Psychology and Aging
2004, Vol. 19, No. 3, 482–494

0882-7974/04/$12.00 DOI: 10.1037/0882-7974.19.3.482

Age-Related Differences in Performance on the Wisconsin Card Sorting Test: A Meta-Analytic Review

Matthew G. Rhodes
Florida State University

Two meta-analyses investigating age-related differences in performance on a popular measure of executive function, the Wisconsin Card Sorting Test (WCST), are reported. The 1st meta-analysis examined age-related changes in performance for the number of categories achieved, and the 2nd meta-analysis examined performance for the number of perseverative errors committed. Results indicated that robust age differences were present on both measures. Further analysis of moderator variables revealed reliable effects of education and test version on both measures, whereas test modality led to marginally significant differences in effect sizes obtained only for the number of categories achieved. Findings are discussed along with current accounts of age differences in performance of the WCST.

The Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948; Heaton, Chelune, Talley, Kay, & Curtiss, 1993) is among the most frequently administered neuropsychological tests (Butler, Retzlaff, & Vanderploeg, 1991). The participant's task on the WCST is to sort a set of response cards on the basis of several dimensions (color, form, and number), which change periodically during the course of sorting. Participants are not instructed as to how to sort the cards but must infer the correct sorting principles through limited feedback from the experimenter, who only tells the participant whether the sort is correct or incorrect. After a particular number of consecutive successful sorts (the criterion varies depending on the version), the sorting principle changes and the participant must adjust accordingly. Scores are tallied along several dimensions, with the number of categories achieved and the number of perseverative errors committed the most commonly reported measures. A participant is scored as having achieved a category when they make a prespecified number of correct sorts for a particular category. For example, in the most common version of the test (Heaton et al., 1993), the participant must make 10 consecutive successful sorts to achieve a category. Perseverative errors occur when the participant sorts according to a category that was formerly correct but is no longer in effect.

Performance on the WCST is assumed to be dependent on an array of cognitive processes, typically labeled *executive functions.* A number of processes have been proposed as hallmarks of executive functioning, including the monitoring and control of behavior, suppression of irrelevant information, reasoning, updating information in working memory, inhibition of prepotent responses, planning, shifting, and control of attention, among others (e.g., Baddeley, 1996; Kane & Engle, 2002; Koechlin, Ody, & Kouneiher, 2003; Miyake et al., 2000; Shimamura, 2000; Waltz et al., 1999). These executive functions are presumably localized in the prefrontal cortex (PFC) or, more specifically, the dorsolateral PFC (e.g., Duncan, 1993, 1995; Goldman-Rakic, 1987; Kane & Engle, 2002; Kimberg & Farah, 1993; Shallice & Burgess, 1991; Waltz et al., 1999). In fact, the popularity of the WCST as a measure of executive function is due largely to evidence indicating that it is sensitive to dysfunction or damage to the PFC (e.g., Drewe, 1974; Milner, 1963).[1]

The WCST is particularly important within the aging literature, as several researchers have suggested that a decline in executive functions may be a primary factor in the cognitive deficits present in aging populations (e.g., Moscovitch & Winocur, 1995; West, 1996; Whelihan & Lesher, 1985). Neurologically, there is compelling evidence that the neural substrates underlying the PFC, the putative seat of executive function, deteriorate more rapidly with age than other cortical regions (see Raz, 2000, for a review). There is also extensive behavioral evidence indicating that older adults perform more poorly than young controls on a number of tasks thought to rely on executive functioning that are likewise sensitive to frontal lobe lesions. For example, older adults show greater interference on incongruent trials of the Stroop task (Houx, Jolles, & Vreeling, 1993), produce fewer words on tests of verbal and semantic fluency (e.g., Howard, 1980; Troyer, Moscovitch, & Winocur, 1997), have greater difficulty on tasks that require planning (S. Daigneault & Braun, 1993), and show impairments in source memory (e.g., Glisky, Polster, & Routhieaux, 1995; McIntyre & Craik, 1987).

Age differences are also apparent on the WCST. For example, S. Daigneault, Braun, and Whitaker (1992) reported that older adults committed significantly more perseverative errors and completed fewer categories than young adults. These findings have been replicated in a number of other studies (e.g., Axelrod & Henry,

I thank Rick Wagner, Colleen Kelley, Katinka Dijkstra, and Michelle Bourgeios for valuable comments and suggestions on an earlier version of this article.

Correspondence concerning this article should be addressed to Matthew G. Rhodes, Department of Psychology, Florida State University, Tallahassee, FL 32306. E-mail: rhodes@psy.fsu.edu

---

[1] This finding has not been entirely consistent (e.g., Anderson, Damasio, Jones, & Tranel, 1991; van den Broek, Bradshaw, & Szabadi, 1993; see Mountain & Snow, 1993, for a review). However, a recent meta-analysis (Demakis, 2003) of 23 studies comparing the WCST performance of patients with frontal versus nonfrontal lesions revealed that frontal patients achieved significantly fewer categories ($d = -0.35$) and made significantly more perseverative errors ($d = -0.32$) than nonfrontal patients. Thus, the WCST does appear to be sensitive to lesion location.

1992; Beatty, 1993; Fristoe, Salthouse, & Woodard, 1997; Parkin & Walter, 1991a), but there are exceptions. Haaland, Vranes, Goodwin, and Garry (1987), for example, found that young-old participants (aged 64–69 years) committed fewer perseverative errors and achieved more categories than younger participants on a 64-card version of the WCST. Boone, Ghaffarian, Lesser, and Hill-Gutierrez (1993) likewise demonstrated that young-old participants (aged 60–69 years) achieved more categories and committed slightly fewer perseverative errors than middle-aged participants (aged 45–49 years). Thus, although the general pattern of evidence indicates that there are significant age-related deficits on the WCST, age-related differences are not always present.

Given that the WCST has become an increasingly popular measure of executive function in aging research (Bryan & Luszcz, 2000), a meta-analysis can serve to synthesize the existing literature. A meta-analysis can also serve two other purposes. First, no comprehensive quantitative review of age-related differences in performance on the WCST currently exists. Given its widespread use as a measure of executive function, a meta-analysis may help to clarify the efficacy of the WCST as an executive task. Second, the meta-analysis technique is ideally suited for examining potential moderator variables that may underlie performance on the WCST. This is particularly important given the variability in participants and the variety of administration methods that are replete in the literature on the WCST.

The Meta-Analysis

A *meta-analysis* is a quantitative summary of a research literature that attempts to describe and test the magnitude of various effect sizes across studies. One can think of an effect size as a standardized score, analogous to a z score, for a dependent variable or variables of interest. Using such effect sizes, one can examine the influence of possible moderator variables and determine where effect sizes reliably converge or differ (Hedges & Olkin, 1985). For the current meta-analysis, effect sizes were derived by comparing the performance of healthy older adults, with an average age of 55 years or older, to younger adults, with an average age between 20 and 35 years, on the two most commonly reported indices of performance on the WCST: the number of categories achieved and the number of perseverative errors committed.

Method

Studies for the current meta-analysis were obtained from a computerized search of PsycINFO, MedLine, and *Dissertation Abstracts* using the keywords *executive function, aging, Wisconsin Card, Modified Card,* and *prefrontal cortex.* Additional searches were conducted in the *Web of Science* citation index and through manual inspection of published articles. To be included in the meta-analysis, articles had to meet several criteria: (a) Articles must have included participants with an average age of 55 years or older and a group of younger participants, with an average age between 20 and 35 years; (b) participants must be free of any neurological or psychiatric disorders and report good health; (c) WCST data were collected as a primary component of the article or were part of a larger neuropsychological screening; and (d) the article included means and standard deviations or enough statistical information (e.g., F, t, p, or r values) to permit the estimation of effect sizes. When studies did not include enough information, the authors were contacted to provide the missing information. If this was unsuccessful, the study was either excluded from the meta-analysis or, if partial data was present, the effect size was conservatively estimated on the basis of information reported in the text.

A total of 34 studies met the inclusion criteria for the first meta-analysis based on the number of categories achieved.[2] These articles were published between 1990 and 2003 and included at least one group of older participants and one group of younger participants. Fifty-three effect sizes were calculated on the basis of a total of 3,049 participants (1,687 older participants and 1,362 younger participants). The average age of the older adults was 71.27 years ($SD = 7.04$) and the average age of the younger adults was 24.50 years ($SD = 3.14$). On the basis of those studies reporting demographic data, younger adults had an average of 14.44 years of education, whereas older adults had an average of 13.37 years of education, $t(71) = 2.61$, $p < .05$. Approximately 55% of the younger participants were women, whereas approximately 58% of the older participants were women.

A total of 34 studies met the inclusion criteria for the second meta-analysis based on the number of perseverative errors committed. These articles were published between 1990 and 2003 and included at least one group of older participants and one group of younger participants. Fifty-five effect sizes were calculated on the basis of a total of 2,923 participants (1,643 older participants and 1,280 younger participants). The average age of the older participants was 71.54 years ($SD = 7.05$), and younger participants averaged 24.77 years ($SD = 3.06$) of age. On the basis of those studies reporting demographic data, younger adults had an average of 14.35 years of education and older adults had an average of 13.39 years of education, $t(74) = 2.26$, $p < .05$. In addition, approximately 55% of the younger participants were women, whereas approximately 58% of the older participants were women.

Listing and Discussion of Coded Variables

Age. The first major variable to be coded was the average age reported for the older participants tested. Age was coded as a categorical variable with three levels: young-old (55–64 years), middle-old (65–74 years), and old-old (75+ years). In cases in which only the age range was available, age was coded on the basis of the central point in the range reported.

Test version. The original version of the WCST (e.g., Milner, 1963), still most typically administered (Heaton et al., 1993), consists of two sets of 64 cards and a maximum of six categories. Categories are completed after 10 consecutive, successful sorts. Heaton's (1981) original scoring rules have been revised over time in an effort to reduce ambiguity, and scoring is currently based on Heaton et al. (1993). Nelson (1976) modified the WCST to ease administration with participants who might become unduly stressed by the length of the task. Nelson's Modified Card Sorting Test (MCST) includes two sets of 24 cards and further reduces sorting requirements by requiring participants to produce only six consecutive successful sorts to complete a category. On the basis of its limited sensitivity to frontal lobe damage, one review of the MCST (de Zubicaray & Ashton, 1996) has suggested that it may be a completely different version of the test than the original WCST. Results from Demakis's (2003) meta-analysis of WCST performance substantiate this conclusion, as, in seven studies, the MCST did not discriminate between patients with frontal versus nonfrontal lesions. Whether this lack of sensitivity applies to aging

---

[2] One may notice that this figure seems to be an underestimate in comparison with the volume of studies administering the WCST that are available. The primary reason for this is that a number of studies that administer the WCST to older adults do not include a young control group and thus do not meet the inclusion criteria. For example, many studies only administer the WCST to different groups of healthy older adults (e.g., Axelrod & Henry, 1992; Loranger & Misiak, 1960). In addition, a number of other studies compare healthy older adults with other older adults with neurological or psychiatric problems, such as Parkinson's disease (e.g., Beatty, Monson, & Goodkin, 1989; Brown & Marsden, 1988; Woods & Tröster, 2003), dementia of the Alzheimer's type (e.g., Bondi et al., 2003), Korsakoff's syndrome (e.g., Joyce & Robbins, 1991), or depression (e.g., Hart et al., 1988; Kinderman, Kalayam, Brown, Burdick, & Alexopoulos, 2000). This issue is revisited in the General Discussion section.

populations is of some question. Other revisions of the original test of have been made, reducing the task to 72 cards (Hart, Kwentus, Wade, & Taylor, 1988; Jenkins & Parson, 1978) and, more recently, to 64 cards (Kongs, Thompson, Iverson, & Heaton, 2000). However, these versions did not meet the inclusion criteria for the current meta-analysis and were not examined. In summary, test version was coded as a categorical variable with two levels: Heaton (WCST) and Nelson (MCST).

Test modality. Computerized administrations of the WCST have become increasingly common (Degl'Innocenti & Backman, 1996; Fristoe et al., 1997; Spencer & Raz, 1994) and may produce results that differ from manual administrations. For example, Feldstein et al. (1999), on the basis of their results from testing younger participants in computerized and manual versions of the WCST, suggested that normative data from manual versions of the WCST were not applicable to computerized versions. Thus, test modality was coded as a categorical variable with two levels: manual or computer based.

Education. Several studies have demonstrated a relationship between education and performance on the WCST or MCST (e.g., Compton, Bachman, & Logan, 1997; G. Daigneault, Joly, & Frigon, 2002; de Zubicaray, Smith, Chalk, & Semple, 1998; Lineweaver, Bondi, Thomas, & Salmon, 1999; Maylor, Moulson, Muncer, & Taylor, 2002). For example, de Zubicaray et al. (1998) reported a modest but significant correlation between education and the number of categories achieved ($r = .37$). Lineweaver et al. (1999) also reported a moderate correlation between education and the number of perseverative errors committed ($r = -.27$). Education was coded as a categorical variable based on the average number of years of education reported by older participants with the following levels: less than 12 years, 12–15 years, greater than 15 years, and unspecified. These levels essentially correspond to an approximately high school level education or less (less than 12 years), some university education (12–15 years), and extensive university education (greater than 15 years).

Method of Analysis

Following Hedges (1994), the fixed-effects model for analyzing categorical variables was used. The first step is to determine whether all effect sizes are homogenous (Q_T). The Q_T statistic has an approximate chi-square distribution with $k - 1$ degrees of freedom, where k is the number of studies. For each categorical variable, a within-group fit statistic (Q_W) was calculated to determine whether studies within groups were homogenous. A significant Q_W would indicate that significant within-groups heterogeneity existed. Given this condition, one may attempt a more detailed analysis of effect sizes and examine variation between groups using the between-groups fit statistic (Q_B). This test is directly analogous to the one-way analysis of variance omnibus F test for variation of group means, and a significant Q_B indicates that the average weighted effect size differed across groups.

Results and Discussion

Measures of effect size based on pooled standard deviations (e.g., Cohen's *d*, Hedges's *g*) are most frequently used in meta-analyses. However, across studies, older adults exhibited significantly more variability in performance than did younger adults, compromising pooled standard deviations as a measure of variability. Table 1 displays descriptive statistics for younger and older adults on the basis of measures of categories achieved and perseverative errors committed for the data analyzed in the current study. Examination of the table confirms the observation that older adults demonstrated considerably more variability, evident in larger mean standard deviations on both measures, than younger adults. In such cases, a pooled measure of variability is inappropriate. Therefore, per the recommendation of Rosenthal (1994), Glass's Δ (Glass, McGaw, & Smith, 1981) was calculated as the estimate of effect size as follows, using only a single measure of variability:

$$(M_{\text{older}} - M_{\text{younger}})/SD_{\text{younger}},$$

where M_{older} is the mean for the older participants, M_{younger} the mean for the younger participants, and SD_{younger} is the standard deviation for younger participants.[3] In the interest of clarity, negative effect sizes for both categories achieved and the number of perseverative errors committed denotes poorer performance by older adults than younger adults on the measure of interest (i.e., fewer categories achieved or more perseverative errors committed). All effect sizes to be reported were weighted as a function of sample size.[4]

Table 1
Descriptive Statistics for the Studies Examined in the Meta-Analysis

Variable	Older participants	Young participants
Categories achieved		
M	3.99	5.58
SD	1.83	1.10
Perseverative errors committed		
M	15.85	6.92
SD	11.44	5.04

Note. The standard deviation refers to the mean of the standard deviations reported across all studies examined.

Given that the measure of variability to be used is based on data from younger participants, one can make the case that effect size estimates may be somewhat inflated. Therefore, a pooled estimate of effect size (Hedges's *g*) was also calculated for comparison. A description of this calculation along with results is provided in the Appendix. Instances in which the metric of effect size to be used in the current study (i.e., Δ) departs from *g* and changes the interpretation of moderator variables are described in the General Discussion section.

---

[3] One can transform data to make the population standard deviations more similar by using log transformations or other methods (e.g., square roots). However, this remedy requires access to the original data set from each study, making it impractical as an alternative.

[4] Specifically, weights (w_i) are inversely proportional to the conditional variance (v_i), which in turn is inversely proportional to the study sample size. For Δ-based effect sizes, the conditional variance is calculated as follows:

$$v_i = [(n_1 + n_2)/(n_1 - n_2)] + [\Delta^2/2(n_2 - 1)],$$

where n_1 is the sample size if the first group, n_2 is the sample size of the second group, and Δ is the effect size. As can be seen, the larger the sample size, the smaller the conditional variance. (In more conventional terms, one can think of the conditional variance as the square of the standard error.) Given that weights are calculated as the inverse of the conditional variance (i.e., $w_i = 1/v_i$), larger weights will be derived from studies with larger sample sizes. Thus, studies with larger sample sizes, which should presumably provide a more accurate estimate of the population effect size, are given more weight than studies with smaller sample sizes (see Shadish & Haddock, 1994, for a more detailed discussion of these issues).

Categories Achieved Model

The average weighted effect size for the number of categories achieved was significant ($\Delta = -1.13$; 95% confidence interval [CI] $= -1.22, -1.04$), indicating that older adults achieved significantly fewer categories than younger adults.[5] (Unless otherwise noted, the alpha level for all statistical tests was set to .05.) In addition, effect sizes were clearly not homogenous, as there was substantial heterogeneity ($Q_T = 220.49$, $p < .00$).

Figure 1 displays a stem-and-leaf plot of mean weighted effect sizes for the number of categories achieved. An examination of the figure indicates that the distribution has a slightly negative skew, with a number of effect sizes clustered on the higher side of the mean (i.e., grouped toward 0). Two other points are relevant with regard to the distribution of effect sizes. First, with only a few exceptions, it appears that there is a distinct cutoff in effect sizes at approximately –.50. This may be indicative of a publication bias against small or even positive effect sizes, suggesting that the population mean weighted effect size may be smaller than that reported here. It is the case that a number of additional studies that administer the WCST to groups of older participants do not include a comparison group of younger participants and thus are not included in the current meta-analysis (see Footnote 2). Nevertheless, an enormous number of studies with effect sizes of zero would have to exist ($k_0 = 9{,}476$) for the average weighted effect size to no longer differ significantly from zero. Second, three outliers appear to be present at values of −4.4, –5.5, and −6.8. These data were included in the analyses to be reported; however, additional analyses were also conducted that excluded all outliers. Those analyses resulted in one substantive departure of note from the data reported (for test modality) that is described in the appropriate section.

Main effect analyses. Table 2 presents a summary of main effect analyses for the number of categories achieved. The test of Q_B may be used to determine whether the levels of a categorically coded variable differ significantly from each other. First, the effect of age (young-old, middle-old, old-old) was significant ($Q_B = 51.59$), with participants achieving fewer categories with increasing age. Specifically, old-old participants achieved the fewest categories ($\Delta = -1.88$; 95% CI $= -2.12, -1.64$; $k = 15$), followed by middle-old participants ($\Delta = -1.08$; 95% CI $= -1.19, -0.97$; $k = 31$), who achieved fewer categories than young-old participants ($\Delta = -0.70$; 95% CI $= -0.92, -0.47$; $k = 7$). Partitioning the within-group fit statistic (Q_W) according to levels of age indicated significant within-group heterogeneity for middle-old ($Q_{W2} = 130.53$) and old-old participants ($Q_{W3} = 27.17$).

-6.		8
-5.		5
-4.		4
-3.		1, 3, 2
-2.		0, 1, 1, 1, 1, 2, 2, 2, 2, 4, 5, 8, 8
-1.		0, 0, 0, 0, 1, 3, 3, 3, 4, 5, 5, 6, 7, 8, 9
-0.		1, 2, 4, 5, 5, 6, 7, 7, 7, 7, 7, 8, 8, 8, 8, 8, 8, 8, 9

Figure 1. Stem-and-leaf plot of weighted effect sizes for the number of categories achieved. The first digit of each effect size is listed in the stem column, with the second digit listed in the leaf column. Thus, the top two entries of the figure are read as −6.8 and –5.5, respectively.

Table 2
Summary of Main Effect Analyses for the Number of Categories Achieved

			95% CI		*Q*	
Variable	*k*	Δ	Lower	Upper	Within	Between
Age						
Young-old	7	−0.70	−0.92	−0.47	11.16	51.59**
Middle-old	31	−1.08	−1.19	−0.97	130.53**	
Old-old	15	−1.88	−2.12	−1.64	27.17*	
Version						
Heaton	40	−1.07	−1.16	−0.97	173.48**	10.23**
Nelson	13	−1.49	−1.72	−1.25	36.74**	
Modality						
Manual	45	−1.17	−1.27	−1.07	186.24**	3.04
Computer based	8	−0.98	−1.17	−0.80	31.61**	
Education						
>12 years	8	−1.64	−1.90	−1.38	26.26**	24.87**
12–15 years	25	−1.04	−1.16	−0.91	60.96**	
<15 years	8	−0.97	−1.14	−0.81	66.93**	
Unspecified	12	−1.49	−1.82	−1.16	41.43**	

Note. k = number of effect sizes; Δ = mean weighted effect size; CI = confidence interval; Q = heterogeneity; Heaton = Wisconsin Card Sorting Test (WCST); Nelson = modified version of WCST.
* $p < .05$. ** $p < .00$.

The effect of test version (Heaton, Nelson) was significant ($Q_B = 10.23$), indicating that the average weighted effect size for the number of categories achieved differed between the two test versions examined. Inspection of these data shows that older adults achieved fewer categories with the Nelson version of the test ($\Delta = -1.49$; 95% CI $= -1.72, -1.25$; $k = 13$) than with the Heaton version ($\Delta = -1.07$; 95% CI $= -1.16, -0.97$; $k = 40$). Therefore, although it may not be as sensitive to frontal lobe impairment as the Heaton version (Demakis, 2003), the Nelson version appears to be sensitive to age differences. Partitioning the within-group fit statistic (Q_W) according to levels of test version indicated significant within-group heterogeneity for both the Nelson ($Q_{W2} = 36.74$) and Heaton ($Q_{W1} = 173.48$) versions.

These data appear to demonstrate that Nelson's (1976) MCST resulted in more substantial age differences than Heaton's (1981; Heaton et al., 1993) WCST version. However, this conclusion must be treated with caution. For example, seven of the effect sizes contributing to the mean weighted effect size for the Heaton version were derived from young-old participants, whereas the Nelson version did not include any participants from the young-old age group. Therefore, a follow-up analysis was conducted that excluded all young-old participants from the data for the Heaton version. Results showed that even after excluding these young-old participants, the Nelson version of the test (see mean weighted effect size above) was more sensitive to age differences than the

---

[5] The mean unweighted effect size ($\Delta = -1.75$) was substantially larger than the mean weighted effect size. Further, the mean unweighted effect size was significantly greater than zero, $t(52) = -9.77$, $p = .00$. In addition, the median unweighted effect size was –1.41. Finally, the mean weighted effect size based on a pooled measure of variability (i.e., Hedges's g) was considerably smaller ($g = -.87$) than the mean weighted Δ-based effect size reported.

Heaton version ($\Delta = -1.15$; 95% CI = –1.26, –1.05; $k = 33$), a difference that was significant ($Q_B = 6.26$, $p < .05$).

The effect of test modality (manual, computer based) was marginally significant ($Q_B = 3.04$, $p = .08$), as effect sizes were somewhat larger when the WCST was administered manually ($\Delta = -1.17$; 95% CI = –1.27, –1.07; $k = 45$) than in a computer-based format ($\Delta = -0.98$; 95% CI = –1.17, –0.80; $k = 8$). Partitioning the within-group fit statistic according to levels of test modality indicated that significant within-group heterogeneity existed for both manual ($Q_{W1} = 186.24$) and computer-based versions ($Q_{W2} = 31.61$). It must be noted that when all outliers were excluded from the data analysis, the difference between the manual ($\Delta = -1.16$; 95% CI = –1.26, –1.06; $k = 43$) and computer-based ($\Delta = -0.92$; 95% CI = –1.11, –0.74; $k = 7$) versions of the WCST was statistically significant ($Q_B = 4.69$, $p < .05$).

Finally, analysis of education (less than 12 years, 12–15 years, more than 15 years, unspecified) indicated that effect sizes differed on the basis of the average number of years of education attained ($Q_B = 24.87$). In particular, older participants who averaged less than 12 years of education ($\Delta = -1.64$; 95% CI = –1.90, –1.38; $k = 8$) achieved fewer categories than older participants, who, on average, had between 12 to 15 years of education ($\Delta = -1.04$; 95% CI = –1.16, –0.91; $k = 25$) or greater than 15 years of education ($\Delta = -0.97$; 95% CI = –1.14, –0.81; $k = 8$). Thus, it appears that education is related to the number of categories achieved, a factor that is included in norms for the WCST (Heaton et al., 1993). Partitioning the within-group fit statistic (Q_W) according to levels of education indicated significant within-group heterogeneity for each age group.

Data for participants whose average level of education was unspecified ($\Delta = -1.49$; 95% CI = –1.82, –1.16; $k = 12$) are not as easily interpreted. Further inspection of these data indicates that this group was disproportionately made up of the older age groups coded in the study. Specifically, six of the effect sizes were based on participants classified as middle-old whereas five effect sizes were based on participants classified as old-old; in only one study were participants classified as young-old. When the oldest group of participants was excluded from the analysis, there was a moderate reduction in the mean weighted effect size ($\Delta = -1.39$; 95% CI = –1.85, –0.93; $k = 7$). Though not a complete account, this suggests that the large effect size evident for participants whose education was not specified may be an artifact of the distribution of age groups within that category.

Hierarchical regression analyses. One potential factor that must be accounted for when considering the effects of age is that it is often confounded with other variables, such as education. For example, according to the most recent census estimates (United States Census Bureau, 2001), approximately 71% of older adults aged 75 years and older have attained 12 or fewer years of education. Further, 64% of older adults between the ages of 65 and 74 years and 54% of older adults between the ages of 55 and 64 years have attained 12 or fewer years of education. Thus, any effect of education in the current meta-analysis may be more firmly rooted in differences in chronological age. To address this question, all variables were examined in a hierarchical weighted least squares regression analysis, with weighting done in the same manner used to calculate effect sizes (Hedges & Olkin, 1985; see also Footnote 4). Such an analysis allows one to determine whether a particular variable increases the proportion of explained variability (Q) by an amount greater than would be expected by chance. Age was first entered in the model, followed by education, which was followed by a block with methodological variables corresponding to test version and test modality. Results showed that age alone accounted for slightly less than a fifth of the variability in effect sizes ($R^2 = .18$). Entering education accounted for an additional 3.6% of the variability in effect sizes ($Q_{\text{CHANGE}} = 8.10$, $p < .05$). Test version and test modality accounted for a small but nonsignificant portion (1.5%) of the variability in effect sizes not explained by age and education ($Q_{\text{CHANGE}} = 3.30$, $p > .20$). The full model therefore accounted for slightly less than a quarter of the variability in effect sizes (i.e., $R^2 = .23$). Education explained a small proportion of the variability in effect sizes beyond that explained by age, whereas test modality and test version made more minimal contributions.

As previously noted, a portion of the effect sizes examined were derived for older adults whose level of education was not specified, potentially masking the relationship between WCST performance and education. Therefore, a second, identical set of hierarchical analyses was conducted that excluded effect sizes for older adults with unspecified levels of education ($k = 12$). Results showed that age alone accounted for a fifth of the variability in effect sizes ($R^2 = .20$). Adding education (with unspecified education excluded) explained an additional 4% of the variability in effect sizes ($Q_{\text{CHANGE}} = 6.82$, $p < .05$). Entering test version and test modality explained an additional 3% of the variability in effect sizes ($Q_{\text{CHANGE}} = 5.84$, $p = .12$). Thus, the full model excluding effect sizes for unspecified levels of education accounted for over a quarter of the variability in effect sizes ($R^2 = .28$).

Discussion. Several conclusions can be drawn from the meta-analysis of the number of categories achieved. Clearly, robust age differences were evident, as older participants achieved fewer categories with increasing age. In addition, education was positively related to age, with participants with less than 12 years of education in particular achieving the fewest categories. These data are thus consistent with the normative data that are currently available (Heaton et al., 1993; Lineweaver et al., 1999; see also Axelrod & Henry, 1992). Hierarchical regression analyses indicated that education contributed a small but significant proportion explained variability to effects sizes in addition to that explained by age.

Of additional importance in the current meta-analysis are data for two other variables that have less conclusive standing in the literature. First, the Nelson version of the test (the MCST) appears to be just as sensitive to age differences for the number of categories achieved, if not more sensitive, than the Heaton (WCST) version. This may be contrasted with recent data indicating that the MCST is not sensitive to other variables, such as lesion location (Demakis, 2003). Second, on the basis of the limited number of computer-based studies available, it appears that the modality of the test (manual vs. computer based) did lead to differences in the number of categories achieved, though this difference was only significant when outliers were excluded. Regression analyses indicated that test version and test modality accounted for only a small proportion of explained variability in effect sizes and are likely far less important than the age of the participant. A second meta-analysis examines whether these findings hold true for the number of perseverative errors committed.

Perseverative Errors Model

Across all studies analyzed, the mean weighted effect size for the number of perseverative errors committed was significant ($\Delta = -1.29$; 95% CI = –1.40, –1.19), indicating that older adults committed significantly more perseverative errors than younger adults.[6] In addition, effect sizes were not homogenous, with considerable heterogeneity present ($Q_T = 398.06$, $p < .00$).

A stem-and-leaf plot of weighted effect sizes for the number of perseverative errors committed is displayed in Figure 2. Examination of the figure indicates that there is a slightly negatively skewed distribution of effect sizes around the mean. The majority of effect sizes are grouped at $\Delta = 1.00$ or larger, but there are several substantially smaller than 1.00, including one positive effect size (i.e., one case of older adults committing fewer errors than younger adults) and one effect size of zero. However, an enormous number of studies with effect sizes of zero would have to exist ($k_0 = 11{,}903$) for the average weighted effect size to no longer differ significantly from zero. It must also be noted that there appear to be two outliers present in the distribution of effect sizes, at values of –7.6 and –14.8. Analyses that excluded these observations resulted in slightly smaller estimates of effect size only for the Nelson version of the test and for participants whose level of education was not specified.[7] Thus, all observations were included in the meta-analysis of perseverative errors committed.

Main effect analyses. Table 3 presents a summary of main effect analyses for the number of perseverative errors committed. The effect of age (young-old, middle-old, old-old) was significant ($Q_B = 58.39$), with age positively related to the number of perseverative errors committed. In particular, old-old participants committed the most perseverative errors ($\Delta = -2.17$; 95% CI = –2.44, –1.90; $k = 17$). They were followed by middle-old participants ($\Delta = -1.25$; 95% CI = –1.37, –1.13; $k = 31$), who committed more perseverative errors than young-old participants ($\Delta = -0.81$; 95% CI = –1.04, –0.58; $k = 7$). Partitioning the within-group fit statistic according to levels of age indicated that significant within-group heterogeneity existed for all groups.

Table 3
Summary of Main Effect Analyses for the Number of Perseverative Errors Committed

Variable	k	Δ	95% CI Lower	95% CI Upper	Q Within	Q Between
Age						
Young-old	7	−0.81	−1.04	−0.58	15.86*	58.39**
Middle-old	31	−1.25	−1.37	−1.13	230.42**	
Old-old	17	−2.17	−2.44	−1.90	93.34**	
Version						
Heaton	41	−1.19	−1.30	−1.09	237.63**	43.02**
Nelson	14	−2.41	−2.76	−2.06	117.35**	
Modality						
Manual	46	−1.30	−1.42	−1.18	331.89**	0.05
Computer based	9	−1.28	−1.47	−1.08	66.06**	
Education						
>12 years	10	−1.70	−1.95	−1.44	89.72**	77.96**
12–15 years	23	−1.11	−1.24	−0.97	125.07**	
<15 years	7	−1.09	−1.28	−0.89	28.87*	
Unspecified	15	−2.91	−3.33	−2.49	76.38**	

Note. k = number of effect sizes; Δ = mean weighted effect size; CI = confidence interval; Q = heterogeneity; Heaton = Wisconsin Card Sorting Test (WCST); Nelson = modified version of WCST.
* $p < .05$. ** $p < .00$.

The effect of test version (Heaton, Nelson) was significant ($Q_B = 43.02$), as age differences were more pronounced for the Nelson version of the test ($\Delta = -2.41$; 95% CI = –2.76, –2.06; $k = 14$) than the Heaton version ($\Delta = -1.19$; 95% CI = –1.30, –1.09; $k = 41$). Partitioning the within-group fit statistic according to levels of test version indicated significant within-group heterogeneity for both the Heaton ($Q_{W1} = 237.63$) and Nelson ($Q_{W2} = 117.35$) versions. Thus, as in the categories achieved analysis, the Nelson version of the WCST was more sensitive to age differences than the Heaton version. A follow-up test was conducted that excluded all young-old participants, in a manner identical to the categories achieved meta-analysis. Results showed that even after excluding all young-old participants, the Nelson version of the test remained more sensitive to age differences than the Heaton version ($Q_B = 35.44$, $p < .00$).

In contrast to the categories achieved model, analysis of the effect of test modality (manual, computer based) revealed that the

-14.	8
-13.	
-12.	
-11.	
-10.	
-9.	
-8.	
-7.	6
-6.	0
-5.	0, 0, 4, 7, 9, 9
-4.	8
-3.	0, 0, 1, 2, 3, 4, 5
-2.	2, 3, 3, 3, 4, 7, 8, 9
-1.	0, 1, 1, 3, 4, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 8, 9
-0.	1, 2, 3, 6, 6, 7, 8, 8, 9, 9, 9
0.	0, 4

Figure 2. Stem-and-leaf plot of weighted effect sizes for the number of perseverative errors committed.

---

[6] Results showed that the mean unweighted effect size ($\Delta = -2.68$) for the number of perseverative errors committed was markedly larger than the mean weighted effect size and was significantly greater than zero, $t(54) = -7.87$, $p = .00$. The median unweighted effect size was 1.90. In addition, the mean weighted effect size based on a pooled measure of variability was significantly smaller ($g = -.83$) than the mean weighted Δ-based effect size reported.

[7] Specifically, the mean weighted effect size for the Nelson version of the test changed from –2.41 to –2.27 (95% CI = –2.63, –1.92; $k = 12$) when outliers were excluded. Likewise, the mean weighted effect size for older adults with unspecified levels of education changed from –2.91 to –2.73 (95% CI = –3.16, –2.31; $k = 13$) when outliers were excluded.

number of perseverative errors committed did not differ between modalities ($Q_B < 1$). Specifically, the mean weighted effect size for manual administration (Δ = –1.30; 95% CI = –1.42, –1.18; $k = 46$) was almost identical to that for computer-based administration (Δ = –1.28, 95% CI = –1.47, –1.08; $k = 9$). Partitioning the within-group fit statistic according to levels of test modality indicated that significant within-group heterogeneity existed for both versions of the test.

Analysis of the effect of education (less than 12 years, 12–15 years, more than 15 years, unspecified) indicated that effect sizes differed on the basis of the average number of years of education attained ($Q_B = 77.96$), with education negatively related to the number of perseverative errors committed. Specifically, participants who averaged less than 12 years of education (Δ = –1.70; 95% CI = –1.95, –1.44; $k = 10$) made more perseverative errors than did participants who averaged between 12 to 15 years of education (Δ = –1.11; 95% CI = –1.24, –0.97; $k = 23$). In addition, participants who averaged between 12 to 15 years of education made slightly fewer errors than participants who averaged over 15 years of education (Δ = –1.09; 95% CI = –1.28, –0.89; $k = 7$). Partitioning the within-group fit statistic (Q_W) according to levels of education indicated that significant within-group heterogeneity existed for each group.

As in the analysis of categories achieved, data for participants whose average level of education was unspecified (Δ = –2.91; 95% CI = –3.33, –2.49; $k = 15$) were most similar to data for participants with the least education. Part of this effect size may be derived from the distribution of ages for participants with unspecified levels of education. Further inspection of these data indicated that one study was classified as having young-old participants, nine studies were classified as having middle-old participants, and five studies were classified as having old-old participants. When old-old participants were excluded from the analysis, the resulting effect size was lowered substantially (Δ = –2.51; 95% CI = –2.97, –2.06; $k = 10$). Thus, this provides a partial account of differences in effect sizes derived for participants whose education is unspecified. Otherwise, it is not entirely clear why effect sizes differed for this group.

Hierarchical regression models. All moderator variables were examined in a hierarchical weighted least squares regression analysis. Age was entered first, followed by education, which was followed by a methodological block with test version and test modality. Results showed that age alone explained approximately a fifth of the variability in effect sizes ($R^2 = .21$). Entering education explained an additional 14% of the variability in effect sizes ($Q_{\text{CHANGE}} = 56.80$, $p < .00$). Finally, entering a block containing test version and test modality explained approximately 3.4% more of the variability in effect sizes ($Q_{\text{CHANGE}} = 13.47$, $p < .01$). Thus, the full model accounted for well over a third of the variability in effect sizes obtained ($R^2 = .38$).

An identical analysis was also conducted that excluded effect sizes derived from older subjects whose level of education was not specified ($k = 15$). Unlike the categories achieved analysis, excluding subjects in this manner did not increase the amount of variability explained by the model. Specifically, results showed that age accounted for almost a quarter of the variability in effect sizes ($R^2 = .24$). Education explained an additional 1.4% of the variability, a value that was not significant ($Q_{\text{CHANGE}} = 3.60$, $p = .17$). Entering test version and test modality accounted for an additional 4% of the variability in effect sizes ($Q_{\text{CHANGE}} = 10.06$, $p < .01$). Therefore, the full model excluding effect sizes for subjects whose level of education was not specified accounted for less than a third of the variability in effect sizes obtained ($R^2 = .29$).

Discussion. Consistent with the previous meta-analysis of categories achieved, analysis of the number of perseverative errors committed revealed that age differences were most evident for the oldest age groups. As well, effect sizes were largest for those age groups with less than 12 years of education attained on average, a finding consistent with the extant normative data (Heaton et al., 1993). Regression analyses also indicated that education contributed substantial unique explained variability beyond that explained by age. As in the meta-analysis of categories achieved, the Nelson version of the test was more sensitive to age differences than the Heaton version. However, in contrast to the categories achieved model, effect sizes did not differ on the basis of test modality. Results from hierarchical regression analyses indicated that these variables accounted for a small portion of explained variability in addition to that explained by age and education.

General Discussion

Findings from the meta-analyses reported demonstrate that the WCST is sensitive to age differences, with results revealing well over a standard deviation difference between younger and older adults on both measures of interest. In particular, the mean weighted effect size for the number of categories achieved was –1.13 and the mean weighted effect size for the number of perseverative errors committed was –1.29. These two measures are highly correlated (Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Lineweaver et al., 1999) and often load on the same factor (Boone, Ponton, Gorsuch, Gonzalez, & Miller, 1998; Koren et al., 1998; Salthouse, Fristoe, & Rhee, 1996). In addition, both categories achieved (e.g., J. R. Crawford, Bryan, Luszcz, Obonsawin, & Stewart, 2000; Glisky et al., 1995; Johnson, De Leonardis, Hashtroudi, & Ferguson, 1995) and perseverative errors committed (e.g., Arbuckle & Gold, 1993; Head, Raz, Gunning-Dixon, Williamson, & Acker, 2002; Raz, Gunning-Dixon, Head, Dupuis, & Acker, 1998) are used with approximately equal consistency as predictors in investigations of age-related changes in cognition or as measures of executive abilities or frontal integrity. The current data suggest that the measure of perseverative errors committed is marginally more sensitive to age differences in comparison with categories achieved and may be a better metric of executive function if a single score from the WCST is to be used.

Moderator Variables

Results from analyses of moderator variables revealed reliable effects of age and education for both measures examined. Regarding age, performance declines were particularly steep for the oldest participants examined (i.e., the old-old group), as they demonstrated close to a two standard deviation difference for the number of categories achieved (Δ = –1.64) and the number of perseverative errors committed (Δ = –1.90) in comparison with younger participants (cf. Axelrod & Henry, 1992; Heaton et al., 1993). In addition, the influence of education was most apparent for those groups of older participants who averaged less than 12 years of education. Participants in this group committed more perseverative errors and achieved fewer categories than participants who averaged more than 12 years of education. Clearly, age is the more powerful predictor of performance. However, both analyses, particularly that for perseverative errors, demonstrated that education did account for a significant portion of explained variability.

Data from the current meta-analysis also showed that effect sizes differed on the basis of test modality only for the number of categories achieved. On the basis of their comparison of the manual version of the WCST with a computerized version, Feldstein et al. (1999) suggested that normative data from the WCST was not applicable to computerized versions. The null result for the effect of test modality for the number of perseverative errors does not necessarily contravene their findings. In particular, Feldstein et al. examined the performance of several groups of younger participants on various computerized versions of the WCST that differed on the basis of the method of input, including conditions using a mouse, keyboard, or touch screen. They reported several departures from the normative data. For example, participants who used a keyboard or touch screen made significantly more errors than did participants in a manual version, whereas performance for participants who used a mouse was quite similar to the manual version. Test modality may thus interact with the method of input. These issues may be further exacerbated when testing older adults, who report less computer usage than younger adults and require more time to acquire computer skills (e.g., Charness, Kelley, Bosman, & Mottram, 2001). Consequently, although the current data provide inconsistent support for the notion that computerized versions of the WCST produce data that differ from manual versions, further investigation is warranted to determine whether manual and computerized versions of the WCST are equivalent measures.

Findings from both meta-analyses also indicated that Nelson's (1976) MCST resulted in more robust age differences than the Heaton (1981; Heaton et al., 1993) WCST version of the test, even when young-old participants were excluded from the analysis. This is particularly important in light of recent work demonstrating that the MCST does not differentiate between frontal and nonfrontal patients (Demakis, 2003), a finding echoed by de Zubicaray and Ashton's (1996) review of the MCST. One possibility is that the MCST may be less reliant on the PFC than the Heaton version of the test.[8] Regardless, although the MCST may be of questionable validity as a tool to diagnose lesion location, it seems to provide more than adequate sensitivity to age differences. The administration advantages of the MCST are of course apparent, as it takes less time to administer and leads to less frustration for the participant. However, it is not clear whether these different versions of the test, with the reduction in cards from 128 to 48 for the MCST, measure precisely the same construct. As previously noted, the MCST requires fewer correct sorts (six) than the WCST before a category is achieved. Grant and Berg (1948), in some of the original work with the WCST, demonstrated that performance improved as the number of correct sorts required to complete a category was increased. Presumably, increasing the number of correct sorts required facilitates learning of the sorting rule. The greater age differences reported for the MCST perhaps reflect this, as 10 sorts (WCST) may facilitate learning of the sorting rules to a greater extent than six sorts (MCST), leading to better performance.

Direct evidence comparing the WCST and MCST is limited. Greve and Smith (1991) administered the MCST and an earlier standardized version of the WCST (Heaton, 1981) to 18 older participants. Participants were administered one version of the test during an initial session and were administered the other version several weeks later. A strong, significant correlation between the number of perseverative errors committed on both tests was obtained ($r = .60$). However, the correlation between the number of categories achieved on the two tests was far weaker ($r = .38$) and did not reach significance. Part of this failure to detect a significant correlation may be due to the fact that a number of participants performed near ceiling on the measure of categories achieved, truncating the variability in scores. It is also apparent that a relatively small number of participants were tested. Thus, more conclusive evidence is needed in the form of direct comparisons between performance on the WCST and the MCST.

Accounting for Age Differences

Despite the ubiquity of the WCST as a measure of executive function in aging populations (Bryan & Luszcz, 2000), there are relatively few explanatory accounts of the source of age-related differences in performance. As previously noted, the WCST initially gained popularity as a diagnostic tool to detect damage to the PFC (Drewe, 1974; Milner, 1963). Given that the PFC deteriorates rapidly with age (Raz, 2000), age-related declines in performance on the WCST may be explained as the result of a concomitant decrease in frontal lobe integrity (cf. West, 1996). Several studies (Head et al., 2002; Raz et al., 1998; Schretlen et al., 2000) have demonstrated that performance on the WCST is correlated with volumetric measures of the PFC. However, this does not necessarily elucidate the underlying cognitive processes that mediate performance.

One possibility is that age-related declines on the WCST are linked to a more fundamental age-related decline in working memory (Baddeley, 1986), a system presumed to be dependent on the PFC (see Kane & Engle, 2002, for a review). Working memory is likely crucial for performance on the WCST given that the participant must keep in mind information about previous sorts while simultaneously processing information to determine the next sort (Dehane & Changeux, 1991; Hartman, Bolton, & Fehnel, 2001; Hartman, Steketee, Silva, Lanning, & Andersson, 2003; Kimberg & Farah, 1993; but see Stratta et al., 1997). Hartman et al. (2001; Experiment 2) have tested this hypothesis by reducing working memory demands on older participants during performance of the WCST. Specifically, older and younger participants were either administered the standard version of the WCST or a modified version that provided explicit visual feedback (i.e., a cardboard arrow labeled *YES* or *NO* was placed above the most recent sort), in addition to the standard oral feedback typically given.[9] Results from the standard administration condition indicated that older participants achieved significantly fewer categories and committed significantly more perseverative errors than younger participants. However, when administered the modified version, age differences were largely eliminated. Therefore, age differences on the WCST may be a function of reduced working

---

[8] I thank an anonymous reviewer for this suggestion.

[9] This condition was not included in the current meta-analysis as it deviated a great deal from the standard administration without any comparable manipulations that might allow it to be analyzed as a moderator variable.

memory, a capacity with well-established age-related deficits (Verhaeghen, Marcoen, & Gossens, 1993).

Alternatively, Salthouse and colleagues (Fristoe et al., 1997; Salthouse et al., 1996) have suggested that age differences in performance on the WCST are the result of a more basic deficit in processing speed (cf. Salthouse, 1996). Evidence for this perspective comes from work demonstrating that a large proportion of age-related variance in performance of the WCST is reduced when measures of processing speed are taken into account. For example, Salthouse et al. (1996) removed the majority of the age-related variability for the number of perseverative errors committed after accounting for measures of perceptual and motor speed. Furthermore, Fristoe et al. (1997) found that age-related variability in the use of feedback during performance of the WCST was accounted for by measures of working memory and speed of processing. In turn, age-related differences in working memory were largely accounted for by speed of processing measures.

Thus, there is some question as to whether age differences on the WCST are the result of reduced working memory, general cognitive slowing, or a combination of the two, as they may be redundant constructs (cf. Charness & Schultetus, 1998). Such questions also leave unresolved the matter of exactly which executive function is tapped by the WCST. On the basis of structural equation modeling, Miyake et al. (2000) have suggested that performance on the WCST reflects a factor termed *Shifting,* which refers to the ability to shift back and forth between multiple tasks. Data from Hartman et al. (2001; see also Kimberg & Farah, 1993) suggest that the WCST also involves processes devoted to updating working memory.

Alternative Estimates of Effect Size and Other Caveats

The estimate of effect size reported in the current study (Δ) used a measure of variability based on the standard deviations of younger participants. This may potentially inflate estimates of effect size, and findings should be interpreted cautiously for this reason. The Appendix provides a more conservative set of effect size estimates using a pooled measure of variability (Hedges's g). Even with a more conservative estimate of effect size, a number of findings remain unchanged. For example, for the categories achieved model, reliable effects of age and education are present. In addition, g-based effect sizes yielded reliable effects of age and no effect of test modality for the perseverative errors model, consistent with findings from the model based on Δ.

However, several differences between effect size estimates were evident. Foremost among these, the effect of test version did not differ for either model with g-based effect sizes. In the case of the categories achieved model, this is less problematic, as the difference between the Heaton (1981) and Nelson (1976) versions approached significance ($p = .09$). However, for the perseverative errors model, the difference between the Heaton and Nelson versions was minimal ($Q_B < 1$). This is perhaps not surprising, given that the measure of perseverative errors is characterized by heightened variability in performance for older adults (see Table 1). Thus, effect sizes based on a pooled estimate of variability may perhaps be unduly influenced by this greater variability. Such an explanation may also provide an account of why education, a reliable moderator variable in normative data for the WCST (e.g., Heaton et al., 1993), was only marginally significant ($p = .06$) for the perseverative errors model with g-based effect sizes. A useful perspective for compromising between these two effect size estimates may be to view Δ-based effect sizes as statistically sound (cf. Rosenthal, 1994) but providing an upper boundary estimate of age differences. Such estimates, at the very least, point to trends in the WCST, including possible differences between the Nelson and Heaton versions of the test. Thus, Δ-based estimates may be evaluated in concert with g-based effect sizes, which, though statistically questionable in the context of the current data, provide a lower boundary estimate of the true effect size.

It is also important to bear in mind that the meta-analyses reported examined only a portion of the literature on the WCST and aging. This follows primarily because a number of studies did not provide a young control group with which to compare age-related differences in performance (see Footnote 2). Clearly, an increase in the number of studies would likewise increase the statistical power of the analyses presented. This would be particularly relevant for examining the effects of test version and modality on WCST performance. However, the basic pattern of age differences would likely remain unaltered, as an enormous number of studies would have to exist (approximately 9,400 in the case of the categories achieved model and approximately 12,000 for the perseverative errors model) to remove what are highly reliable age differences.

Summary and Conclusion

Overall, the current study indicates that age differences are robust on the two most common measures of performance on the WCST, the number of categories achieved and the number of perseverative errors committed. These differences are most pronounced for the number of perseverative errors committed and are particularly acute for the oldest age groups. Furthermore, these data revealed reliable effects of test version and education. The effect of test modality was inconsistent, as effect sizes differed only for the measure of categories achieved. Accounts of age-related differences on the WCST are less clear, but the data indicate that it may tap executive functions, such as updating working memory and shifting. In summary, the WCST is a complex measure of executive function characterized by reliable and robust age differences.

References

References marked with an asterisk indicate studies included in both meta-analyses. References marked with two asterisks indicate studies included in only the categories achieved meta-analysis, and references marked with three asterisks indicate studies included in only the perseverative errors committed meta-analysis.

Anderson, S. W., Damasio, H., Jones, R. D., & Tranel, D. (1991). Wisconsin Card Sorting Test performance as a measure of frontal damage. *Journal of Clinical and Experimental Neuropsychology, 13,* 909–922.

Arbuckle, T. Y., & Gold, D. P. (1993). Aging, inhibition, and verbosity. *Journal of Gerontology: Psychological Sciences, 5B,* P225–P232.

Axelrod, B. N., & Henry, R. R. (1992). Age-related performance on the Wisconsin Card Sorting, Similarities, and Controlled Oral Word Association Tests. *Clinical Neuropsychologist, 6,* 16–26.

Baddeley, A. (1986). *Working memory.* London: Clarendon Press.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A,* 5–28.

*Beatty, W. W. (1993). Age differences on the California Card Sorting Test: Implications for the assessment of problem solving by the elderly. *Bulletin of the Psychonomic Society, 31,* 511–514.

Beatty, W. W., Monson, N., & Goodkin, D. E. (1989). Access to semantic memory in Parkinson's disease and multiple sclerosis. *Journal of Geriatric Psychiatry and Neurology, 2,* 153–162.

Bondi, M. W., Houston, W. S., Salmon, D. P., Corey-Bloom, J., Katzman, R., Thal, L. J., et al. (2003). Neuropsychological deficits associated with Alzheimer's disease in the very-old: Discrepancies in raw vs. standardized scores. *Journal of the International Neuropsychological Society, 9,* 783–795.

Boone, K. B., Ghaffarian, S., Lesser, I. M., & Hill-Gutierrez, E. (1993). Wisconsin Card Sorting Test performance in healthy, older adults: Relationship to age, sex, education, and IQ. *Journal of Clinical Psychology, 49,* 54–60.

Boone, K. B., Ponton, M. O., Gorsuch, R. L., Gonzalez, J. J., & Miller, B. L. (1998). Factor analysis of four measures of prefrontal lobe functioning. *Archives of Clinical Neuropsychology, 13,* 585–595.

Brown, R. G., & Marsden, C. D. (1988). An investigation of the phenomenon of "set" in Parkinson's disease. *Movement Disorders, 2,* 152–161.

Bryan, J., & Luszcz, M. A. (2000). Measurement of executive function: Considerations for detecting adult age differences. *Journal of Clinical and Experimental Neuropsychology, 22,* 40–55.

*Bryan, J., & Luszcz, M. A. (2001). Adult age differences in self-ordered pointing task performance: Contributions from working memory, executive function, and speed of information processing. *Journal of Clinical and Experimental Neuropsychology, 23,* 608–619.

Butler, M., Retzlaff, P., & Vanderploeg, R. (1991). Neuropsychological test usage. *Professional Psychology: Research and Practice, 22,* 510–512.

Charness, N., Kelley, C. L., Bosman, E. A., & Mottram, M. (2001). Word-processing training and retraining: Effects of adult age, experience, and interface. *Psychology and Aging, 16,* 110–127.

Charness, N., & Schultetus, R. S. (1998, April). *Eye movements and digit symbol performance.* Poster presented at the Cognitive Aging Conference, Atlanta, GA.

Compton, D. M., Bachman, L. D., & Logan, J. A. (1997). Aging and intellectual ability in young, middle-aged, and older educated adults: Preliminary results from a sample of college faculty. *Psychological Reports, 81,* 79–90.

*Crawford, J. R., Bryan, J., Luszcz, M. A., Obonsawin, M. C., & Stewart, L. (2000). The executive decline hypothesis of cognitive aging: Do executive deficits qualify as differential deficits and do they mediate age-related memory decline? *Aging, Neuropsychology, and Cognition, 7,* 9–31.

*Crawford, S., & Channon, S. (2002). Dissociation between performance on abstract tests of executive function and problem solving in real-life-type situations in normal aging. *Aging and Mental Health, 6,* 12–21.

Daigneault, G., Joly, P., & Frigon, J. Y. (2002). Executive functions in the evaluation of accident risk of older drivers. *Journal of Clinical and Experimental Neuropsychology, 24,* 221–238.

Daigneault, S., & Braun, C. M. J. (1993). Working memory and the self-ordered pointing task: Further evidence of early prefrontal decline in normal aging. *Journal of Clinical and Experimental Neuropsychology, 15,* 881–895.

*Daigneault, S., Braun, C. M., & Whitaker, H. A. (1992). Early effects of normal aging on preservative errors. *Developmental Neuropsychology, 8,* 99–114.

*Davis, H. P., Cohen, A., Gandy, M., Colombo, P., VanDusseldorp, G., Simolke, N., & Romano, J. (1990). Lexical priming deficits as a function of age. *Behavioral Neuroscience, 104,* 288–297.

*Degl'Innocenti, A., & Backman, L. (1996). Aging and source memory: Influences of intention to remember and associations with frontal lobe tests. *Aging, Neuropsychology, and Cognition, 3,* 307–319.

Dehane, S., & Changeux, J. (1991). The Wisconsin Card Sorting Test: Theoretical analysis and modeling in a neuronal network. *Cerebral Cortex, 1,* 62–79.

Demakis, G. J. (2003). A meta-analysis review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology, 17,* 255–264.

de Zubicaray, G., & Ashton, R. (1996). Nelson's (1976) Modified Card Sorting Test: A review. *The Clinical Neuropsychologist, 10,* 245–254.

de Zubicaray, G. I., Smith, G. A., Chalk, J. B., & Semple, J. (1998). The Modified Card Sorting Test: Test–retest stability and relationships with demographic variables in a healthy older adult sample. *British Journal of Clinical Psychology, 37,* 457–466.

Drewe, E. A. (1974). The effect of type and area of brain lesion on Wisconsin Card Sorting Test performance. *Cortex, 10,* 159–170.

Duncan, J. (1993). Selection of input and goal in the control of behavior. In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control. A tribute to Donald Broadbent* (pp. 53–71). Oxford, England: Oxford University Press.

Duncan, J. (1995). Attention, intelligence, and the frontal lobes. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 721–733). Cambridge, MA: MIT Press.

*Fabiani, M., & Friedman, D. (1995). Changes in brain activity patterns in aging—The novelty oddball. *Psychophysiology, 32,* 579–594.

*Fabiani, M., & Friedman, D. (1997). Dissociations between memory for temporal order and recognition memory in aging. *Neuropsychologia, 35,* 129–141.

Feldstein, S. N., Keller, F. R., Portman, R. E., Durham, R. L., Klebe, K. J., & Davis, H. P. (1999). A comparison of computerized and standard versions of the Wisconsin Card Sorting Test. *Clinical Neuropsychologist, 13,* 303–313.

*Fisk, J. E., & Sharp, C. (2002). Syllogistic reasoning and cognitive ageing. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 55A,* 1273–1293.

*Fristoe, N. M., Salthouse, T. A., & Woodard, J. L. (1997). Examination of age-related deficits on the Wisconsin Card Sorting Test. *Neuropsychology, 11,* 428–436.

*Fucetola, R., Seidman, L. J., Kremen, W. S., Faraone, S. V., Goldstein, J. M., & Tsuang, M. T. (2000). Age and neuropsychologic function in schizophrenia: A decline in executive abilities beyond that observed in healthy volunteers. *Biological Psychiatry, 48,* 137–146.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Glisky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9,* 229–235.

Goldman-Rakic, P. S. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In F. Plum (Ed.), *Handbook of physiology*: *The nervous system* (Vol. 5, pp. 373–417). Bethesda, MD: American Physiological Society.

Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of reinforcement and ease of shifting to new responses in a Weigel-type card-sorting problem. *Journal of Experimental Psychology, 38,* 404–411.

Greve, K. W., & Smith, M. C. (1991). A comparison of the Wisconsin Card Sorting Test with the Modified Card Sorting Test for use with older adults. *Gerontology and Geriatrics Education, 11,* 57–65.

Haaland, K., Vranes, L., Goodwin, J., & Garry, P. (1987). Wisconsin Card Sorting Test performance in a healthy elderly population. *Journal of Gerontology, 42,* 345–346.

Hart, R. P., Kwentus, J. A., Wade, J. B., & Taylor, J. R. (1988). Modified Wisconsin Sorting Test in elderly normal, depressed, and demented patients. *The Clinical Neuropsychologist, 2,* 49–56.

*Hartman, M., Bolton, E., & Fehnel, S. E. (2001). Accounting for age differences on the Wisconsin Card Sorting Test: Decreased working memory, not inflexibility. *Psychology and Aging, 16,* 385–399.

Hartman, M., Steketee, M. C., Silva, S., Lanning, K., & Andersson, C. (2003). Wisconsin Card Sorting Test performance in schizophrenia: The role of working memory. *Schizophrenia Research, 63,* 201–217.

***Head, D., Raz, N., Gunning-Dixon, F., Williamson, A., & Acker, J. D. (2002). Age-related differences in the course of cognitive skill acquisition: The role of regional cortical shrinkage and cognitive resources. *Psychology and Aging, 17,* 72–84.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual.* Odessa, FL: Psychological Assessment Resources.

*Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded.* Odessa, FL: Psychological Assessment Resources.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 286–299). New York: Russell Sage Foundation.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Houx, P. J., Jolles, J., & Vreeling, F. W. (1993). Stroop interference: Aging effects assessed with the Stroop Color–Word Test. *Experimental Aging Research, 19,* 209–224.

Howard, D. V. (1980). Category norms: A comparison of the Battig and Montague (1960) with the response of adults between the ages of 20 and 80. *Journal of Gerontology, 35,* 225–231.

*Isingrini, M., & Vazou, F. (1997). Relation between fluid intelligence and frontal lobe functioning in older adults. *International Journal of Aging and Human Development, 45,* 99–109.

Jenkins, R. L., & Parson, O. A. (1978). Cognitive deficits in male alcoholics as measured by a modified Wisconsin Card Sorting Test. *Alcohol Technical Reports, 7,* 76–83.

**Johnson, M. K., De Leonardis, D. M., Hashtroudi, S., & Ferguson, S. A. (1995). Aging and single versus multiple cues in source monitoring. *Psychology and Aging, 10,* 507–517.

Joyce, E. M., & Robbins, T. W. (1991). Frontal lobe function in Korsakoff and non-Korsakoff alcoholics: Planning and spatial working memory. *Neuropsychologia, 29,* 709–723.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and Review, 9,* 637–671.

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General, 4,* 411–428.

Kinderman, S. S., Kalayam, B., Brown, G. G., Burdick, K. E., & Alexopoulos, G. S. (2000). Executive functions and P300 latency in elderly depressed patients and control subjects. *American Journal of Psychiatry, 8,* 57–65.

*Kirkby, K. C., Montgomery, I. M., Badcock, R., & Daniels, B. A. (1995). A comparison of age-related deficits in memory and frontal lobe function following oral lorazepam administration. *Journal of Psychopharmacology, 9,* 319–325.

*Klein, C., Fischer, B., Hartnegg, K., Heiss, W. H., & Roth, M. (2000). Optomotor and neuropsychological performance in old age. *Experimental Brain Research, 135,* 141–154.

**Kliegel, M., Ramuschkat, G., & Martin, M. (2003). Exekutive funktionen und prospektive gedächtnisleistung im alter—Eine differentielle analyse von ereignisund zeitbasierter prospektiver gedächtnisleistung [Executive functions and prospective memory performance in old age: An analysis of event-based and time-based prospective memory]. *Zeitschrift fuer Gerontologie und Geriatrie, 36,* 35–41.

Koechlin, E., Ody, C., & Kouneiher, F. (2003, November 14). The architecture of cognitive control in the human prefrontal cortex. *Science, 302,* 1181–1185.

Kongs, S. K., Thompson, L. L., Iverson, G. L., & Heaton, R. K. (2000). *Wisconsin Card Sorting Test—64 card version: Professional manual.* Odessa, FL: Psychological Assessment Resources.

Koren, D., Seidman, L. J., Harrison, R. H., Lyons, M. J., Kremen, W. S., Caplan, B., et al. (1998). Factor structure of the Wisconsin Card Sorting Test: Dimensions of deficit in schizophrenia. *Neuropsychology, 12,* 289–302.

*Kramer, A. F., Humphrey, D. G., Larish, J. F., Logan, G. D., & Strayer, D. L. (1994). Aging and inhibition: Beyond a unitary view of inhibitory processing in attention. *Psychology and Aging, 9,* 491–512.

*Laicona, M., Inzaghi, M. G., De Tanti, A., & Capitani, E. (2000). Wisconsin Card Sorting Test: A new global score, with Italian norms, and its relationship with the Weigl sorting test. *Neurological Sciences, 21,* 279–291.

Lineweaver, T. T., Bondi, M. W., Thomas, R. G., & Salmon, D. P. (1999). A normative study of Nelson's (1976) modified version of the Wisconsin Card Sorting Task in healthy older adults. *Clinical Neuropsychologist, 13,* 328–347.

Loranger, A. W., & Misiak, H. (1960). The performance of aged females on five non-language tests of intellectual functions. *Journal of Clinical Psychology, 16,* 189–191.

*MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function, and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging, 17,* 598–609.

*Maylor, E. A., Moulson, J. M., Muncer, A.-M., & Taylor, L. A. (2002). Does performance on theory of mind tasks decline in old age? *British Journal of Psychology, 93,* 465–485.

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian Journal of Psychology, 41,* 175–192.

Milner, B. (1963). Effects of different brain lesions on card sorting. *Archives of Neurology, 9,* 90–100.

*Mitchell, D. B., & Bruss, P. J. (2003). Age differences in implicit memory: Conceptual, perceptual, or methodological? *Psychology and Aging, 18,* 807–822.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.

Moscovitch, M., & Winocur, G. (1995). Frontal lobes, memory, and aging. *Annals of the New York Academy of Sciences, 769,* 119–150.

Mountain, M. A., & Snow, W. G. (1993). Wisconsin Card Sorting Test as a measure of frontal pathology. *Clinical Neuropsychologist, 7,* 108–118.

*Nagahama, Y., Fukuyama, H., Yamauchi, H., Katsumi, Y., Magata, Y., Shibasaki, H., & Kimura, J. (1997). Age-related changes in cerebral blood flow activation during a card sorting test. *Experimental Brain Research, 114,* 571–577.

Nelson, H. E. (1976). A modified card sorting test sensitive to frontal lobe defects. *Cortex, 12,* 313–324.

**Newson, R. S., Kemps, E. B., & Luszcz, M. A. (2003). Cognitive mechanisms underlying decrements in mental synthesis in older adults. *Aging, Neuropsychology, and Cognition, 10,* 28–43.

*Parkin, A. J., & Java, R. I. (1999). Deterioration of frontal lobe function in normal aging: Influences of fluid intelligence versus perceptual speed. *Neuropsychology, 13,* 539–545.

*Parkin, A. J., & Walter, B. M. (1991a). Aging, short-term memory, and frontal dysfunction. *Psychobiology, 19,* 175–179.

*Parkin, A. J., & Walter, B. M. (1991b). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging, 7,* 290–298.

***Perfect, T. J., & Dasgupta, Z. R. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory & Cognition, 25,* 849–858.

Raz, N. (2000). Aging of the brain and its impact on cognitive performance: Integration of structural and functional findings. In F. I. M. Craik & T. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 1–90). Mahwah, NJ: Erlbaum.

Raz, N., Gunning-Dixon, F. M., Head, D., Dupuis, J. H., & Acker, J. D. (1998). Neuroanatomical correlates of cognitive aging: Evidence from structural magnetic resonance imaging. *Neuropsychology, 12,* 95–114.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

*Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General, 132,* 566–594.

*Salthouse, T. A., Fristoe, N. M., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology, 10,* 272–285.

Schretlen, D., Pearlson, G. D., Anthony, J. C., Aylward, E. H., Augustine, A. M., Davis, A., et al. (2000). Elucidating the contributions of processing speed, executive ability, and frontal lobe volume to normal age-related differences in fluid intelligence. *Journal of the International Neuropsychological Society, 6,* 52–61.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114,* 727–741.

Shimamura, A. P. (2000). The role of the prefrontal cortex in dynamic filtering. *Psychobiology, 28,* 207–218.

*Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology, 14,* 299–309.

*Spencer, W. D., & Raz, N. (1994). Memory for facts, source, and context: Can frontal lobe dysfunction explain age-related differences? *Psychology and Aging, 9,* 149–159.

Stratta, P., Daneluzzo, E., Prosperinin, P., Bustini, M., Mattei, P., & Rossi, A. (1997). Is Wisconsin Card Sorting Test performance related to "working memory" capacity? *Schizophrenia Research, 27,* 11–19.

*Trott, C. T., Friedman, D., Ritter, W., Fabiani, M., & Snodgrass, J. G. (1999). Episodic priming and memory for temporal source: Event-related potentials reveal age-related differences in prefrontal functioning. *Psychology and Aging, 14,* 390–413.

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology, 11,* 138–146.

United States Census Bureau. (2001, June). *The older population in the United States: March 2000.* Retrieved December 17, 2003, from http://www.census.gov/population/www/socdemo/age/ppl-147.html

van den Broek, M. D., Bradshaw, C. M., & Szabadi, E. (1993). Utility of the Modified Wisconsin Card Sorting Test in neuropsychological assessment. *British Journal of Clinical Psychology, 32,* 333–343.

Verhaeghen, P., Marcoen, A., & Gossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journal of Gerontology: Psychological Sciences, 48B,* P157–P171.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10,* 119–125.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120,* 272–292.

Whelihan, W. M., & Lesher, E. L. (1985). Neuropsychological changes in frontal function with aging. *Developmental Neuropsychology, 1,* 371–380.

***Winocur, G., Moscovitch, M., & Stuss, D. T. (1996). Explicit and implicit memory in the elderly: Evidence for double dissociation involving medial temporal- and frontal-lobe functions. *Neuropsychology, 10,* 57–65.

Woods, S. P., & Tröster, A. I. (2003). Prodromal frontal/executive dysfunction predicts incident dementia in Parkinson's disease. *Journal of the International Neuropsychological Society, 9,* 17–24.

Appendix

g-Based Effect Size Model for Categories Achieved

This appendix describes the calculation and results for a pooled estimate of effect size. The effect size estimate used (Δ) in the current study was based on a single measure of variability derived from younger participants. This perhaps provides a more liberal estimate of effect size than the true population effect size. Thus, Hedges's *g* was also calculated to provide a more conservative estimate of effect size based on pooled standard deviations. The general formula for Hedges's *g* is as follows:

$$g = \frac{M_e - M_c}{\sqrt{\frac{(n_e - 1)s_e^2 + (n_c - 1)s_c^2}{n_e + n_c - 2}}},$$

where M_e is the mean of the experimental group (i.e., older participants), M_c is the mean of the control group (i.e., younger participants), n_e is the number of participants in the experimental group, n_c is the number of participants in the control group, s_e is the standard deviation of the experimental group, and s_c is the standard deviation of the control group. A key feature of this estimate for present purposes is that the denominator pools standard deviations from both groups. The weighting function (w_i) for *g*-based effect sizes is identical to that for Δ-based effect sizes, calculated as the inverse of the conditional variance (v_i)—that is, $1/v_i$. However, the conditional variance for *g* is calculated slightly differently, as follows:

$$v_i = [(n_1 + n_2)/(n_1 - n_2)] + [g^2/2(n_1 + n_2 - 2)],$$

where n_1 is the sample size of the first group, n_2 is the sample size of the second group, and *g* is the effect size estimate. Results and analyses of *g*-based effect sizes are presented in Tables A1 and A2.

(*Appendix continues*)

Table A1

Summary of Main Effect Analyses for the Number of Categories Achieved for g-Based Effect Sizes

			95% CI		*Q*	
Variable	*k*	*g*	Lower	Upper	Within	Between
Age						
Young-old	7	−0.58	−0.79	−0.37	2.17	22.00**
Middle-old	31	−0.83	−0.93	−0.74	47.86*	
Old-old	15	−1.21	−1.38	−1.03	13.14	
Version						
Heaton	40	−0.84	−0.92	−0.75	54.67	2.87
Nelson	13	−1.01	−1.20	−0.83	27.63*	
Modality						
Manual	45	−0.85	−0.94	−0.77	74.55*	1.57
Computer based	8	−0.97	−1.14	−0.80	3.85	
Education						
<12 years	8	−1.30	−1.51	−1.09	14.75	22.41**
12–15 years	25	−0.79	−0.91	−0.68	23.28	
>15 years	8	−0.73	−0.88	−0.58	10.53	
Unspecified	12	−0.97	−1.20	−0.74	14.20	

Note. k = number of effect sizes; g = mean weighted effect size; CI = confidence interval; Q = heterogeneity. Heaton = Wisconsin Card Sorting Test (WCST); Nelson = modified version of WCST.

* $p < .05$. ** $p < .00$.

Table A2

Summary of Main Effect Analyses for the Number of Perseverative Errors Committed for g-Based Effect Sizes

			95% CI		Q	
Variable	*k*	*g*	Lower	Upper	Within	Between
Age						
Young-old	7	−0.61	−0.82	−0.40	2.89	14.84**
Middle-old	31	−0.80	−0.90	−0.70	63.94**	
Old-old	17	−1.10	−1.27	−0.94	16.86	
Version						
Heaton	41	−0.83	−0.92	−0.74	79.32**	0.40
Nelson	14	−0.90	−1.08	−0.71	18.81	
Modality						
Manual	46	−0.82	−0.91	−0.73	84.13**	0.78
Computer based	9	−0.90	−1.07	−0.74	13.62	
Education						
<12 years	10	−1.03	−1.21	−0.84	18.84*	7.52
12–15 years	23	−0.78	−0.90	−0.66	48.34**	
>15 years	7	−0.73	−0.91	−0.56	6.07	
Unspecified	15	−0.95	−1.14	−0.75	17.76	

Note. k = number of effect sizes; g = mean weighted effect size; CI = confidence interval; Q = heterogeneity. Heaton = Wisconsin Card Sorting Test (WCST); Nelson = modified version of WCST.

* $p < .05$. ** $p < .00$.

Received December 29, 2003
Revision received March 24, 2004
Accepted March 29, 2004 ■