


Review

A Meta-Analysis of Relationships between Measures of Wisconsin Card Sorting and Intelligence

Bruno Kopp ^{1,*} , Natasha Maldonado ¹, Jannik F. Scheffels ¹, Merle Hendel ¹ and Florian Lange ^{1,2}

¹ Department of Neurology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany; n.maldonado@posteo.de (N.M.); scheffels.jannik@mh-hannover.de (J.F.S.); merle_hendel@t-online.de (M.H.); florian.lange@kuleuven.be (F.L.)

² Behavioral Engineering Research Group, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium

* Correspondence: kopp.bruno@mh-hannover.de; Tel.: +49-511-5322439; Fax: +49-511-5323115

Received: 14 November 2019; Accepted: 26 November 2019; Published: 29 November 2019



Abstract: The Wisconsin Card Sorting Test (WCST) represents a widely utilized neuropsychological assessment technique for executive function. This meta-analysis examined the discriminant validity of the WCST for the assessment of mental shifting, considered as an essential subcomponent of executive functioning, against traditional psychometric intelligence tests. A systematic search was conducted, resulting in 72 neuropsychological samples for the meta-analysis of relationships between WCST scores and a variety of intelligence quotient (IQ) domains. The study revealed low to medium-sized correlations with IQ domains across all WCST scores that could be investigated. Verbal/crystallized IQ and performance/fluid IQ were indistinguishably associated with WCST scores. To conclude, the WCST assesses cognitive functions that might be partially separable from common conceptualizations of intelligence. More vigorous initiatives to validate putative indicators of executive function against intelligence are required.

Keywords: Wisconsin Card Sorting Test; intelligence; executive function; shifting; meta-analysis; psychometrics; validity

1. Introduction

Psychological functions of the frontal lobes of the human brain remain enigmatic despite decades of research (see [1–5] for overviews). A substantial part of this research was based on the Wisconsin Card Sorting Test (WCST), which was originally developed in the 1940s [6,7]. The WCST requires sorting cards and using feedback to shift between different task rules. It consists of cards depicting simple geometric figures that vary in color, shape, and number. Examinees have to sort cards in accordance with one of three viable rules, i.e., according to the color, shape, or number of the depicted object(s). In order to identify the currently valid sorting rule, examinees have to rely on verbal feedback, which is provided by the examiner on each trial. Positive feedback indicates that cards were matched according to the correct rule on the current trial, whereas negative feedback indicates that the applied rule was invalid. The examiner changes the task rule after a number of successively correct card sorts have been conducted by the examinee. In this regard, the WCST bears similarities to task-switching paradigms that are often utilized in experimental psychology ([8] for overview; [9–11]). Popular WCST scores include (1) the number of completed runs of correct card sorts (usually referred to as ‘categories’), (2) the number of perseverative errors or responses, (3) the number of non-perseverative errors, (4) the number of failures to maintain a rule (or ‘set’), and (5) the number of total errors (see [12,13] for overviews).

The WCST was introduced to neuropsychology in the 1950s for assessing higher visual functions [14]. Based on this study, it was originally thought that performance on the WCST was sensitive to (traumatic) posterior brain lesions. However, Milner's seminal study [15] revealed that the presence of (massive, unilateral) prefrontal excisions was associated with frequently occurring rule perseverations on the WCST, despite preserved indicators of intelligence. (Milner's publication [15] does not designate the specific intelligence test (e.g., FSIQ, VIQ, PIQ) that was utilized for the quantification of intelligence. Given that the patients were doubtlessly studied in the 1950s, and given the information provided in the related paper [16], the most probable interpretation of 'the IQ' in Milner's study is that this acronym represents the IQ that was obtained from the Wechsler–Bellevue Intelligence Scale [17]. A longstanding criticism of intelligence tests such as the early versions of the Wechsler batteries (up to WAIS-R) was that there was disproportionate emphasis on measures of Gc. By contrast, these early versions of the Wechsler batteries had weaker representation of measures of Gf.) Hence, this study laid the ground for the proposition that frontal lobe functions can be assessed behaviorally by means of the WCST [18–20]. In the decades to follow, the idea that the WCST measures frontal lobe functions was successively replaced by the conception of the WCST as a test of executive function (EF), thereby relaxing the otherwise strong constraints about the neuroanatomical substrates of WCST performance (see [21–25] for reviews). EFs encompass higher cognitive functions, usually defined as a set of domain-general cognitive control mechanisms supporting goal-directed behavior (e.g., [26]), but their exact nature remains a matter of debate [27–30]. The unity/diversity model of EF, which represents a well-validated individual differences model of EF, proposes that specific EF factors of updating and shifting exist next to a general factor that is involved in all EF tasks [31–34]. Applying an early version of their model, Miyake and colleagues [35] showed that the number of perseverative errors committed on the WCST specifically reflected individual differences in the shifting factor.

Since the time of Milner's study [15], the WCST has received several modifications (most notably by [36]) and multiple standardizations [37–39]. The availability of standardized test versions, as well as the prevalent acceptance of the EF construct, may have contributed to widespread dissemination of the WCST in clinical neuropsychology. The WCST is currently the most popular assessment instrument for EF [13]. Behavioral performance on the WCST is commonly interpreted in terms of mental shifting, a process which represents an important subcomponent of EF, and ensures cognitive flexibility in accordance with task requirements.

The dissociation between WCST performance and intelligence that has been reported in Milner's study [15] might also have contributed to the widely held belief that EF represents a psychological construct that is separable from intelligence (e.g., [26]). While a detailed discussion of the facets of human intelligence goes beyond the scope of this article, a few remarks seem justified here. David Wechsler once defined intelligence as the "the global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" [40] (p. 3). From this definition, it is evident that intelligence and EF share substantial conceptual overlap.

Psychological science of the 20th century evidenced a controversy about the most reasonable theoretical model of intelligence [41]. Spearman initially identified a single general ability that he named *g* (for "general factor" [42,43], but see [44]). Meanwhile, a consensus regarding the dimensionality of intelligence has only been achieved insofar as most researchers agree with the assumption of a hierarchical structure of cognitive abilities that underlie intelligence, with *g* at its highest level. Cattell [45] distinguished two types of cognitive abilities that are relevant for intelligence in a revision of Spearman's concept of *g*. Fluid intelligence (*Gf*) was hypothesized as the ability to solve novel problems by using reasoning, and crystallized intelligence (*Gc*) was hypothesized as a knowledge-based ability that was heavily dependent on education. After Horn [46] identified a number of broad cognitive abilities in a revision of the *Gf-Gc* theory, Carroll [47] proposed a hierarchical model with three levels, which is now known as the CHC (Cattell–Horn–Carroll) model of intelligence [48]. The bottom level consists of highly specialized, task-specific abilities. The second level consists of a number of broad cognitive abilities, including *Gf* and *Gc*. Carroll accepted Spearman's concept of *g* as a representation

of the highest level, affecting performance on any particular test solely via its influence on identified broad cognitive abilities [47]. The CHC model of intelligence forms the basis of many contemporary cognitive test batteries [49].

Regardless of one's preferred theoretical model of intelligence, the most widely utilized tests of intelligence are the Wechsler Adult Intelligence Scale (WAIS) and the Wechsler Intelligence Scale for Children (WISC). The initial version of the WAIS was released in 1955 [50], followed by the WAIS-R (1981) [51], WAIS-III (1997) [52], and WAIS-IV (2008) [53]. The initial version of the WISC was released in 1949 [54], followed by the WISC-R (1974) [55], WISC-III (1991) [56], and WISC-IV (2003) [57]. Apart from an estimate of general intelligence (i.e., Full Scale IQ, FSIQ), Wechsler tests were often used to obtain sub-scores of Verbal intelligence (VIQ) and Non-verbal Performance intelligence (PIQ). The concepts of VIQ and PIQ are closely related to the CHC abilities Gc and Gf, respectively. (The verbal VIQ and the non-verbal PIQ represent concepts that are a little bit broader than Gc and Gf. In the case of the WAIS-R, the VIQ includes the subtests {information, comprehension, arithmetic, digit span, similarities, and vocabulary}, while the PIQ includes the subtests {picture arrangement, picture completion, block design, object assembly, and digit symbol}. This is similar for the WAIS-III because here the subtests that comprise the WAIS-III VIQ, which are labeled verbal comprehension {vocabulary, similarities, information, comprehension} and working memory {arithmetic, digit span, letter-number sequencing}, confer to Gc plus Gsm (short-term memory; see [58], Table 5). The subtests that comprise the WAIS-III PIQ, which are labeled perceptual organization {picture completion, block design, matrix reasoning} and processing speed {digit-symbol coding, symbol search}, confer to Gf, Gv (visuospatial abilities) plus Gs (processing speed; see [58], Table 5)).

Other IQ tests focus more directly on the assessment of CHC-compatible broad cognitive abilities. For example, the National Adult Reading Test [59] (NART) is often used to assess Gc in clinical neuropsychology, under the assumption that this education-dependent facet of intelligence is relatively insensitive to neurological alterations, and can thus serve as an estimate of the premorbid level of crystallized intelligence [26]. Crawford et al. [60] found that the NART predicted 72% of WAIS-VIQ variance, but only 33% of the WAIS-PIQ variance. Raven's Progressive Matrices [61] (RPM) and the Culture Fair Test [62] (CFT) are often considered as quintessential measures of fluid intelligence (e.g., [43]). The RPM and the CFT are also closely related to the PIQ since both tests utilize non-verbal materials. In the remainder of this article, we have thus considered the NART as a proxy for the VIQ (both rather focusing on the assessment of Gc), and the RPM/CFT as proxies for the PIQ (all rather focusing on the assessment of Gf).

A number of authors have tried to unify intelligence and neuropsychological assessment based on the CHC model [58,63–71]. For example, Jewsbury et al. [65] showed that popular neuropsychological EF tests were subsumable under CHC broad cognitive abilities based on factor analytic methods, although particular EF tests were related to distinct CHC constructs. Most importantly in the present context, WCST perseverative errors were found to be related to GvGf, i.e., visuospatial (Gv) and fluid (Gf) facets of intelligence in that study (see also [64]). The conclusion that WCST performance and fluid intelligence are highly correlated was also corroborated by a neuropsychological study: Roca et al. [72] showed that when patients suffering from frontal lobe lesions and controls were matched on the CFT, i.e., on a measure of Gf, the frequency of WCST total errors no longer differed between these groups. The authors took these data to suggest that the unique variance in WCST performance was negligible once the variance that this measure shared with fluid intelligence was accounted for. The conclusion drawn by Roca and colleagues [72] lies in obvious conflict with Milner's [15] assertion that the WCST allows for the detection of frontal dysfunctions in the absence of noticeable declines in intelligence.

The question to what degree WCST performance is separable from measures of intelligence is of vital importance for the concept of EF. EF would be an unnecessary psychological construct if discriminant validation of WCST scores against indicators of intelligence should fail. Cronbach characterized the issue in the following words:

“To defend the proposition that a test measures a certain variable defined by a theory, one looks basically for two things. The first is *convergence* of indicators. [. . .] The second kind of evidence is *divergence* of indicators that are supposed to represent different constructs. If a test is said to measure “ability to reason with numbers,” it should not rank pupils in the order a test of sheer computation gives, because the computation test cannot reasonably be interpreted as a reasoning test. The test interpretation should also be challenged if the correlation with a test of verbal reasoning is very high, because this would suggest that general reasoning ability accounts for the ranking, so that specialized ability to reason with numbers is an unnecessary concept.” ([73], p. 144; italics in the original text)

According to Cronbach’s example, the construct of numerical reasoning would be unnecessary in the case that discriminant validation against computational abilities or verbal reasoning should fail. In general, any worthwhile cognitive construct (e.g., numerical reasoning) requires discriminant validation against related cognitive constructs (e.g., computational ability, general reasoning). We referred to this prerequisite of designing an evidence-based cognitive architecture as ‘Cronbach’s hurdle’. Of importance for the present study, putative EF tests (such as the WCST) had to demonstrate discriminant validity against measures of intelligence in order for the EF construct to take the hurdle. The provision of empirical support for discriminant validity of the WCST has been a relatively neglected topic [74]. Some of the few exceptions to that rule were discussed above in detail. These studies had their methodological grounding in factor analytic methods [64,65,68–70], in regression methods [71], or in neuropsychological patient studies [15,72].

The present meta-analysis complements the hitherto available evidence with regard to the discriminant validity of WCST scores against intelligence. For that purpose, our meta-analysis focused on the correlations between popular WCST scores (i.e., number of categories, frequency of various types of errors) and a variety of IQ domains (i.e., FSIQ, PIQ, VIQ). Individual studies often fail to obtain reliable estimates of correlations due to insufficient sample sizes [75]. By pooling data from these studies using meta-analytical techniques, one cannot only arrive at more reliable correlation estimates, but also examines potential origins of between-study variability in the strength of these correlations [76]. Thus, the present meta-analysis of correlations between WCST scores and IQ domains informed the ongoing discussion (a) about the construct validity of the WCST, and, by way of this, (b) about the overlap between EF and intelligence in a more general sense.

2. Materials and Methods

2.1. Search Strategy

A systematic literature review was conducted in 2017 by MH and updated in July 2018 by NM. We searched for records including the term “card sort” in combination with any of the following keywords regarding intelligence domains and tests: “intelligence”, “iq”, “fsiq”, “viq”, “piq”, “WAIS”, “WISC”, “normative”, “progressive matrices”, “Raven’s matrices” and “Raven’s”. PubMed (705 studies), Science Direct (326 studies), Web of Science (741 studies) and, in addition, the Compendium of Neuropsychological Tests [77] (10 studies) yielded a total of (1782 studies) hits for these combinations of search keywords (Figure 1). First of all, double appearances (861 studies) were excluded. Thirty-five additional papers were published in languages other than English and therefore had to be excluded. We screened the titles and abstracts of the remaining records and excluded studies that did not involve an assessment of original data from the WCST and intelligence domains (e.g., reviews or meta-analyses). 182 studies of the left over 844 studies remained inaccessible via local university libraries or open access.

In total, we accessed 662 full texts, and we checked whether the data reported in these papers included correlation coefficients for the relationship between any scores of WCST performance and any domains of intelligence. At this step, studies were excluded when it became apparent that they did not administer the WCST, or when they did not report data from the WCST and at least one domain of intelligence. Papers that only reported test statistics for group difference involving WCST scores

and intelligence domains, without reports of correlative relationships between WCST and intelligence, were also excluded at this step.

Of the remaining 92 studies, 45 did not report correlations between WCST scores and intelligence domains and were therefore excluded. The studies that had to be excluded at this final step either reported the results of multivariate statistics (e.g., regression analyses or factor analyses) that did not allow for the estimation of bivariate correlations, or they did not include a measure that could be utilized for estimating the FSIQ, VIQ, or PIQ domains. Forty-seven studies remained for the final meta-analysis.

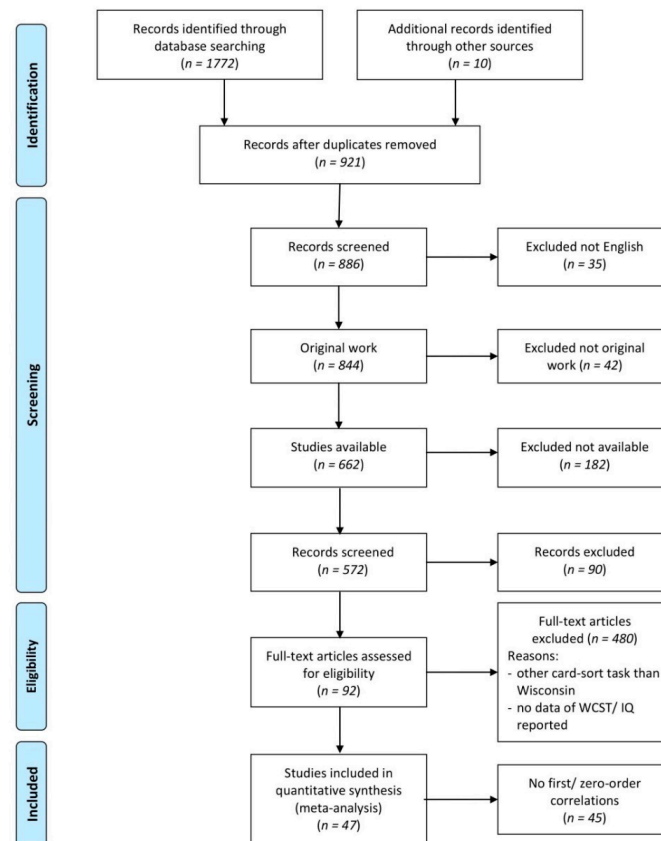


Figure 1. Flow chart depicting the selection of articles for our meta-analysis.

2.2. Data Extraction and Coding

2.2.1. WCST Scores

The extracted studies reporting correlations between WCST performance and intelligence reported a large variety of different WCST scores. To guarantee adequate statistical power for all analyses, we decided to focus on the WCST scores that had been reported in at least five independent studies. All those excluded (such as conceptual level responses or numbers of trials required to complete the first category) were found to be reported in a maximum of two studies. Analyzed scores included:

1. the number of categories completed (correct sequences of 6 or 10 consecutive correct matches to the criterion sorting category; the sequence length depends on the test version)
2. the frequency of perseverative errors or responses (persisting to respond to an incorrect stimulus characteristic)
3. the frequency of non-perseverative errors (errors that are not considered as perseverative errors)
4. the frequency of failures to maintain the set (e.g., when five or more consecutive correct matches are made, followed by at least one error prior to successfully completing the category) and

5. the frequency of total errors.

We did not distinguish between absolute and relative scores of the different error types (e.g., between the number and the percentage of perseverative errors). These scores are typically highly correlated and pooling data across these two types of measures allowed for a more powerful analysis of the relationship between the respective facet of WCST performance and intelligence. When a study reported both absolute and relative figures for a particular error type, we extracted the correlations involving the absolute figures. Similarly, to avoid redundancy and increase statistical power, we selected only one perseveration score for each study and did not further distinguish between perseverative errors and perseverative responses (see [21]). When multiple scores of perseveration were reported, we extracted the correlations involving perseverative errors [11]. One included study distinguished between two types of perseverative errors and we averaged correlation coefficients across both types to extract a single score representing perseveration for this study. Some studies did not report the total frequency of errors, but the total frequency of correct responses. For those studies, we changed the sign of the correlation coefficients involving the total frequency of correct responses to obtain an estimate for the correlation coefficients involving the total frequency of errors.

2.2.2. IQ Domains

We distinguished between three domains of intelligence, verbal intelligence (VIQ), performance intelligence (PIQ), and full-scale intelligence (FSIQ). With regard to VIQ, most studies reported correlations involving VIQ scores from a version of the WAIS or WISC. When studies reported only correlations involving VIQ subdomains, this information was used to estimate the correlation between VIQ and WCST scores. When multiple VIQ subdomains were reported (e.g., similarities and information), we computed average correlations across them. When only a single VIQ subdomain was reported (e.g., vocabulary) we took its correlation as the best estimate of the correlation between VIQ and WCST performance. Vocabulary tests that are used to assess premorbid intelligence (i.e., the NART and the MWT [78]) were also considered to be measures of VIQ as was the Ammons Quick Test [79].

Similarly, values for the PIQ category were obtained by extracting reported Wechsler PIQ aggregate scores or by estimating PIQ based on the reported Wechsler PIQ subdomains. The remaining measures in the PIQ category included the CFT (and its matrix subtest), the RPM, the Shipley Abstraction test [80], the matrices subtest of the Stanford–Binet Intelligence Scale [81], and the reasoning subtest of Thurstone’s Primary Mental Abilities [82].

One study reported two indices of VIQ and two other studies reported two indices of PIQ. For these studies, we extracted the correlations involving the more common domain (i.e., the one that was more frequently reported in the other included studies) and conducted robustness tests using the alternative domain.

The FSIQ category consisted exclusively of FSIQ aggregate scores obtained from the intelligence tests of Wechsler and Kaufman [83]. Some studies reported data from established short versions of a Wechsler test and others created ad hoc short versions by combining scores from subtests of the VIQ and PIQ domains (e.g., Vocabulary and Block Design). They were all considered as FSIQ domains for the present set of analyses.

2.3. Correlation Coefficients

Given the selection of five WCST scores and three domains of intelligence, we extracted a theoretical maximum of 15 correlation coefficients per independent study sample. Some studies reported multiple correlation coefficients per measure combination as a result of investigating this correlation in independent subgroups (e.g., patients vs. control participants). For these studies, we extracted correlation coefficients separately for every independent sample of participants. Most of the included studies reported Pearson’s *r* or Spearman’s *rho* correlation coefficients, one study reported Kendall’s *tau*, and another study reported a mix of parametric and non-parametric correlations (see Table 1).

Table 1. Overview of the studies included in the meta-analysis.

First Author	Year	N	Sample	% Fem	Age (M)	Age (SD)	WCST Version	FSIQ	VIQ	PIQ	Cor
Ardila [84]	2000	50	children (healthy)	0.0	14.4	1.0	Heaton, 1981	WISC-R	WISC-R	WISC-R	
Bird [85]	2004	90	adults (healthy)	62.2	57.0	8.3	Nelson, 1976		NART		mix
Boone [86]	1998	250	adults (healthy and patients, psy & neuro)	46.0	55.5	15.5	Heaton, 1981		WAIS-R	WAIS-R	
Chien [87]	2009	99	adults (healthy)	0.0	20.2	0.6	Heaton, 1993 (c)	WAIS-R			
Cianchetti [88]	2007	101	children (healthy)	52.5	4.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	113	children (healthy)	50.4	5.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	119	children (healthy)	47.1	6.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	161	children (healthy)	52.8	7.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	94	children (healthy)	52.1	8.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	98	children (healthy)	50.0	9.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	119	children (healthy)	50.4	10.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	122	children (healthy)	48.4	11.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	100	children (healthy)	50.0	12.0	0.0	Nelson, 1976			RPM	P
Cianchetti	2007	99	children (healthy)	48.5	13.0	0.0	Nelson, 1976			RPM	P
Crawford [89]	2000	123	adults (healthy)	61.0	39.4	13.4	Nelson, 1976	WAIS-R	WAIS-R		P
Crawford	1999	90	adults (healthy)	55.6	72.8	6.5	Nelson, 1976	WAIS-R (s)			P
Davis [90]	2000	62	adults (healthy)	51.6	20.3	1.5	Heaton et al., 1993		WAIS-III (s)	PMA (s)	P
de Zubicaray [91]	1998	36	adults (healthy)	66.7	70.1	5.6	Nelson, 1976	WAIS-R	WAIS-R	WAIS-R	S
Dieci [92]	1997	88	adults (healthy and patients, psy)	28.4	27.3	7.0	Heaton, 1981	WAIS-R	WAIS-R	WAIS-R	S
Dolan [93]	2002	60	adults (patients, psy)	0.0	29.8	6.6	Heaton, 1981		NART		S
Evans [94]	2016	192	children (healthy)	52.1	12.4	1.8	Heaton et al., 1993	WASI (s)			P
Giovagnoli [95]	2001	26	adults (patients, neuro)		36.8	10.9	Nelson, 1976			RPM	S
Giovagnoli	2001	21	adults (patients, neuro)		33.3	11.2	Nelson, 1976			RPM	S
Giovagnoli	2001	18	adults (patients, neuro)		36.6	13.4	Nelson, 1976			RPM	S
Giovagnoli	2001	15	adults (patients, neuro)		41.4	9.8	Nelson, 1976			RPM	S
Giovagnoli	2001	14	adults (patients, neuro)		30.7	8.8	Nelson, 1976			RPM	S
Giovagnoli	2001	18	adults (patients, neuro)		32.6	12.2	Nelson, 1976			RPM	S
Giovagnoli	2001	30	adults (patients, neuro)		35.2	14.3	Nelson, 1976			RPM	S
Giovagnoli	2001	23	adults (patients, neuro)		35.6	13.4	Nelson, 1976			RPM	S
Giovagnoli	2001	36	adults (healthy)		36.1	10.7	Nelson, 1976			RPM	S
Golden [96]	1998	112	adults (patients, neuro)	48.2	37.4	13.3	Heaton, 1981	WAIS-R	WAIS-R	WAIS-R	P
Han [97]	2016	180	adolescents (healthy and patients, psy)	49.5	13.7	1.5	Heaton, 1981 (c)	K-BIT (s)			
Heinrichs [98]	1990	56	adults (patients, neuro)	30.4	43.8	13.6	Heaton, 1981	WAIS-R			P
Ilonen [99]	2000	27	adults (patients, psy)	63.0	33.0	13.6	Heaton et al., 1993	WAIS-R			S
Isingrini [100]	1997	35	adults (healthy)	57.1	35.5	7.6	Nelson, 1976		WAIS (s)	CM	
Isingrini	1997	72	adults (healthy)	48.6	80.6	8.6	Nelson, 1976		WAIS (s)	CM	
Keefe [101]	1994	54	adults (healthy and patients, psy)	59.3	34.8	10.5	Heaton, 1981		WAIS-R (s,v)	WAIS-R (s)	P
Kilincaslan [102]	2010	39	children (healthy and patients, psy)	15.4	12.2	2.7	Heaton, 1993 (c)	WISC-R	WISC-R	WISC-R	
Lee [103]	2009	39	adults (patients, psy)	51.3	32.4	7.2	Heaton, 1993 (c)	WAIS-R (s)			P
Lee	2009	33	adults (healthy)	57.6	29.0	8.9	Heaton, 1993 (c)	WAIS-R (s)			P
Lehto [104]	2003	51	children (healthy)	41.2	9.2	0.3	Heaton, 1981			RPM	P
Lehto	2003	40	adults (healthy)	62.5	30.1	9.6	Heaton, 1981			RPM	P

Table 1. Cont.

First Author	Year	N	Sample	% Fem	Age (M)	Age (SD)	WCST Version	FSIQ	VIQ	PIQ	Cor
Lehto [105]	2004	46	children (healthy)	43.5	12.5	0.3	Heaton, 1981			RPM	P
LeMonda [106]	2012	44	children (patients, psy)	22.7	8.1	1.0	Heaton et al., 1993			WISC-R (s)/SB4(s)	
Lichtenstein [107]	2018	226	adults and adolescents (healthy and patients, psy)	35.4	13.6	2.6	Heaton, et al. 1993	WISC-III & WISC-IV			P
Lineweaver [108]	1999	229	adults (healthy)	57.6	69.1	8.6	Nelson, 1976	WAIS-R (s)	WAIS-R (s)	WAIS-R (s)	S
Liss [109]	2001	21	children (patients, psy)	14.3	9.2	0.3	Heaton et al., 1993	n/a	n/a	n/a	P
Liss	2001	34	children (patients, psy)	29.4	9.1	0.1	Heaton et al., 1993	n/a	n/a	n/a	P
Lucey [110]	1997	38	adults (healthy and patients, psy)	47.4	38.0	11.5	Heaton, 1981		NART		P
Minshew [111]	2002	90	adults and adolescents (patients, psy)		21.4	9.7	Heaton et al., 1993	WAIS-R			
Minshew	2002	107	adults and adolescents (healthy)		21.2	9.8	Heaton et al., 1993	WAIS-R			
Mullane [112]	2007	30	children (healthy and patients, psy)	26.7	8.8	1.2	Heaton et al., 1993	WISC-III (s)			P
Nestor [113]	2015	81	adults (healthy)		40.8	9.1	Heaton, 1981	WAIS-III	WAIS-III	WAIS-III	P
Obonsawin [114]	1999	146	adults and adolescents (healthy)	47.3	40.3	14.0	Nelson, 1976	WAIS-R	WAIS-R	WAIS-R	K
Obonsawin [115]	2002	123	adults (healthy)	38.2	40.3	14.0	Nelson, 1976	WAIS-R			P
Owashi [116]	2009	27	adults (patients, psy)	55.6	41.5	10.1	Heaton, 1993 (c)	WAIS-R (s)			S
Perry [117]	1998	71	adults (patients, psy)	60.6	34.2	8.7	Heaton, 1981		WAIS-R (s,v)		P
Roca [118]	2012	31	adults (healthy and patients, neuro)		60.6	8.0	Nelson, 1976			RCPM	
Roca [72]	2010	74	adults (healthy and patients, neuro)		49.9	12.6	Nelson, 1976			CFT (s)	P
Rossell [119]	2003	78	adults (patients, psy)	0.0	33.7	8.5	Heaton, 1981		NART		P
Salthouse [120]	1996	259	adults (healthy)	63.3	51.4	18.4	Heaton et al., 1993			WAIS-R (s)/SA	P
Schiebener [121]	2015	112	children (healthy)	52.7	13.6	3.4	Nelson, 1976 (c)			RPM	P
Shura [122]	2016	205	adults (patients, psy & neuro)	10.8	34.9	9.1	Heaton, 1981 (c)		WAIS-III (s)		P
South [123]	2007	19	children (patients, psy)	26.3	14.9	2.7	Grant, 1948	n/a	n/a	n/a	P
South	2007	18	children (healthy)	38.9	14.1	2.9	Grant, 1948	n/a	n/a	n/a	P
Steingass [124]	1994	101	adults (patients, psy)	21.9	50.5	8.1	Nelson, 1976	WAIS (s)	WAIS (s)/MWT-B (v)	WAIS (s)	P
Sweeney [125]	1991	44	adults (patients, psy)	40.9	28.5	8.6	Heaton, 1981		AQT (v)		
Syngelaki [126]	2009	70	adults and adolescents (healthy and patients, psy)	0.0	16.3	1.5	Heaton, 2005 (c)	WASI (s)			P
Taconnat [127]	2007	81	adults (healthy)	51.9	66.0	8.2	Heaton et al, 1993			CFT	P
Whiteside [128]	2016	304	adults (patients, neuro)	54.9	45.1	13.4	Heaton et al., 1993		WAIS-III (s,v)		P
Yasuda [129]	2014	33	adults and adolescents (patients, psy)	39.4	26.1	11.5	Kashima et al., 1987 (c)	WAIS-III	WAIS-III	WAIS-III	P
Yasuda	2014	33	adults and adolescents (healthy)	39.4	26.8	9.6	Kashima et al., 1987 (c)	WAIS-III	WAIS-III	WAIS-III	P

Note: % fem = percent female participants in the sample, age (M) = mean age in the sample, age (SD) = standard deviation of participants' age, WCST = Wisconsin Card Sorting Test, FSIQ = full scale IQ, VIQ = verbal IQ, PIQ = performance IQ, cor = correlation coefficient, psy = psychiatric, neuro = neurological, (c) = computerized, (s) = short form/subscales, (v) = vocabulary test, WISC = Wechsler Intelligence Scale for Children, WAIS = Wechsler Adult Intelligence Scale, WASI = Wechsler Abbreviated Scale of Intelligence, R(C)PM = Raven's (Colored) Progressive Matrices, NART = National Adult Reading Test, PMA = Primary Mental Abilities, K-BIT = Kaufman Brief Intelligence Test, CM = Cattell's Matrices, SB4 = Stanford-Binet Intelligence Scale, CFT = Cattell Culture Fair Intelligence Test, AQT = Ammons Quick Test, Shipley Abstraction Test, MWT-B = Multiple Choice Word Test-B, P = Pearson's *r*, S = Spearman's *rho*, K = Kendall's *tau*, mix = parametric and non-parametric correlations.

We did not invert the sign of the extracted correlation coefficients, that is, positive correlations between IQ and the number of completed WCST categories and negative correlations between IQ and WCST error scores indicated that WCST performance improved with increasing IQ. When a study did not report the size of a correlation coefficient, but only that this coefficient was not statistically significant, we excluded this coefficient from our analyses. However, we ran additional robustness analyses to test whether our results changed when these coefficients were included as correlations of $r = 0$.

2.4. Basic Meta-Analysis

Mean effect sizes and confidence intervals for the relationships between the five selected WCST scores and three selected intelligence domains were calculated using the random-effects model method (with inverse variance weights) proposed by Hedges and Vevea [130] and implemented by Field and Gillett [76]. Heterogeneity of effect sizes were examined using Cochran's Q and the I^2 index [131]. By comparing Cochran's Q (estimated under fixed-effect assumptions) to a χ^2 distribution, we tested whether heterogeneity among studies was significant. The I^2 index served as an estimate of between-study heterogeneity in true effect sizes, with I^2 values of about 25%, 50% and 75% indicating low, moderate, and high heterogeneity, respectively [131].

2.5. Moderator Analyses

We examined whether the strength of the correlation between domains of intelligence and WCST performance were moderated by the sample and study characteristics that were routinely reported in neuropsychological studies on the WCST-IQ relationship. To guarantee a minimum statistical power for these analyses, we focused on the correlations involving the two most frequently reported WCST scores, that is, the number of completed WCST categories and the frequency of perseverations. We extracted the following variables as potential moderators of the WCST-IQ relationship: (1) the mean age of participants, (2) the standard deviation of participants' age, (3) the proportion of female participants in the sample, (4) clinical status, (5) the WCST version used, (6) the WCST administration method used, and (7) the intelligence test used.

Demographic variables (mean age, standard deviation of age as well as proportion of female participants) were treated as continuous predictors. For illustrative purposes, we also created three groups of studies including participants from different age categories (mean age <18 years, 18–50 years, >50 years) and repeated our analyses with mean age as a categorical predictor. Some studies reported WCST and IQ data from a sample that was smaller than the sample for which they provided demographic data (i.e., not all participants completed all neuropsychological measures). For these studies, we estimated the age and sex distribution of the sample of interest (i.e., the sample underlying the computation of WCST-IQ correlations) by extracting the demographic data for the total sample. When studies did not provide the standard deviation of participants' age, standard deviations were estimated from range data (minimum, maximum) according to the procedure described by Wan, Wang, Liu, and Tong [132].

Regarding the moderating role of clinical status, we tested whether correlations between IQ domains and WCST scores were stronger in samples of patients with psychiatric or neurological disorders than they were in samples of healthy participants. Some studies reported correlation coefficients from mixed samples of patients and healthy controls. These studies were excluded for the analysis of the moderating role of clinical status.

With regard to the WCST version used in the individual studies, we distinguished (a) between the Heaton version [37,38] and the Nelson version [36] of the test and (b) between computerized and non-computerized WCST versions. IQ domains were contrasted based on their comprehensiveness. With regard to FSIQ, we distinguished between studies that used an abbreviated version (established or *ad hoc*) of an FSIQ test (i.e., an aggregate of a subset of subtests) and studies that used full version FSIQ tests. With regard to VIQ tests, we distinguished between pure vocabulary tests (NART, MWT-B,

Ammons Quick Test, Wechsler vocabulary subtests) and more comprehensive VIQ tests (i.e., aggregates of at least two Wechsler subtests). With regard to PIQ tests, we distinguished between culture-reduced (matrices) tests (Cattell's matrices, CFT, RPM, RCPM) and Wechsler scores (aggregated across at least two subtests).

The relationship between potential moderators and the size of the WCST-IQ correlation was examined using separate weighted multiple regression analyses (random-effects model with inverse variance weights [76]). Continuous moderator variables (i.e., mean age, standard deviation of age, proportion of female participants) were z-transformed to facilitate comparisons between regression coefficients.

2.6. Publication Bias Analysis

The Begg and Mazumdar rank correlation test was calculated as implemented in the syntax by Field and Gillett [76] to examine the relationship between effect sizes and their standard errors. A positive correlation between these two variables was indicative of a small-study effect (i.e., the tendency for studies with smaller samples to produce larger effect size estimates). Such an overrepresentation of small studies with large effect sizes can be the result of publication bias [133] and it would likely contribute to an overestimation of the true effect size. Note that this logic only applied when the effect size in question was positive (e.g., as expected for the correlation between WCST categories and IQ tests). With regard to negative average effect sizes (e.g., as expected for the correlation between WCST error scores and IQ domains), negative correlations between effect sizes and standard errors were indicative of small-study effects.

2.7. Partial Correlations

Partial correlations were used to examine a) the age-corrected relationship between WCST performance and IQ and b) the IQ-corrected relationship between age and WCST performance. Partial correlations were either directly extracted from the publication or calculated based on zero-order correlations using <http://vassarstats.net/> [134]. Only 11 studies reported the information necessary to be included into one of our meta-analyses on partial correlations and we decided to pool these data across WCST scores and IQ domains. For this purpose, the former was recoded so that larger values indicated better performance. When partial correlations involving multiple WCST scores were reported, the average partial correlation across them was extracted.

3. Results

The results of our random-effects meta-analysis of the correlation between IQ domains and WCST performance is displayed in Table 2. An inspection of Table 2 reveals that most of the analyzed WCST scores were significantly related to all IQ domains. With the notable exception of weak correlations between IQ and WCST failures to maintain set, the size of WCST-IQ correlations ranged from small-to-medium ($r = 0.19$) to medium-to-large ($r = 0.44$) [135]. Correlations appeared to be stronger when they involved (a) a general (i.e., categories, total errors) versus more specific (i.e., perseverations, non-perseverative errors) WCST performance score, and (b) a general (i.e., FSIQ) versus more specific (i.e., VIQ, PIQ) domain of IQ. Note, however, that most of the 95% confidence intervals surrounding the corresponding effect sizes overlapped substantially.

Table 2. Results of the meta-analyses of correlations between WCST performance and intelligence.

IQ Domain	Statistic	Categories	Perseverations	NPE	FMS	TE
	Number of samples (<i>k</i>)	20	25	6	6	11
	Significant correlations (%)	70	76	50	0	64
FSIQ	Total <i>N</i>	1533	2049	664	553	710
	Average effect size <i>r</i>	0.44	−0.39	−0.29	−0.05	−0.42
	[95% CI]	[0.36, 0.51]	[−0.45, −0.33]	[−0.46, −0.11]	[−0.14, 0.03]	[−0.51, −0.31]
	<i>Q</i>	63.31 *	50.41 *	26.56 *	1.34	22.02 *
	<i>I</i> ²	68.41	48.42	73.64	0	45.50
	τ _{Begg & Mazumar}	−0.08	−0.26	0.07	−0.07	0.15
	<i>p</i> _{Begg & Mazumar}	0.626	0.076	0.851	0.851	0.529
	Number of samples (<i>k</i>)	19	24	6	4	11
	Significant effects (%)	74	71	67	0	64
	VIQ	Total <i>N</i>	1755	2071	546	260
Average effect size <i>r</i>		0.33	−0.31	−0.30	−0.02	−0.37
[95% CI]		[0.26, 0.39]	[−0.36, −0.26]	[−0.44, −0.16]	[−0.15, 0.10]	[−0.45, −0.29]
<i>Q</i>		37.08 *	30.99	14.33 *	3.03	17.01
<i>I</i> ²		46.06	19.33	51.15	0	29.45
τ _{Begg & Mazumar}		−0.20	−0.01	−0.20	1	0.18
<i>p</i> _{Begg & Mazumar}		0.234	0.941	0.573	−	0.435
Number of samples (<i>k</i>)		28	42	17	14	22
Significant effects (%)		75	52	53	14	73
PIQ		Total <i>N</i>	2506	3256	1784	1386
	Average effect size <i>r</i>	0.34	−0.29	−0.19	−0.08	−0.36
	[95% CI]	[0.27, 0.39]	[−0.34, −0.24]	[−0.27, −0.11]	[−0.13, −0.02]	[−0.42, −0.29]
	<i>Q</i>	69.75 *	88.87 *	44.76 *	14.70	62.07 *
	<i>I</i> ²	58.42	51.61	59.79	0	62.95
	τ _{Begg & Mazumar}	0.05	0.17	0.14	−0.11	−0.11
	<i>p</i> _{Begg & Mazumar}	0.693	0.121	0.433	0.584	0.498

Note: Significant correlations (%) = percentage of included correlation coefficients with a 95% confidence interval excluding 0. * significant heterogeneity at $\alpha = 0.05$. WCST = Wisconsin Card Sorting Test, NPE = non-perseverative errors, FMS = failures to maintain set, TE = total errors, FSIQ = full scale IQ, VIQ = verbal IQ, PIQ = performance IQ.

Robustness analyses revealed that average correlation coefficients increased slightly (by $r = 0.002$ to $r = 0.021$) when alternative IQ tests were included (see Section 2.2.2) and decreased slightly (by $r = 0.009$ to $r = 0.023$) when coefficients that were described as non-significant were included as coefficients of $r = 0$. The negligible magnitude of these changes suggested that the results displayed in Table 2 were largely invariant to the analytical choices we made when extracting effect sizes from individual studies.

Rank correlation tests did not find any of the significant WCST-IQ correlations to be significantly affected by small-study effects. In combination with the funnel plots displayed in Figure 2, these results suggest that meta-analytical correlation coefficients are unlikely to be substantially overestimated due to publication bias. This notion receives further support from the observation that many of the publications included in this meta-analysis reported non-significant correlations between WCST scores and IQ domains (see Table 2).

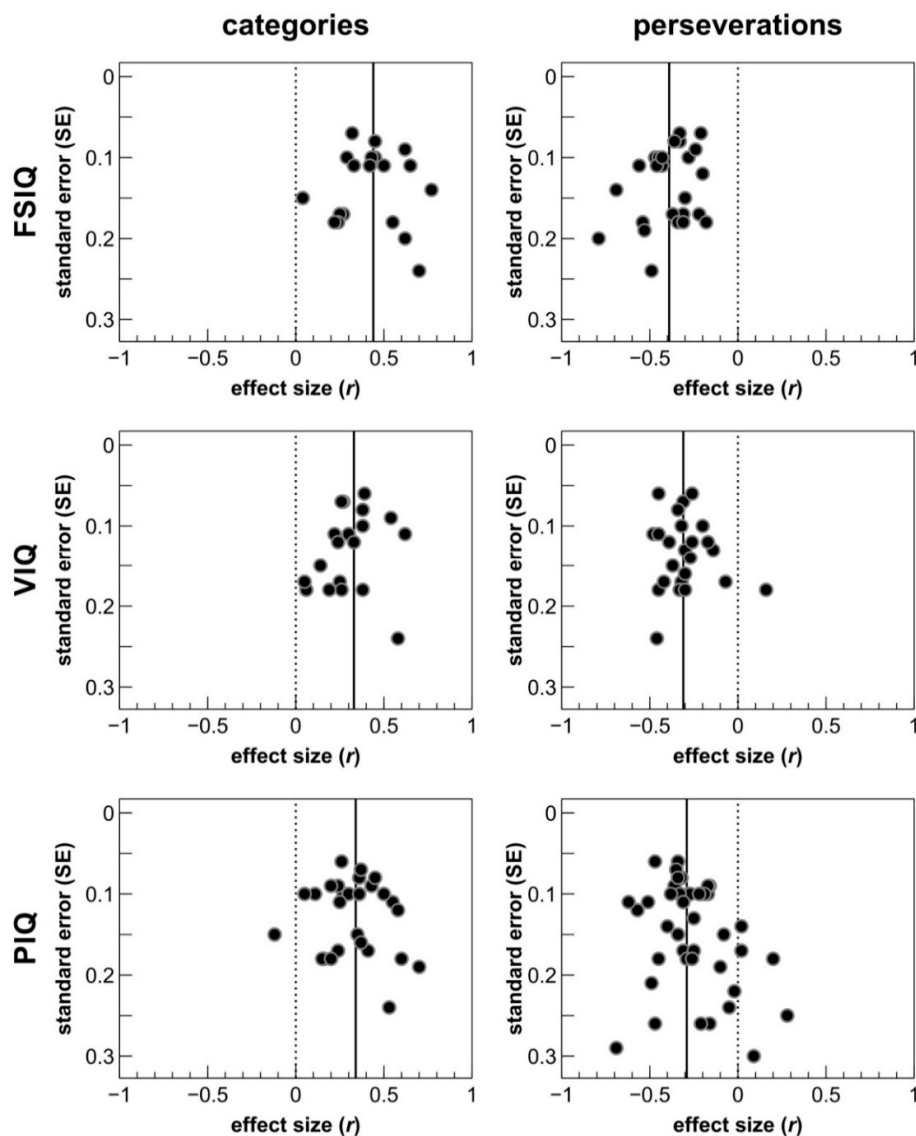


Figure 2. Funnel plots illustrating the relationships between effect sizes and their standard errors. Straight vertical lines indicate the average meta-analytical effect size obtained for the respective correlation between a WCST score (categories, perseverations) and intelligence domain (FSIQ, VIQ, PIQ). Dotted vertical lines display zero effects for comparison. WCST = Wisconsin Card Sorting Test, FSIQ = full scale IQ, VIQ = verbal IQ, PIQ = performance IQ.

3.1. Moderator Analyses

Effect-size heterogeneity was moderate (i.e., around $I^2 = 50\%$) for most of the analyzed WCST-IQ correlations. These results indicated that the size of these correlations may vary as a function of sample or study characteristics. Our moderator analyses (see Table 3) identified some of the factors that contributed to heterogeneity in the size of WCST-IQ correlations.

Table 3. Results of the meta-regression analyses conducted to examine the role of potential moderators of WCST-intelligence relationships.

Moderator	Categories					Perseverations				
	β	95% CI	df	<i>t</i>	<i>p</i>	β	95% CI	df	<i>t</i>	<i>p</i>
Continuous moderators										
FSIQ										
Mean age	0.00	[−0.09, 0.11]	17	0.15	0.880	0.00	[−0.08, 0.07]	22	−0.21	0.837
SD age	0.07	[−0.05, 0.18]	17	1.15	0.268	−0.04	[−0.11, 0.04]	22	−0.93	0.362
Percent female	0.01	[−0.09, 0.12]	14	0.27	0.795	0.01	[−0.06, 0.09]	19	0.35	0.733
VIQ										
Mean age	−0.02	[−0.09, 0.06]	16	−0.49	0.633	0.00	[−0.06, 0.06]	21	0.64	0.949
SD age	0.06	[−0.03, 0.15]	16	1.35	0.195	−0.03	[−0.10, 0.04]	21	−0.73	0.471
Percent female	0.01	[−0.06, 0.08]	15	0.34	0.742	0.05	[−0.04, 0.07]	20	0.57	0.573
PIQ										
Mean age	0.04	[−0.01, 0.10]	25	1.60	0.122	−0.07 *	[−0.11, −0.02]	39	−3.05	0.004
SD age	0.04	[−0.02, 0.11]	25	1.33	0.197	−0.03	[−0.08, 0.01]	39	−1.38	0.174
Percent female	0.00	[−0.09, 0.10]	23	0.06	0.951	−0.01	[−0.10, 0.08]	29	−0.25	0.801
Categorical moderators	χ^2		df		<i>p</i>	χ^2		df		<i>p</i>
FSIQ										
Age group	1.14		2		0.566	2.89		2		0.236
Clinical status	2.60		1		0.107	4.76 *		1		0.029
WCST version	0.01		1		0.911	1.16		1		0.282
WCST administration	2.71		1		0.100	2.33		1		0.127
IQ test type	0.33		1		0.567	2.93		1		0.087
VIQ										
Age group	1.81		2		0.405	0.48		2		0.787
Clinical status	0.26		1		0.609	0.00		1		0.972
WCST version	0.30		1		0.586	1.37		1		0.242
WCST administration	2.78		1		0.095	0.83		1		0.363
IQ test type	0.25		1		0.617	0.02		1		0.888
PIQ										
Age group	2.893		2		0.235	13.77 *		2		0.001
Clinical status	3.431		1		0.064	0.18		1		0.669
WCST version	0.41		1		0.522	3.78		1		0.052
Categorical moderators	χ^2		df		<i>p</i>	χ^2		df		<i>p</i>
PIQ WCST administration	0.42		1		0.518	0.06		1		0.810
IQ test type	0.03		1		0.858	2.55		1		0.110

Note: * significant moderator at $\alpha = 0.05$. WCST = Wisconsin Card Sorting Test, FSIQ = full scale IQ, VIQ = verbal IQ, PIQ = performance IQ.

First, we found the correlation between WCST perseverations and PIQ domains to depend on the mean age of participants in the sample, $\beta = -0.07$, $t(39) = -3.05$, $p = 0.004$. Correlations were markedly stronger in samples with a mean age above 50 years, $k = 7$, $r = -0.44$, 95% CI $[-0.51, -0.35]$, $I^2 = 57\%$, than they were in samples of younger adults (18–50 years), $k = 18$, $r = -0.24$, 95% CI $[-0.33, -0.13]$, $I^2 = 49\%$, or in samples of children and adolescents (below 18 years), $k = 17$, $r = -0.25$, 95% CI $[-0.30, -0.19]$, $I^2 = 4\%$, as shown in Figure 3.

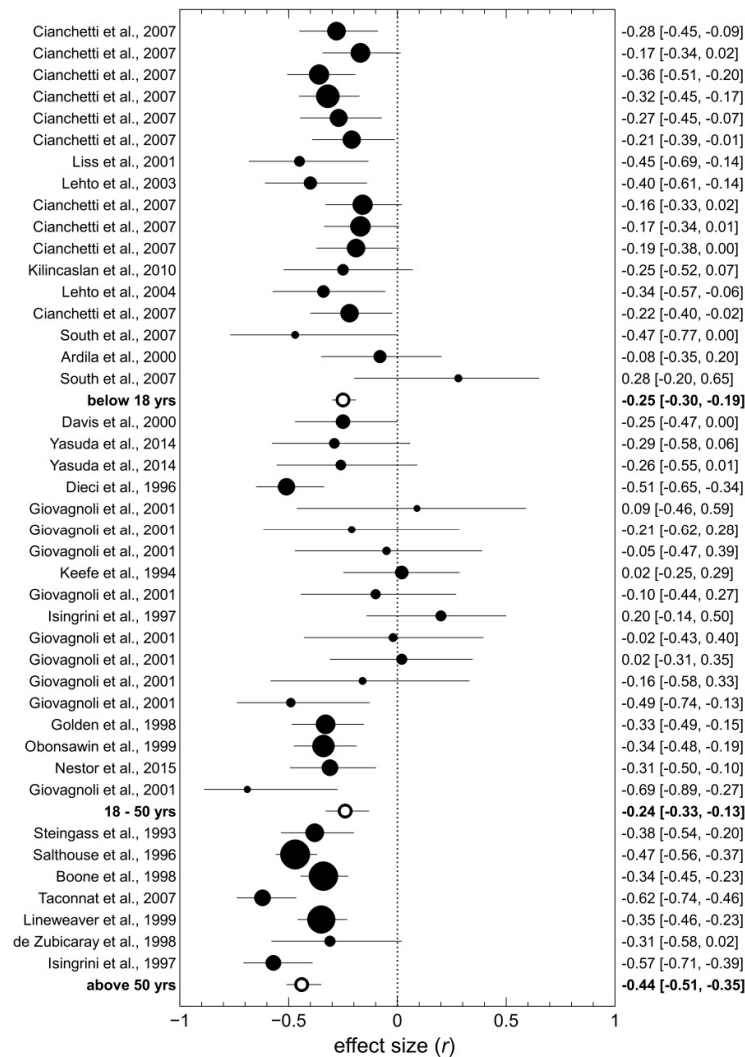


Figure 3. Effect sizes of the correlation between WCST perseverations and performance IQ as a function of participants' mean age. Effect sizes are sorted from youngest (top) to oldest (bottom) sample. The size of black-filled circles is proportional to sample size. White-filled circles indicate average effect sizes across age groups. WCST = Wisconsin Card Sorting Test.

Second, moderator analyses revealed the size of the correlation between WCST perseverations and FSIQ domains to be related to participants' clinical status, $\chi^2(1) = 4.76$, $p = 0.029$. Correlations were stronger in samples of patients, $k = 9$, $r = -0.49$, 95% CI $[-0.60, -0.36]$, $I^2 = 62\%$, than they were in samples of healthy individuals, $k = 11$, $r = -0.34$, 95% CI $[-0.39, -0.27]$, $I^2 = 17\%$.

Finally, we did not find the size of WCST-IQ correlations to depend significantly on the WCST version or IQ test administered in the original studies. This lack of effect-size difference cannot be considered conclusive given the small number of studies involved in these comparisons. However, it is worth noting that irrespective of the administered test versions, WCST-IQ correlations remained significant. For both categories and perseverations, we found exclusively significant correlations

with all IQ domains for both Heaton versions, $|r| = 0.33\text{--}0.46$, all $p < 0.001$, and Nelson versions, $|r| = 0.26\text{--}0.45$, all $p < 0.001$, of the WCST. Similarly, all WCST-IQ correlations were substantial when involving traditional, non-computerized WCST variants, $|r| = 0.29\text{--}0.47$, all $p < 0.001$, as well as in the small number of studies involving computerized WCST versions, $|r| = 0.23\text{--}0.34$, all $p < 0.026$. With regard to IQ test variants, comprehensive Wechsler-type measures of FSIQ, VIQ, and PIQ were found to correlate significantly with WCST categories and perseverations, $|r| = 0.30\text{--}0.43$, all $p < 0.001$. Correlations were also substantial ($p < 0.001$) in the smaller samples of studies administering short-version FSIQ tests, categories: $r = 0.37$; perseverations: $r = -0.31$, vocabulary VIQ tests as indicators of premorbid intelligence, categories: $r = 0.27$; perseverations: $r = -0.29$, or culture-reduced (matrices) PIQ tests, categories: $r = 0.32$, 95%, perseverations: $r = -0.27$.

3.2. Partial Correlation Analyses

Meta-analysis of partial correlation coefficients revealed a significant relationship between WCST and IQ performance when controlling for age, $k = 10$, $r = 0.36$, 95% CI [0.25, 0.46], $I^2 = 70\%$. A rank correlation test did not find this relationship to be significantly affected by small-study effects, $r = 0.20$, $p = 0.421$. The correlation between age and WCST performance was not significant when controlling for IQ, $k = 9$, $r = -0.04$, 95% CI [-0.24, 0.17], $I^2 = 91\%$. Of note, this overall null-correlation resulted from a positive IQ-corrected relationship between age and WCST performance in samples of children and adolescents, $k = 3$, $r = 0.35$, 95% CI [0.15, 0.52], $I^2 = 67\%$, and a negative IQ-corrected relationship between age and WCST performance in adult samples, $k = 6$, $r = -0.22$, 95% CI [-0.32, -0.11], $I^2 = 49\%$.

4. Discussion

The present meta-analysis examined discriminant validity of WCST scores against common domains of intelligence. We found robust, low to medium-sized correlations between WCST performance and IQ across all WCST scores and IQ domains that we investigated. Solely the average correlations between WCST failures to maintain set and IQ amounted to coefficients very close to zero. Average correlations between WCST non-perseverative errors and IQ were higher ($|r| = 0.19\text{--}0.30$), and correlations between the most commonly utilized WCST scores (number of categories, total errors, perseverations) and IQ were generally the highest. Average correlations between these WCST scores and FSIQ were somewhat higher ($|r| = 0.39\text{--}0.44$) than those between them and VIQ ($|r| = 0.31\text{--}0.37$) and PIQ ($|r| = 0.29\text{--}0.36$), respectively. Taken together, the present meta-analysis revealed modest correlations between most of the WCST scores and IQ domains, based on sample sizes that varied between $N = 260$ and $N = 3256$.

4.1. Discriminant Validity of the WCST

If one thinks about the observed correlations in terms of the proportion of the variance in WCST scores that is predictable from IQ, the calculation of *r squared* reveals about 0 (failures to maintain set) to 19 (categories) percent shared variance, leaving 81 to 100 percent unique WCST variance, i.e., variance unaccounted for by common measures of intelligence. Our findings therefore suggest that WCST and FSIQ, VIQ, PIQ represent partially separable measures of cognitive abilities. One possibility to account for these results lies in the referral to the unity/diversity model of EF that we shortly presented in the Introduction. According to the latest revision of the model, performance on EF tasks can be accounted for by a general EF factor, an updating-specific factor, and a shifting-specific factor. Both the general EF factor and the updating-specific factor seem to share substantial variance with measures of intelligence. In contrast, the shifting factor, which has been shown to underlie perseverative errors on the WCST [35], appears to be largely unrelated to intelligence [31,136].

A second possibility to account for these results is grounded in the distinction between measures and constructs, and the argument that less than perfect measurement reliability attenuates the actual correlations that may exist at the construct level [137,138]. Yet, even when we corrected for potential attenuation of the correlations that might result from imperfect reliabilities of WCST scores and IQ

domains, the unique portion of WCST variance surpassed the portion of WCST variance that was shared with variance in IQ. As an example, consider the strongest correlation that we found in our meta-analysis, that is, the correlation between WCST categories and FSIQ ($r_{xy} = 0.44$). Based on a quite conservative estimate for the reliability of the IQ tests that were studied ($r_{xx} = 0.90$), and a computed average reliability of WCST categories (about $r_{yy} = 0.60$; see Tables A1 and A2 in the Appendix A), this observed correlation can be ‘disattenuated’ to yield an estimated correlation of $r_{xtyt} = 0.60$ at the construct level. These numbers imply 36 percent shared variance between WCST categories and FSIQ at the construct level in comparison to 19 percent that shared variance at the measurement level (the Appendix A provides more details for disattenuation analyses). While these numbers indicate that the proportion of shared variance at the construct level might be about two times higher than the proportion of shared variance at the measurement level, they also indicate that—at best—about one third of the variability in mental shifting (i.e., the presumed construct behind performance on the WCST) can be accounted for by variability in the common construct of intelligence. The results from these disattenuation analyses hence suggest that the mental shifting construct is not identical to the commonly presumed intelligence construct. We conclude that WCST scores and IQ domains represent partially independent constructs, even if the attenuation of the correlations between them is taken into account. These results support the idea that, despite conceptual overlap between accepted definitions of intelligence and executive functioning, the WCST assesses cognitive abilities that are partially separable from intelligence.

One of the reviewers asked for a report of internal reliability indices of a WCST version since our literature review concerning this matter was fruitless, as discussed above (the literature reports only indices of test-retest reliability). The question that arises here is whether the assumption of WCST internal reliability estimates of at least $r_{yy} = 0.60$ is defensible because otherwise the construct-level correlation between WCST scores and IQ domains would surpass the calculated estimate of $r_{xtyt} = 0.60$. Note, that we recently published internal reliability estimates of WCST scores, and that these estimates fell well above $r = 0.90$, suggesting that $r_{yy} = 0.60$ yields a well-defensible estimate for the upper bound of construct-level correlations between WCST scores and IQ domains [9].

Hence, the WCST surpasses Cronbach’s hurdle of sufficient discriminant validity against common intelligence tests, thereby justifying the utilization of the mental shifting construct in the context of the WCST. Viewed from this broader perspective, the overlap between EF (or at least, a more specific EF shifting factor, represented by the WCST) and intelligence (represented by the most popular IQ tests) seems to corroborate earlier conclusions that were drawn from individual differences studies [31,35].

4.2. Differential Associations between WCST Scores and IQ Domains

The correlations between distinct WCST scores and the FSIQ showed some differences that may be worth a short comment. The average correlation between WCST categories and FSIQ was highest ($r = 0.44$), followed by the correlations between WCST total errors ($r = -0.42$), WCST perseverations ($r = -0.39$), WCST NPE ($r = -0.29$) and FSIQ, with the lowest correlations (close to zero) between WCST FMS and FSIQ ($r = -0.05$). Hence, differential correlations between WCST categories, total errors, perseverations and FSIQ appear negligible. Comparable correlations with FSIQ suggested that these WCST scores reflected, at least to some degree, a common cognitive ability (e.g., shifting, general EF, or general understanding of the task). This common cognitive ability seemed to be less relevant for the other two WCST scores that we analyzed (i.e., NPE and FMS), as suggested by lower correlations with FSIQ. Given the interdependencies of WCST scores, it seems plausible that a deficit which relates to an increase in the number of committed perseverations also increases total errors and decreases WCST categories, while leaving NPE and FMS more or less unaffected.

Alternatively, the differential correlations between two groups of WCST scores (categories, total errors, perseverations vs. NPE, FMS) and FSIQ might be related to differences in their psychometric properties. Specifically, the first group of WCST scores may on average provide more reliable estimates in comparison with the second group (see Appendix A). In this case, the attenuation effect that was

discussed above in detail would more strongly affect the second group of WCST scores than the first group. Hence, it is possible that all analyzed WCST measures are indicators of the ability that causes WCST performance to correlate with FSIQ, but that these indicators simply differ in how reliably they assess this ability.

The issue of differential associations between WCST scores and IQ remains to be substantiated by future studies. This research should not only be concerned with differential associations between specific WCST scores and intelligence, but it should also be concerned with their reliability. As we have seen, inferences about validity hinge upon knowledge about the reliability that can be assumed for the scores under consideration.

The correlations between WCST scores and distinct IQ domains followed the pattern $r_{\text{WCST-FSIQ}} > r_{\text{WCST-VIQ}} = r_{\text{WCST-PIQ}}$ (see Table 2). The FSIQ advantage over verbal and non-verbal domains was of little surprise, given that the FSIQ yielded more reliable estimates (with an average split-half reliability across age groups of 0.98; cited following [13]) compared to the VIQ (with an average split-half reliability across age groups of 0.97 [13]), and compared to the PIQ (with an average split-half reliability across age groups of 0.94 [13]). Given the fact that the FSIQ comprised both, verbal and non-verbal domains, this difference in reliability was probably due to the positive effects of additional items on scale reliability [139]. The attenuation effect therefore exerted a slightly stronger influence on the WCST-VIQ and WCST-PIQ correlations compared to the WCST-FSIQ correlation.

WCST-VIQ and WCST-PIQ correlations were almost identical, perhaps with a marginal advantage for the verbal compared to the non-verbal domains. This finding may be interpreted as indicating that the predictability of WCST variance based on verbal domains, which are more closely related to Gc, does not differ from the predictability of WCST variance based on non-verbal domains, which are more closely related to Gf. The irrelevance of the contents of the particular IQ domains on the WCST-IQ correlations seems to be at odds with the common assumption that EF in general, and WCST scores in particular, are preferentially correlated with Gf, while the correlation with Gc is negligible [43,72,140]. However, reliability coefficients were slightly higher for the verbal domains compared to the non-verbal ones (see above; also, estimates of the NART test-retest reliability exceeded 0.95 [60], whereas reliability estimates for the RPM and CFT seemed to lie in the interval between 0.80 and 0.90; see [62] (cited after [141]) and [13], but see [142] for RPM test-retest reliability estimates > 0.90). Given the available reliability estimates, the predictability of WCST variance from verbal IQ domains may be actually lower than the predictability of WCST variance from non-verbal IQ domains, but this difference may be counteracted by differential degrees of attenuation due to slight differences in verbal and non-verbal IQ reliabilities.

We suggest conducting specifically designed studies for a better understanding of the relationship between WCST scores and tests of crystallized and fluid intelligence. Notice that the conductance of such a study would be of theoretical importance since the enigma of unimpaired IQ in patients with frontal lobe damage and WCST deficits [15] was explained by shortcomings in the utilized intelligence tests, and specifically, insufficient measurement of fluid intelligence at that time [140]. Other researchers emphasized the association between the frontal lobes and fluid intelligence [43,72,143]. Tranel et al. [140], however, did not find evidence for an association between the presence of frontal lesions and decrements in the Matrix Reasoning subtest of the WAIS-III, rendering the issue controversial to what degree Gf mediated the association between frontal lesions and EF. Later work from this group [144,145] and other groups [43] pointed to the direction that general intellectual abilities draw on a circumscribed, albeit distributed, network in frontal and parietal lobes. A deeper review of the neural mechanisms of intelligence would go beyond the scope of this article, and the interested reader is referred to the relevant literature [43,146–148].

4.3. The Role of Moderator Variables

We also found considerable effect-size heterogeneity, and our moderator analyses could only account for small portions of this heterogeneity. Despite the fact that substantial portions of the

study-by-study variability remain unexplained, our moderator analyses led to the identification of two potentially interesting moderators (i.e., age, clinical status). First, the correlations between WCST perseverations and non-verbal IQ were markedly stronger in samples with a mean age above 50 years than they were in samples of younger adults (18–50 years) or in samples of children and adolescents. Second, correlations between WCST perseverations and the FSIQ were larger in patient samples than in control samples. One account of these moderator effects holds generalized cognitive decline responsible, which could occur in some of the studied individuals, regardless of whether one looks at samples of healthy elderly people or at clinically relevant disease states. Generalized cognitive decline in some elderly or ill individuals should increase inter-individual variability on WCST scores and in IQ in these samples as compared to the usually more homogenous samples of young and healthy participants. To the extent that this cognitive decline is generalized, additional variance in WCST scores and in IQ domains would be shared, thereby increasing the correlation between them in elderly and ill samples. However, these moderator effects should be interpreted with caution, due to the low statistical power of the comparisons that we were able to carry out. Cautiousness is also suggested by the fact that these relationships did not occur consistently across all WCST-IQ combinations.

The number of data available for the analysis of partial, age-corrected WCST-IQ correlations was severely limited. Yet, there seemed to exist substantial WCST-IQ correlations even if the age of the participants was partialled out of the raw WCST-IQ correlations ($r = 0.36$). Note that these partial correlations might be larger for some WCST-IQ combinations. However, we had to pool the data across all available WCST scores and IQ domains, a procedure which implied that the partial correlation might have been attenuated through the inclusion of WCST scores that showed weaker correlations with IQ (i.e., WCST FMS).

4.4. Limitations of the Present Meta-Analysis and Suggestions for Future Studies

The opportunity to obtain more conclusive evidence from the present meta-analysis was restrained by a number of obstacles: First, most of the studies that we analyzed comprised relatively small sample sizes. The effect-size estimates that we could provide were based on quite a variable N s, ranging from $N = 260$ (WCST FMS-VIQ) to $N = 3256$ (WCST perseverations-PIQ). Although our bias analyses did not suggest that the size of the correlations covaried with the within-study standard errors (see Figure 2), the presence of these strongly unbalanced N s calls for more systematic studies. These studies should not only explore plausible associations (i.e., WCST perseverations-Gf, WCST perseverations-PIQ), but also less plausible associations (WCST FMS-Gc, WCST FMS-VIQ) with identical rigor in order to facilitate valid inferences about differential associations.

Second, a lack of conceptual sharpness with regard to the structure of intelligence that was apparent in many studies prevented firm conclusions about differential associations of theoretical importance between WCST scores and crystallized and fluid intelligence, respectively. Neither EF, nor intelligence, can be considered as a unitary construct. In the case of EF, the available evidence points into the direction that WCST scores load on an isolable mental shifting factor [35,149]. Yet, it remains a viable possibility that mental shifting abilities share substantial portions of variance with a hitherto not identified cognitive ability that also underlies performance on common intelligence tests. As an example, studies of ‘relational’ reasoning that manipulated the number of to-be-integrated relations revealed that frontal lobes are critical for the integration of multiple relations [150–154]. Successful WCST performance requires a comparable integration of information from multiple dimensions. Future studies should be more specifically designed toward theory-driven research questions, and they should primarily be concerned with isolating more circumscribed facets of WCST performance and intelligence alike.

Third, the role of reliability in understanding construct validity seems to be underrated. In fact, Schmidt and Hunter’s methods of meta-analysis [155] could not be applied due to the currently less than satisfactory knowledge about reliabilities of the WCST scores. No information about internal reliability was available in any of the published studies, and the data about test-retest reliability are by

and large discouraging [9]. The apparently low reliability of WCST scores is certainly below levels considered acceptable in clinical practice (but see [9]), supporting the idea to merge them to a global index of executive functioning (as in Schretlen's modified WCST version [39]). Further research should try to circumvent the reliability limitations of the WCST scores by appropriate efforts at improving this useful assessment technique. Future studies of WCST-IQ associations should preferably be conducted using psychometrically matched [156,157] measures of crystallized and fluid intelligence.

Taken together, these considerations imply the necessity to conduct large-scale, theory-driven, and psychometrically sound studies of EF-IQ associations in both healthy and clinical samples. We consider these studies of importance because construct validity increasingly gains importance in the field [74,158–161]. This is particularly true with regard to the discriminant validity of neuropsychological tests that are assumed to assess EF as a construct separable from standard conceptualizations of intelligence. The demand for construct validation, in particular the assurance of discriminant validity, is due to the progressively disregarded role of criterion validity, where evidence for a frontal localization of dysexecutive symptoms serves as a major criterion for test validity.

5. Conclusions

The analyzed data revealed low to medium-sized associations between WCST scores and IQ, which were not simply mediated by effects of age, suggesting that the two types of assessment shared portions of variance to a modest degree. The major finding of our meta-analysis was that, at best, about one third of the variability in the EF facet measured by the WCST (i.e., mental shifting [35,149]) could be accounted for by variability in common indicators of intelligence.

Our conclusion is that, despite conceptual overlap between accepted definitions of intelligence and executive functioning, the WCST assesses cognitive abilities that are partially separable from intelligence. Furthermore, our findings provide little evidence, if at all, for differential associations between distinct WCST scores and IQ. Although some of them (categories, total errors, perseverations) showed somewhat stronger associations with IQ than others (non-perseverative errors, failures to maintain set), it remains to be seen whether this differential association should be attributed to their contents, or alternatively, to their psychometric properties. WCST scores were associated in comparable strength with verbal and non-verbal IQ domains, thus not supporting the idea that the former were preferably associated with non-verbal/fluid intelligence. Overall, our meta-analysis represents a step toward evidence-based neuropsychology [162,163] by shedding light on the hitherto understudied discriminant validity of a widely used test of executive functions.

Author Contributions: Conceptualization, B.K.; methodology, B.K. and F.L.; software, F.L.; validation, B.K., N.M., J.F.S., M.H. and F.L.; formal analysis, F.L.; investigation, N.M. and M.H.; resources, B.K.; data curation, N.M., M.H. and F.L.; writing—original draft preparation, B.K. and F.L.; writing—review and editing, B.K., J.F.S. and F.L.; visualization, N.M., M.H. and F.L.; supervision, B.K.; project administration, B.K., N.M., J.F.S., M.H. and F.L.; funding acquisition, B.K. and F.L.

Funding: This research was funded by a research grant from the Karlheinz-Hartmann-Foundation, Hannover, Germany (awarded to B.K.) and by the FWO and European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement (No.665501; awarded to F.L.).

Acknowledgments: The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

The available reliability studies of common WCST measures are summarized in Table A1. It is worth noting that our compilation of reliability studies does not pretend being exhaustive because our search for suitable studies was not systematic. Inspection of Table A1 reveals that estimates of WCST internal consistency remained completely unavailable.

All studies relied on repeated administration of the WCST, estimating test-retest-reliabilities with variable retest periods. Nine studies examined the administration of a WCST version in non-clinical populations (cumulative $N = 568$), implying that the five studies of clinical populations achieved a cumulative N of only 120 patients. In addition, the diagnoses under consideration were quite heterogeneous (traumatic brain injury, sleep apnoe syndrome, autism, learning disability). Very little is known about test-retest reliability when the WCST is administered in clinical populations that are of major interest for neuropsychologists. In that regard, one has to rely on the estimates from those two studies that looked at patients who suffered from traumatic brain injury (cumulative $N = 57$ [164,165]). However, the results from these two studies can hardly be considered as being convergent.

Many different coefficients were considered across these studies (Pearson's r , Spearman's ρ , Kendall's τ , intra-class correlations, generalizability coefficients). Most of the studies were based on quite small sample sizes. Charter [166] showed that precise reliability estimates presuppose sample sizes of $N \geq 400$. In order to obtain reliable estimates of WCST test-retest reliabilities, we averaged across all available studies, thereby ignoring the methodological differences between them, including the population under study, the duration of the retest period, the WCST version, and the type of coefficient under consideration. Given that reliability is the result of a test administered under particular circumstances to the sample under study, we also report averaged WCST test-retest reliability coefficients separately for samples (non-clinical, clinical) and test versions. The results are shown in Table A2.

Table A1. A summary of published studies on the WCST reliability.

Study	Year	Population	N	Type of Reliability	Test-Retest Interval	WCST Version	WCST Measure	Coefficient	Type of Correlation
Basso et al. [167]	1999	Non-clinical	50	Test-retest	12 months	Heaton et al. 1993	CAT	0.54	6
							TE	0.50	6
							PE, P	0.52, 0.50	6
							FMS	-0.02	6
Bird et al. [85]	2004	Non-clinical	90	Test-retest	1 month	Nelson 1976	TE	0.34	1
							PE	0.38	1
Bowden et al. [168]	1998	Non-clinical	75	Test-retest('alternate' forms)	Same day	Heaton et al. 1981	CAT	0.60	2
							TE	0.51	2
							PE, P	0.32, 0.30	2
							NPE	0.43	2
de Zubicaray et al. [91]	1998	Non-clinical	36	Test-retest	7.5 months	Nelson 1976	CAT	0.28	2
							TE	0.36	2
							PE	0.27	2
							NPE	0.38	2
							FMS	0.49	2
Greve et al. [164]	2002	TBI	34	Test-retest	66 weeks	Heaton et al. 1993	CAT	0.53	3
							TE	0.82	2
							PE, P	0.80, 0.78	2
							NPE	0.50	2
							FMS	0.26	3
Heaton et al. [38]	1993	Non-clinical	46	Test-retest	33 days	Heaton et al. 1993	TE	0.71	5
							PE, P	0.52, 0.53	5
							NPE	0.72	5
Ingram et al. [169]	1998	Sleep apnoe patients	29	Test-retest	12 days	Computerized WCST	CAT	0.70	2
							TE	0.79	2
							PE, P	0.83, 0.79	2
							FMS	0.50	2
Lineweaver et al. [108]	1999	Non-clinical	142	Test-retest	24 months	Nelson 1976	CAT	0.56	1
							PE	0.64	1
							NPE	0.46	1

Table A1. Cont.

Study	Year	Population	N	Type of Reliability	Test-Retest Interval	WCST Version	WCST Measure	Coefficient	Type of Correlation
Ozonoff [169]	1995	Autistic children	17	Test-retest	30 months	Standard WCST	TE P	0.94 0.93	5 5
		Learning disabled children and adolescents	17	Test- retest	30 months	Standard WCST	TE P	0.90 0.94	5 5
Paolo et al. [170]	1996	Non-clinical	87	Test-retest	12 months	Heaton et al. 1981	CAT	0.65	3
							TE	0.66	3
							PE, P	0.65, 0.63	3
							NPE	0.55	3
							FMS	0.13	
Steinmetz et al. [171]	2010	Non-clinical	22	Test-retest	Same day	Heaton et al. 1993	TE	0.68	6
							PE	0.72	6
							FMS	0.16	6
Tate et al. [165]	1998	Non-clinical	20	Test-retest	8 months	Heaton et al. 1993	CAT	0.88	2
							TE	0.79	4
							PE, P	0.72, 0.68	4
							NPE	0.74	4
							FMS	-0.04	4
		TBI	23	Test-retest	10 months	Heaton et al. 1993	CAT	0.29	2
							TE	0.39	4
					PE, P	0.34, 0.33	4		
					NPE	0.32	4		
					FMS	-0.32	4		

Note. CAT = categories; TE = total errors; PE, P = perseveration errors, perseverations; NPE = non-perseverative errors; FMS = failures to maintain set; TBI = traumatic brain injury; 1 = Pearson's *r*, 2 = Spearman's *rho*, 3 = Kendall's *tau*, 4 = intra-class coefficient, 5 = generalizability coefficient, 6 = unspecified; *Study-wise *N* served as the weighting factor.

Several observations deserve comments: First, average weighted test-retest reliabilities of WCST number of categories, total errors, and perseverations appeared to be very similar, and these coefficients achieved average values close to 0.60. Non-perseverative errors seem to show slightly lower average test-retest reliabilities (about 0.50). Failures to maintain set achieved a very low average test-retest reliability of about 0.15.

Table A2. Summary statistics of WCST reliability estimates.

Overall and Sample-Specific	WCST Measure	Cumulative N	Unweighted Average	Weighted Average *
all (14) available samples	CAT	496	0.56	0.55
	TE	546	0.65	0.58
	PE, P	688	0.61	0.56
	NPE	463	0.51	0.50
	FMS	301	0.15	0.16
non-clinical (9) samples	CAT	410	0.59	0.58
	TE	426	0.57	0.53
	PE, P	568	0.52	0.52
	NPE	406	0.55	0.51
	FMS	215	0.14	0.14
clinical (5) samples	CAT	86	0.51	0.52
	TE	120	0.77	0.76
	PE, P	120	0.76	0.77
	NPE	57	0.41	0.43
	FMS	86	0.15	0.19
Test Versions (excl. cWCST)	WCST Measure	Cumulative N	Unweighted Average	Weighted Average *
Heaton et al. version (10) samples	CAT	345	0.58	0.59
	TE	391	0.69	0.65
	PE, P	391	0.64	0.57
	NPE	285	0.54	0.53
	FMS	236	0.03	0.06
Nelson version (3) samples	CAT	178	0.42	0.50
	TE	126	0.35	0.35
	PE, P	268	0.43	0.50
	NPE	178	0.42	0.44
	FMS	36	0.49	0.49

Note. cWCST = computerized WCST; CAT = categories; TE = total errors; PE, P = perseverations; NPE = non-perseverative errors; FMS = failures to maintain set; * Study-wise N served as the weighting factor.

Second, reliability estimates appeared to be higher in clinical ($0.7 \leq r \leq 0.8$) in comparison to non-clinical ($0.5 \leq r \leq 0.55$) samples on two WCST measures (i.e., 'number of total errors', 'perseverations'). The apparently low reliability of WCST measures often brings the argument forward that the WCST should not be utilized in clinical practice. However, these numbers suggest that appropriate studies should be conducted to reappraise the relatively good reliability of WCST perseverations in clinical samples because this measure represents the theoretically and clinically most relevant WCST measure.

Third, there was a general tendency toward lower test-retest reliability coefficients when retest periods were short: The Pearson correlation between the duration of that period (in days) and the reported coefficient for WCST perseverations amounted to $r = 0.57$ ($p < 0.05$ with $N = 14$ duration-coefficient pairings), indicating that procedural learning during the first test administration affects task performance on the second test administration over brief retest periods. This finding supports the idea that the WCST is a one-shot test that can only be administered once with any particular patient over brief time periods [13].

Fourth, with the remarkable exception of ‘failures to maintain set’, the Heaton and colleagues’ WCST versions [37,38] achieved slightly higher test-retest reliability coefficients compared to the Nelson version [36]. This putative version effect, if true, should probably be attributed to test length [139] because Heaton and colleagues’ WCST versions utilize 64 or 128 trials, whereas Nelson’s version is based on 48 trials.

Table A3 shows construct-level WCST-IQ correlations for a set of observed IQ-WCST correlations ($0.05 \leq r_{xy} \leq 0.50$, reflecting the range of observed coefficients that we found in our meta-analysis). Construct-level WCST-IQ correlations result from the application of Spearman’s attenuation formula [137]

$$r_{x_i y_i} = \frac{r_{xy}}{\sqrt{r_{xx} r_{yy}}} \quad (\text{A1})$$

with $r_{x_i y_i}$ representing a construct-level correlation, r_{xy} representing an observed, measurement-level correlation, and r_{xx} and r_{yy} representing reliabilities of variables x (IQ; $0.90 \leq r_{xx} \leq 0.99$) and y (WCST; $0.10 \leq r_{yy} \leq 0.60$), respectively.

Table A3. Construct-level correlations, for different values of observed correlations (0.05, 0.1, 0.15, . . . , 0.5), IQ reliability (0.90, 0.95, 0.99), and WCST reliability (0.1, 0.2, 0.3, 0.4, 0.5, 0.6).

Observed Correlation	IQ Reliability	WCST Reliability					
		0.10	0.20	0.30	0.40	0.50	0.60
0.05	0.90	0.167	0.118	0.096	0.083	0.075	0.068
	0.95	0.162	0.115	0.094	0.081	0.073	0.066
	0.99	0.159	0.112	0.092	0.079	0.071	0.065
0.10	0.90	0.333	0.236	0.192	0.167	0.149	0.136
	0.95	0.324	0.229	0.187	0.162	0.145	0.132
	0.99	0.318	0.225	0.183	0.159	0.142	0.130
0.15	0.90	0.500	0.354	0.289	0.250	0.224	0.204
	0.95	0.487	0.344	0.281	0.243	0.218	0.199
	0.99	0.477	0.337	0.275	0.238	0.213	0.195
0.20	0.90	0.667	0.471	0.385	0.333	0.298	0.272
	0.95	0.649	0.459	0.375	0.324	0.290	0.265
	0.99	0.636	0.449	0.367	0.318	0.284	0.259
0.25	0.90	0.833	0.589	0.481	0.417	0.373	0.340
	0.95	0.811	0.574	0.468	0.406	0.363	0.331
	0.99	0.795	0.562	0.459	0.397	0.355	0.324
0.30	0.90	1.00	0.707	0.577	0.500	0.447	0.408
	0.95	0.973	0.688	0.562	0.487	0.435	0.397
	0.99	0.953	0.674	0.550	0.477	0.426	0.389
0.35	0.90	1.00	0.825	0.674	0.583	0.522	0.476
	0.95	1.00	0.803	0.656	0.568	0.508	0.464
	0.99	1.00	0.787	0.642	0.556	0.497	0.454
0.40	0.90	1.00	0.943	0.770	0.667	0.596	0.544
	0.95	1.00	0.918	0.749	0.649	0.580	0.530
	0.99	1.00	0.899	0.734	0.636	0.569	0.519
0.45	0.90	1.00	1.00	0.866	0.750	0.671	0.612
	0.95	1.00	1.00	0.843	0.730	0.653	0.596
	0.99	1.00	1.00	0.826	0.715	0.640	0.584
0.50	0.90	1.00	1.00	0.962	0.833	0.745	0.680
	0.95	1.00	1.00	0.937	0.811	0.725	0.662
	0.99	1.00	1.00	0.917	0.795	0.711	0.649

Note. IQ = intelligence quotient; WCST = Wisconsin Card Sorting Test.

References

1. Fuster, J. *The Prefrontal Cortex*, 5th ed.; Academic Press: New York, NY, USA, 2015.
2. Miller, B.L.; Cummings, J.L. *The Human Frontal Lobes: Functions and Disorders*, 3rd ed.; The Guilford Press: New York, NY, USA, 2018.
3. Passingham, R.E.; Wise, S.P. *The Neurobiology of the Prefrontal Cortex*; Oxford University Press: Oxford, MS, USA, 2012.
4. Stuss, D.T.; Knight, R.T. *Principles of Frontal Lobe Function*; Oxford University Press: Oxford, MA, USA, 2013.
5. Szczepanski, S.M.; Knight, R.T. Insights into Human Behavior from Lesions to the Prefrontal Cortex. *Neuron* **2014**, *83*, 1002–1018. [[CrossRef](#)] [[PubMed](#)]
6. Grant, D.A.; Berg, E. A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *J. Exp. Psychol.* **1948**, *38*, 404–411. [[CrossRef](#)] [[PubMed](#)]
7. Berg, E.A. A Simple Objective Technique for Measuring Flexibility in Thinking. *J. Gen. Psychol.* **1948**, *39*, 15–22. [[CrossRef](#)] [[PubMed](#)]
8. Grange, J.; Houghton, G. *Task Switching and Cognitive Control*; Oxford University Press: New York, NY, USA, 2014.
9. Kopp, B.; Lange, F.; Steinke, A. The Reliability of the Wisconsin Card Sorting Test in Clinical Practice. *Assessment* **2019**, 1–16. [[CrossRef](#)] [[PubMed](#)]
10. Lange, F.; Seer, C.; Müller, D.; Kopp, B. Cognitive caching promotes flexibility in task switching: Evidence from event-related potentials. *Sci. Rep.* **2015**, *5*, 1–12. [[CrossRef](#)] [[PubMed](#)]
11. Lange, F.; Brückner, C.; Knebel, A.; Seer, C.; Kopp, B. Executive dysfunction in Parkinson's disease: A meta-analysis on the Wisconsin Card Sorting Test literature. *Neurosci. Biobehav. Rev.* **2018**, *93*, 38–56. [[CrossRef](#)]
12. Mitrushina, M.; Boone, K.B.; Razani, J.; D'Elia, L.F. *Handbook of Normative Data for Neuropsychological Assessment*, 2nd ed.; Oxford University Press: New York, NY, USA, 2005.
13. Sherman, E.; Tan, J.; Hrabok, M. *A Compendium of Neuropsychological Tests. Fundamentals of Neuropsychological Assessment and Test Reviews for Clinical Practice*, 4th ed.; Oxford University Press: Oxford, MS, USA, 2020.
14. Teuber, H.K.; Battersby, W.S.; Bender, M.B. Performance of complex visual tasks after cerebral lesions. *J. Nerv. Ment. Dis.* **1951**, *114*, 413–429.
15. Milner, B. Effects of Different Brain Lesions on Card Sorting. *Arch. Neurol.* **1963**, *9*, 90–100. [[CrossRef](#)]
16. Scoville, W.B.; Milner, B. Loss of recent memory after bilateral hippocampal lesions. *J. Neurol. Neurosurg. Psychiatry* **1957**, *20*, 11–21. [[CrossRef](#)]
17. Wechsler, D. *The Measurement of Adult Intelligence*; Williams and Wilkins: Baltimore, MA, USA, 1939.
18. Luria, A.R. Two kinds of motor perseveration in massive injury of the frontal lobes. *Brain* **1965**, *88*, 1–10. [[CrossRef](#)]
19. Pribram, K.H. The Primate Frontal Cortex—Executive of the Brain. In *Psychophysiology of the Frontal Lobes*; Pribram, K.H., Luria, A.R., Eds.; Academic Press: New York, NY, USA, 1973; pp. 293–314.
20. Teuber, H.L. Unity and diversity of frontal lobe functions. *Acta Neurobiol. Exp. (Wars)*. **1972**, *32*, 615–656. [[PubMed](#)]
21. Demakis, G.J. A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology* **2003**, *17*, 255–264. [[CrossRef](#)] [[PubMed](#)]
22. Lange, F.; Seer, C.; Salchow, C.; Dengler, R.; Dressler, D.; Kopp, B. Meta-analytical and electrophysiological evidence for executive dysfunction in primary dystonia. *Cortex* **2016**, *82*, 133–146. [[CrossRef](#)] [[PubMed](#)]
23. Lange, F.; Seer, C.; Müller-Vahl, K.; Kopp, B. Cognitive flexibility and its electrophysiological correlates in Gilles de la Tourette syndrome. *Dev. Cogn. Neurosci.* **2017**, *27*, 78–90. [[CrossRef](#)]
24. Lange, F.; Kip, A.; Klein, T.; Müller, D.; Seer, C.; Kopp, B. Effects of rule uncertainty on cognitive flexibility in a card-sorting paradigm. *Acta Psychol. (Amst)*. **2018**, *190*, 53–64. [[CrossRef](#)]
25. Nyhus, E.; Barceló, F. The Wisconsin Card Sorting Test and the cognitive assessment of prefrontal executive functions: A critical update. *Brain Cogn.* **2009**, *71*, 437–451. [[CrossRef](#)]
26. Lezak, M.D.; Howieson, D.B.; Bigler, E.D.; Tranel, D. *Neuropsychological Assessment*, 5th ed.; Press, O.U., Ed.; Oxford University Press: Oxford, MS, USA, 2012.
27. Banich, M.T. Executive Function: The Search for an Integrated Account. *Curr. Dir. Psychol. Sci.* **2009**, *18*, 89–94. [[CrossRef](#)]

28. Diamond, A. Executive Functions. *Annu. Rev. Psychol.* **2013**, *64*, 135–168. [[CrossRef](#)]
29. Miller, E.K.; Cohen, J.D. An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* **2001**, *24*, 167–202. [[CrossRef](#)]
30. Suchy, Y. Executive Functioning: Overview, Assessment, and Research Issues for Non-Neuropsychologists. *Ann. Behav. Med.* **2009**, *37*, 106–116. [[CrossRef](#)]
31. Friedman, N.P.; Miyake, A.; Young, S.E.; DeFries, J.C.; Corley, R.P.; Hewitt, J.K. Individual differences in executive functions are almost entirely genetic in origin. *J. Exp. Psychol. Gen.* **2008**, *137*, 201–225. [[CrossRef](#)] [[PubMed](#)]
32. Friedman, N.P.; Miyake, A.; Altamirano, L.J.; Corley, R.P.; Young, S.E.; Rhea, S.A.; Hewitt, J.K. Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Dev. Psychol.* **2016**, *52*, 326–340. [[CrossRef](#)] [[PubMed](#)]
33. Miyake, A.; Friedman, N.P. The Nature and Organization of Individual Differences in Executive Functions. *Curr. Dir. Psychol. Sci.* **2012**, *21*, 8–14. [[CrossRef](#)]
34. Friedman, N.P.; Miyake, A. Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex* **2017**, *86*, 186–204. [[CrossRef](#)]
35. Miyake, A.; Friedman, N.P.; Emerson, M.J.; Witzki, A.H.; Howerter, A.; Wager, T.D. The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cogn. Psychol.* **2000**, *41*, 49–100. [[CrossRef](#)]
36. Nelson, H.E. A Modified Card Sorting Test Sensitive to Frontal Lobe Defects. *Cortex* **1976**, *12*, 313–324. [[CrossRef](#)]
37. Heaton, R.K. *Wisconsin Card Sorting Test (WCST)*; Psychological Assessment Resources: Odessa, Ukraine, 1981.
38. Heaton, R.K.; Chelune, G.; Talley, J.L.; Kay, G.G.; Curtis, G. *Wisconsin Card Sorting Test Manual: Revised and Expanded*; Psychological Assessment Resources: Odessa, Ukraine, 1993.
39. Schretlen, D.J. *Modified Wisconsin Card Sorting Test (m-WCST)*; Psychological Assessment Resources: Lutz, Germany, 2010.
40. Wechsler, D. *Measurement of Adult Intelligence*, 3rd ed.; Williams and Wilkins: Baltimore, MD, USA, 1944.
41. Deary, I.J. Intelligence. *Annu. Rev. Psychol.* **2012**, *63*, 453–482. [[CrossRef](#)]
42. Spearman, C. “General intelligence”, objectively determined and measured. *Am. J. Psychol.* **1904**, *15*, 201–292. [[CrossRef](#)]
43. Duncan, J. *How Intelligence Happens*; Yale University Press: London, UK, 2010.
44. Jensen, A.R. *The G Factor: The Science of Mental Ability*; Praeger: New York, NY, USA, 1998.
45. Cattell, R.B. Some theoretical issues in adult intelligence testing. *Psychol. Bull.* **1941**, *38*, 592.
46. Horn, J.L.; Cattell, R.B. Refinement and test of the theory of fluid and crystallized general intelligences. *J. Educ. Psychol.* **1966**, *57*, 253–270. [[CrossRef](#)]
47. Carroll, J.B. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*; Oxford University Press: New York, NY, USA, 1993.
48. McGrew, K.S. CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence* **2009**, *37*, 1–10. [[CrossRef](#)]
49. Schneider, W.J.; McGrew, K.S. The Cattell–Horn–Carroll Model of Intelligence. In *Contemporary Intellectual Assessment: Theories, Tests, and Issues*; Flanagan, D.P., Harrison, P.L., Eds.; Guilford Press: New York, NY, USA, 2012; pp. 99–144.
50. Wechsler, D. *Wechsler Adult Intelligence Scale: Manual*; The Psychological Corporation: New York, NY, USA, 1955.
51. Wechsler, D. *Wechsler Adult Intelligence Scale–Revised*; The Psychological Corporation: New York, NY, USA, 1981.
52. Wechsler, D. *Wechsler Adult Intelligence Scale (WAIS-III): Administration and Scoring Manual*; The Psychological Corporation: San Antonio, TX, USA, 1997.
53. Wechsler, D. *Wechsler Adult Intelligence Scale (WAIS-IV)*; The Psychological Corporation: San Antonio, TX, USA, 2008.
54. Wechsler, D. *Wechsler Intelligence Scale for Children (WISC): Manual*; The Psychological Corporation: New York, NY, USA, 1949.

55. Wechsler, D. *Manual of the Wechsler Intelligence Scale for Children—Revised*; The Psychological Corporation: New York, NY, USA, 1974.
56. Wechsler, D. *WISC-III Wechsler Intelligence Scale for Children: Manual*; The Psychological Corporation: San Antonio, TX, USA, 1991.
57. Wechsler, D. *Wechsler Intelligence Scale for Children (WISC-IV)*; The Psychological Corporation: San Antonio, TX, USA, 2003.
58. Jewsbury, P.A.; Bowden, S.C.; Duff, K. The Cattell–Horn–Carroll Model of Cognition for Clinical Assessment. *J. Psychoeduc. Assess.* **2017**, *35*, 547–567. [[CrossRef](#)]
59. Nelson, H.E. *The National Adult Reading Test (NART): Test Manual*; NFER-Nelson: Windsor, UK, 1982.
60. Crawford, J.R.; Parker, D.M.; Stewart, L.E.; Besson, J.A.O.; De Lacey, G. Prediction of WAIS IQ with the National Adult Reading Test: Cross-validation and extension. *Br. J. Clin. Psychol.* **1989**, *28*, 267–273. [[CrossRef](#)]
61. Raven, J.C.; Court, J.; Raven, J. *Manual for Raven’s Progressive Matrices*; H. K. Lewis: London, UK, 1976.
62. Cattell, R.B.; Cattell, A.K. *Measuring Intelligence with the Culture Fair Tests*; Institute for Personality and Ability Testing: Champaign, IL, USA, 1973.
63. Floyd, R.G.; Bergeron, R.; Hamilton, G.; Parra, G.R. How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan Executive Function System and the Woodcock–Johnson III tests of cognitive abilities. *Psychol. Sch.* **2010**, *47*, 721–738. [[CrossRef](#)]
64. Hoelzle, J.B. *Neuropsychological Assessment and the Cattell–Horn–Carroll (CHC) Cognitive Abilities Model*; University of Toledo: Toledo, Spain, 2008.
65. Jewsbury, P.A.; Bowden, S.C.; Strauss, M.E. Confirmatory factor analysis of executive function models: A Cattell–Horn–Carroll based reanalysis integrating the switching, inhibition, and updating model of executive function with the Cattell–Horn–Carroll model. *J. Exp. Psychol. Gen.* **2016**, *145*, 220–245. [[CrossRef](#)] [[PubMed](#)]
66. Roberds, E.L. Evaluating the Relationship Between CHC Factors and Executive Functioning, Ball State University. Available online: <http://cardinalscholar.bsu.edu/handle/123456789/199524> (accessed on 2 May 2015).
67. Salthouse, T.A. Relations Between Cognitive Abilities and Measures of Executive Functioning. *Neuropsychology* **2005**, *19*, 532–545. [[CrossRef](#)]
68. van Aken, L.; Kessels, R.P.C.; Wingbermühle, E.; Wiltink, M.; van der Heijden, P.T.; Egger, J.I.M. Exploring the incorporation of executive functions in intelligence testing: Factor analysis of the WAIS-III and traditional tasks of executive functioning. *Int. J. Appl. Psychol.* **2014**, *4*, 73–80.
69. van Aken, L.; Kessels, R.P.C.; Wingbermühle, E.; van der Veld, W.M.; Egger, J.I.M. Fluid intelligence and executive functioning more alike than different? *Acta Neuropsychiatr.* **2016**, *28*, 31–37. [[CrossRef](#)]
70. van Aken, L.; van der Heijden, P.T.; van der Veld, W.M.; Hermans, L.; Kessels, R.P.C.; Egger, J.I.M. Representation of the Cattell–Horn–Carroll Theory of Cognitive Abilities in the Factor Structure of the Dutch–Language Version of the WAIS-IV. *Assessment* **2017**, *24*, 458–466. [[CrossRef](#)]
71. van Aken, L.; van der Heijden, P.T.; Oomens, W.; Kessels, R.P.C.; Egger, J.I.M. Predictive Value of Traditional Measures of Executive Function on Broad Abilities of the Cattell–Horn–Carroll Theory of Cognitive Abilities. *Assessment* **2017**, *26*, 1375–1385. [[CrossRef](#)]
72. Roca, M.; Parr, A.; Thompson, R.; Woolgar, A.; Torralva, T.; Antoun, N.; Manes, F.; Duncan, J. Executive function and fluid intelligence after frontal lobe lesions. *Brain* **2010**, *133*, 234–247. [[CrossRef](#)] [[PubMed](#)]
73. Cronbach, L.J. *Essentials of Psychological Testing*, 3rd ed.; Harper & Row: New York, NY, USA, 1970.
74. Jewsbury, P.A.; Bowden, S.C. Construct Validity has a Critical Role in Evidence-Based Neuropsychological Assessment. In *Neuropsychological Assessment in the Age of Evidence-Based Practice*; Oxford University Press: New York, NY, USA, 2017; pp. 33–63.
75. Schönbrodt, F.D.; Perugini, M. At what sample size do correlations stabilize? *J. Res. Pers.* **2013**, *47*, 609–612. [[CrossRef](#)]
76. Field, A.P.; Gillett, R. How to do a meta-analysis. *Br. J. Math. Stat. Psychol.* **2010**, *63*, 665–694. [[CrossRef](#)] [[PubMed](#)]
77. Strauss, E.; Sherman, E.M.S.; Spreen, O. *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*; Oxford University Press: New York, NY, USA, 2006.

78. Lehrl, S. *Mehrfach-Wortschatz-Intelligenztest (MWT-B): Handbuch*; Straube: Erlangen, Germany, 1977.
79. Ammons, R.B.; Ammons, C.H. The Quick Test (QT): Provisional Manual. *Psychol. Rep.* **1962**, *11*, 111–118. [[CrossRef](#)]
80. Zachary, R.A.; Shipley, W.C. *Shipley Institute of Living Scale: Revised Manual*, 4th ed.; Western Psychological Services: Los Angeles, CA, USA, 1986.
81. Thorndike, R.L. *The Stanford-Binet Intelligence Scale: Guide for Administering and Scoring*; The Riverside Publishing Company: Chicago, IL, USA, 1986.
82. Thurstone, L.; Thurstone, T. *SRA Primary Mental Abilities Test*; Science Research Associates: Chicago, IL, USA, 1949.
83. Kaufman, A.S.; Kaufman, N.L. *Kaufman Brief Intelligence Test: Manual*; American Guidance Service: Circle Pines, MN, USA, 1990.
84. Ardila, A.; Pineda, D.; Rosselli, M. Correlation Between Intelligence Test Scores and Executive Function Measures. *Arch. Clin. Neuropsychol.* **2000**, *15*, 31–36. [[CrossRef](#)]
85. Bird, C.M.; Papadopoulou, K.; Ricciardelli, P.; Rossor, M.N.; Cipolotti, L. Monitoring cognitive changes: Psychometric properties of six cognitive tests. *Br. J. Clin. Psychol.* **2004**, *43*, 197–210. [[CrossRef](#)]
86. Boone, K.B.; Pontón, M.O.; Gorsuch, R.L.; González, J.J.; Miller, B.L. Factor Analysis of Four Measures of Prefrontal Lobe Functioning. *Arch. Clin. Neuropsychol.* **1998**, *13*, 585–595. [[CrossRef](#)]
87. Chien, C.C.; Huang, S.F.; Lung, F.W. Maximally efficient two-stage screening: Determining intellectual disability in Taiwanese military conscripts. *J. Multidiscip. Healthc.* **2009**, *2*, 39–44.
88. Cianchetti, C.; Corona, S.; Foscoliano, M.; Contu, D.; Sannio-Fancello, G. Modified Wisconsin Card Sorting Test (MCST, MWCST): Normative Data in Children 4–13 Years Old, According to Classical and New Types of Scoring. *Clin. Neuropsychol.* **2007**, *21*, 456–478. [[CrossRef](#)]
89. Crawford, J.R.; Bryan, J.; Luszcz, M.A.; Obonsawin, M.C.; Stewart, L. The Executive Decline Hypothesis of Cognitive Aging: Do Executive Deficits Qualify as Differential Deficits and Do They Mediate Age-Related Memory Decline? *Aging, Neuropsychol. Cogn.* **2000**, *7*, 9–31. [[CrossRef](#)]
90. Davis, R.N.; Nolen-Hoeksema, S. Cognitive Inflexibility Among Ruminators and Nonruminators. *Cognit. Ther. Res.* **2000**, *24*, 699–711. [[CrossRef](#)]
91. de Zubizaray, G.I.; Smith, G.A.; Chalk, J.B.; Semple, J. The Modified Card Sorting Test: Test-retest stability and relationships with demographic variables in a healthy older adult sample. *Br. Psychol. Soc.* **1998**, *37*, 457–466. [[CrossRef](#)]
92. Dieci, M.; Vita, A.; Silenzi, C.; Caputo, A.; Comazzi, M.; Ferrari, L.; Ghiringhelli, L.; Mezzetti, M.; Tenconi, F.; Invernizzi, G. Non-selective impairment of Wisconsin Card Sorting Test performance in patients with schizophrenia. *Schizophr. Res.* **1997**, *25*, 33–42. [[CrossRef](#)]
93. Dolan, M.; Millington, J.; Park, I. Personality and Neuropsychological Function in Violent, Sexual and Arson Offenders. *Med. Sci. Law* **2002**, *42*, 34–43. [[CrossRef](#)] [[PubMed](#)]
94. Evans, L.D.; Kouros, C.D.; Samanez-Larkin, S.; Garber, J. Concurrent and Short-Term Prospective Relations among Neurocognitive Functioning, Coping, and Depressive Symptoms in Youth. *J. Clin. Child Adolesc. Psychol.* **2016**, *45*, 6–20. [[CrossRef](#)]
95. Giovagnoli, A.R. Relation of sorting impairment to hippocampal damage in temporal lobe epilepsy. *Neuropsychologia* **2001**, *39*, 140–150. [[CrossRef](#)]
96. Golden, C.J.; Kushner, T.; Lee, B.; McMorro, M.A. Searching for the meaning of the category test and the wisconsin card sort test: A Comparative analysis. *Int. J. Neurosci.* **1998**, *93*, 141–150. [[CrossRef](#)]
97. Han, G.; Helm, J.; Iucha, C.; Zahn-Waxler, C.; Hastings, P.D.; Klimes-Dougan, B. Are Executive Functioning Deficits Concurrently and Predictively Associated with Depressive and Anxiety Symptoms in Adolescents? *J. Clin. Child Adolesc. Psychol.* **2016**, *45*, 44–58. [[CrossRef](#)]
98. Heinrichs, R.W. Variables associated with Wisconsin Card Sorting Test performance in neuropsychiatric patients referred for assessment. *Neuropsychiatry, Neuropsychol. Behav. Neurol.* **1990**, *3*, 107–112.
99. Ilonen, T.; Taiminen, T.; Lauerma, H.; Karlsson, H.; Helenius, H.Y.M.; Tuimala, P.; Leinonen, K.-M.; Wallenius, E.; Salokangas, R.K.R. Impaired Wisconsin Card Sorting Test performance in first-episode schizophrenia: Resource or motivation deficit? *Compr. Psychiatry* **2000**, *41*, 385–391. [[CrossRef](#)]
100. Isingrini, M.; Vazou, F. Relation between Fluid Intelligence and Frontal Lobe Functioning in Older Adults. *Int. J. Aging Hum. Dev.* **1997**, *45*, 99–109. [[CrossRef](#)] [[PubMed](#)]

101. Keefe, R.S.E.; Silverman, J.M.; Lees Roitman, S.E.; Harvey, P.D.; Duncan, M.A.; Alroy, D.; Siever, L.J.; Davis, K.L.; Mohs, R.C. Performance of nonpsychotic relatives of schizophrenic patients on cognitive tests. *Psychiatry Res.* **1994**, *53*, 1–12. [[CrossRef](#)]
102. Kilincaslan, A.; Motavalli Mukkaddes, N.; Kücükayzici, G.S.; Gürvit, H. Assessment of executive/attentional performance in Asperger's disorder. *Turkish J. Psychiatry* **2010**, *21*, 289–299.
103. Lee, S.J.; Lee, H.-K.; Kweon, Y.-S.; Lee, C.T.; Lee, K.-U. The Impact of Executive Function on Emotion Recognition and Emotion Experience in Patients with Schizophrenia. *Psychiatry Investig.* **2009**, *6*, 156–162. [[CrossRef](#)] [[PubMed](#)]
104. Lehto, J.E.; Elorinne, E. Gambling as an Executive Function Task. *Appl. Neuropsychol.* **2003**, *10*, 234–238. [[CrossRef](#)]
105. Lehto, J.E. A Test for Children's Goal-Directed Behavior: A Pilot Study. *Percept. Mot. Skills* **2004**, *98*, 223–236. [[CrossRef](#)]
106. LeMonda, B.C.; Holtzer, R.; Goldman, S. Relationship between executive functions and motor stereotypies in children with autistic disorder. *Res. Autism Spectr. Disord.* **2012**, *6*, 1099–1106. [[CrossRef](#)]
107. Lichtenstein, J.D.; Erdodi, L.A.; Rai, J.K.; Mazur-Mosiewicz, A.; Flaro, L. Wisconsin Card Sorting Test embedded validity indicators developed for adults can be extended to children. *Child Neuropsychol.* **2018**, *24*, 247–260. [[CrossRef](#)]
108. Lineweaver, T.T.; Bondi, M.W.; Thomas, R.G.; Salmon, D.P. A Normative Study of Nelson's (1976) Modified Version of the Wisconsin Card Sorting Test in Healthy Older Adults. *Clin. Neuropsychol.* **1999**, *13*, 328–347. [[CrossRef](#)]
109. Liss, M.; Fein, D.; Allen, D.; Dunn, M.; Feinstein, C.; Morris, R.; Waterhouse, L.; Rapin, I. Executive Functioning in High-functioning Children with Autism. *J. Child Psychol. Psychiatry* **2001**, *42*, 261–270. [[CrossRef](#)]
110. Lucey, J.V.; Burness, C.E.; Costa, D.C.; Gacinovic, S.; Pilowsky, L.S.; Ell, P.J.; Marks, I.M.; Kerwin, R.W. Wisconsin Card Sorting Task (WCST) errors and cerebral blood flow in obsessive-compulsive disorder (OCD). *Br. J. Med. Psychol.* **1997**, *70*, 403–411. [[CrossRef](#)] [[PubMed](#)]
111. Minshew, N.J.; Meyer, J.; Goldstein, G. Abstract reasoning in autism: A dissociation between concept formation and concept identification. *Neuropsychology* **2002**, *16*, 327–334. [[CrossRef](#)] [[PubMed](#)]
112. Mullane, J.C.; Corkum, P.V. The Relationship Between Working Memory, Inhibition, and Performance on the Wisconsin Card Sorting Test in Children With and Without ADHD. *J. Psychoeduc. Assess.* **2007**, *25*, 211–221. [[CrossRef](#)]
113. Nestor, P.G.; Nakamura, M.; Niznikiewicz, M.; Levitt, J.J.; Newell, D.T.; Shenton, M.E.; McCarley, R.W. Attentional Control and Intelligence: MRI Orbital Frontal Gray Matter and Neuropsychological Correlates. *Behav. Neurol.* **2015**, *2015*, 1–8. [[CrossRef](#)] [[PubMed](#)]
114. Obonsawin, M.C.; Crawford, J.R.; Page, J.; Chalmers, P.; Low, G.; Marsh, P. Performance on the Modified Card Sorting Test by normal, healthy individuals: Relationship to general intellectual ability and demographic variables. *Br. J. Clin. Psychol.* **1999**, *38*, 27–41. [[CrossRef](#)]
115. Obonsawin, M.; Crawford, J.R.; Page, J.; Chalmers, P.; Cochrane, R.; Low, G. Performance on tests of frontal lobe function reflect general intellectual ability. *Neuropsychologia* **2002**, *40*, 970–977. [[CrossRef](#)]
116. Owashi, T.; Iwanami, A.; Nakagome, K.; Higuchi, T.; Kamijima, K. Thought Disorder and Executive Dysfunction in Patients with Schizophrenia. *Int. J. Neurosci.* **2009**, *119*, 105–123. [[CrossRef](#)]
117. Perry, W.; Braff, D.L. A multimethod approach to assessing perseverations in schizophrenia patients. *Schizophr. Res.* **1998**, *33*, 69–77. [[CrossRef](#)]
118. Roca, M.; Manes, F.; Chade, A.; Gleichgerrcht, E.; Gershanik, O.; Arévalo, G.G.; Torralva, T.; Duncan, J. The relationship between executive functions and fluid intelligence in Parkinson's disease. *Psychol. Med.* **2012**, *42*, 2445–2452. [[CrossRef](#)]
119. Rossell, S.L.; Coakes, J.; Shapleske, J.; Woodruff, P.W.R.; David, A.S. Insight: Its relationship with cognitive function, brain volume and symptoms in schizophrenia. *Psychol. Med.* **2003**, *33*, 111–119. [[CrossRef](#)]
120. Salthouse, T.A.; Fristoe, N.; Rhee, S.H. How localized are age-related effects on neuropsychological measures? *Neuropsychology* **1996**, *10*, 272–285. [[CrossRef](#)]
121. Schiebener, J.; García-Arias, M.; García-Villamisar, D.; Cabanyes-Truffino, J.; Brand, M. Developmental changes in decision making under risk: The role of executive functions and reasoning abilities in 8- to 19-year-old decision makers. *Child Neuropsychol.* **2015**, *21*, 759–778. [[CrossRef](#)] [[PubMed](#)]

122. Shura, R.D.; Miskey, H.M.; Rowland, J.A.; Yoash-Gantz, R.E.; Denning, J.H. Embedded Performance Validity Measures with Postdeployment Veterans: Cross-Validation and Efficiency with Multiple Measures. *Appl. Neuropsychol. Adult* **2016**, *23*, 94–104. [CrossRef] [PubMed]
123. South, M.; Ozonoff, S.; McMahon, W.M. The relationship between executive functioning, central coherence, and repetitive behaviors in the high-functioning autism spectrum. *Autism* **2007**, *11*, 437–451. [CrossRef]
124. Steingass, H.-P.; Sartory, G.; Canavan, A.G.M. Chronic alcoholism and cognitive function: General decline or patterned impairment? *Pers. Individ. Dif.* **1994**, *17*, 97–109. [CrossRef]
125. Sweeney, J.A.; Keilp, J.G.; Haas, G.L.; Hill, J.; Weiden, P.J. Relationships between medication treatments and neuropsychological test performance in schizophrenia. *Psychiatry Res.* **1991**, *37*, 297–308. [CrossRef]
126. Syngelaki, E.M.; Moore, S.C.; Savage, J.C.; Fairchild, G.; Van Goozen, S.H.M. Executive Functioning and Risky Decision Making in Young Male Offenders. *Crim. Justice Behav.* **2009**, *36*, 1213–1227. [CrossRef]
127. Tacconat, L.; Clarys, D.; Vanneste, S.; Bouazzaoui, B.; Isingrini, M. Aging and strategic retrieval in a cued-recall test: The role of executive functions and fluid intelligence. *Brain Cogn.* **2007**, *64*, 1–6. [CrossRef]
128. Whiteside, D.M.; Kealey, T.; Semla, M.; Luu, H.; Rice, L.; Basso, M.R.; Roper, B. Verbal Fluency: Language or Executive Function Measure? *Appl. Neuropsychol. Adult* **2016**, *23*, 29–34. [CrossRef]
129. Yasuda, Y. Cognitive inflexibility in Japanese adolescents and adults with autism spectrum disorders. *World J. Psychiatry* **2014**, *4*, 42. [CrossRef]
130. Hedges, L.V.; Vevea, J.L. Fixed-and random-effects models in meta-analysis. *Psychol. Methods* **1998**, *3*, 486–504. [CrossRef]
131. Higgins, J.P.T.; Thompson, S.G.; Deeks, J.J.; Altman, D.G. Measuring inconsistency in meta-analyses. *BMJ* **2003**, *327*, 557–560. [CrossRef] [PubMed]
132. Wan, X.; Wang, W.; Liu, J.; Tong, T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med. Res. Methodol.* **2014**, *14*, 135. [CrossRef] [PubMed]
133. Carter, E.C.; McCullough, M.E. Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Front. Psychol.* **2014**, *5*, 1–11. [CrossRef]
134. VassarStats: Website for Statistical Computation. Available online: <http://vassarstats.net/> (accessed on 16 April 2018).
135. Cohen, J. A power primer. *Psychol. Bull.* **1992**, *112*, 115–159. [CrossRef]
136. Friedman, N.P.; Miyake, A.; Robinson, J.L.; Hewitt, J.K. Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Dev. Psychol.* **2011**, *47*, 1410–1430. [CrossRef]
137. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [CrossRef]
138. Hunter, J.E.; Schmidt, F.L. Cumulative research knowledge and social policy formulation: The critical role of meta-analysis. *Psychol. Public Policy, Law* **1996**, *2*, 324–347. [CrossRef]
139. Nunnally, J.C.; Bernstein, I.H. *Psychometric Theory (Vol. 3)*; McGraw-Hill: New York, NY, USA, 1994.
140. Tranel, D.; Manzel, K.; Anderson, S.W. Is the Prefrontal Cortex Important For Fluid Intelligence? A Neuropsychological Study Using Matrix Reasoning. *Clin. Neuropsychol.* **2008**, *22*, 242–261. [CrossRef]
141. Nenty, H.J.; Dinero, T.E. A Cross-Cultural Analysis of the Fairness of the Cattell Culture Fair Intelligence Test Using the Rasch Model. *Appl. Psychol. Meas.* **1981**, *5*, 355–368. [CrossRef]
142. Burke, H.R. Raven's Progressive Matrices: Validity, Reliability, and Norms. *J. Psychol.* **1972**, *82*, 253–257. [CrossRef]
143. Duncan, J.; Seitz, R.J.; Kolodny, J.; Bor, D.; Herzog, H.; Ahmed, A.; Newell, F.N.; Emslie, H. A Neural Basis for General Intelligence. *Science (80-.)*. **2000**, *289*, 457–460. [CrossRef] [PubMed]
144. Gläscher, J.; Tranel, D.; Paul, L.K.; Rudrauf, D.; Rorden, C.; Hornaday, A.; Grabowski, T.; Damasio, H.; Adolphs, R. Lesion Mapping of Cognitive Abilities Linked to Intelligence. *Neuron* **2009**, *61*, 681–691. [CrossRef] [PubMed]
145. Gläscher, J.; Rudrauf, D.; Colom, R.; Paul, L.K.; Tranel, D.; Damasio, H.; Adolphs, R. Distributed neural system for general intelligence revealed by lesion mapping. *Proc. Natl. Acad. Sci.* **2010**, *107*, 4705–4709. [CrossRef]
146. Basten, U.; Hilger, K.; Fiebach, C.J. Where smart brains are different: A quantitative meta-analysis of functional and structural brain imaging studies on intelligence. *Intelligence* **2015**, *51*, 10–27. [CrossRef]

147. Deary, I.J.; Penke, L.; Johnson, W. The neuroscience of human intelligence differences. *Nat. Rev. Neurosci.* **2010**, *11*, 201–211. [[CrossRef](#)] [[PubMed](#)]
148. Euler, M.J. Intelligence and uncertainty: Implications of hierarchical predictive processing for the neuroscience of cognitive ability. *Neurosci. Biobehav. Rev.* **2018**, *94*, 93–112. [[CrossRef](#)] [[PubMed](#)]
149. Lange, F.; Seer, C.; Kopp, B. Cognitive flexibility in neurological disorders: Cognitive components and event-related potentials. *Neurosci. Biobehav. Rev.* **2017**, *83*, 496–507. [[CrossRef](#)]
150. Christoff, K.; Prabhakaran, V.; Dorfman, J.; Zhao, Z.; Kroger, J.K.; Holyoak, K.J.; Gabrieli, J.D.E. Rostrolateral Prefrontal Cortex Involvement in Relational Integration during Reasoning. *Neuroimage* **2001**, *14*, 1136–1149. [[CrossRef](#)]
151. Kroger, J.K.; Sabb, F.W.; Fales, C.L.; Bookheimer, S.Y.; Cohen, M.S.; Holyoak, K.J. Recruitment of Anterior Dorsolateral Prefrontal Cortex in Human Reasoning: A Parametric Study of Relational Complexity. *Cereb. Cortex* **2002**, *12*, 477–485. [[CrossRef](#)]
152. Morrison, R.G.; Krawczyk, D.C.; Holyoak, K.J.; Hummel, J.E.; Chow, T.W.; Miller, B.L.; Knowlton, B.J. A Neurocomputational Model of Analogical Reasoning and its Breakdown in Frontotemporal Lobar Degeneration. *J. Cogn. Neurosci.* **2004**, *16*, 260–271. [[CrossRef](#)] [[PubMed](#)]
153. Waltz, J.A.; Knowlton, B.J.; Holyoak, K.J.; Boone, K.B.; Mishkin, F.S.; de Menezes Santos, M.; Thomas, C.R.; Miller, B.L. A System for Relational Reasoning in Human Prefrontal Cortex. *Psychol. Sci.* **1999**, *10*, 119–125. [[CrossRef](#)]
154. Waltz, J.A.; Knowlton, B.J.; Holyoak, K.J.; Boone, K.B.; Back-Madruga, C.; McPherson, S.; Masterman, D.; Chow, T.; Cummings, J.L.; Miller, B.L. Relational Integration and Executive Function in Alzheimer’s Disease. *Neuropsychology* **2004**, *18*, 296–305. [[CrossRef](#)] [[PubMed](#)]
155. Schmidt, F.L.; Hunter, J.E. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*, 3rd ed.; Sage publications: London, UK, 2015.
156. Chapman, L.J.; Chapman, J.P. Problems in the measurement of cognitive deficits. *Psychol. Bull.* **1973**, *79*, 380–385. [[CrossRef](#)] [[PubMed](#)]
157. Chapman, L.J.; Chapman, J.P. The measurement of differential deficit. *J. Psychiatr. Res.* **1978**, *14*, 303–311. [[CrossRef](#)]
158. Greve, K.M.; Ingram, F.; Bianchini, K.J.; Stanford, M.S. Latent Structure of the Wisconsin Card Sorting Test in a Clinical Sample. *Arch. Clin. Neuropsychol.* **1998**, *13*, 597–609. [[CrossRef](#)]
159. Greve, K.W.; Stickle, T.R.; Love, J.M.; Bianchini, K.J.; Stanford, M.S. Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Arch. Clin. Neuropsychol.* **2005**, *20*, 355–364. [[CrossRef](#)]
160. Paolo, A.; Tröster, A.I.; Axelrod, N.; Koller, W.C. Construct validity of the WCST in normal elderly and persons with Parkinson’s disease. *Arch. Clin. Neuropsychol.* **1995**, *10*, 463–473. [[CrossRef](#)]
161. Sánchez-Cubillo, I.; Periáñez, J.A.; Arover-Roig, D.; Rodríguez-Sánchez, J.M.; Ríos-Lago, M.; Tirapu, J.; Barceló, F. Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *J. Int. Neuropsychol. Soc.* **2009**, *15*, 438–450. [[CrossRef](#)]
162. Bowden, S.C. *Neuropsychological Assessment in the Age of Evidence-Based Practice*; Oxford University Press: New York, NY, USA, 2017.
163. Chelune, G.J. Evidence-Based Practices in Neuropsychology. In *Neuropsychological Assessment in the Age of Evidence-Based Practice*; Bowden, S.C., Ed.; Oxford University Press: New York, NY, USA, 2017; pp. 155–181.
164. Greve, K.W.; Love, J.M.; Sherwin, E.; Mathias, C.W.; Houston, R.J.; Brennan, A. Temporal Stability of the Wisconsin Card Sorting Test in a Chronic Traumatic Brain Injury Sample. *Assessment* **2002**, *9*, 271–277. [[CrossRef](#)]
165. Tate, R.L.; Perdices, M.; Maggiotto, S. Stability of the Wisconsin Card Sorting Test and the Determination of Reliability of Change in Scores. *Clin. Neuropsychol.* **1998**, *12*, 348–357. [[CrossRef](#)]
166. Charter, R.A. Sample Size Requirements for Precise Estimates of Reliability, Generalizability, and Validity Coefficients. *J. Clin. Exp. Neuropsychol.* **1999**, *21*, 559–566. [[CrossRef](#)] [[PubMed](#)]
167. Basso, M.R.; Bornstein, R.A.; Lang, J.M. Practice Effects on Commonly Used Measures of Executive Function Across Twelve Months. *Clin. Neuropsychol.* **1999**, *13*, 283–292. [[CrossRef](#)]
168. Bowden, S.C.; Fowler, K.S.; Bell, R.C.; Whelan, G.; Clifford, C.C.; Ritter, A.J.; Long, C.M. The Reliability and Internal Validity of the Wisconsin Card Sorting Test. *Neuropsychol. Rehabil.* **1998**, *8*, 243–254. [[CrossRef](#)]

169. Ingram, F.; Greve, K.W.; Ingram, P.T.F.; Soukup, V.M. Temporal stability of the Wisconsin Card Sorting Test in an untreated patient sample. *Br. J. Clin. Psychol.* **1999**, *38*, 209–211. [[CrossRef](#)] [[PubMed](#)]
170. Paolo, A.M.; Axelrod, B.N.; Tröster, A.I. Test-Retest Stability of the Wisconsin Card Sorting Test. *Assessment* **1996**, *3*, 137–143. [[CrossRef](#)]
171. Steinmetz, J.-P.; Brunner, M.; Loarer, E.; Houssemand, C. Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin Card Sorting Test? *Psychol. Assess.* **2010**, *22*, 199–202. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).