been scored and $I_R = C/CW$ for time per item. A higher ratio score means less interference from incongruent words when naming the colors in the color–word condition. The great advantage of the ratio score is that it is not sensitive to transformations like from number of items per unit of time to time per item and vice versa. This is illustrated in Table 1: Either as number of items per unit or as time per item, ratios were 0.50 for patients and 0.75 for control participants. Note that the ratio score has in fact been applied occasionally in past research (Bondi et al., 2002; Saunders, 1980; Weir, Bruun, & Barber, 1997). The ratio score does not correct for base-word reading directly, but base-word reading can be entered as a covariate in the analysis of group effects on the ratio score. Moreover, ratio scores, much more than the difference score, correct for the extent that higher interference reflects general slowing, and this slowing is often more pronounced in patients. Therefore, it reveals only abnormalities in interference that exceed the contribution of general slowing. This can also be viewed as a limitation that may not always be desirable (e.g., Henik & Salo, 2004): If general slowing does not affect the extent of interference, the "true" interference level may be underestimated when using ratio scores.[1]

An alternative for the ratio score to quantify Stroop interference would be analysis of covariance (ANCOVA), with the color score as a covariate in the analysis of group effects on the color–word score (see Capitani, Laiacona, Barbarotto, & Cossa, 1999). It can be argued that ANCOVA, like the ratio method, corrects for general slowing to a considerable extent. By using a regression coefficient (b) derived from a normative sample, individual interference scores can be estimated as $CW - b \times C$. However, as the use of subtraction already suggests, ANCOVA is also sensitive to data transformation (this can be readily demonstrated by conducting ANCOVAs on time-per-item data as well as on the transforms to items-per-time scores; the latter and the former ANCOVAs will yield different F and p values for the group effect).

Aims of the Current Review

Previous meta-analyses have addressed specifically either children or adults with ADHD, excluded computerized Stroop studies, and included studies using invalid quantification methods to assess Stroop interference. This may have resulted in different interpretations of the data. Hence, an exhaustive analysis of the existing literature dealing with these issues is still wanted to more validly determine whether patients with ADHD are impaired in Stroop interference control. We performed a meta-analysis (Lipsey & Wilson, 2001) on Stroop difference scores (I_D) that covers both child and adult studies as well as all Stroop versions but excludes quantification by Golden interference scores. A disadvantage of meta-analysis is that ratio scores cannot be used as a method to evaluate group differences in interference because standard deviations (*SD*s) of ratio scores are usually not provided and because effect sizes for ratio scores cannot be derived from means and *SD*s for each of the two Stroop conditions. To investigate the difference in ratio interference scores between individuals with and without ADHD, we applied a quantitative test for consistency across studies (see Frazier et al., 2004, for a similar approach in relation to full scale IQ). Additionally, we examined what the use of different methods to quantify interference in the Stroop implies for the interpretation of real data by investigating the consistency across studies separately for each quantification method (i.e., I_R, I_{DT}, and I_{DN}) and by exploring the consistency across methods.

Method

Literature Search

Studies were gathered by searching the PsychInfo and PubMed databases for the terms *Stroop* and *ADHD* and by checking references in review articles (Barkley, Grodzinsky, & DuPaul, 1992; Frazier et al., 2004; Hervey et al., 2004; Homack & Riccio, 2004; Nigg, 2001; Sergeant, Geurts, & Oosterlaan, 2002; Van Mourik et al., 2005) and in the retrieved articles. The present search was completed in January 2006. Inclusion criteria were (a) inclusion of a control group, (b) *Diagnostic and Statistical Manual of Mental Disorders (DSM)-III*, *DSM-III-R*, or *DSM-IV* diagnosis of ADHD (American Psychiatric Association, 1980, 1987, 1994), (c) the Stroop test containing a color–word condition and an appropriate control condition (e.g., a color condition), and (d) reporting means and standard deviations of the scores in the incongruent-color–word and control-color-naming conditions.

A total of 48 studies was found that compared performance in the Stroop test between patients who have ADHD (or are hyperactive) and normal control participants. In 39 studies, the diagnosis ADHD was based on *DSM-III*, *DSM-III-R*, or *DSM-IV* criteria. Of the 39 studies, 14 studies were excluded for several reasons: The Stroop test did not include a control condition, color or color–word scores were not provided, or the data were ambiguous. Another 9 studies were excluded because they provided only Golden's age-adjusted *t* scores (Golden, 1978). Attempts were made to locate authors of studies that did not report color or color–word scores and studies that reported only Golden's interference scores, leading to additional data on 3 studies (Goldberg et al., 2005; Rucklidge & Tannock, 2002; Scheres et al., 2004). Eventually, 6 studies that provided time for each card or reaction per item and 12 studies that used the Golden Stroop test, in which items completed in 45 s are the variable of interest, were included in the meta-analysis. The consistency analysis included all studies from the meta-analysis plus 1 study (Carlson, Lahey, & Neeper, 1986); the latter study did not provide standard deviations and therefore could not be included in the meta-analysis. Demographics and the characteristics of each of the studies included are shown in Table 2.

Meta-Analysis

Mean and standard deviation for the interference score for each study were calculated with the means and standard deviations for the incongruent-color–word and the control-color-naming condi-

[1] The hypothesis of general slowness modulating the Stroop interference effect was tested with a database on (computerized) Stroop interference as assessed in 32 healthy adults (Lansbergen, Kenemans, & Van Hell, in press). For each participant, RT per item was computed separately for each of four RT quantiles (Rouder & Speckman, 2004). An analysis of variance with factors quantile (4) and Stroop condition (2; congruent vs. incongruent) revealed significant increase in Stroop interference with increasing slowness, indicating that the size of Stroop interference reflects the general efficiency of information processing. So, in this experimental group, general slowness modulates the Stroop effect, and to correct for it, ratio scores should be used.

Table 2
Demographics and Test Characteristics of the Included Studies in the Review

Author	*n*		Age range (in years)	Mean age (in years)		% male/% female		Computer/card	Vocal/manual	Blocked/mixed	Variable	Significance
	ADHD	Control		ADHD	Control	ADHD	Control					
Bush et al., 1999*	8	8	22–47	36.6	37.3	62.5/37.5	62.5/37.5	comp	manual	blocked	RT (in ms) per item	unknown
Carlson, Lahey, & Neeper, 1986	35	16	7–11	10.4	9.6	74.3/25.7	81.3/18.7	card	vocal	blocked	Time (in s) for ? items	*ns*
Carter, Krener, Chaderjian, Northcutt, & Wolfe, 1995	19	19	9–12	10.6	10.6	73.7/26.3	78.9/21.1	comp	vocal	mixed	RT (in ms) per item	*sign*
Gaultney, Kipp, Weinstein, & McNeill, 1999*	29	29	8–15	11.0	10.7	72.4/27.6	55.2/44.8	card	vocal	blocked	Time for 30 items	*ns*
Goldberg et al., 2005	20	19	8–12	9.8	10.0	90.0/10.0	68.4/31.6	card	vocal	blocked	No. of items in 45 s	*ns*
Golden & Golden, 2002*	43	43	6–15	9.8	9.9	81.4/18.6	81.4/18.6	card	vocal	blocked	No. of items in 45 s	*ns*
Houghton et al., 1999*	94	28	6–13	10.0	10.2	58.5/41.5	53.6/46.4	card	vocal	blocked	No. of items in 45 s	*ns*
Johnson et al., 2001*	56	38	20–63	33.3	40.8	71.4/28.6	63.2/36.8	card	vocal	blocked	No. of items in 45 s	unknown
Lufi, Cohen, & Parish-Plass, 1990*	29	20	9–16	12.7	12.7	100.0/0	100.0/0	card	vocal	blocked	No. of items in 45 s	*sign*
Nigg, Blaskey, Huang-Pollock, & Rappley, 2002*	64	41	7–12	10.0	10.1	73.4/26.6	58.5/41.5	card	vocal	blocked	No. of items in 45 s	*ns*
Nigg, Butler, Huang-Pollock & Henderson, 2002*	22	21	> 18	23.1	21.6	45.5/54.5	42.9/57.1	comp	vocal	mixed	RT (in ms) per item	*ns*
Rapport, Van Voorhis, Tzelapis, & Friedman, 2001*	35	32	> 18	32.9	33.2	68.6/31.4	59.4/40.6	card	vocal	blocked	No. of items in 45 s	*sign*
Riordan et al., 1999*	21	15	> 18	31.8	36.5	81.0/19.0	80.0/20.0	card	vocal	blocked	No. of items in 45 s	unknown
Rucklidge & Tannock, 2002*	35	37	13–16	15.2	15.0	57.1/42.9	48.6/51.4	card	vocal	blocked	No. of items in 45 s	*ns*
Scheres et al., 2004	18	20	8–12	9.3	9.9	100.0/0	100.0/0	card	vocal	blocked	Time for 100 items	*sign*
Seidman, Biederman, Faraone, Weber, & Ouellette, 1997*	118	99	9–22	14.5	15.3	100.0/0	100.0/0	card	vocal	blocked	No. of items in 45 s	*sign*
Seidman, Biederman, Weber, Hatch, & Faraone, 1998*	64	73	19–59	36.3	40.1	51.6/48.4	45.2/54.8	card	vocal	blocked	No. of items in 45 s	*ns*
Silverstein et al., 1995	17	17	> 18	36.0	31.0	52.9/47.1	47.1/52.9	card	vocal	blocked	Time (in s) for ? items	unknown
Walker, Shores, Trollor, Lee, & Sachdev, 2000*	30	30	17–50	25.8	25.8	83.3/16.7	66.7/33.3	card	vocal	blocked	No. of items in 45 s	*ns*

Note. Asterisks indicate studies included in the meta-analysis but cited only in the Web supplement. *sign* = significant differences in interference between attention-deficit/hyperactivity disorder (ADHD) patients and healthy control participants; *ns* = no significant difference in interference between participants with ADHD and control participants; RT = reaction time.

tions. Mean interference scores were computed by subtracting the control-condition score from the incongruent-color–word score (when the dependent variable of interest was time for each card or reaction time [RT] per item) or by reverse subtraction (when the dependent variable was number of items in 45 s) separately for the ADHD and control group. Standard deviations of the interference scores were calculated with the following formula:

$$SD_{\mathrm{int}} = \sqrt{2\times\{[(SD_{\mathrm{cont}})^2+(SD_{\mathrm{inc}})^2]/2\} \times (1-r)}, \qquad (2)$$

where SD_{cont} is the pooled standard deviation (i.e., the square root of the weighted average of the two variances) across the ADHD group and the control group for each control-condition score, SD_{inc} is the pooled standard deviation across the ADHD group and the control group for each incongruent-condition score, and r is the Pearson coefficient of correlation ($r = .954$) between performance in the control condition and the performance in the incongruent condition as derived from an extant data set from adult individuals (Kenemans, Wieleman, Zeegers, & Verbaten, 1999). It was assumed that the correlation between the control and incongruent condition in this data set approaches the correlation between these two conditions in the studies used in this meta-analysis.

The meta-analysis was conducted according to Lipsey and Wilson (2001). The standardized mean difference effect size was computed for each study as the mean interference score for the ADHD group minus the mean interference score for the control group divided by the pooled standard deviation of the interference score. This effect size score has been shown to be upwardly biased when based on a small sample size. Therefore, a correction was performed using Hedges' formula (Lipsey & Wilson, 2001). The effect size can be interpreted relative to three criterion levels: small (0.2), medium (0.5), and large (0.8) (Cohen, 1988).

Statistical analyses were performed to test whether the mean effect size did significantly differ from 0 (for statistical details, see Lipsey & Wilson, 2001). Effects were considered significant when alpha was lower than 5%. Homogeneity analyses were performed to test whether the various effect sizes all estimate the same population effect size. The traditional Q statistic may be considered insufficiently valid because the outcome depends on the number of studies that are included. An alternative measure to quantify heterogeneity is I^2, which is independent from the number of studies (Higgins & Thompson, 2002). I^2 estimates the percentage of total variation among studies due to heterogeneity rather than to sample errors. I^2 can be interpreted relative to three criterion levels: low (25%), moderate (50%), and high (75%). When effect size distribution was moderately heterogeneous ($I^2 >$ 50%), moderator variables were evaluated to assess whether they could explain the excess in effect size variability. The impact of a moderator variable was tested by applying a Q statistic. When within-group heterogeneity was still moderate ($I^2 > 50\%$), the random effects model effect sizes instead of the fixed effects model effect sizes were applied for each category within the moderator variable.

Moderator Variables

The variables that were evaluated as potential moderators were age (< 17 years for children vs. ≥ 17 years for adults) and outcome (time/RT vs. number of items in 45 s). Of the 18 studies included in the meta-analysis, 9 included children and 8 included adults (1 study included both age groups). In 12 studies, the dependent variable was the number of items in 45 s (here labeled as item studies); in 6 studies, it was reaction per item or time on each card (here labeled as time studies). In addition to the moderator variables of age and outcome, other causes may explain the heterogeneity across studies such as variation in Stroop test parameters (e.g., mixed vs. blocked design, computerized vs. noncomputerized Stroop tests, vocal vs. manual responses) and patient samples. Unfortunately, too few studies used mixed design, computerized Stroop tests, or manual responses to examine these possible moderator variables. Also, most studies did not report the subtype of the patients with ADHD who were included.

Consistency Analysis

The consistency analysis was conducted on ratio interference scores calculated for 12 studies that measured number of items in 45 s (item studies) and 7 studies that measured time or RT (time studies). Of these 7 time studies, 3 measured RT per item and 4 provided time for each card. Consistency for ratio interference score (I_R) was assessed by conducting a paired-samples t test between pairs of ratio scores for participants with ADHD and ratio scores for control participants.

Additionally, 17 ADHD studies were used to examine what the use of different methods, that is, I_R, I_{DT}, and I_{DN}, to quantify interference in the Stroop implies for the interpretation of real data (2 time studies were excluded because they did not describe the Stroop test in sufficient detail to transform the data into number of items in 45 s: Carlson et al., 1986, and Silverstein, Como, Palumbo, West, & Osborn, 1995). I_{DN} is defined as the difference between the color and color–word score based on number of items in 45 s, I_{DT} is the difference between the color and color–word score based on reaction time for each item. For item studies and time-for-each-card studies, I_{DT} was calculated after transforming color and color–word scores into time for each item ($I_{DT} = CW - C$). For time studies, I_{DN} was calculated after transforming color and color–word scores into number of items in 45 s ($I_{DN} = C - CW$). Consistency across studies for each interference score (I_R, I_{DT}, and I_{DN}) was assessed by paired-sample t tests. Furthermore, consistency was compared across methods by transforming I_R, I_{DT}, and I_{DN} to Z scores for each group separately. These Z scores were entered in a multivariate analysis with group (ADHD vs. control) and method (I_R, I_{DT}, and I_{DN}) as within-subjects factors.

Results

Table 3 shows the mean values and standard deviations per condition and per group for all studies included in this review. Table 3 also presents the calculated difference interference scores, the calculated standard deviations of the interference score, and the effect sizes for the 18 studies used in the meta-analysis. Note that 12 of the 18 effect sizes were positive, indicating that healthy control participants are more resistant to interference as compared with patients with ADHD. Supplementary materials and a complete list of the articles included in this meta-analysis can be found on the Web.

Individual raw effect sizes ranged from −0.95 to 2.31, yielding a mean effect size of 0.28 ($SD = 0.91$) and median of 0.19. The

Table 3
Control-Color-Naming Condition, Incongruent-Color–Word Condition, Difference Interference Score, and Effect Size for Both Attention-Deficit/Hyperactivity Disorder (ADHD) and Control Groups

	ADHD			Control				
Author	Mean (*SD*) control condition	Mean (*SD*) incongruent condition	Mean difference score	Mean (*SD*) control condition	Mean (*SD*) incongruent condition	Mean difference score	*SD* interference score	Effect size
Bush et al., 1999*	748.0 (104.0)	801.0 (135.0)	53.0	691.0 (42.0)	720.0 (51.0)	29.0	27.7	0.82
Carlson, Lahey, & Neeper, 1986	23.23	41.31		17.87	36.67			
Carter, Krener, Chaderjian, Northcutt, & Wolfe, 1995	1,025.3 (166.1)	1,172.1 (241.8)	146.8	939.7 (282.8)	1,013.1 (301.9)	73.4	76.9	0.93
Gaultney, Kipp, Weinstein, & McNeill, 1999*	16.3 (5.3)	29.5 (11.3)	13.3	16.8 (4.5)	28.7 (7.7)	11.9	2.3	0.61
Goldberg et al., 2005	48.45 (9.59)	27.18 (6.47)	21.3	51.42 (9.09)	27.74 (7.29)	23.7	2.5	−0.95
Golden & Golden, 2002*	49.3 (14.6)	26.5 (9.7)	22.8	54.3 (13.9)	32.8 (7.0)	21.5	3.6	0.36
Houghton et al., 1999*	46.6 (11.5)	25.2 (8.9)	21.4	52.4 (10.8)	28.8 (8.7)	23.6	3.1	−0.70
Johnson et al., 2001*	74.6 (11.4)	42.3 (9.2)	32.3	77.5 (10.5)	46.0 (14.0)	31.5	3.4	0.22
Lufi, Cohen, & Parish-Plass, 1990*	55.9 (9.4)	28.3 (7.1)	27.6	63.25 (11.9)	37.3 (9.9)	26.0	2.9	0.54
Nigg, Blaskey, Huang-Pollock, & Rappley, 2002*	43.5 (9.9)	23.6 (6.9)	19.9	50.3 (8.9)	29.8 (9.3)	20.5	2.7	−0.22
Nigg, Butler, Huang-Pollock, & Henderson, 2002*	721.0 (145.0)	792.0 (179.0)	71.0	620.0 (93.0)	689.0 (101.0)	69.0	40.9	0.05
Rapport, Van Voorhis, Tzelpis, & Friedman, 2001*	70.3 (15.6)	41.0 (10.5)	29.3	80.3 (10.4)	48.5 (6.8)	31.8	3.5	−0.72
Riordan et al., 1999*	65.4 (11.7)	36.3 (10.0)	29.1	79.3 (15.2)	49 (17.9)	30.3	4.1	−0.28
Rucklidge & Tannock, 2002	69.83 (3.38)	45.43 (10.45)	24.4	79.70 (12.10)	52.35 (11.91)	27.4	3.3	−0.89
Scheres et al., 2004	109.5 (26.08)	207.33 (63.94)	97.8	86.95 (28.77)	151.45 (67.65)	64.5	15.3	2.13
Seidman, Biederman, Faraone, Weber, & Oullette, 1997*	58.3 (13.9)	32.6 (11.3)	25.7	63.4 (13.6)	38.3 (11.1)	25.1	3.8	0.16
Seidman, Biederman, Weber, Hatch, & Faraone, 1998*	72.7 (12.8)	40.9 (11.6)	31.8	72.1 (11.0)	41.8 (9.3)	30.3	3.4	0.44
Silverstein, Como, Palumbo, West, & Osborn, 1995	60.0 (12.9)	119.0 (25.3)	59.0	52.0 (6.6)	99.0 (16.4)	47.0	5.1	2.31
Walker, Shores, Troller, Lee, & Sachdev, 2000*	60.5 (8.6)	33.4 (6.5)	27.1	72.7 (12.5)	46.0 (8.1)	26.7	2.8	0.14

Note. Asterisks indicate studies used in the meta-analysis but cited only in the Web supplement.

average weighted effect size was 0.24 (95% confidence interval, .12, .36; combined $Z = 3.85$; $p < .001$). Although the effect size is significantly larger than zero, according to the criteria of Cohen (1988), it is small. Homogeneity analyses yielded a significant result, $Q(17) = 86.25$, $p < .001$ ($I^2 = 80.29\%$), indicating considerable variance in effect size across studies. Further analyses tested whether the two moderator variables of age and outcome differentiated systematically between studies with larger or smaller effect sizes. Age group did not explain the excess effect size variability (effect size = 0.32 for studies on children and 0.20 for studies on adults), but outcome variable did differentiate studies with larger and smaller effect sizes, $Q(1) = 28.16$, $p < .001$. However, the pooled (over item and time studies) within-groups variance still indicated heterogeneity, $Q(16) = 58.09$, $p < .001$ ($I^2 = 72.5\%$), reflecting that the variability in effect size was due to both systematic differences between studies and an additional random component. So, the random effects model effect sizes were adopted, revealing a significant effect size for time studies (1.11; 95% confidence interval .42, 1.80; combined $Z = 3.14$; $p = .002$), but a nonsignificant effect size for item studies (–0.007).

Further examination of the data revealed two outliers with extremely high effect sizes of 2.13 (Scheres et al., 2004) and 2.31 (Silverstein et al., 1995). Post hoc meta-analyses were performed without these studies and without the 2 studies with the highest negative effect sizes of −0.95 and −0.89 (Goldberg et al., 2005; Rucklidge & Tannock, 2002). The raw mean effect size was 0.17 ($SD = 0.51$) and the median was 0.19. The post hoc analysis yielded a significant weighted effect size of 0.17 (95% confidence interval, .05, .30; combined $Z = 2.76$; $p = .006$) as well as significant heterogeneity, $Q(13) = 34.39$, $p < .001$ ($I^2 = 62.2\%$). Again, age was not a significant moderator variable (effect size = 0.27 for studies on children and 0.11 for studies on adults), whereas outcome variable was a significant moderator variable, $Q(1) = 5.88$, $p = .015$. The pooled within-groups variance was still significantly heterogeneous, $Q(12) = 28.51$, $p = .005$ ($I^2 = 57.9\%$). The random effects model effect sizes yielded a mean effect size of 0.06 for the 10 item studies (95% confidence interval =.18, .30; combined $Z = 0.51$; $p = .612$) and of 0.55 for the 4 time studies (95% confidence interval, .16, .95; combined $Z = 2.76$; $p = .006$).

Consistency Analysis

Table 4 presents the calculated ratio score (I_R) per group for all studies used in the consistency analysis. Figure 1 presents the difference in mean ratio scores between the ADHD group and control group for item studies (number of items in 45 s, No. 1–No. 12) and for time studies (time or RT; Time 1–Time 7). As shown

Table 4
Calculated Interference Scores for Both ADHD and Control Groups and for Each Included Study

Author	Study (see figures)	I_{DN} (= C − CW)		I_{DT} (= CW − C)		I_R = (CW/C or C/CW)	
		ADHD	Control	ADHD	Control	ADHD	Control
Bush et al., 1999*	Time 1	3.98	2.62	53.00	29.00	0.93	0.96
Carlson, Lahey, & Neeper, 1986	Time 2					0.56	0.49
Carter, Krener, Chaderjian, Northcutt, & Wolfe, 1995	Time 3	5.50	3.47	146.80	73.40	0.87	0.93
Gaultney, Kipp, Weinstein, & McNeill et al., 1999*	Time 4	37.22	33.17	442.67	395.00	0.55	0.59
Goldberg et al., 2005	No. 1	21.27	23.68	726.84	747.06	0.56	0.54
Golden & Golden, 2002*	No. 2	22.8	21.5	785.33	543.22	0.54	0.60
Houghton et al., 1999*	No. 3	21.4	23.6	819.05	703.99	0.54	0.55
Johnson et al., 2001*	No. 4	32.3	31.5	459.86	398.03	0.57	0.59
Lufi, Cohen, & Parish-Plass, 1990*	No. 5	27.6	26.0	784.53	496.59	0.51	0.59
Nigg, Blaskey, Huang-Pollock, & Rappley, 2002*	No. 6	19.9	20.5	872.30	615.44	0.54	0.59
Nigg, Butler, Huang-Pollock, & Henderson et al., 2002*	Time 5	5.60	7.27	71.00	69.00	0.91	0.90
Rapport, Van Voorhis, Tzelpis, & Friedman et al., 2001*	No. 7	29.3	31.8	457.45	367.44	0.58	0.60
Riordan et al., 1999*	No. 8	29.1	30.3	551.13	350.69	0.56	0.62
Rucklidge & Tannock, 2002	No. 9	24.4	27.35	346.11	294.98	0.65	0.66
Scheres et al., 2004	Time 6	19.39	22.04	978.30	645.00	0.53	0.57
Seidman Biederman, Faraone, Weber, & Oullette, 1997*	No. 10	25.7	25.1	608.5	465.15	0.56	0.60
Seidman, Biederman, Weber, Hatch, & Faraone, 1998*	No. 11	31.8	30.3	481.26	452.43	0.56	0.58
Silverstein, Como, Palumbo, West, & Osborn, 1995	Time 7					0.50	0.53
Walker, Shores, Troller, Lee, & Sachdev, 2000*	No. 12	27.1	26.7	603.51	359.28	0.55	0.63

Note. Asterisks indicate studies used in the meta-analysis but cited only in the Web supplement. I_{DN} = calculated difference interference scores in number of items named in 45 s; I_{DT} = calculated difference interference scores in time (milliseconds) per item; I_R = calculated ratio interference scores; C = color score; CW = color–word score; ADHD = attention-deficit/hyperactivity disorder.

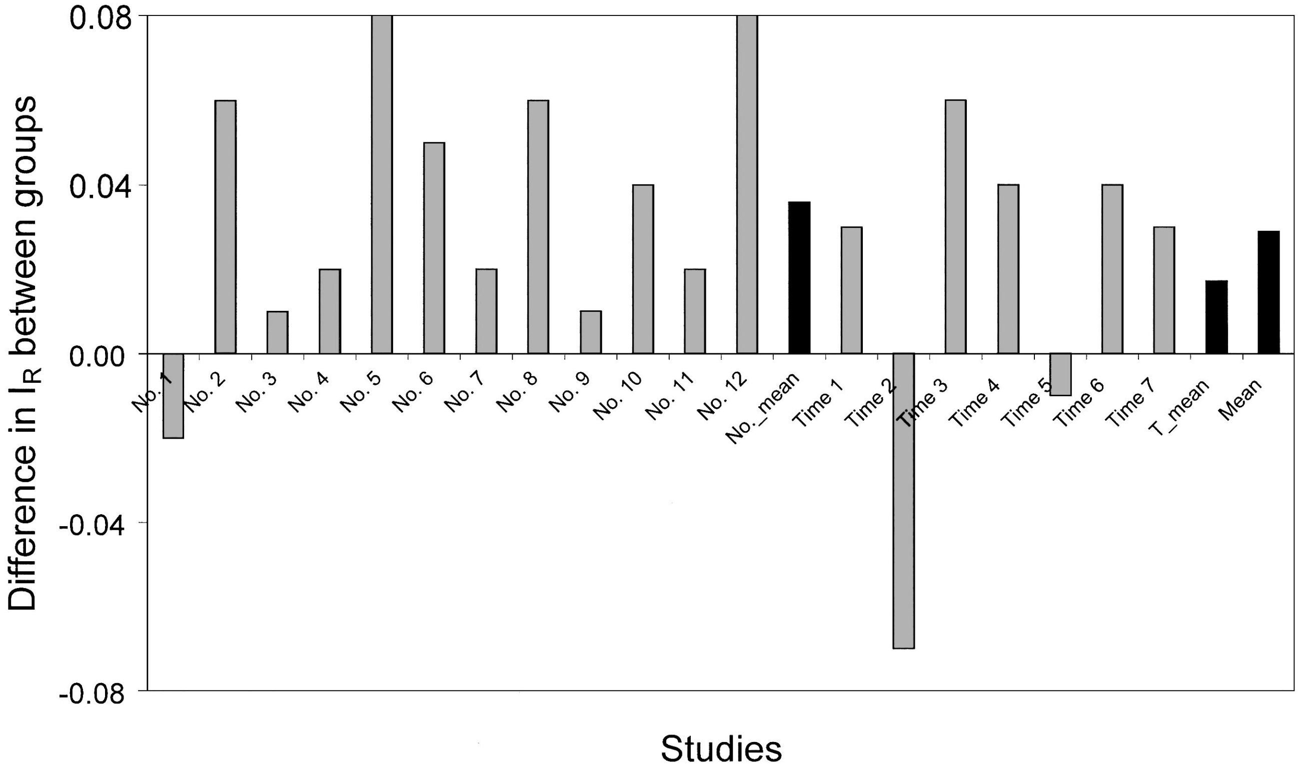

Figure 1. Differences between patients with attention-deficit/hyperactivity disorder and control participants in ratio scores (I_R) for 12 studies that measured number of items in 45 s (No. 1–No. 12) and for 7 studies that provided time or reaction time (Time 1–Time 7). No._mean = average difference across No. 1–No. 12; T_mean = average difference across Time 1–Time 7; Mean = average difference across all studies.

in Figure 1, 5 of 7 time studies and 11 of 12 item studies found larger ratio scores in control participants as compared with participants with ADHD, indicating more resistance to interference in healthy control participants. The difference between the ADHD group and the control group with respect to the mean ratio scores across studies was significant, $t(18) = 3.28$, $p = .004$; 0.61 and 0.64 for the ADHD group and the control group, respectively.

For the 12 item and 5 time studies (see Figures 1, 2, and 3: No. 1–No. 12; Time 1 and Time 3–Time 6; Carlson et al., 1986, and Silverstein et al., 1995, were excluded), comparative paired-samples *t* tests were conducted for I_{DN}, I_{DT}, and I_R. Table 4 presents mean calculated Stroop interference scores for patients with ADHD and control participants separately for each study. Whereas I_{DN} is greater in ADHD relative to control in 8 of 17 studies (Figure 2), group differences in I_R and I_{DT} are mostly consistent across studies (Figures 1 and 3). Consistency analyses revealed significant consistency across studies for I_{DT} and I_R, $t(16) = 4.78$, $p < .01$; $t(16) = -4.82$, $p < .01$, respectively, but not for I_{DN}, $t(16) = -0.30$, $p = .77$. Note that calculated ratio scores are identical for both scoring methods but confirm the time-per-item results.

Furthermore, consistency across measures was compared by transforming I_R, I_{DT}, and I_{DN} to *Z* scores for each group separately. The *F* test for the interaction between group (ADHD vs. control) and method (I_R, I_{DT}, I_{DN}) was indeed significant, $F(2, 32) = 13.55$, $p < .001$, indicating a difference in sensitivity between the three interference scores.

Finally, 13 of 19 included studies reported word (W) scores, and consistency analysis indicated that the hypothesis of a difference between groups can be accepted with more than 95% certainty, $t(12) = 5.19$, $p < .001$, with patients being slower readers than control participants.

Discussion

This review addresses several shortcomings of previous meta-analyses and reviews. Previous meta-analyses have addressed specifically either children or adults with ADHD. Moreover, they excluded computerized Stroop studies. The present meta-analysis took into account the literature on the difference in interference between individuals with and without ADHD, covering both age groups as well as all versions of the Stroop test, including computerized Stroop tests. An even more important problem of previous meta-analyses involves the inclusion of studies that used invalid quantification methods. In the majority of studies on Stroop interference in ADHD, interference was quantified as either the difference score or Golden's interference score, and this has yielded an inconsistent pattern of results with regard to the question of abnormal interference in ADHD. We argued that Golden's method and the difference score are problematic, and a ratio score may be preferred. Golden's method is not adequate with respect to controlling the contribution of base speed of word reading, resulting in underestimation of interference. The difference score, but

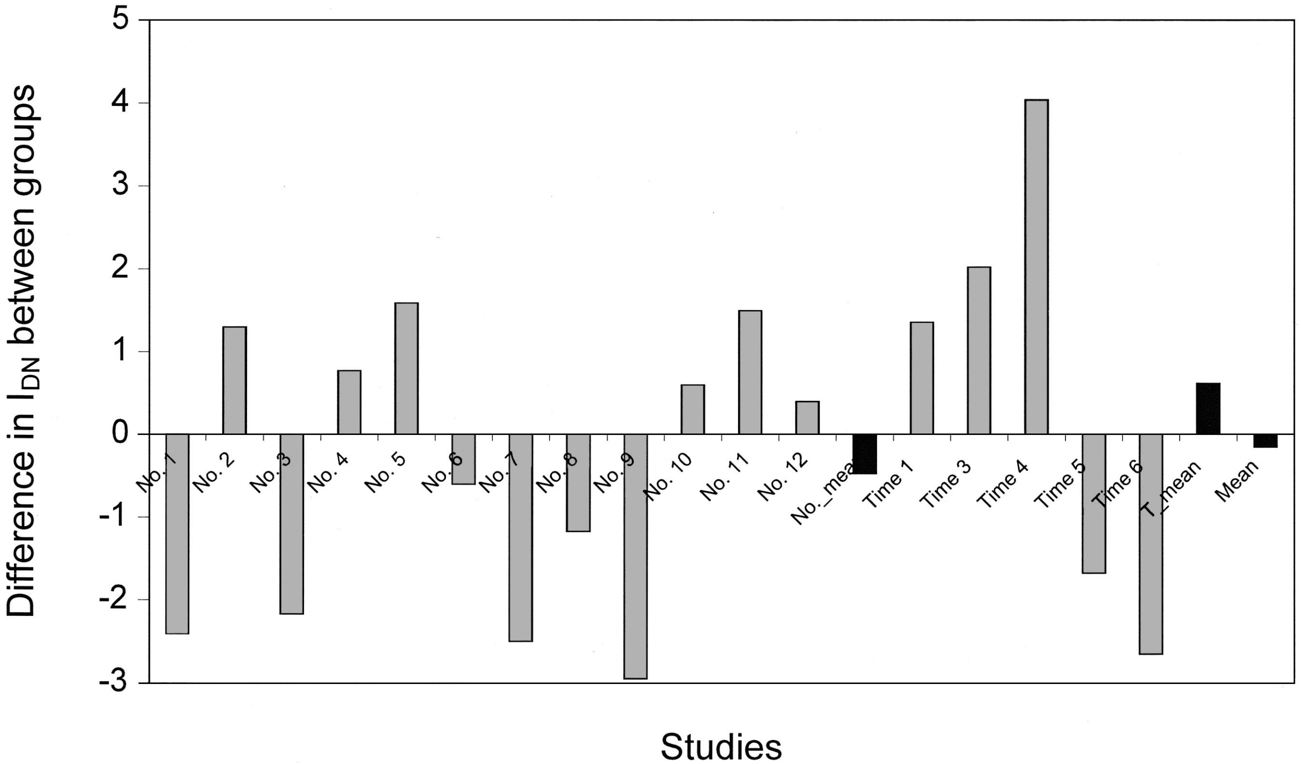

Figure 2. Differences between patients with attention-deficit/hyperactivity disorder and control participants in number-of-items difference scores (I_{DN}) for 12 studies that measured number of items in 45 s (No. 1–No. 12) and for 5 studies that provided time or reaction time (Time 1, Time 3–Time 6). No._mean = average difference across difference No. 1–No. 12; T_mean = average difference across time studies; Mean = average difference across all studies.

not the ratio score, is sensitive to the transformation from number of items per unit of time into time per item and vice versa.

The present meta-analyses excluded studies using Golden's interference score. Two meta-analyses, 1 of which included 18 studies and 1 with four outliers removed, found significant effect sizes (0.24 and 0.17, respectively), with more interference for the ADHD than for the control group. The mean effect size in the original meta-analysis may be considered small relative to the effect size of Homack and Riccio (2004; effect size = 0.75). However, to a large extent this may be due to the use of the difference score in individual studies to quantify interference instead of using the reported Golden's interference scores for Golden studies (as in Homack & Riccio, 2004). Van Mourik et al. (2005) already reported a large difference in effect sizes for Golden's method and the difference interference score (0.40 and −0.003), confirming our suggestion that different methods to quantify interference in the Stroop affect the interpretation of real data. Very conspicuously, the presently observed effect sizes were larger for time studies (1.11 and 0.55 for meta-analysis with and without four outliers, respectively) than for item studies (−0.007 and 0.06). A possible explanation relates exactly to the transformation-sensitivity problem that was noted for the difference score. As exemplified in Table 1, but also apparent in real data when Figures 2 and 3 are compared, group differences that are significant with time scores readily disappear after transformation to number-of-items scores. Regarding the moderator variable age, no significant difference in effect size was found between children and adults with ADHD, indicating that the increase in interference is independent of age. In a meta-analysis of stop-task studies, Lijffijt et al. (2005) found that the deficiency in stopping did hardly exceed that in general reaction time in children, whereas it clearly did so in adults. Stroop interference scores, if validly determined, account for differences in general reaction time, but differences in interference between ADHD and control are equal for child and adult samples.

Consistency analyses of ratio interference scores revealed more interference for the ADHD than for the control group, confirming the results of the meta-analysis. Only in 3 of 19 studies was the ratio score higher in the ADHD group than in the control group. The clear results from the consistency analysis also contradict the possible concern that the correlation between performance in the two Stroop conditions, as assumed within the meta-analysis, was based on an adult sample and may not be evidently generalized to child populations.

A consistency analysis with respect to the difference between ADHD and control groups in base-word reading as assessed by performance on the word task indicated that across studies the ADHD groups were systematically slower in base-word reading. This result provides further evidence for a higher Stroop interfer-

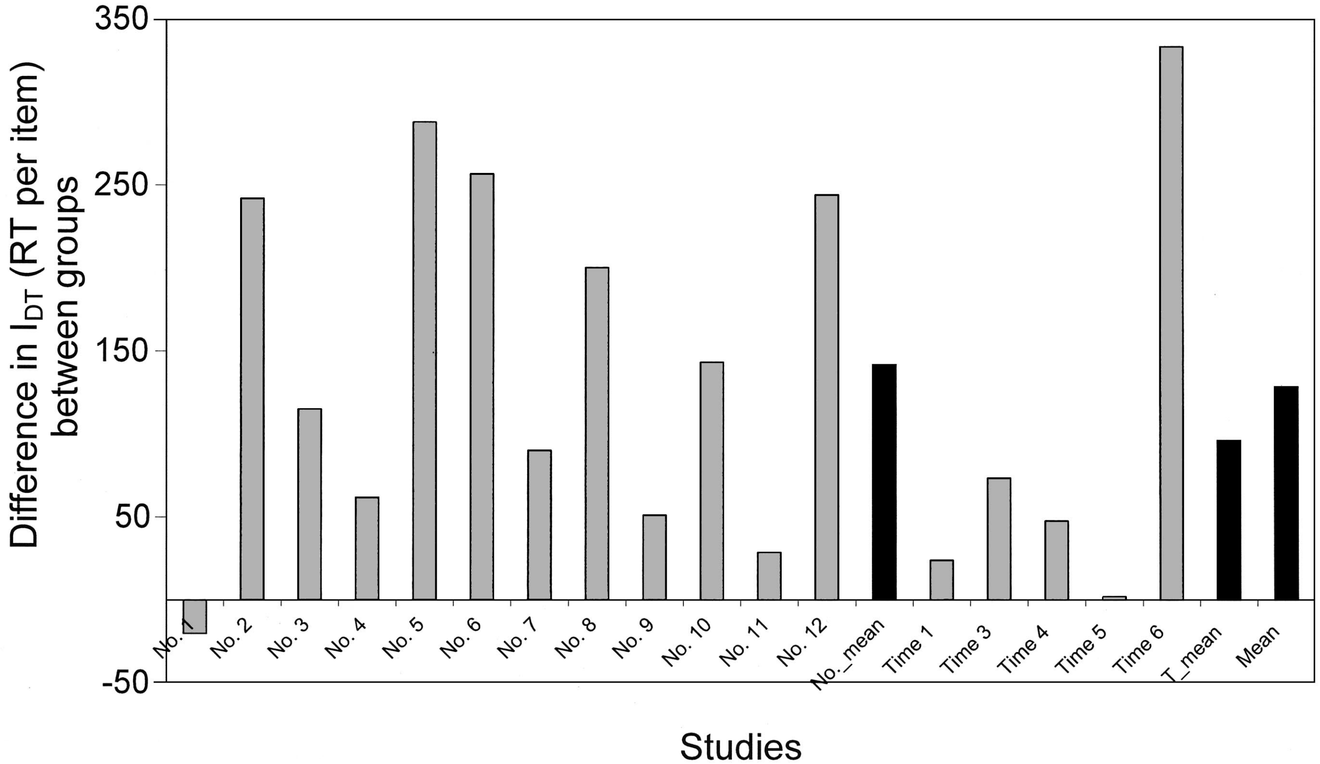

Figure 3. Differences between patients with attention-deficit/hyperactivity disorder and control participants in time difference scores (I_{DT}) for 12 studies that measured number of items in 45 s (No. 1–No. 12) and for 5 studies that provided time or reaction time (Time 1, Time 3–Time 6). RT = reaction time; No._mean = average difference across difference No. 1–No. 12; T_mean = average difference across time studies; Mean = average difference across all studies.

ence in ADHD because the "true" interference level is underestimated in slow word readers without controlling for slow word reading.

As noted, the differences between ADHD and control in Stroop interference, as assessed by ratio score, are significant but generally small. For example, mean ratio interference scores across 19 studies were 0.64 for control and 0.61 for ADHD. For a control-condition score of 80 items in 45 s, this would result in 51.2 items in the incongruent condition for control groups and in only 2 items less (48.8) for ADHD groups. Therefore, it is important to determine which version of the Stroop test is the most sensitive with respect to the group difference in interference. The present meta-analysis revealed higher effect sizes for studies that used time per card or per item as an outcome variable rather than number of items per unit of time. Moreover, computerized Stroop tests (see the relatively high effect size from Carter, Krener, Chaderjian, Northcutt, & Wolfe, 1995) may be more sensitive than standard or Golden Stroop tests. An additional argument for using computerized Stroop tests is the tightly controlled experimental situation, resulting more easily in significance because of a relative low variance of interference scores. The card versions of the Stroop test are problematic for several reasons. First, performance may be influenced by the presence of distractors in the area surrounding the target. Second, color and color–word scores in card versions of the Stroop test are rated by observers, resulting in relatively inaccurate scoring. Furthermore, whereas reaction time as recorded in computerized Stroop tests is truly the time needed to respond to a color word, time as recorded in card versions includes the time needed to make errors and correct for errors. An additional advantage of a computerized Stroop test with a mixed design is the possibility to incorporate congruent trials beside incongruent and neutral trials. In this task version, participants cannot revert to word reading rather than color naming in the congruent condition, enabling the quantification of interference as well as facilitation (neutral vs. congruent). Future research should elucidate further factors that determine the sensitivity of Stroop tests in detecting differences in interference between patient and control groups. One possible example involves the manipulation of the relative probabilities of the control and incongruent stimuli, respectively, in mixed designs (Carter et al., 1995; Tzelgov, Henik, & Berger, 1992). The percentage of incongruent stimuli is negatively related to the amount of interference in the Stroop test (Tzelgov et al., 1992), which has been interpreted as reflecting less strategic control over interference when incongruent stimuli are rare. It could well be the case that in such circumstances control over interference would deteriorate even more in pathological samples, especially those suffering from problems of behavioral inhibition. Furthermore, future research should quantify interference in the Stroop test by several methods and compare the results, particularly in studies that compare two groups that differ in the general speed of processing.

The conclusion of more interference in the Stroop test in ADHD, relative to control, is also consistent with results from related conflict tasks. For example, Jonkman et al. (1999) reported increased interference from distractor letters flanking a centrally presented target letter in children with ADHD relative to control participants. The present study also complements the earlier reviews on reduced stopping performance in ADHD (Frazier et al., 2004; Hervey et al., 2004; Lijffijt et al., 2005; Oosterlaan et al., 1998) and increased commission-error rates in the CPT (Frazier et al., 2004; Hervey et al., 2004; Losier et al., 1996). The effect sizes of stopping behavior and commission errors varied between 0.54 and 0.85 and between 0.55 and 0.73, respectively. Together, these tasks largely cover the spectrum of behavioral inhibition, which has been hypothesized to be the core symptom of ADHD (especially for the combined and hyperactive subtypes; Barkley, 1997). The stop paradigm and the CPT primarily reflect the ability to suppress ongoing behavior and the ability to suppress a very potent (but not yet initiated) response. The Stroop test assesses in particular the ability to control interference from alternative response tendencies that compete with the one that is adequate given the context. The present analysis strongly suggests that also this aspect of behavioral inhibition is consistently disturbed in ADHD.

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

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

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

Barkley, R. A. (1997). *ADHD and the nature of self-control.* New York: Guilford Press.

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.

Bondi, M. W., Serody, A. B., Chan, A. S., Eberson-Shumate, S. C., Delis, D. C., Hansen, L. A., et al. (2002). Cognitive and neuropathologic correlates of Stroop color–word test performance in Alzheimer's disease. *Neuropsychology, 16,* 335–543.

Capitani, E. M., Laiacona, M. M., Barbarotto, R. M., & Cossa, F. M. (1999). How can we evaluate interference in attentional tests? A study based on bi-variate non-parametric tolerance limits. *Journal of Clinical Experimental Neuropsychology, 21,* 216–228.

Carlson, C. L., Lahey, B. B., & Neeper, R. (1986). Direct assessment of the cognitive correlates of attention deficit disorders with and without hyperactivity. *Journal of Psychopathology and Behavioral Assessment, 8,* 69–86.

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

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

Corkum, P. V., & Siegel, L. S. (1993). Is the continuous performance task a valuable research tool for use with children with attention-deficit-hyperactivity disorder? *Journal of Child Psychology and Psychiatry, 34,* 1217–1239.

Frazier, T. W., Demaree, H. A., & Youngstrom, E. A. (2004). Meta-analysis of intellectual and neuropsychological test performance in attention-deficit/hyperactivity disorder. *Neuropsychology, 18,* 543–555.

Goldberg, M. C., Mostofsky, S. H., Cutting, L. E., Mahone, E. M., Astor, B. C., Denckla, M. B., et al. (2005). Subtle executive impairment in children with autism and children with ADHD. *Journal of Autism and Developmental Disorders, 35,* 279–293.

Golden, C. J. (1978). *The Stroop color and word test: A manual for clinical and experimental uses.* Chicago: Stoelting Co.

Golden, C. J., & Freshwater, S. M. (2002). *The Stroop color and word test: A manual for clinical and experimental uses.* Chicago: Stoelting Co.

Henik, A., & Salo, R. (2004). Schizophrenia and the Stroop effect. *Behavioral and Cognitive Neuroscience Reviews, 3,* 42–59.

Hervey, A. S., Epstein, J. N., & Curry, J. F. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: A meta-analytic review. *Neuropsychology, 18,* 485–503.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539–1558.

Homack, S., & Riccio, C. A. (2004). A meta-analysis of the sensitivity and specificity of the Stroop color and word test with children. *Archives of Clinical Neuropsychology, 19,* 725–743.

Jonkman, L. M., Kemner, C., Verbaten, M. N., Van Engeland, H., Kenemans, J. L., Camfferman, G., et al. (1999). Perceptual and response interference in children with attention-deficit hyperactivity disorder and the effects of methylphenidate. *Psychophysiology, 36,* 419–429.

Kenemans, J. L., Wieleman, J. S. T., Zeegers, M., & Verbaten, M. N. (1999). Caffeine and Stroop interference. *Pharmacology, Biochemistry and Behavior, 63,* 589–598.

Lansbergen, M. M., Van Hell, E., & Kenemans, J. L. (in press). Impulsivity and conflict in the Stroop task: An ERP study. *Journal of Psychophysiology*.

Lijffijt, M., Kenemans, J. L., Verbaten, M. N., & van Engeland, H. (2005). A meta-analytic review of stopping performance in ADHD: Deficient inhibitory motor control? *Journal of Abnormal Psychology, 114,* 216–222.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

Losier, B. J., McGrath, P. J., & Klein, R. M. (1996). Error patterns on the continuous performance test in non-medicated and medicated samples of children with and without ADHD: A meta-analytic review. *Journal of Child Psychology and Psychiatry, 37,* 971–978.

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

Oosterlaan, J., Logan, G. D., & 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.

Riccio, C. A., Reynolds, C. R., & Lowe, P. A. (2001). *Clinical applications of continuous performance tests: Measuring attention and impulsive responding in children and adults.* New York: Wiley.

Rouder, J. N., & Speckman, P. L. (2004). An evaluation of the Vincentizing method of forming group-level response time distributions. *Psychonomic Bulletin & Review, 11,* 419–427.

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.

Saunders, D. R. (1980). Definition of Stroop interference in volunteers and non-volunteers. *Perceptual and Motor Skills, 51,* 343–354.

Scheres, A., Oosterlaan, J., Geurts, H., Morein-Zamir, S., Meiran, N., Schut, H., et al. (2004). Executive functioning in boys with ADHD: Primarily an inhibition deficit? *Archives of Clinical Neuropsychology, 19,* 569–594.

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.

Silverstein, S. M., Como, P. G., Palumbo, D. R., West, L. L., & Osborn, L. M. (1995). Multiple sources of attentional dysfunction in adults with Tourette's syndrome: Comparison with attention deficit-hyperactivity disorder. *Neuropsychology, 9,* 157–164.

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

Tzelgov, J., Henik, A., & Berger, J. (1992). Controlling Stroop effects by manipulating expectations for color words. *Memory & Cognition, 20,* 727–735.

Van Mourik, R., Oosterlaan, J., & Sergeant, J. (2005). The Stroop revisited: A meta-analysis of interference control in AD/HD. *Journal of Child Psychology and Psychiatry, 46,* 150–165.

Weir, C., Bruun, C., & Barber, T. (1997). Are backward words special for older adults? *Psychology and Aging, 12,* 145–149.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57,* 1336–1346.

Received December 9, 2004
Revision received October 23, 2006
Accepted October 30, 2006 ■