



Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility

Stephanie Miles¹ · Caitlin A. Howlett² · Carolyn Berryman³ · Maja Nedeljkovic¹ · G. Lorimer Moseley² · Andrea Phillipou^{1,4,5,6}

Accepted: 22 January 2021 / Published online: 22 March 2021
© The Psychonomic Society, Inc. 2021

Abstract

The Wisconsin Card Sorting Test (WCST) is a popular neurocognitive task used to assess cognitive flexibility, and aspects of executive functioning more broadly, in research and clinical practice. Despite its widespread use and the development of an updated WCST manual in 1993, confusion remains in the literature about how to score the WCST, and importantly, how to interpret the outcome variables as indicators of cognitive flexibility. This critical review provides an overview of the changes in the WCST, how existing scoring methods of the task differ, the key terminology and how these relate to the assessment of cognitive flexibility, and issues with the use of the WCST across the literature. In particular, this review focuses on the confusion between the terms ‘perseverative responses’ and ‘perseverative errors’ and the inconsistent scoring of these variables. To our knowledge, this critical review is the first of its kind to focus on the inherent issues surrounding the WCST when used as an assessment of cognitive flexibility. We provide recommendations to overcome these and other issues when using the WCST in future research and clinical practice.

Keywords Cognitive flexibility · Set-shifting · Executive function · WCST · Neurocognitive measures · Neurocognitive assessment

Cognitive flexibility refers to the mental ability that enables us to effectively adjust to changing task and/or environmental demands (Deák, 2003; Scott, 1962) and is thought to arise from the interaction between higher-order executive functions

(Dajani & Uddin, 2015). Further, cognitive flexibility is a critical cognitive function that allows an individual to switch their cognitive strategies, consider two or more aspects of an object, idea, or complex situation simultaneously, and appropriately adapt behavioural strategies (Bilgin, 2009; Dennis & Vander Wal, 2010; Diamond, 2013). There are many strategies by which researchers and clinicians assess cognitive flexibility, one of which is the use of neurocognitive tests.

The most commonly used neurocognitive test for assessing cognitive flexibility is the Wisconsin Card Sorting Test (WCST; Berg, 1948; Johnco, Wuthrich, & Rapee, 2014; Tchanturia et al., 2012). Originally popular as an assessment of frontal lobe damage (Barceló & Knight, 2002; Milner, 1963), by 2005, the WCST was the seventh most frequent test implemented by clinical neuropsychologists (Rabin, Barr, & Burton, 2005). The WCST yields upwards of seven variables. ‘Perseverative responses’ and ‘perseverative errors’ are the most commonly used to assess cognitive flexibility (Baker, Georgiou-Karistianis, et al., 2018a; Baker, Gibson, Georgiou-Karistianis, & Giummarra, 2018b; Dickson, Ciesla, & Zelic, 2017; Garcia-Willingham, Roach, Kasarskis, & Segerstrom, 2018; Gelonch, Garolera, Valls, Rosselló, & Pifarré, 2016; Wollenhaupt et al., 2019), but

Stephanie Miles and Caitlin A. Howlett are Co-first authors

G. Lorimer Moseley and Andrea Phillipou are Co-last-authors

✉ Stephanie Miles
smiles@swin.edu.au

¹ Centre for Mental Health, Swinburne University of Technology, PO Box 218, Hawthorn, VIC 3122, Australia

² Innovation, Implementation & Clinical Translation (IMPACT) in Health, Allied Health and Human Performance, University of South Australia, Adelaide, Australia

³ Brain Stimulation, Imaging and Cognition Research Group, School of Medicine, The University of Adelaide, Adelaide, Australia

⁴ Department of Mental Health, St Vincent’s Hospital, Melbourne, Australia

⁵ Department of Psychiatry, The University of Melbourne, Melbourne, Australia

⁶ Department of Mental Health, Austin Health, Melbourne, Australia

‘number of categories completed’, ‘failure-to-maintain-set’, ‘trials to complete the first category’ and ‘non-perseverative errors’ have also been used (e.g. Abbate-Daga, Buzzichelli, Marzola, Amianto, & Fassino, 2014; Aloï et al., 2015; Bischoff-Grethe et al., 2013; Dickson et al., 2017; Gelonch et al., 2016; Tchanturia et al., 2012; Wollenhaupt et al., 2019; Zmigrod, Rentfrow, & Robbins, 2018). This has led to criticism of the WCST for having too many outcome variables (Figuroa & Youmans, 2013), that are not clearly linked to particular cognitive domains (Greve, Stickler, Love, Bianchini, & Stanford, 2005). Moreover, various definitions and interchanging of variables, inconsistency in how to obtain the variables, and unclear, incomplete, or misleading reporting clouds the field. Here we highlight several critical problems with the current use of the WCST and recommend solutions for implementation in research and clinical practice.

An overview of the Wisconsin Card Sorting Test

The WCST is a card matching task. For each trial, a response card is placed above four multidimensional stimulus cards (see Fig. 1 for an example of the WCST). The cards presented in the task vary on three dimensions—colour (red, blue, yellow, green), form (circles, triangles, stars, crosses), and number (one, two, three, four). For each ‘trial’, participants ‘match’ the response card to one of the four stimulus cards, without specific instruction by the administrator. The sorting rule is the dimension on which the card needs to be correctly matched, and the participant identifies the sorting rule through a process of trial and error. For example, a response card with two blue triangles can be matched according to colour (blue), form (triangle), or number (two). After each response, the participant receives feedback (i.e. ‘correct’ or ‘incorrect’) that

can be used to establish the correct sorting rule. Typically, the sorting rule changes without warning after ten correct responses in a row (this is called ‘completing a category’) and the participant must ‘start again’ to establish the new sorting rule for the following category. In the standard administration of the task, the order of the sorting rule is colour, form, number, colour, form, number (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The WCST terminates when either (i) all six categories are completed, or (ii) 128 trials are completed. The WCST was first developed as a manually administered task using physical cards; however, it has since been adapted to a computerised format (Heaton & PAR Staff, 2008), and is widely used in both physical and electronic forms.

Changes in the WCST since its development

The WCST was first developed by Berg (1948) to assess flexibility in thinking, but has since seen some changes (see Eling, Derckx, & Maes, 2008, for a full historical and conceptual review). Briefly, the original WCST (Berg, 1948) required participants to make five correct responses in a row before the sorting rule changed (it is now ten), and nine categories were completed (it is now six). The original WCST consisted of 60 stimulus cards (Berg, 1948); Grant and Berg (1948) modified the task to include 64 stimulus cards, which were repeated if the participant used all 64 cards before completing the task (i.e. completing all categories). The original WCST had no limits on the number of trials taken to complete the task (Berg, 1948); there are now a maximum of 128 trials (Heaton et al., 1993); however, shorter and perhaps more practical versions of the WCST have been developed and are commonly used in clinical settings (i.e. the WCST-64 card version; Greve, 2001). Berg (1948) considered only the average number of errors made and the number of categories

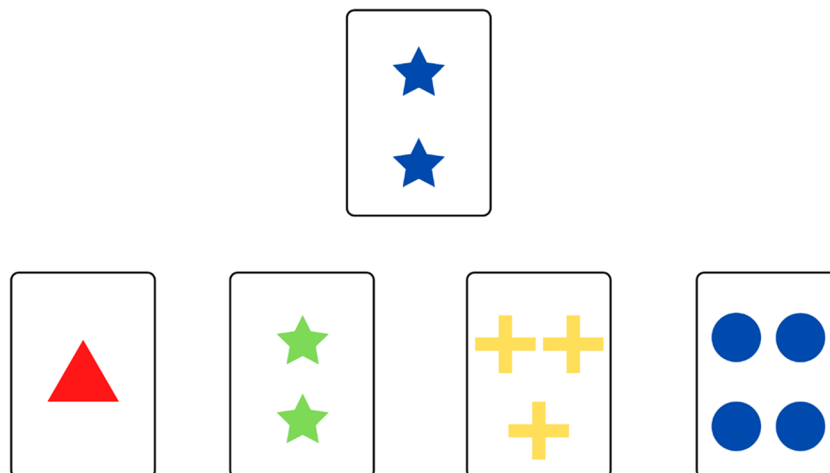


Fig. 1 An example of the WCST display. The figure depicts a response card with two blue stars. The stimulus cards are from left to right; one red triangle, two green stars, three yellow crosses, and four blue circles. In the

manual administration, the four stimulus cards would be laid on a table in front of the participant, and the participant is handed a deck of multidimensional response cards to sort

completed during the WCST; Grant and Berg (1948) added perseverative and non-perseverative errors as key outcome variables for the WCST and provided basic scoring information for the task.

Despite the test's development over time, its popularity, and widespread use, little to no standardisation of the administration, scoring, and terminology existed until Heaton, Chelune, Talley, Kay, and Curtiss (1981) published a standardised manual. Even then, the administration and scoring rules remained unclear to many (Flashman, Horner, & Freides, 1991; Greve, 1993), and supplementary, more transparent scoring guidelines were developed (e.g. Axelrod, Goldman, & Woodard, 1992; Flashman et al., 1991). An updated WCST manual was published in 1993 (Heaton et al., 1993), providing clarification for how to score perseverative responses, one of the key variables used today to assess cognitive flexibility.

Inconsistent terminology and unclear definitions

There are inconsistencies in the WCST terminology used throughout the literature, and the definitions used to operationalise these terms vary between studies. These inconsistencies have caused significant confusion for both researchers and clinicians who use the WCST, and have led to further discrepancies, bespoke scoring approaches, and a lack of transparency in reporting administration and scoring. Many sources of confusion exist. For example, Heaton et al. (1993) used '*dimension*' to refer to the characteristics of the cards by which they can be sorted (i.e. colour, form, number, or other); WCST studies have variably used *attribute*, *characteristic*, *criterion*, *rule*, *category*, or *principle* in lieu of dimension. The term '*category*' refers to the discrete sections of the WCST (Grant & Berg, 1948; Heaton et al., 1993). There are six categories, which follow the order colour, form, number, colour, form, number. The category determines the correct sorting rule. Given that the dimensions are repeated twice, categories are sometimes referred to in terms of their numerical order (e.g. first category).

Confusion and problems in scoring the WCST

The 'perseverated-to' principle

The 'perseverated-to' principle, a key scoring principle of the WCST, remains a poorly understood concept. The perseverated-to principle was formally conceptualised by Heaton et al. (1993) after supplementary scoring guidelines for the WCST were developed by Flashman et al. (1991). The perseverated-to principle was described by Heaton et al.

(1993) in conjunction with a definition of perseveration, which may have caused confusion between the perseverated-to principle and the meaning of perseverative responses. In an attempt to clarify the key scoring terms of the WCST, we have provided definitions in Table 1 that are in line with the standardised manual (Heaton et al., 1993). The perseverated-to principle is established in the first category of the WCST after the first unambiguous incorrect response. Subsequent responses that match the perseverated-to principle are scored as perseverative. However, an ambiguous response must also meet the criteria for the 'sandwich rule', which specifies that an ambiguous response preceded and followed by an unambiguous response must match the dimension of the unambiguous response, thus demonstrating a perseverative response pattern.¹ Following the completion of a category (i.e. when ten correct responses are made in a row), the previously *correct sorting rule* becomes the *perseverated-to principle* in the next category (Heaton et al., 1993). For example, if the correct sorting rule in the previous category was colour and the correct sorting rule for the current category is form, colour becomes the perseverated-to principle (see Fig. 2, column (a) for an example). According to Heaton et al. (1993), the perseverated-to principle can also change within a category. Following from the previous example, if the participant makes three sequential unambiguous errors according to number within the form category (during which colour is the initial perseverated-to-principle), then the dimension 'number' becomes the new perseverated-to principle (see Fig. 2, column (a) for an example). It is important to note that in this situation the new perseverated-to principle does not come into effect until the second unambiguous error, and thus the second unambiguous error (not the first) is the first response to be scored as perseverative.

Perseverative responses and perseverative errors: Two key variables to assess cognitive flexibility

Perseverative responses are persistent responses made by a participant on the basis of an incorrect (previous or novel) stimulus dimension (Heaton et al., 1993; McCallum, 2017). For example, a participant may attempt to sort a card based on its form, receive feedback that this response is incorrect, and yet continue to make the same mistake on the following card. Typically, perseverative responses are incorrect (i.e. the response does not match the correct sorting rule), and are termed perseverative errors (Grant & Berg, 1948). However, a perseverative response may also be correct because of the

¹ The sandwich rule was first conceptualised by Grant & Berg (1948), fully developed by Flashman et al. (1991) and later incorporated into the updated WCST manual by Heaton et al. (1993)

Table 1 Key WCST scoring terminology derived from the revised WCST manual (Heaton et al., 1993)

Correct - the response given matches the correct sorting rule	Incorrect - the response given does not match the correct sorting rule
Correct sorting rule - the dimension that participants are meant to be sorting the cards by, i.e. the rule that is in effect for that category	Perseverated-to principle - the incorrect sorting dimension which a participant is repeatedly responding to (e.g. form, when the correct sorting rule is colour). The perseverated-to principle applies to only one dimension at a time (i.e. colour or form or number)
Perseverative responses - when the dimension the participant is sorting by matches the perseverated-to-principle, i.e. the participant continually responds using an incorrect dimension even after receiving feedback that this response is incorrect. Participants cannot persevere on the <i>other</i> dimension	Non-perseverative responses - when the dimension the participant is using to sort does not match the perseverated-to-principle
Ambiguous - when the stimulus card chosen matches the response card on two or more dimensions (e.g. colour and form). By definition, responses that match the <i>other</i> dimension are ambiguous. Figure 1 provides an illustration of a potentially ambiguous response. The response card matches the second stimulus card on both form and number (two stars)	Unambiguous - when the stimulus card chosen matches the response card on only one dimension (e.g. colour). Figure 1 provides an illustration of a potentially unambiguous response, in that the response card matches the last stimulus card on colour only (blue)

potential for ambiguous responses and the above-mentioned sandwich rule. On some trials of the WCST, the chosen stimulus card will match the response card on multiple dimensions (e.g. the response card ‘four green circles’ matches the stimulus card ‘four blue circles’ on both number and form). These sorts of responses are considered ambiguous because the administrator cannot be certain of which dimension the participant is using to sort the card. Typically, an inspection of the unambiguous responses before and after an ambiguous response can indicate the sorting rule that a participant is using.

According to the sandwich rule, an ambiguous response that follows and is followed by an unambiguous perseverative error, and matches the perseverated-to principle, is scored as a perseverative response (see Fig. 2, column (b) for an example of this). Thus, even in cases wherein the ambiguous response matches the correct sorting rule, the response is scored as perseverative, i.e. it is a perseverative response, but not a perseverative error. To summarise: all perseverative errors are perseverative responses, but not all perseverative responses are perseverative errors (see Fig. 3). The terms have been

(a)	(b)	(c)
Number sequentially correct	Number sequentially correct	Number sequentially correct
Response	Response	Response
Perseverative response	Perseverative response	Perseverative response
Correct sorting rule: colour, followed by form	Correct sorting rule: number, followed by form	Correct sorting rule: colour, followed by form
10 <u>C</u> F N O	10 C F <u>N</u> O	8 <u>C</u> F N O
<u>C</u> F N O ^{ptp} p	C F <u>N</u> O ^{ptp} p	9 <u>C</u> F <u>N</u> O
<u>C</u> F N O p	C F <u>N</u> O p	10 <u>C</u> F N O
C F <u>N</u> O	1 C <u>F</u> <u>N</u> O p	<u>C</u> F N O ^{ptp} p
C F <u>N</u> O ^{ptp} p	C F <u>N</u> O p	<u>C</u> F N O p
C F <u>N</u> O p	C F <u>N</u> O p	C F <u>N</u> O
C F <u>N</u> O p	1 C <u>F</u> N O	1 C <u>F</u> N O

Fig. 2 Examples of various scoring scenarios for the WCST according to the Heaton et al. (1993) scoring manual. *Note:* Each main column comprises three scoring columns—Number sequentially correct, Response, and Perseverative response. Each column is divided by a line at the point of category and rule change. Column (a) provides an illustration of the perseverated-to principle after a category change and how the perseverated-to principle can change within a category after three unambiguous responses; column (b) provides an illustration of how to score

perseverative responses according to the sandwich rule when there is an ambiguous response; and column (c) provides an illustration of how to score perseverative errors. The number in the correct column signifies the number of correct sequential responses; a bolded and underlined letter indicates which dimensions the participant’s response matches; C = colour; F = form; N = number; O = other (i.e. the response did not match on any of the three key dimensions); p = perseverative response; ptp = indicates the establishment of the perseverated-to-principle

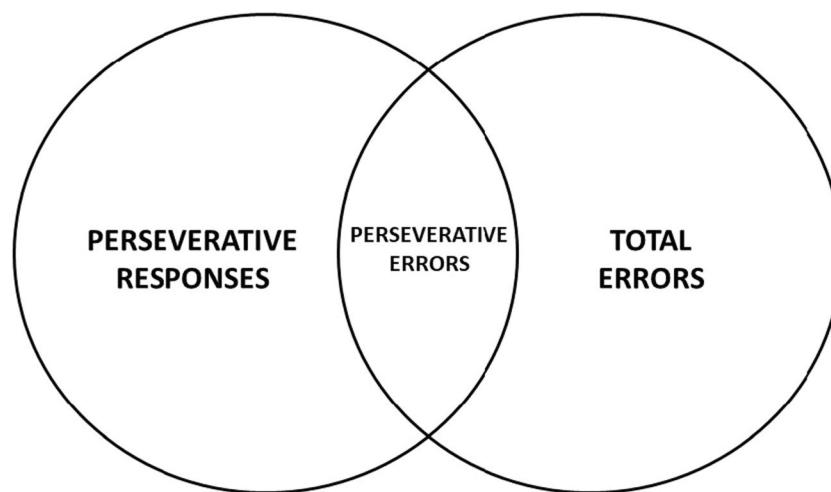


Fig. 3 A Venn diagram illustrating the relationship between perseverative responses and errors made on the WCST

sometimes used interchangeably (e.g. Gelonch et al., 2016; Øverås, Kapstad, Brunborg, Landrø, & Lask, 2015; Strauss, Sherman, & Spreen, 2006), illustrating the potential for erroneous results for both variables. This problem even infiltrates the revised WCST manual (Heaton et al., 1993), which uses the terms interchangeably and fails to highlight whether perseverative responses, perseverative errors, or both, should be reported.

Grant and Berg (1948) and Heaton et al. (1993) have different methods for scoring perseverative responses and perseverative errors. Grant and Berg (1948) specified that perseverative responses occurred after a rule change and were responses that matched the sorting rule of the previous category. Yet, the more contemporary definition provided by the revised WCST manual specifies that perseverative responses are those that match the perseverated-to principle (Heaton et al., 1993). Thus, a perseverative response (or perseverative error) can occur at any point in the task, including before a rule change and during the first category of the task (Heaton et al., 1993). Problematically, both scoring methods continue to be used in current research, which is likely to create difficulties in comparing results across studies.

When describing the number of categories achieved on the WCST, Strauss et al. (2006) explain that ‘scores can range from 0 for the subject who never gets the idea at all to 6’ (p. 528–529). Paradoxically, a participant who never learns the correct sorting rule (i.e. they fail to achieve ten correct responses in a row), and thus never advances from the first category, will receive a score of zero perseverative errors under the Grant and Berg (1948) scoring method. Considered to reflect high cognitive flexibility, a score of zero perseverative errors in this situation would be not only imprecise but contradictory. In these instances, the Grant and Berg (1948) scoring method provides an erroneous impression of cognitive flexibility. Typically, in instances where the participant has failed to complete the first

category and the Grant and Berg (1948) method has been used, if the participant’s responses are examined in greater detail, it becomes evident that the participant has perseverated within the first failed category. For example, the participant may have sorted according to an incorrect dimension several times in a row before testing another dimension, thus demonstrating cognitive rigidity. Consequently, in these cases, the Heaton et al. (1993) scoring method may be more appropriate to capture all instances of perseverative responses. Although the number of categories completed may not be of importance when considering other WCST variables (e.g. failure-to-maintain-set), it is evident that the number of categories completed in conjunction with the scoring method used have an influence on the number or percentage of perseverative responses scored.

Perhaps in response to the problems outlined, some authors implement their own scoring method. The implementation of different scoring techniques in some papers has led to discrepancies in the total number/percentage of perseverative responses scored. For example, according to the 1993 WCST manual, if the first response after the completion of a category (i.e. the first trial in which a new sorting rule is in effect) is an unambiguous error that matches the previous sorting rule, it is marked as a perseverative error (see Fig. 2, column (c); Heaton et al., 1993). However, in Channon (1996), these responses were not scored as perseverative errors ‘[because] the sorting principle changed without warning and they had no means of knowing this in advance’ (p. 109). In both examples, the sorting rule changes without warning at the end of a category. The two scoring techniques represent differences in the approach to, and understanding of, perseverative responses. Indeed, Channon’s (1996) method of scoring would afford up to six fewer perseverative responses than Heaton et al. (1993), making studies using these different methods incomparable. There are many cases of studies implementing their own scoring techniques for the WCST or failing to specify the

scoring method at all (e.g. Perpiñá, Segura, & Sanchez-Reales, 2017; Van Eylen et al., 2011). The inconsistencies in scoring methods present challenges when comparing results and generalising findings.

Using the WCST to assess cognitive flexibility

Although the WCST is widely accepted as an assessment of cognitive flexibility (Cragg & Chevalier, 2012; Dennis & Vander Wal, 2010; Figueroa & Youmans, 2013), neurocognitive tasks inevitably assess other executive functions (Buchsbaum, Greer, Chang, & Berman, 2005; Miyake, Emerson, & Friedman, 2000). Importantly, the WCST is a complex task that requires the use of multiple executive functions including attention, memory, and implicit learning (Buchsbaum et al., 2005; Cepeda, Kramer, & Gonzalez de Sather, 2001; Friederich & Herzog, 2011; Wu et al., 2014). Consequently, when the WCST is used to assess cognitive flexibility, the influence of other executive functions should be considered and accounted for where possible.

Typically, perseverative responses and/or perseverative errors are presented as indicators of cognitive flexibility (e.g. Baker, Georgiou-Karistianis, et al., 2018a; Baker, Gibson, et al., 2018b; Dickson et al., 2017; Garcia-Willingham et al., 2018; Gelonch et al., 2016; Wollenhaupt et al., 2019). However, some research has also used other variables (e.g. the number of categories completed, the number of trials taken to complete the first category, non-perseverative errors, and failure-to-maintain-set) as indicators of cognitive flexibility (e.g. Abbate-Daga et al., 2014; Aloï et al., 2015; Bischoff-Grethe et al., 2013; Dickson et al., 2017; Gelonch et al., 2016; Tchanturia et al., 2012; Wollenhaupt et al., 2019; Zmigrod et al., 2018). The extent to which some of these variables assess cognitive flexibility remains under debate; we recommend the reader refer to Figueroa and Youmans (2013) for discussion of the failure-to-maintain-set variable and its appropriateness as a measure of cognitive flexibility or distractibility. Further, despite the consensus that perseverative responses and/or perseverative errors are indicative of cognitive flexibility, there is no empirical evidence that conclusively verifies that these variables assess this construct. Rather, there is a general acceptance that a pattern of repetitive incorrect responding suggests rigidity and an inability to adapt to change. Comparing the WCST variables, particularly perseverative responses and/or perseverative errors, to other accepted neurocognitive tasks commonly used to assess cognitive flexibility (e.g. the Trail Making Test (TMT); Reitan, 1958) would be a first step in establishing the validity of these outcomes as markers of cognitive flexibility. Previous studies have identified a moderate to non-existent relationship between the cognitive flexibility outcomes of the WCST and the TMT (Chaytor, Schmitter-Edgecombe, & Burr, 2006;

Herbrich, Kappel, Winter, & van Noort, 2019; Kortte, Horner, & Windham, 2002; O'donnell, Macgregor, Dabrowski, Oestreicher, & Romero, 1994; Pignatti & Bernasconi, 2013; Van Autreve, De Baene, Baeken, van Heeringen, & Vervaeke, 2013). However, it is noteworthy that these studies have correlated diverse outcomes of these tests, and some of the reported variables may not be appropriate for assessing cognitive flexibility (i.e. TMT – Part B; Vall & Wade, 2015). Establishing consensus on which WCST and TMT variables to use when assessing cognitive flexibility will enable the field to compare findings across studies to elucidate the construct validity of such neurocognitive tests. In addition, leveraging advances in the study of cellular, circuit-level, and whole-brain imaging to uncover the neural correlates of cognitive flexibility is necessary to move the field forward (Lie, Specht, Marshall, & Fink, 2006; Specht, Lie, Shah, & Fink, 2009; Yuan & Raz, 2014).

Recommendations and conclusions

Establishing consensus on how to best define the WCST key terms and variables should be made a priority to promote the standardisation of assessments and comparability of results across studies. We recommended that perseverative responses and perseverative errors be defined and reported separately. Given that the Heaton et al. (1993) method of scoring is perhaps better able to assess perseverative responses in the first category of the WCST, we recommend that the Heaton et al. (1993) method be used to score the WCST. However, we acknowledge the complexity of this scoring system and recognise that training may be required before administrators are confident in scoring and interpreting any WCST data using the Heaton et al. (1993) method. Automated scoring that is facilitated by a computer program may be the best choice for a novice administrator. The use of a computerised version of the WCST (ideally the Heaton and PAR Staff (2008) program) reduces the opportunity for human error and misinterpretation of the scoring instructions described by Heaton et al. (1993). A non-commercial open-source version could be utilised in research and in clinical practice. In this instance, we recommend that the scoring code conform to the scoring methods outlined in the Heaton et al. (1993) manual. In instances where other computerised versions are implemented, manually inspecting the raw data to ensure that there are no atypical summary scores is an appropriate precaution to take.

To reduce the complexity of scoring the WCST, alternate versions of the test have been developed with all ambiguous cards removed (e.g. the Modified Card Sorting Test; Nelson, 1976). However, more research is needed to examine the extent to which the scores from the original WCST (e.g. perseverative errors and perseverative responses) relate to the scores

Table 2 A checklist of recommendations for using and reporting the WCST

Topic	Checklist item
Points of consideration before using the WCST as an assessment of cognitive flexibility	
Administration	Decide on which version of the WCST will be used, i.e. manual administration or computerised version.
Scoring	Decide on which scoring method will be used. Where possible, we recommend that that Heaton et al. (1993) administration and scoring be implemented. Any adaptation of the WCST should be appropriately justified and empirically supported.
Outcome variables	Establish primary outcome variables, i.e. perseverative responses (sum and/or percentage of total trials) and perseverative errors (sum and/or percentage of total trials).
Items to report in a manuscript	
Administration and scoring	Explicitly state which scoring method is used. Where possible, we recommend that that Heaton et al. (1993) administration and scoring be implemented. Any adaptation of the WCST should be appropriately justified and empirically supported.
Defined terms	Explicitly state whether the task is a manual administration or computerised version. <i>Dimension</i> : the characteristics of the cards by which they can be sorted (i.e. colour, form, number, or other). Avoid using the following terms: attribute, characteristic, criterion, rule, category, or principle. <i>Category</i> : refers to the discrete sections of the WCST. <i>Perseverative response</i> : persistent responses made by a participant on the basis of an incorrect (previous or novel) stimulus dimension. Perseverative responses can be correct or incorrect. <i>Perseverative error</i> : perseverative responses that are incorrect.
Variables for assessing cognitive flexibility	Report appropriate variables when assessing cognitive flexibility, i.e. perseverative responses (sum and/or percentage of total trials) and perseverative errors (sum and/or percentage of total trials). Avoid reporting WCST variables which may assess other executive functions as indicators of cognitive flexibility (e.g. failure-to-maintain-set, non-perseverative errors). These WCST variables can be reported if required (e.g. when assessing other domains of executive functioning), providing they are not described as indicators of cognitive flexibility.

of the modified versions. The reduction in the total number of cards presented due to the removal of ambiguous cards creates fewer opportunities for a participant to initiate cognitive flexibility. Relatedly, the maximum number of possible perseverative responses in modified versions of the WCST is lower than in the original WCST. Hence, raw scores on the modified WCST may not be directly comparable with raw scores on the original WCST, potentially leading to further confusion in the literature, concerns surrounding validity, and problems with study comparability. We recommend that future research investigate the validity of the different outcomes of the modified WCST to provide clarification on whether removing ambiguous cards from the WCST is an appropriate step to improve scoring simplicity.

In conclusion, the WCST is undoubtedly a popular neurocognitive task that has been widely used by both researchers and clinicians since its inception. Despite the wide implementation of the WCST within the field of neuropsychology, this task is not without its limitations. The inconsistent scoring of the WCST is a major source of confusion for users and creates challenges in interpreting and comparing

findings. There has also been a lack of consensus in the key WCST terminology and its corresponding definitions. Further, the outcome variables for assessing cognitive flexibility vary among studies. We recommend that users of the WCST follow the recommendations described above and presented in Table 2. Specifically, it is fundamental that authors who are considering using the WCST are transparent about the format of the task, cite the chosen scoring method, and report both perseverative responses and perseverative errors when using this task as an index to assess cognitive flexibility so as to better capture this latent construct.

Funding SM and CAH are supported by Research Training Program (RTP) scholarships (Australian Government, Department of Education, Skills and Employment). CB is supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship [ID 1127155]. GLM is supported by a NHMRC Leadership Investigator Grant [ID 1178444]. AP is supported by a NHMRC Project Grant [ID 1159953]. These funding bodies had no role in the analysis or interpretation of the data, writing of the manuscript, or the decision to submit the paper for publication.

Declarations

Conflicting Interests GLM has received support from: ConnectHealth UK, Seqirus, Kaiser Permanente, Workers' Compensation Boards in Australia, Europe and North America, AIA Australia, the International Olympic Committee, Port Adelaide Football Club, Arsenal Football Club. Professional and scientific bodies have reimbursed him for travel costs related to presentation of research on pain at scientific conferences/symposia. He has received speaker fees for lectures on pain and rehabilitation. He receives book royalties from NOIgroup publications, Dancing Giraffe Press & OPTP for books on pain and rehabilitation. CB has received support from Workers' Compensation Boards in Australia. All remaining authors (SM, CAH, MN, and AP) declare no potential conflicts of interest (personal or financial) with respect to the research, authorship, and/or publication of this manuscript.

References

- Abbate-Daga, G., Buzzichelli, S., Marzola, E., Amianto, F., & Fassino, S. (2014). Clinical investigation of set-shifting subtypes in anorexia nervosa. *Psychiatry Research*, 219(3), 592–597. <https://doi.org/10.1016/j.psychres.2014.06.024>
- Aloi, M., Rania, M., Caroleo, M., Bruni, A., Palmieri, A., Cauteruccio, M. A., ... Segura-García, C. (2015). Decision making, central coherence and set-shifting: A comparison between binge eating disorder, anorexia nervosa and healthy controls. *BMC Psychiatry*, 15(1), 1–10. <https://doi.org/10.1186/s12888-015-0395-z>
- Axelrod, B. N., Goldman, R. S., & Woodard, J. L. (1992). Interrater reliability in scoring the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, 6(2), 143–155.
- Baker, K. S., Georgiou-Karistianis, N., Lampit, A., Valenzuela, M., Gibson, S. J., & Giummarra, M. J. (2018a). Computerised training improves cognitive performance in chronic pain: A participant-blinded randomised active-controlled trial with remote supervision. *Pain*, 159(4), 644–655. <https://doi.org/10.1097/j.pain.0000000000001150>
- Baker, K. S., Gibson, S. J., Georgiou-Karistianis, N., & Giummarra, M. J. (2018b). Relationship between self-reported cognitive difficulties, objective neuropsychological test performance and psychological distress in chronic pain. *European Journal of Pain*, 22(3), 601–613. <https://doi.org/10.1002/ejp.1151>
- Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, 40(3), 349–356. [https://doi.org/10.1016/S0028-3932\(01\)00110-5](https://doi.org/10.1016/S0028-3932(01)00110-5)
- Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, 39, 15–22.
- Bilgin, M. (2009). Developing a cognitive flexibility scale: Validity and reliability studies. *Social Behavior and Personality: An International Journal*, 37(3), 343–353. <https://doi.org/10.2224/sbp.2009.37.3.343>
- Bischoff-Grethe, A., McCurdy, D., Grenesko-Stevens, E., Irvine, L. E., Wagner, A., Yau, W. Y. W., ... Kaye, W. H. (2013). Altered brain response to reward and punishment in adolescents with anorexia nervosa. *Psychiatry Research - Neuroimaging*, 214(3), 331–340. <https://doi.org/10.1016/j.psychres.2013.07.004>
- Buchsbaum, B. R., Greer, S., Chang, W.-L., & Berman, K. F. (2005). Meta-analysis of neuroimaging studies of the Wisconsin Card-Sorting Task and component processes. *Human Brain Mapping*, 25(1), 35–45. <https://doi.org/10.1002/hbm.20128>
- Cepeda, N. J., Kramer, A. F., & Gonzalez de Sather, J. C. M. (2001). Changes in executive control across the life span: Examination of task-switching performance. *Developmental Psychology*, 37(5), 715–730. <https://doi.org/10.1037/0012-1649.37.5.715>
- Channon, S. (1996). Executive dysfunction in depression: The Wisconsin Card Sorting Test. *Journal of Affective Disorders*, 39(2), 107–114. [https://doi.org/10.1016/0165-0327\(96\)00027-4](https://doi.org/10.1016/0165-0327(96)00027-4)
- Chaytor, N., Schmitter-Edgecombe, M., & Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Archives of Clinical Neuropsychology*, 21(3), 217–227. <https://doi.org/10.1016/j.acn.2005.12.002>
- Cragg, L., & Chevalier, N. (2012). The processes underlying flexibility in childhood. *Quarterly Journal of Experimental Psychology*, 65(2), 209–232. <https://doi.org/10.1080/17470210903204618>
- Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, 38(9), 571–578. <https://doi.org/10.1016/j.tins.2015.07.003>
- Deák, G. O. (2003). The development of cognitive flexibility and language abilities. *Advances in Child Development and Behavior*, 31, 271–327. [https://doi.org/10.1016/S0065-2407\(03\)31007-9](https://doi.org/10.1016/S0065-2407(03)31007-9)
- Dennis, J. P., & Vander Wal, J. S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research*, 34(3), 241–253. <https://doi.org/10.1007/s10608-009-9276-4>
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dickson, K. S., Ciesla, J. A., & Zelic, K. (2017). The role of executive functioning in adolescent rumination and depression. *Cognitive Therapy and Research*, 41(1), 62–72. <https://doi.org/10.1007/s10608-016-9802-0>
- Eling, P., Derckx, K., & Maes, R. (2008). On the historical and conceptual background of the Wisconsin Card Sorting Test. *Brain and Cognition*, 67(3), 247–253. <https://doi.org/10.1016/j.bandc.2008.01.006>
- Figueroa, I. J., & Youmans, R. J. (2013). Failure to maintain set: A measure of distractibility or cognitive flexibility? *Proceedings of the Human Factors and Ergonomics Society*, 57(1), 828–832. <https://doi.org/10.1177/1541931213571180>
- Flashman, L. A., Horner, M. D., & Freides, D. (1991). Note on scoring perseveration on the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, 5(2), 190–194.
- Friederich, H.-C., & Herzog, W. (2011). Cognitive-behavioural flexibility in anorexia nervosa. In R. A. H. Adan & W. H. Kaye (Eds.), *Behavioral neurobiology of eating disorders (Current topics in behavioral neurosciences)* (6th ed., pp. 111–123). Berlin: Springer-Verlag. https://doi.org/10.1007/978-3-642-01083-3_8
- Garcia-Willingham, N. E., Roach, A. R., Kasarskis, E. J., & Segerstrom, S. C. (2018). Self-regulation and executive functioning as related to survival in motor neuron disease: Preliminary findings. *Psychosomatic Medicine*, 80(7), 665–672. <https://doi.org/10.1097/PSY.0000000000000602>
- Geloch, O., Garolera, M., Valls, J., Rosselló, L., & Pifarré, J. (2016). Executive function in fibromyalgia: Comparing subjective and objective measures. *Comprehensive Psychiatry*, 66, 113–122. <https://doi.org/10.1016/j.comppsy.2016.01.002>
- Grant, D. A., & Berg, E. A. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, 38(4), 404–411.
- Greve, K. W. (1993). Can perseverative responses on the Wisconsin Card Sorting Test be scored accurately? *Archives of Clinical Neuropsychology*, 8(6), 511–517. [https://doi.org/10.1016/0887-6177\(93\)90051-2](https://doi.org/10.1016/0887-6177(93)90051-2)
- Greve, K. W. (2001). The WCST-64: A standardized short-form of the Wisconsin Card Sorting Test. *Clinical Neuropsychologist*, 15(2), 228–234. <https://doi.org/10.1076/clin.15.2.228.1901>

- Greve, K. W., Stickler, T. R., Love, J. M., Bianchini, K. J., & Stanford, M. S. (2005). Latent structure of the Wisconsin Card Sorting Test: A confirmatory factor analytic study. *Archives of Clinical Neuropsychology*, 20(3), 355–364. <https://doi.org/10.1016/j.acn.2004.09.004>
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1981). *Wisconsin Card Sorting Test: Manual* (1st ed.). Odessa: Psychological Assessment Resources.
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test manual: Revised and expanded*. Odessa: Psychological Assessment Resources.
- Heaton, R. K., & PAR Staff. (2008). *Wisconsin Card Sorting Test: Computer Version 4—Research Edition*. PAR.
- Herbrich, L. R., Kappel, V., Winter, S. M., & van Noort, B. M. (2019). Executive functioning in adolescent anorexia nervosa: Neuropsychology versus self- and parental-report. *Child Neuropsychology*, 25(6), 816–835. <https://doi.org/10.1080/09297049.2018.1536200>
- Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). Reliability and validity of two self-report measures of cognitive flexibility. *Psychological Assessment*, 26(4), 1381–1387. <https://doi.org/10.1037/a0038009>
- Kortte, K. B., Horner, M. D., & Windham, W. K. (2002). The Trail Making Test, Part B: Cognitive flexibility or ability to maintain set? *Applied Neuropsychology*, 9(2), 106–109. https://doi.org/10.1207/S15324826AN0902_5
- Lie, C.-H., Specht, K., Marshall, J. C., & Fink, G. R. (2006). Using fMRI to decompose the neural processes underlying the Wisconsin Card Sorting Test. *NeuroImage*, 30(3), 1038–1049. <https://doi.org/10.1016/j.neuroimage.2005.10.031>
- McCallum, R. S. (2017). *Handbook of nonverbal assessment*. <https://doi.org/10.1007/978-3-319-50604-3>
- Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of Neurology*, 9(1), 90–100. <https://doi.org/10.29085/9781856048767.060>
- Miyake, A., Emerson, M. J., & Friedman, N. P. (2000). Assessment of executive functions in clinical settings: Problems and recommendations. *Seminars in Speech and Language*, 21(2), 169–183.
- Nelson, H. E. (1976). A Modified Card Sorting Test sensitive to frontal lobe defects. *Cortex*, 12(4), 313–324. [https://doi.org/10.1016/S0010-9452\(76\)80035-4](https://doi.org/10.1016/S0010-9452(76)80035-4)
- O'Donnell, J. P., Macgregor, L. A., Dabrowski, J. J., Oestreich, J. M., & Romero, J. J. (1994). Construct validity of neuropsychological tests of conceptual and attentional abilities. *Journal of Clinical Psychology*, 50(4), 596–600. [https://doi.org/10.1002/1097-4679\(199407\)50:4<596::AID-JCLP2270500416>3.0.CO;2-S](https://doi.org/10.1002/1097-4679(199407)50:4<596::AID-JCLP2270500416>3.0.CO;2-S)
- Øverås, M., Kapstad, H., Brunborg, C., Landrø, N. I., & Lask, B. (2015). Are poor set-shifting abilities associated with a higher frequency of body checking in anorexia nervosa? *Journal of Eating Disorders*, 3(1), 1–8. <https://doi.org/10.1186/s40337-015-0053-3>
- Perpiñá, C., Segura, M., & Sanchez-Reales, S. (2017). Cognitive flexibility and decision-making in eating disorders and obesity. *Eating and Weight Disorders*, 22(3), 435–444. <https://doi.org/10.1007/s40519-016-0331-3>
- Pignatti, R., & Bernasconi, V. (2013). Personality, clinical features, and test instructions can affect executive functions in eating disorders. *Eating Behaviors*, 14(2), 233–236. <https://doi.org/10.1016/j.eatbeh.2012.12.003>
- Rabin, L. A., Barr, W. B., & Burton, L. A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, 20(1), 33–65. <https://doi.org/10.1016/j.acn.2004.02.005>
- Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills*, 8, 271–276.
- Scott, W. A. (1962). Cognitive complexity and cognitive flexibility. *Sociometry*, 25(4), 405–414. <https://doi.org/10.2307/2785779>
- Specht, K., Lie, C.-H., Shah, N. J., & Fink, G. R. (2009). Disentangling the prefrontal network for rule selection by means of a non-verbal variant of the Wisconsin Card Sorting Test. *Human Brain Mapping*, 30(5), 1734–1743. <https://doi.org/10.1002/hbm.20637>
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). Oxford University Press.
- Tchanturia, K., Davies, H., Roberts, M., Harrison, A., Nakazato, M., Schmidt, U., ... Morris, R. (2012). Poor cognitive flexibility in eating disorders: Examining the evidence using the Wisconsin Card Sorting Task. *PLoS ONE*, 7(1). <https://doi.org/10.1371/journal.pone.0028331>
- Vall, E., & Wade, T. D. (2015). Trail Making Task performance in inpatients with anorexia nervosa and bulimia nervosa. *European Eating Disorders Review*, 23(4), 304–311. <https://doi.org/10.1002/erv.2364>
- Van Autreve, S., De Baene, W., Baeken, C., van Heeringen, C., & Vervaeke, M. (2013). Do restrictive and bingeing/purging subtypes of anorexia nervosa differ on central coherence and set shifting? *European Eating Disorders Review*, 21(4), 308–314. <https://doi.org/10.1002/erv.2233>
- Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: Explaining the inconsistencies? *Research in Autism Spectrum Disorders*, 5(4), 1390–1401. <https://doi.org/10.1016/j.rasd.2011.01.025>
- Wollenhaupt, C., Wilke, L., Erim, Y., Rauh, M., Steins-Loeber, S., & Paslakis, G. (2019). The association of leptin secretion with cognitive performance in patients with eating disorders. *Psychiatry Research*, 276, 269–277. <https://doi.org/10.1016/j.psychres.2019.05.001>
- Wu, M., Brockmeyer, T., Hartmann, M., Skunde, M., Herzog, W., & Friederich, H.-C. (2014). Set-shifting ability across the spectrum of eating disorders and in overweight and obesity: A systematic review and meta-analysis. *Psychological Medicine*, 44(16), 3365–3385. <https://doi.org/10.1017/S0033291714000294>
- Yuan, P., & Raz, N. (2014). Prefrontal cortex and executive functions in healthy adults: A meta-analysis of structural neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 42, 180–192. <https://doi.org/10.1016/j.neubiorev.2014.02.005>
- Zmigrod, L., Rentfrow, P. J., & Robbins, T. W. (2018). Cognitive underpinnings of nationalistic ideology in the context of Brexit. *Proceedings of the National Academy of Sciences of the United States of America*, 115(19), E4532–E4540. <https://doi.org/10.1073/pnas.1708960115>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.