Assessment of executive functions in adults: A systematic review and empirical investigation

Gloria Pizzamiglio, BSc (Hons)

Submitted in partial fulfilment of the requirements for the degree of

Doctorate in Clinical Psychology

School of Health and Wellbeing

College of Medical, Veterinary and Life Sciences

University of Glasgow

September 2023

Table of Contents

List of tables

List of Figures

Acknowledgments

Firstly, I would like to express my heartfelt gratitude to all the clinicians who took the time to participate in our study amidst the demands of their busy work schedule. Your contributions have been invaluable, and I am delighted to finally be able to share all that I have learned from you over the past few months.

Huge thanks to my supervisor, Dr Jessica Fish, for providing tremendous amounts of support and guidance over the last three years. You have been so incredibly flexible and generous with your time, and always full of kind and encouraging words which kept me going in the toughest of times. I would also like to acknowledge and give my warmest thanks to my placement and field supervisor, Dr Luke Williams. Your support, kindness, and understanding of all the many ups and downs of this journey have been key to the completion of this thesis (and semi-retention of my sanity!). I am extremely grateful to Professor Jon Evans for all his precious advice during the different stages of my doctoral training, including but not limited to this research project. Special thanks also go to the clinicians and colleagues who helped with the survey piloting for their feedback and insightful comments, and to Greg for all the prompt input as second reviewer.

I thank my family and friends whose support and encouragement have been a constant source of strength throughout this journey. I am beyond excited to be able to spend more time with you all. To Owen, I cannot thank you enough for all that you have done for me over the last year and especially in these last few months. Your love and understanding have made all the difference, as have all our Gelatessa breaks and caipirinha Fridays. I truly could not have done this without you by my side.

Lastly, to Glyn. As I approach the end of my clinical training, I can’t help but think that life sometimes has a ruthless way of disrupting plans and leading people to the destination that, perhaps, they were meant to reach all along. You have had such a profound influence on the path that I have taken and on the ideal of the scientist-practitioner I can only strive to become. Your infectious curiosity, nurturing nature, and quest for making our clinical and research work truly relevant to the people that we serve will undoubtedly continue to guide and shape the rest of my career. I could not thank you enough.

Chapter 1

A systematic review of the association between BADS scores and DEX ratings in adults with an acquired brain injury

Prepared in accordance with the author requirements for *The Clinical Neuropsychologist* journal.

https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=ntcn20#word-limits

Abstract

Objectives: This systematic review examined the relationship between BADS scores and DEX ratings, as provided by patients (DEX-Self), significant others (DEX-SO) and health care professionals (DEX-HCP), in individuals with acquired brain injury (ABI). It also aimed to evaluate which BADS scores show the strongest associations with DEX ratings.

Method: A systematic search was conducted on 5th September 2023 on CINAHL, PsycINFO, MEDLINE, Embase and Web of Science Core Collection databases for peer-reviewed studies written in English with adults (≥18 years) with non-progressive, progressive, or a mix of non-progressive and progressive ABI (npABI, pABI, mABI, respectively), which tested the association between BADS and DEX scores. Grey literature and studies using composite DEX scores were excluded. Risk of bias was assessed using the AXIS tool. A narrative synthesis was performed.

Results: Twelve of the 15 studies included found statistically significant correlations between at least one BADS score and DEX ratings. DEX-SO and DEX-HCP ratings more reliably predicted BADS performance than DEX-Self ratings. The BADS score that consistently showed medium to strong associations with DEX ratings across all three ABI groups was the Total Profile Score, although the number of studies which found this association was small and the quality of the evidence mixed.

Conclusions: Firm conclusions about the association between BADS and DEX scores are difficult to draw due to the limitations in the current evidence; however, some support was found for the ecological validity of the BADS battery as a whole to assess EF in adults with ABI. Future studies investigating this relationship will benefit from larger and more homogeneous samples.

Introduction

Acquired brain injury (ABI) refers to any brain damage that occurs after birth which may affect brain structure and function (Elbaum, 2019). ABI encompasses both progressive (e.g., neurodegenerative diseases) and non-progressive (e.g., tumour, infection, physical injury, etc.) conditions, both of which can have profound effects on an individual's physical, cognitive, and emotional functioning (Coetzer, Daisley, Newby, & Weatherhead, 2013). Particularly, deficits in executive functions (EF), a broad set of skills necessary for goal-directed behaviours and social functioning, are a common clinical feature in individuals with ABI (Mueller & Dollaghan, 2013). However, accurate assessment of EF is challenged by the limited ability of traditional EF tests to predict functional competence in day-to-day tasks (Burgess, Alderman, Evans, Emslie, & Wilson, 1998).

To address these limitations in ecological validity (i.e., how well test performance can predict performance in real-world settings), EF tasks more analogous to those that may occur in real-life situations have been developed, including the Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, Alderman, Burgess, Emslie, & Evans, 1996). The BADS comprises six subtests that evaluate different components of EF, including planning, practical problem-solving, multitasking and inhibition. Each subtest yields a profile score (PS; range=0-4), with the battery total profile score (TPS) ranging from 0 to 24; higher scores on the BADS indicate better performance. The Dysexecutive Questionnaire (DEX) also forms part of the BADS battery and comprises 20 questions designed to assess various domains of EF, including cognition, behaviour, motivation, and personality/emotions. Respondents rate the frequency of these behaviours on a five-point Likert scale ranging from “never” (0) to “very often” (4), with higher scores indicating greater EF impairment. The DEX comes in two versions: one for self-ratings and one for others’ ratings (ideally completed by a relative or carer who knows the patient well). As part of the BADS validation study in a sample of individuals with ABI, no significant correlations were found between any of the BADS subtests and the DEX self-ratings (DEX-Self), a finding that was explained as a clinical marker of reduced deficits awareness in patients with executive dysfunction (Wilson et al., 1996). Therefore, the authors used the DEX significant others’ (DEX-SO) ratings to assess the ecological validity of the BADS. Significant negative correlations were found between DEX-SO scores, all six BADS subtests and the BADS TPS, which the authors presented as evidence of the battery’s ecological validity.

Since this initial validation work, several other studies have been conducted to assess the relationship between BADS and DEX scores in different clinical and non-clinical populations; however, findings in individuals with ABI have been mixed and have generally failed to

replicate the results of the original study to the same extent (Wood & Liossi, 2006). The use of the DEX as a valid measure of EF against which the ecological validity of the BADS can be assessed was initially questioned (Bennett, Ong, & Ponsford, 2005); however, the DEX has since been confirmed to have adequate concurrent validity when correlated with other EF questionnaires, making it a valid measure for assessing EF in individuals with brain injuries (Boelen, Spikman, Rietveld, & Fasotti, 2009). Some authors have argued that the ecological validity of neuropsychological tests may vary depending on the type of population assessed and the level of injury or disease severity (Chaytor & Schmitter-Edgecombe, 2003), which may explain the inconsistent findings given the heterogeneity of samples included. Others discussed potential factors which may account for some of the unexplained variance in DEX-SO ratings, including the extent to which the patient's deficits negatively impact the rater, the rater's stage of adjustment to, and acceptance of, their loved one's injury and deficits (Renison, Ponsford, Testa, Richardson, & Brownfield, 2012). To address these issues, it has been suggested that health care professionals (HCPs), such as neuropsychologists and occupational therapists, may provide more reliable ratings than significant others (Bennett et al., 2005); however, rating accuracy may still vary widely depending on the setting HCPs work in (e.g., community vs inpatient rehabilitation), how much contact they have had with the patient, and lack of first-hand knowledge of the patient's pre-injury functioning.

Overall, the extent of the ecological validity of the BADS remains unclear, as does the question of who may provide more reliable ratings on the DEX. Despite the mixed empirical evidence available, the BADS remains widely used in different clinical populations as an ecologically valid measure of EF (Boyle et al., 2023).

Rationale

Given the variation in study findings in ABI individuals and the heterogeneity in methodologies and samples included, it would be of value to systematically synthesise current evidence and assess which conclusions can (or cannot) be drawn based on existent studies and their quality. We anticipate that the findings of this review will be of relevance to both clinical practitioners and researchers, who may welcome up-to-date, evidence-based information on the ecological validity of the BADS, as measured by the DEX, as an EF measure in ABI populations.

Objectives

This systematic review aimed to examine the statistical association between BADS scores and DEX ratings in individuals with ABI by answering the following questions:

1) What is the extent of association between BADS scores and DEX ratings provided by patients (DEX-Self), significant others (DEX-SO) or health care professionals (DEX-HCP) in adults with ABI?
2) Which, if any, BADS score(s) show the strongest significant associations with DEX ratings in adults with ABI?

As the association between BADS and DEX may vary depending on ABI subtypes, the above questions will be explored grouping ABI patients by non-progressive (npABI), progressive (pABI), or mixed groups with progressive or non-progressive conditions (mABI), providing enough studies (at least two per group) are available.

Methods

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) updated guidelines (Page et al., 2021; Appendix 1.1). A protocol for this review was registered prospectively on the International Prospective Register of Systematic Reviews (PROSPERO) database (registration number: CRD42023459706).

Eligibility criteria

Inclusion criteria

- Adult participants (≥18 years) with a diagnosis of npABI or pABI of any severity or stage. Studies on adult samples which did not specify the participants' age-range were included;
- Studies which reported information on the statistical association (correlation and/or regression analyses) between:
    - at least one of the six BADS subtests and/or the BADS Total Profile Score (TPS)
    - at least one of these DEX ratings: DEX-Self, DEX-SO, and/or DEX-HCP;
- Any quantitative observational design, including mixed-methods, providing correlational and/or regression analyses were included;
- Studies published in a peer-reviewed journal;
- Studies written in English.

Exclusion criteria

- Studies that only correlated the BADS scores with a "DEX difference" rating (obtained by subtracting patient ratings from informant ratings, which can be used to quantify patients' unawareness);
- Secondary data (i.e., systematic reviews or meta-analyses) and book chapters;
- Grey literature.

Information sources

The following databases were independently searched from 1st January 1996 to 5th September 2023: Cumulative Index to Nursing and Allied Health Literature (CINAHL; EBSCOhost), PsycINFO (EBSCOhost), MEDLINE (Ovid), Embase (Ovid), and Web of Science Core Collection. Backward and forward citation searching of included studies was conducted using these databases and Google Scholar. Articles not retrieved using these databases were requested via the University of Glasgow Library Services (*n* = 5).

Search strategy

The search strategy (included in Appendix 1.2) was developed using an iterative process with guidance from experienced University librarians. Potentially limiting terms related to the population (i.e., adults, types of ABI) were omitted to maximise sensitivity. Sensitivity was tested by conducting scoping reviews of key papers, and search terms were modified accordingly. Searches were restricted to papers published in English from 1996, the year the BADS was first published.

Selection process

The first author was primarily responsible for the selection process. All search results were first exported and de-duplicated within EndNote X9 (Clarivate Analytics, Philadelphia), and were subsequently uploaded to the Rayyan web-tool (https://www.rayyan.ai/). Potential duplicates identified by Rayyan were checked and removed. The remaining records were screened by title and abstract using a screening tool developed for this review (Appendix 1.3). Three authors were contacted via e-mail and one via ResearchGate to obtain further information to determine eligibility. Two responded confirming ineligibility (one was an abstract for a poster presentation, while the other paper only included correlations using

combined data from patients and controls). No responses were obtained by either the first or the senior author of one paper. A second reviewer (DP) independently screened 10% of the records at the stage of title and abstract screening and full-text screening. There was 'substantial' agreement (96.47% agreement; Cohen's k=.62) between the raters at the title and abstract screening stage, with discrepancies due to the lead reviewer's tendency to err on the side of over-inclusion when in doubt. 'Perfect' agreement was found at the full-text screening stage (100%; k=1).

Data collection process and data items

Two data extraction forms were developed for this review. One form included data on studies and participants' characteristics (study design, country, patients' ABI type, sample size, mean injury/disease duration and severity, age, gender, education, estimated IQ, DEX-SO and DEX-HCP information). The second form included statistical data on the association between BADS scores and DEX ratings (BADS score type, DEX mean scores, correlation and regression type, correlation/regression coefficients, p-values and/or significance levels, confidence intervals and effect sizes, where available). Nine authors were contacted to obtain information not available in their articles for completeness of reporting (e.g., clarification on analyses and statistical results only described qualitatively) within a two-week timeframe. one provided all the requested information (Koerts et al., 2012); four were unable to provide the data (Chan & Manly, 2002; Evans et al., 1997; Grech et al., 2016; Knight et al., 2002); two agreed to look at the data but were not able to provide the information within the timeframe required (Burgess et al., 1998; Norris & Tate, 2000); in two cases, attempts to contact authors yielded no response (Channon & Crawford, 1999; Boelen et al., 2009).

Study risk of bias assessment

The Appraisal tool for Cross-Sectional Studies (AXIS) tool was used to assess risk of bias in the included papers (Downes, Brennan, Williams, & Dean, 2016). This 20-item validated tool addresses three key areas of evaluation: quality of reporting (seven items), quality of study design (seven items), and possible introduction of biases in the study (six items). Each item is rated using a "Yes" (1), "No" (0), or "Do not Know" (DK; 0) except for items 13 and 19 which are reverse scored (i.e., "Yes" = 0; "No" = 1).

The AXIS tool does not include a score cut-off for study quality; rather, it allows a degree of flexibility for raters to evaluate the study's characteristics cumulatively. In the present review this tool was used to report strengths and limitations in each key domain to help assess the

robustness of the evidence presented. Of note, for the methods, results and discussion sections, ratings focused on the information relevant to this review only, rather than on all the analyses and data presented in the studies.

A second reviewer (GH) independently rated 60% (*n* = 9) of the papers included in this review, and initial agreement was “fair” (68.33%). On inspection of the ratings, it was noted that some of the variance was due to the interchangeable use of “No” and “DK” for missing information; when this was taken into account, agreement improved to “moderate” levels (78.33%). Any remaining discrepancies were resolved through discussions between the two reviewers and the lead reviewer’s supervisor, and it was agreed that the lead reviewer’s ratings would be used in this review.

Synthesis methods

A narrative synthesis of the review findings was performed following the three stages recommended by Petticrew and Roberts (2006):

1. Organisation of studies into logical categories depending on the review questions;
2. Within study analysis, which should include a description of the findings and quality ratings;
3. Cross-study synthesis, which should provide an overall summary of any common patterns and inconsistencies in findings across papers, accounting for differences in study quality and other study characteristics which may compromise generalisability.

Systematic tabulation of individual studies was used to display results and relevant characteristics of each study. Studies were grouped within tables based on ABI type, as follows: 1) npABI, 2) pABI, and 3) mABI. To overcome challenges associated with comparison of correlations across heterogenous groups, this review referred to guidelines from Cohen (1988) which categorise r coefficients values between -.10 to .29 as small or weak, between .30 to .49 as medium or moderate, and between .50 to 1.0 as large or strong (irrespective of the positive or negative direction of correlations).

The assessment of reporting bias and confidence in the body of evidence presented in this review was guided by the AXIS tool, as previously described. No studies were excluded based on high risk of bias or missing statistical data. However, these limitations were acknowledged in the synthesis.

Results

Study selection

The database searches yielded 3,355 results. Following de-duplication, titles and abstracts of 1,977 articles were screened; of these, 191 articles were sought for full-text retrieval. A further 13 articles were examined in full after being identified through citation searching. The BADS validation study, which was published in the BADS test manual, was also accessed. A total of 15 studies were included in the review (see Figure 1.1 for the PRISMA flow chart). One study reviewed in full appeared to meet the inclusion criteria; however, it did not specify if DEX-Self or DEX-SO ratings were used for the correlational analyses with BADS scores and attempts to contact the authors yielded no response. The study was therefore excluded from this review.

Study characteristics

Table 1.1 summarises the studies and participants' key characteristics, ordered by type of ABI. All studies tested the association between at least one BADS score and DEX ratings (cross-sectional and correlational design). Six studies involved patients with npABI, five included patients with pABI, and four comprised a mixed group of patients with npABI and patients with pABI (mABI). Ten studies were conducted in Europe (six in the UK, two in the Netherlands, one in Greece and one in Spain); one study was conducted across the UK and the USA, two in Australia, one in Hong Kong and one in Japan. The total combined sample across all studies was 862 (npABI = 289; pABI = 351; mABI = 222). However, in some studies only a sub-sample of participants were included in the correlational analyses due to lack of DEX ratings for some participants (see Table 1.2). A total of 788 participants completed the DEX-Self (npABI = 289; pABI = 299; mABI = 200); 599 of the participants' significant others completed the DEX-SO (npABI = 143; pABI = 286; mABI = 170); 153 patients in the npABI group had health care professionals complete the DEX-HCP.

As in some studies demographic and ABI characteristics were only available for the full sample (rather than the sub-sample included in correlational analyses; see Tables 1.1. and 1.2 for details), data ranges and means presented in the following narrative synthesis are only from studies that provided this information for the exact sample included in the analyses, across all three groups. The youngest and oldest participants were 18 (Emmanouel, Mouza, Kessels, & Fasotti, 2014) and 87 (Fukuta & Mori, 2018), respectively. The male ratio was higher in all the npABI and mABI studies except for one (Emmanouel et

al., 2014). There was a higher ratio of females in studies in the pABI group aside from one study, which included PD patients (Koerts et al., 2012). Mean disease or injury duration ranged from 3.5 months (Chan & Manly, 2002) to 14.8 years (Grech et al., 2016). Disease or injury severity was variable within and across the samples, and ranged from mild (Chan & Manly, 2002) to very severe (Knight, Alderman, & Burgess, 2002).

Risk of bias in studies

Individual study ratings are reported in Appendix 1.4. No study met all the quality criteria; two studies met >75% of the criteria, nine studies met 50-75% and four met 40-50%. See "Reporting biases and certainty of evidence" section in this review for a more detailed summary of the quality appraisal results.

Results of individual studies

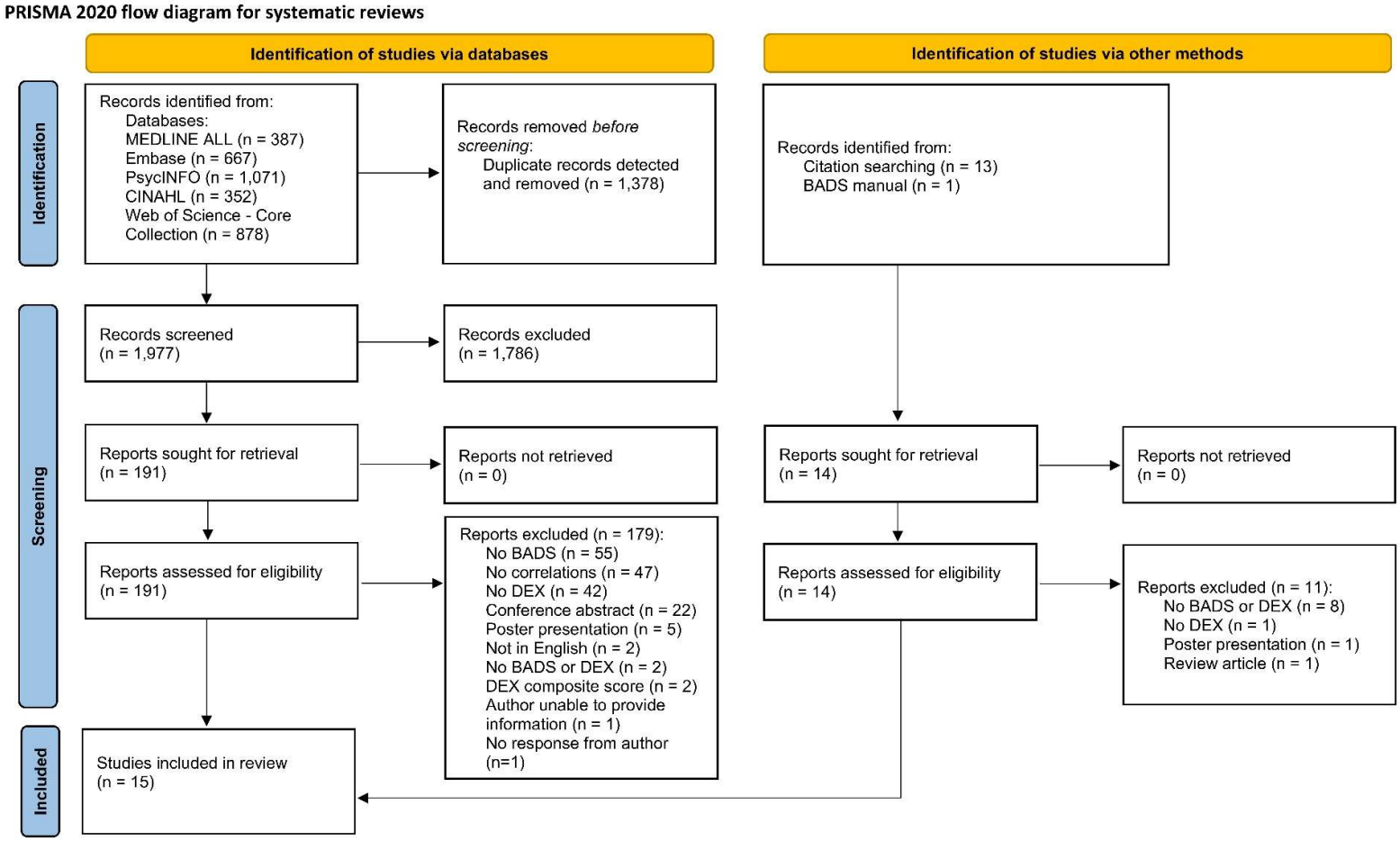
Summary statistics for each study can be found in Table 1.2, arranged by ABI type. Mean and standard deviations (SD) for DEX ratings were not reported in three of the studies. No study reported standard errors or confidence intervals for correlational analyses.

Results of cross-study syntheses

Characteristics of contributing studies

Twelve of the 15 studies reviewed reported statistically significant correlations between at least one BADS score and one of the three DEX ratings. ABI aetiology varied within and between samples. Patients with traumatic brain injuries (TBI) were included in all ten studies in the npABI and mABI groups. The mABI group comprised studies on patients with multiple sclerosis (MS; *n* = 3), Parkinson's disease (PD; *n* = 1) and mild neurocognitive disorder (NCD; *n* = 1) caused by either Alzheimer's disease (AD) or Lewy body dementia (LBD). Two studies in the mABI group included largely overlapping samples (Burgess et al., 1998; Wilson et al., 1996). Aside from three studies in the pABI group, all patients rated themselves, on average, lower on the DEX than did their significant others or their HCPs. Sample size varied widely, and caution should be taken to interpret results of the three studies employing a small sample (<30 patients; Channon & Crawford, 1999; Knight et al., 2002; Norris & Tate, 2000) as statistically significant findings may be falsely positive and/or moderate correlations may not reach statistical significance in underpowered studies. Correlational studies are generally reported to require a minimum of 30 subjects, as smaller samples may give inaccurate estimates of the degree of relationship between two variables (Fraenkel, Wallen, & Hyun, 2012).

Figure 1.1 *PRISMA Flow Chart*



From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

Table 1.1

Study and sample characteristics.

Authors, year, country, design	ABI type (n)	Mean Injury / disease duration (SD) and severity	Age (years) Mean (SD), Range	Gender (M, F)	Education (years) Mean (SD)	Estimated IQ Mean (SD)	DEX-SO information	DEX-HCP information
Group 1: Non-progressive ABI								
Boelen et al. (2009) The Netherlands Cross-sectional	ABI (total sample=81)* TBI (34), LH stroke (18) RH stroke (13), subcortical strokes (4), different aetiologies (tumour, encephalitis, hypoxia; 12).	At least six months (median=20 months) post-injury. Median PTA duration (available for n=21): 16 days.	42.83 (13.75)	55, 26	7-point Dutch classification system (median=5.0)	sGIT IQ: 112.55 (17.32)	Someone who knew the patient well (usually the partner).	A professional (usually a neuropsychologist), who had interviewed the patient and their proxy, but who was not further involved in this study. The DEX-HCP was in some instances filled in after a short acquaintance with the patients.
Chan & Manly (2002) Hong Kong Cross-sectional	TBI (30) Damage in the frontal lobes (33%), temporal lobes (10%) and	Time since injury (months): M=3.45 (SD=38.16), median=15.	38.07 (6.78) Age-range for inclusion: 20-55.	23, 7	M= 10.96 (SD = 2.78)	Used educational level as an approximate IQ estimate.	Somebody who knew the patient well and who, ideally, lived in the same home.	

	multiple regions (6.7%). Half of the group had no detectable abnormalities on CT scan.	Mild to moderate TBI: Median GCS: 15, range= 8-15. Median LOC duration and PTA (days): 0, range=0–1 day.						
Emmanouel et al. (2014) Greece Cross-sectional	ABI total sample (52) AL group (30) with frontal/frontal-temporal damage (cortical or sub-cortical): Trauma (predominantly prefrontal damage): LH (6), RH (7). Stroke: haemorrhagic - basal ganglia (LH=3; RH=2), RH ischemic (3). Lobectomy (tumour	Time since injury (months): M=11.5 (S=8.508), range=6-46 months. Coma duration for TBI and haemorrhagic stroke patients (days): M=15.83 (SD=4.99), range=7-24 All patients were in-patients with severe motor and sensory impairments at the time of onset (TBI, stroke) or after	AL: 41.47 (15.40) PL: 48 (10.73) Range=18-61	22, 28 (no gender data reported for 2 participants)	AL: 12.97 (2.07) PL: 12.5 (2.11)	Raven's Standard Progressive Matrices (IQ): AL: 103.53 (6.458) PL: 105.86 (5.222)		Professional therapists (mostly physiotherapists and speech-therapists) who were familiar with the patients, having worked with them for several months.

	surgery): LH (3), RH (6) – PL group (22) with parietal/parietal-occipital damage: Trauma: LH (1), RH (3). Stroke: LH ischaemic (1), RH haemorrhagic (2), RH ischaemic (7). Lobectomy (tumour surgery): LH (5), RH (3).	surgery (tumours).						
Evans et al. (1997) UK Cross-sectional	ABI (35) Lesion location: damage was more likely to be anterior-frontal/ temporal, than more posterior. Aetiology: head injury, stroke, viral	Injury severity: moderate to severe range.	40.4 (15.49)			NART IQ: 105.0 (12.58)	A relative or carer.	

	encephalitis, and tumour.							
Knight et al. (2002) UK Cross-sectional	ABI (20) TBI (12), cerebrovascular accident (5) or a combination of the two (3).	Time since injury (months): M=80.9 (SD=62.6). Time since admission to hospital (months) M=21.2 (SD=18.6). ABI severity was determined (from case notes) to be either severe or very severe for all participants.	35.6 (11.3) Range=20-53	17, 3		NART-R FSIQ: 100.8 (10.7) WAIS-R FSIQ: 84.4 (11.1)		Rehabilitation staff members who knew the participant well.
Weddell & Wood (2016) UK Cross-sectional	TBI (71) Breakdown: NHS patients (40), ML patients (31) Aetiology: (% NHS / % ML): vehicle accident (45%	Injury duration (months): NHS: M=64.5, range=5-215 ML: M=49.3, range=10-210 Severe TBI (n=63),	NHS: 39.8, range=19-58 ML: 33.8, range=19-63	52, 19 Breakdown: NHS: 30, 10 ML: 22, 9		No estimated IQ but WAIS-III scores available: Mean Verbal Comprehension (SD):		

	/ 87.1%), fall (30% / 9.6%), assault (17.5% / 3.2%), other (7.5% / 0.0%).	moderately severe TBI (n=8) Mean GCS (range): NHS: 9.0 (3-15) ML: 8.7 (3-15) Mean PTA (days): NHS: 26.8, range=0.02-108 ML: 28.5, range=0.7-108				NHS: 93.5 (12.2) ML: 87.8 (14.6) Mean Perceptual Organization (SD): NHS: 104.5 (17.3) ML: 96.0 (16.3)		
Group 2: Progressive ABI								
Cerezo García et al. (2015) Spain Cross-sectional	MS (total sample=100)* Type: RRMS (74); Progressive forms (26), including progressive relapsing, SPMS and primary progressive	Time of evolution (months): M=119.29 (SD=78.29) EDSS: M=2.7 (SD=1.93)	40.57 (10.55) Range=22-66	31, 69	Education level (%): elementary (33%), middle (27%), high (46.7%)		A family member or care giver.	

Fukuta & Mori (2018) Japan Cross-sectional	Mild NCD (32) Cause of NCD: AD (19), LBD (13). Absence of vascular lesions was one of the inclusion criteria	MMSE score: 25.18 (3.03), range=21-30	74.96 (7.57) Range=60-87	7, 25			Caregivers.	
Grech et al. (2016) Australia Cross-sectional	MS (107) Type: RRMS (83) SPMS (24)	Duration since diagnosis (years): M=9.82 (SD=7.46), range=0.3-31.2 Duration since symptom onset (years): M=14.77 (SD=9.23), range=1.0-44.5 EDSS (available for n=70): M=2.90 (SD=2.31), range=0.0-8.0.	48.8 (11.1) Range=26.2-74.5	24, 83	Highest education level (n): secondary (30), TAFE (college) education (21), undergraduate (34), postgraduate (22).		Significant other nominated by the patients.	

Koerts et al. (2012) The Netherlands Cross-sectional	Idiopathic PD (43)	Duration (years): M=5.1, (SD=4.1) UPDRS part III (motor severity): M=24.6, (SD=8.8) H&Y: M=2.2, (SD=0.6)	63.7 (8.6) Range=47-77	24, 19	7-point Dutch classification system: M= 5.2 (SD=1.1)		Partners, (58%), children (19%), close friends or family members (12%), relationship unknown (11%). All were caregivers of patients.	
Preston et al. (2013) UK Cross-sectional	MS (69) Type: RRMS (37), SPMS (32)	Disease duration (years): RRMS: M=8.4 (SD=7) SPMS: M=14.3 (SD=10) Median EDSS scores (range): RRMS: 3.5 (1.0-6.0) SPMS: 6.0 (4.0-8.0)	49 (9) All patients were ≥18	19, 50	13 (3)		Someone nominated by patients who knew them well. 57% chose not to disclose their relationship to the participant. Breakdown of remaining respondents: spouses/partners (57%), parents (12%), other family members (14%), friends/colleagues (17%).	

							N.B. the above data refers to both healthy controls and patients, as the pattern was reported to be consistent across both groups.	
Group 3: Non-progressive and progressive ABI								
Burgess et al. (1998) UK Cross-sectional	ABI (total sample=92)* Head injuries (59%), dementia (primarily AD or frontal lobe dementia; 13%), cerebrovascular accidents (8.5%), encephalitis (6.5%), a range of other conditions (e.g., anoxia and carbon monoxide poisoning,	Patients with mild head injury were excluded.	38.5* (15.1)			NART reading IQ: 103.2 (13.0) WAIS-R FSIQ: 92.1 (15.5)	Someone who knew the patient well (usually either a relative or carer).	

	gunshot wound, Korsakoff syndrome; 13%).							
Channon & Crawford (1999) UK Cross-sectional	ABI AL (16) AL group: unilateral LH (5) or RH (11) lesions involving frontal lobes (anterior group) Aetiology information was provided for the whole group of patients (n=25): vascular damage (15), head injury (5), abscess (2), tumours (2) and sclerosis (1).	Minimum time post-injury for inclusion: at least three months	41.63 (12.65) Age-range for inclusion: 18-70	11, 5	13.00 (2.07)	NART IQ: 108.50 (10.15) TROG score: 78.19 (1.76)	A relative or friend who knew the patients well; professional staff (one occupational therapist, one care worker) gave ratings for two lesion participants in the absence of family or friends.	
Norris & Tate (2000) Australia Cross-sectional	ABI (total sample=36)* TBI (19), MS (17)	PTA duration available for n=15 (days): M=60.47, (SD=53.92), range=10	39.36 (10.75)	21, 15		NART (FSIQ): 101.71, (10.56)	A close relative. The authors noted that this group was extremely heterogeneou	

		days-6 months. DS scale (degree of motor disability): M=4.06, (SD=1.89), range 1–6.					s. “It was difficult to be confident that all relatives were completing the DEX with similar degrees of awareness, understanding, and appreciation of the participant’s everyday functioning.”	
Wilson et al. (1996) UK and USA Cross-sectional	ABI (78) Only breakdown for the whole sample (92) available: closed head injury (59%), encephalitis (6.5%), dementia (13%), stroke (8.5%).		38.8 (15.7) Range: 19-76			Median NART FSIQ=103.2 (13.2), range=73-136. Mean WAIS-R FSIQ=92.8 (15.7), range=53-136).	Significant others (carers or relatives that have close, daily contact with the patients).	

Note. **Abbreviations in table legend**: n=sample size. M=mean. SD=standard deviation. ABI=Acquired Brain Injury. DEX=Dysexecutive Questionnaire. DEX-SO=DEX completed by a significant other. DEX-HCP=DEX completed by a health care professional. **Abbreviations in table content**: AD=Alzheimer's Disease. AL=Anterior Lesion. GCS=Glasgow Coma Scale; it ranges between 3 (lowest) and 15 (highest). TBI is classified as mild (13–15 GCS), moderate (9–12 GCS), and severe (3–8 GCS). LBD=Lewy Body Dementia. LH=Left Hemisphere. LOC=Loss of Consciousness. ML=Medico-Legal. MS=Multiple Sclerosis. NCD=Neurocognitive Disorder. NHS=National Health Service. PD=Parkinson’s Disease. PL=Posterior Lesion. PTA=Post-Traumatic Amnesia. RH=Right Hemisphere. RRMS=Relapsing Remitting MS. SPMS=Secondary Progressive MS. TBI=Traumatic Brain Injury. **Measures**: 7-point Dutch education scale

ranges from 1 (elementary school not finished) to 7 (university degree). DS=Disease Steps scale; it ranges from 0 (functionally normal without limitation on activity) to 6 (confined to a wheelchair). EDSS=Expanded Disability Status Scale; it ranges from 0 (normal neuro exam) to 10 (death due to MS). FSIQ=Full Scale IQ. H&Y=Hoehn and Yahr scale; it ranges from 1 (minimal or no functional disability) to 5 (confinement to bed or wheelchair unless aided). MMSE=Mini Mental State Examination. NART=National Adult Reading Test (NART-R=revised). sGIT=Groninger Intelligence Test shortened version. TROG=Test for Reception of Grammar. UPDRS=Unified Parkinson's Disease Rating Scale; part III scores range from 0–132 (≤32=mild and ≥59=severe). WAIS=Wechsler Adult Intelligence Scale (WAIS-R=revised). * indicates that only a sub-sample of the total sample reported was used in correlational analyses, and that the information reported in this table refers to the total sample size unless stated otherwise. Please refer to Table 1.2 for the sample size for each correlation for each study.

Table 1.2

DEX scores and correlational data of included studies.

Study (patient group, BADS score type, statistical analyses)	**DEX (n)** **Mean Score (SD)**	**Action Program (AP)**	**Key Search (KS)**	**Modified Six Elements (MSE)**	**Rule Shift Cards (RSC)**	**Zoo Map (ZM)**	**Temporal Judgement (TJ)**	**BADS Total Profile Score (TSP)**
Group 1: Non-progressive ABI								
Boelen et al. (2009) ABI BADS RS (number of errors for RSC) Two-tailed Spearman rank-order correlations	DEX-Self (81) 31.93 (13.56)	-.087	-.019	-.015	**.274***	.192	.097	
	DEX-SO (78) 31.77 (14.88)	-.133	-.048	-.020	**.245***	.072	-.036	
	DEX-HCP (81) 34.88 (12.30)	-.137	-.042	-.161	-.044	-.014	-.042	

Chan & Manly (2002) Mild-to-moderate TBI BADS PS Pearson correlations	DEX-Self (30) 32.57 (16.21)			Unclear				
	DEX-SO (30) 38.50 (15.21) DEX-SO Factor 1 (Inhibition) DEX-SO Factor 2 (Intentionality) DEX-SO Factor 3 (Executive Memory)			n.s. -.28 (p=.12) **-.38*** -.05 (p=.79)				
Emmanouel et al. (2014) ABI BADS RS Two-tailed non-parametric Spearman correlations Two standard multiple regression analyses:	DEX-Self (52) AL (n=30): 6.50, (5.45). PL (22): 5.55 (4.7)	**-.298***	.052	.044	.075	.039 -.086 (Zoo Map 2 only)		.072
	DEX-HCP (52) AL (n=30): 24.20 (11.81). PL (n=22): 9.73 (6.67) <u>Regressions</u>	**-.500**** (p^2=.25) *β*=-.176 *t*=-1.169	**-.496**** (p^2=.24) *β*=-.239 *t*=-1.547	**-.661**** (p^2=.43) ***β*=-.496** ***t*=-2.825**	**-.507**** (p^2=.26) *β*=.008 *t*=.052	**-.508**** **-.505**** (Zoo Map 2 only) (p^2=.26) *β*=.314 *t*=1.349		**-.555**** (p^2=.30) ***β*= -.379** ***t*=-2.771**

-Model 1: BADS (TPS), EDT, TQT as predictors of DEX-HCP. -Model 2: BADS subtests as predictors of DEX-HCP.		*p*=.249	*p*=.129	***p*=.007*** Model 2 - Full model was statistically significant [Adjusted R^2=.413, F(6.45)=6.971, p<0.0005, effect size Cohen's f^2=0.7]. Only the MSE was a unique predictor of Model 2.	*p*=.959	*p*=.184 *β*=-.235 *t*=-1.201 *p*=.236 (Zoo Map 2 only)		***p*=.008*** Model 1 - Full model was statistically significant [Adjusted R^2=.304, F(3.48)=8.432, p<0.0005, effect size Cohen's f^2=0.436]. Only the TSP was a unique predictor of Model 1.
Evans et al. (1997) ABI BADS PS and n of rule breaks for MSE No information on type of correlation used.	DEX-Self (35) 27.59 (14.77)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.
	DEX-SO (35) 33.97 (15.85)	-.326 (p=.068)	-.302 (p=.099)	**-.430*** (p=.013) (PS) **.355*** (p=.0498) (number of rule breaks made)	**-.374*** (p=.037)	**-.388*** (p=.025)	-.270 (p=.143)	**-.566**** (p=.001)

Knight et al. (2002) ABI BADS PS Pearson correlations ^ *"While several of the tests comprising the BADS correlated with DEX-O, the best and most consistent coefficients were achieved with the Zoo Map task."*	DEX-Self (20) 22.10 (13.67)							
	DEX-HCP (20) 37.10 (12.63) DEX-HCP Factor 1 (Inhibition) DEX-HCP Factor 2 (Intentionality) DEX-HCP Factor 3 (Executive Memory) DEX-HCP Factor 4 (Positive Affect) DEX-HCP Factor 5 (Negative Affect)	Unclear^	Unclear^	Unclear^	Unclear^	**-.68**** (p=.001) **-.59**** (p=.006) **-.45*** (p=.044) **-.58**** (p=.007) **-.49*** (p=.028) Unclear^	Unclear^	Unclear^
Weddell & Wood (2016) TBI	DEX-Self (71) NHS (n=40): 34.5 (15.8)					**-.33****		

BADS PS Two-tailed Pearson correlation	ML (n=31): 36.4 (15.7)							
Group 2: Progressive ABI								
Cerezo García et al. (2015) MS BADS PS Partial correlations controlling for BDI and STAI scores	DEX-Self (48) 25.40 (14.44), median=24.50					n.s.	n.s.	
	DEX-SO (35) 19.37 (10.05), median=17.00					n.s.	n.s.	
Fukuta & Mori (2018) Mild NCD BADS PS Spearman correlations	DEX-Self (32) 8.34 (8.19)	-.10	.24	.03	.01	.19	-.01	.04
	DEX-SO (32) 16.09 (16.92)	-.42	-.23	-.27	-.41	-.34	-.06	**-.49***
Grech et al. (2016) MS	DEX-Self (107) (M and SD not available)	-.01		.06		.11		

BADS PS Two-tailed correlations (no information on type)	DEX-SO (107) (M and SD not available)	-.06		-.18		-.05		
Koerts et al. (2012) PD BADS PS Spearman correlations	DEX-Self (43) 19.98 (11.65)					-.012 (p=.938)		
	DEX-SO (43) 17.30 (11.91)					.127 (p=.423)		
Preston et al. (2013) MS BADS PS for all individual tests; age-corrected standardised score for TPS Spearman rank-order correlations	DEX-Self (68+1) (a paid carer responded on behalf of one of the patients) 26 (16)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	.099
	DEX-SO (69) 20 (15)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	-.118
	DEX-SO (69) - Temporal sequencing item (under Factor 3: Executive Memory)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	**-.306***

Group 3: Non-progressive and progressive ABI								
Burgess et al. (1998) ABI BADS PS Pearson correlations	DEX-Self (79) (M and SD not available)			.02				
	DEX-SO (79) (M and SD not available) DEX-SO Factor 1 (Inhibition) DEX-SO Factor 2 (Intentionality) DEX-SO Factor 3 (Executive Memory) DEX-SO Factor 4 (Positive Affect) DEX-SO Factor 5 (Negative Affect)			**.40***** **.24*** (p=0.4) **.46***** n.s. n.s n.s.				

Channon & Crawford (1999) ABI – AL group BADS PS Pearson correlations	DEX-Self (16) 17.50 (9.08)	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	
	DEX-SO/HCP (11/2; tot n=13) 25.08 (16.49)	n.s.	n.s.	n.s.	**-.56** (no p-value provided)	n.s.	n.s.	
Norris & Tate (2000) ABI BADS PS Two-tailed Spearman rank-order correlations	DEX-SO (27) (M and SD not available)	.25	.06	.34	–.22	**.38***	–.12	.28
Wilson et al. (1996) ABI	DEX-Self (78) 27.21 (14.48), range=2-59	-.13	-.26	.02	-.07	-.05	-.24	-.19
BADS PS Pearson correlations Three stepwise regression models to determine the best predictors of each DEX-factor	DEX-SO (78) 32.85 (15.98), range=9-67	**-.37****	**-.31****	**-.40*****	**-.45*****	**-.46*****	**-.40*****	**-.62*****
	DEX-SO Factor 1 (Behaviour)	**-.45*****	**-.32****	**-.36****	**-.41****	**-.32****	**-.42****	**-.58*****
	DEX-SO Factor 2 (Cognition)	-.27	-.27	**-.36****	**-.38****	**-.57****	-.23	**-.55****

Variables entered in the equation (score type): BADS (TPS), CET (error score), mWSCT (n of categories achieved, total errors and percentage of perseverative errors made), NART predicted FSIQ; WAIS-FSIQ, and age. BADS TSP was the only significant predictor on all three models.	DEX-SO Factor 3 (Emotion)	-.06	-.22	-.18	-.26	-.27	-.26	**-.32****
	<u>Regression:</u>							
	DEX-SO Factor 1 (Behaviour)	-	-	-	-	-	-	**Multiple R=.59, adjusted-R^2=.33,** F=20.3, p<.0001***
	DEX-SO Factor 2 (Cognition)	-	-	-	-	-	-	**Multiple R=.49, adjusted-R^2=.22, F=12.06, p<.0013****
	DEX-SO Factor 3 (Emotion)	-	-	-	-	-	-	**Multiple R=.34, adjusted-R^2=.09, F=4.83, p<.0341***

Note. * p < .05. **p < .01. ***p < .001. The strength of significant correlations was colour coded as small, medium, or large using Cohen's (1988) guidelines. **Abbreviations**: n.s.=correlation not significant. n=sample size. SD=Standard Deviation. AL=Anterior Lesion. PL=Posterior Lesion. **Measures**: BADS PS=BADS Profile Score. BADS RS=BADS Raw Score. BDI=Beck's Depression Inventory. CET=Cognitive Estimates Test (Shallice & Evans, 1978). DEX=Dysexecutive Questionnaire. DEX-Self=DEX completed by the patient. DEX-SO=DEX completed by a significant other. DEX-HCP=DEX completed by a health care professional. EDT=Everyday Descriptions Task. STAI=State/Trait Anxiety Inventory. TQT=Twenty Questions Test. mWSCT=Wisconsin Card Sorting Test - modified version (Nelson, 1976).

Question 1: investigating the associations between BADS scores and DEX ratings

a) Associations between BADS and DEX-Self

npABI

Five of the six studies in this group performed correlations between at least one BADS score and the DEX-Self; however, information on the statistical association was unclear in one study, which reported no significant correlation between MSE and DEX-Self in the discussion with no information included in the results section (Chan & Manly, 2002). Three studies, all with a sample size of >50 participants, found significant correlations between DEX-Self and the Rule Shift Cards (RSC; Boelen et al., 2009), Action Program (AP; Emmanouel et al., 2014) and Zoo Map (ZM; Weddell & Wood, 2016) subtests. Correlations were small to medium in strength and were in the predicted direction based on the type of score used (number of errors, raw scores, and profile scores, respectively), so that lower performance on the BADS was associated with higher scores on the DEX-Self. The study by Evans, Chua, McKenna, and Wilson (1997) had a smaller sample size (n=35) and did not find any significant correlations between any of the BADS scores and DEX-Self scores.

pABI

All five studies in this group performed correlations between at least one BADS score and DEX-Self ratings. No significant correlations were found in any of the studies. Sample size varied widely between studies, from 32 to 107 participants. Mean DEX scores were not available for one of the studies; however, available scores for three of the remaining four studies in this group indicate that MS and PD patients generally rated themselves higher than their significant others on the DEX. The only study in this group reporting higher mean scores for DEX-SO compared to DEX-Self assessed patients with mild NCD caused by either AD or LBD (Fukuta & Mori, 2018).

mABI

Three of the four studies in this group performed relevant correlations, and no significant correlations were found between the DEX-Self and any of the BADS scores. Sample sizes ranged from 16 to 79.

b) Associations between BADS and DEX-SO

npABI

Three studies performed relevant correlations. DEX-SO and RSC scores were significantly correlated (Boelen et al., 2009; small effect size) as were DEX-SO Factor 2 (Intentionality) and Modified Six Elements' (MSE) scores (Chan & Manly, 2002; medium effect size). Evans et al. (1997) found moderate correlations between DEX-SO and RSC, ZM, MSE, and a strong correlation with the BADS TPS. All correlations were in the expected direction.

pABI

All five studies in this group correlated the two measures, and two found significant associations of medium strength between BADS TPS and DEX-SO (Fukuta & Mori, 2018), and BADS-TPS and DEX-SO temporal sequencing question item (under Factor 3 – Executive Memory; Preston, Hammersley, & Gallagher, 2013). There was no significant correlation between total DEX-SO scores and BADS TPS in Preston et al.'s (2013) study, and it is worth noting that Bonferroni corrections were not applied for the correlations between the 20 individual DEX items and BADS scores, increasing the risk of type I error.

mABI

All four studies in this group found significant correlations between BADS and DEX-SO scores. As discussed, Burgess et al. (1998) and Wilson et al. (1996) presented data from overlapping samples, and both found medium-strength correlations between the MSE and DEX-SO. Of note, the correlation in Burgess' study was reported as positive but interpreted as being in the expected direction (i.e., better performance on the BADS is associated with higher impairments reported on the DEX); it seems likely that the authors reversed the direction of the correlation when presenting it in the table for consistency of comparison with other EF measures. Both authors also correlated DEX Factors with BADS subtests. Burgess et al. (1998) reported a five-factor structure following factor analysis of DEX-SO ratings, and the MSE significantly correlated with Factor 1 (Inhibition), although the effect size was small. A medium-strength correlation was instead found between the MSE and Factor 2 (Intentionality). Wilson et al. (1996) identified a three-factor structure in the DEX-SO, with Factor 1 (Behaviour) and Factor 2 (Cognition) showing moderate correlations with MSE scores. The authors also found moderate correlations between all the other five BADS subtests and the DEX-SO total score, and DEX-SO Factor 1 scores. Strong correlations between the BADS TPS and DEX-SO, DEX-SO Factor 1 and DEX-SO Factor 2 were also found. DEX-SO Factor 2 also correlated with RSC (with a medium effect size) and the ZM (large effect size). Only the BADS TPS was found to be significantly, but only moderately, correlated with DEX-SO Factor 3 (Emotion). The BADS TPS was also found to be the only significant predictor of all three DEX-SO Factors using stepwise regression models with other variables (including other EF measures and FSIQ scores).

The other two studies in this group had a significantly smaller sample size (<30), which as discussed may increase the risk of type I and type II error. Channon and Crawford's (1999) correlation between DEX-SO and RSC was found to be significant with a medium effect size; however, it was computed with only 13 participants, and it is also worth noting that two of the 13 DEX-SO ratings were given by HCPs due to absence of significant others. Norris and Tate's (2000) sample was also small (*n* = 27), with a moderate, significant correlation found between DEX-SO and ZM only. Interestingly, this was not in the predicted direction, with better performance on the BADS associated with higher impairment in EF as rated by significant others. Given the limitations in these two studies, however, it is difficult to ascertain whether these statistically significant findings reflect a true effect.

c) Associations between BADS and DEX-HCP

DEX-HCP ratings were only used in the npABI group, in three of the six studies. Two of these studies reported significant correlations. In Emmanouel et al.'s (2014) study, all BADS subtests (but Temporal Judgment, which was not included in this study) and BADS TPS were significantly correlated with DEX-HCP ratings, with effect sizes ranging from medium to large. Additionally, the authors found the BADS TPS to be the only unique predictor of DEX-HCP scores using a multiple regression model which included other EF tests. The MSE subtest was also found to be a unique predictor of DEX-HCP scores, when included in a multiple regression model with all other BADS subtests and EF tests. Knight et al. (1999) also found a strong significant correlation between ZM and DEX-HCP scores, and moderate to significant correlations between four of the five DEX factors identified by Burgess et al. (1998). The authors reported that several of the other BADS subtests correlated with the DEX-HCP; however, they did not include any further information (quantitative or qualitative) in their paper and were unable to provide this information when contacted due to the raw data no longer being available in accordance with GDPR policies. Of note, their sample size was the smallest within the npABI group (*n* = 20), and as such the likelihood of type I and type II errors should not be overlooked. Conversely, no significant correlations were found between the six BADS scores and DEX-HCP scores in Boelen et al.'s (2009) study, which included a larger sample. The authors reported that in some instances the DEX-HCP was completed by professionals after a relatively short meeting with the patient, which is in contrast with Emmanouel et al.'s (2014) and Knight et al.'s (2002) studies where DEX-HCP raters were reported to know the patients well.

Question 2: investigating which BADS score(s) show the strongest significant associations with DEX ratings

Overall, ZM was the subtest most frequently included in correlational analyses across all groups, followed by the MSE and AP subtests. The BADS TPS was the least commonly included score across the studies reviewed. The proportions of significant correlations (independent of DEX rating type) are presented in Table 1.3. However, these findings may be biased by the small sample size included in some of the studies, as previously discussed. To address this, Table 1.4 presents the same data without the three studies with sample of <30 patients.

Additionally, if threshold levels of significance were adjusted for multiple comparisons (using a Bonferroni correction) based on how many BADS scores were correlated with DEX ratings, only Emmanouel et al.'s (2014), Evans et al.'s (1997; only BADS TPS) and Weddell and Woods' (2016) results would retain significance in the npABI group. No correlations would remain statistically significant in the pABI group after multiple-correction adjustments, while correlations in both studies in the mABI group would still be significant. Overall, only five of the sufficiently powered studies included in the review would retain significant correlations following multiple-comparison corrections. Of these, the strongest significant correlation found most consistently (*n* = 3) is between DEX ratings (DEX-SO in two studies, DEX-HCP in one study) and BADS TPS.

Table 1.3

Number of significant correlations found across all the studies which performed correlational analyses between BADS and any DEX ratings (total n presented in parentheses), by ABI group.

BADS Score	npABI	pABI	mABI	Total Proportion	%Total Proportion
AP	1 (3)	0 (3)	1 (3)*	2 (9)	22%
KS	1 (3)	0 (2)	1 (3)*	2 (8)	25%
MSE	3 (4)	0 (3)	2 (4)*	5 (11)	45%
RSC	3 (3)	0 (2)	2 (3)*	5 (8)	63%
ZM	4 (5)*	0 (5)	2 (3)*	6 (13)	46%
TJ	0 (2)	0 (3)	1 (3)*	1 (8)	13%
BADS TPS	2 (2)	2 (2)	1 (2)*	5 (6)	83%

Note. * indicates that the count included underpowered (<30 subjects) studies (n=3).

Table 1.4

Number and proportion of significant correlations found across all studies which included a sample of >30 patients, by ABI group.

BADS Score	npABI	pABI (n=5)	mABI (n=4)	Total Proportion	% Total Proportion
AP	1 (3)	0 (3)	1 (1)	2 (7)	29%
KS	1 (3)	0 (2)	1 (1)	2 (6)	33%
MSE	3 (4)	0 (3)	2 (2)*	5 (9)*	56%*
RSC	3 (3)	0 (2)	1 (1)	4 (6)	67%
ZM	3 (4)	0 (5)	1 (1)	4 (10)	40%
TJ	0 (2)	0 (3)	1 (1)	1 (6)	17%
BADS TPS	2 (2)	2 (2)	1 (1)	5 (5)	100%

Note. * indicates that two of these studies used overlapping samples.

Reporting biases and certainty of evidence

A summary of the AXIS tool quality ratings for all included papers is presented in Table 1.5.

Quality of reporting

All studies reported clearly defined aims, target population(s), and basic study data. The most common limitation was the lack of clearly stated significance thresholds or precision estimates in the methods section (*n* = 9), although p-values thresholds were usually provided when presenting statistical results. Studies published between 1996 and 2002 obtained the lowest scores in this domain, which may be indicative of the lower or less clear reporting standards required at the time.

Study design quality

All studies employed an appropriate design and relevant outcome variables to address their aims, and the discussions and conclusions drawn were appropriate to the results reported in all studies but one. The most common methodological limitations included lack of sample size calculations (*n* = 13), use of convenience samples (*n* = 13) and possible conflicts of interest (*n* = 8). Lack of information on the attainment of ethical approval or informed consent was also a noteworthy issue in 8 of the studies, likely due to poor reporting although it is difficult to make accurate inferences.

Introduction of bias

All studies presented internally consistent results. The main contributors to low scores in this domain were lack of representative samples and lack of information about non-respondents in all studies. It was also unclear whether the BADS and DEX translations used in three studies had been previously validated for use, while authors of one study reported translating the measures themselves, introducing potential threats to validity.

Table 1.5

Summary of quality ratings for each group of studies.

ABI Group	Quality of reporting (max=7)	Study design quality (max=7)	Introduction of bias (max=6)	Total AXIS score (max=20)
npABI				
Boelen et al. (2009)	6	4	2	12
Chan & Manly (2002)	4	4	2	10
Emmanouel et al. (2014)	5	4	1	10
Evans et al. (1997)	3	3	2	8
Knight et al. (2002)	4	3	2	9
Weddell & Wood (2016)	7	6	3	16
pABI				
Cerezo García et al. (2015)	6	5	1	12
Fukuta & Mori (2018)	7	4	1	12
Grech et al. (2016)	7	5	2	14
Koerts et al. (2012)	7	4	1	12
Preston et al. (2013)	7	7	2	16
mABI				
Burgess et al. (1998)	3	3	2	8
Channon & Crawford (1999)	5	5	2	12
Norris & Tate (2000)	5	3	2	10
Wilson et al. (1996)	3	3	2	8

Discussion

This review assessed the ecological validity of the BADS by examining correlational analyses conducted between its scores and DEX ratings across studies in adults with ABI. Statistically significant correlations were also compared to determine which BADS scores, if any, more consistently showed the strongest association with DEX ratings. Fifteen peer-reviewed studies were identified as eligible and were systematically reviewed.

Association between BADS and DEX ratings

Overall, 12 of the 15 studies included in this review found statistically significant correlations between at least one BADS score and DEX ratings.

Thirteen studies tested the association between BADS and DEX-Self; however, only three out of six studies in the npABI group found weak to moderate correlations (all in the expected direction) between BADS subtests and DEX-Self ratings. This is consistent with prior research which indicates that patients with ABI tend to underrate their own EF deficits due to reduced awareness (Amanzio, Bartoli, Cipriani, & Palermo, 2020), highlighting the limited value of DEX-Self ratings in the assessment of EF unless used in conjunction with more objective measures or others' ratings.

Twelve studies performed correlations between DEX-SO and BADS scores, with nine studies across all three groups finding at least one significant correlation. Correlations varied in strength within and across studies and were all in the expected direction apart from one study (Norris & Tate, 2000) in which higher scores on the ZM (i.e., better performance) were associated with higher scores on the DEX (i.e., greater EF impairment). Of note, the mean DEX scores were not reported in the study, and the exact breakdown of ABI type (TBI and MS) was also not provided for the sub-sample of participants included in the correlational analyses. All these factors and the study's small sample may potentially account for this unusual finding. Overall, a significant proportion of the correlations computed with DEX-SO ratings led to significant results, which is in line with previous evidence and current recommendations to gather information from significant others when assessing EF in ABI adults.

The DEX-HCP was only used in three of the fifteen studies included in this review. All three studies belonged to the npABI group; of these, two found significant correlations in the predicted direction with moderate to large effect sizes. In the study that did not find any significant correlations, some of the HCPs who completed the ratings had met the patients only briefly during a clinical interview, while HCPs in the other two studies were reported to know the patients well. This could potentially explain the discrepancy in findings, as the rater's familiarity with the patient is one of the many factors that likely influences the reliability of responses. While a very small number of studies employed the DEX-HCP, making it difficult to draw generalisable conclusions, it is a promising result and may be relevant for consideration by clinicians who work with ABI populations.

The three studies which did not find any significant correlations were all in the pABI group; two studies included patients with MS (Cerezo García, Plasencia, & Benito, 2015; Grech et al., 2016) and one with PD (Koerts et al., 2012). The pABI was also the only group with studies (n = 3) that reported lower DEX-SO than DEX-Self mean scores. Additionally, DEX

mean scores (Self and SO) were noted to be, on average, lower than in the npABI and mABI groups. The overall pattern of scores and correlation results seems to be suggestive of a difference between patients with MS or PD and patients with other ABI subtypes. Although caution should be used when making inferences based on this small number of studies, these results provide tentative evidence that patients with MS or PD may generally have fewer, or different, EF deficits; it is also possible that their deficits do not affect their daily functioning to the same extent. Moreover, the difficulties in daily functioning reported by these patients or by their significant others using the DEX do not show associations with performance on BADS subtests. This may indicate that the executive profile in these populations may differ to that of npABI, and pABI conditions such as NCD. As the DEX validation study was conducted using a mix of npABI and pABI (which only included dementias), it could be argued that the BADS subtests may not be sensitive to EF deficits in MS and PD. Another potential explanation may be that the negative impact of depression, anxiety, or disability level may influence or distort patients' perceptions of their deficits and daily functioning, so that they notice more deficits than their significant others, which would be in line with findings that have previously been reported in both PD (Marino et al., 2009) and MS (Middleton, Denney, Lynch, & Parmenter, 2006). On the other hand, reduced deficit-awareness noted in other ABI groups may be a potential protective factor against affective symptoms, in line with recent findings in the literature (Azocar, Livingston, & Huntley, 2021; Perry & Coetzer, 2020).

Strongest BADS predictor of DEX ratings

The magnitude of significant correlations between BADS scores and DEX ratings varied widely within and across studies. Only studies with >30 patients were examined to answer this question in line with minimum sample size requirements for correlational analyses. The most commonly included subtest across studies was ZM, which yielded disappointing results as significant correlations were only found in 40% of the studies that tested its relationship with the DEX. However, it should be noted that these results are driven by the pABI group, where none of the five studies found significant correlations. If only npABI and mABI groups are considered, the percentage of positive correlations increases to 80%. This finding indicates that ZM might be a more sensitive measure in patients with npABI. Findings for the second most commonly included subtest, MSE, showed a similar pattern to those for ZM, as did findings for the third most frequently assessed subtest (RSC). The only BADS score that consistently showed medium to strong associations with DEX-SO and DEX-HCP ratings across all three ABI groups is the TPS, although the number of total studies is small (n = 5),

and even smaller if multiple-comparison corrections are applied (*n* = 3). These findings suggest that BADS subtests by themselves may have relatively low sensitivity, especially in pABI populations. Individual BADS subtests may only tap EF skills assessed by some of the questions or domains of the DEX, which may explain the variation in findings; on the other hand, administration of the whole BADS battery may have adequate predictive value in all ABI subtypes.

Limitations of evidence

All studies included in this review were found to have a degree of risk of bias, which has important implications for the above inferences. Generalisability of the results may be limited by the use of relatively small, convenience samples in most studies and lack of power calculations in all studies but two. The heterogeneity of ABI presentations may have led to higher individual variability within and across ABI samples and subtypes, with the lack of homogeneous samples decreasing effect sizes. While this systematic review used a pragmatic approach to categorising patients under three broad ABI subtypes, it was not possible to have more specific categories of conditions due to the limited and heterogeneous data currently available. Two thirds of the studies were conducted in European countries, which may limit their generalisability to other counties and cultures. The use of potentially not validated translations of the BADS and DEX in some of the studies may also pose threats to the validity of their results.

Strengths and limitations of review process

This review has both strengths and limitations which should be acknowledged. The PRISMA guidelines were adhered to throughout the review process, from protocol development to this write-up. The lead reviewer made efforts to contact authors to gather relevant data, which contributed to a more comprehensive and accurate presentation of findings. A validated quality appraisal measure (AXIS tool), rigorously developed using a Delphi process with three rounds of consensus, was chosen for risk of bias assessments. This tool allowed to evaluate each study's reporting quality, design quality and risk of bias.

Limitations include the review process being primarily conducted by the lead reviewer, with only a proportion of the screening process and risk of bias assessments being second-rated, leading to potential researcher bias. However, the use of PRISMA guidelines, structured screening tools and regular supervision discussions with the senior author helped mitigate this risk to some extent. Only peer-reviewed articles written in English were considered in

the review; this resulted in the exclusion of otherwise potentially eligible studies and may have increased the risk of publication and language bias. Grey literature was not included as part of this review, which may also increase publication bias; however, this decision was informed by known challenges associated with it including low reproducibility of the retrieval process, less clearly defined quality standards of the evidence, and time constraints (Benzies, Premji, Hayden, & Serrett, 2006; Brietzke, Gomes, Gerchman, & Freire, 2023).

Implications and recommendations for future research

These findings have important implications for the clinical assessment of EF in adults with ABI and for future research on this topic. Clinicians should be cautious when only selecting individual BADS subtests to assess EF, due to possible limitations in their predictive validity when used on their own. Relying solely on patients' self-reports may also be problematic, due to the under or over-reporting of deficits in ABI populations which may lead to inaccurate information about their impairments. Reports given by significant others or by HCPs who know the patients well were found to be more accurate, although clinicians should be aware of possible factors that may impact on the reliability of these scores. Overall, each of these three DEX ratings may yield unique and important contributions to the evaluation of EF, some of which cannot be measured by more objective EF tests. Clinicians may consider gathering all three ratings, where possible, or at the very least ratings from patients and their significant others which should be evaluated both qualitatively and quantitatively. Subtracting DEX-Self ratings from DEX-SO, or DEX-HCP, ratings may also be useful as a measure of insight as suggested by Burgess et al. (1998). The administration of the whole BADS battery is also recommended; however, it is more time intensive and its use may not be feasible in the limited time available to some clinicians. Generally, both BADS and DEX should be employed clinically, where possible, to assess EF; their interpretation should be guided by clinical interview, observations and evidence-based information available for the specific patient population assessed, as differences were noted between ABI subtypes.

Future research should employ more robust studies to evaluate the ecological validity of the BADS, using a combination of different questionnaire measures and ensuring sample size requirements are met. Although researchers in this field are often limited to convenience samples, they should consider including more homogeneous groups whenever possible to increase the generalisability of their results. This will help address some of the substantial limitations in the current available literature and provide further clarification regarding the ecological validity of the BADS. Another direction for future research would be to consider individual EF processes assessed by each of the BADS subtests and evaluate their

influence on specific aspects of day-to-day functioning, by examining the association with DEX factors related to those processes as well as with other EF questionnaires which assess similar domains. The impact of mood and injury severity on BADS and DEX scores should also be further explored.

Conclusions

This systematic review highlighted common patterns and inconsistencies in the relationship between BADS scores and DEX ratings, taking into account the potential influence of the studies' characteristics and quality on the results presented. Although the evidence available is limited, some support was found for the ecological validity of the BADS battery as a whole in adults with ABI, as assessed by DEX others' ratings. The evidence for individual BADS subtests as predictors of DEX ratings is mixed and less clear, largely due to the heterogeneity of the samples included and considerable risk of bias in the studies assessed. More sufficiently powered studies exploring the association between BADS and DEX scores are needed to improve the currently limited understanding of its ecological validity.

References

Amanzio, M., Bartoli, M., Cipriani, G. E., & Palermo, S. (2020). Executive Dysfunction and Reduced Self-Awareness in Patients With Neurological Disorders. A Mini-Review. *Frontiers in Psychology, 11*, 1697-1697. doi:10.3389/fpsyg.2020.01697

Azocar, I., Livingston, G., & Huntley, J. (2021). The Association Between Impaired Awareness and Depression, Anxiety, and Apathy in Mild to Moderate Alzheimer's Disease: A Systematic Review. *Frontiers in psychiatry, 12*, 633081-633081. doi:10.3389/fpsyt.2021.633081

Bennett, P. C., Ong, B., & Ponsford, J. (2005). Measuring executive dysfunction in an acute rehabilitation setting: Using the dysexecutive questionnaire (DEX). *Journal of the International Neuropsychological Society, 11*(4), 376-385. doi:10.1017/S1355617705050423

Benzies, K. M., Premji, S., Hayden, K. A., & Serrett, K. (2006). State-of-the-Evidence Reviews: Advantages and Challenges of Including Grey Literature. *Worldviews on evidence-based nursing, 3*(2), 55-61. doi:10.1111/j.1741-6787.2006.00051.x

Boelen, D. H., Spikman, J. M., Rietveld, A. C., & Fasotti, L. (2009). Executive dysfunction in chronic brain-injured patients: assessment in outpatient rehabilitation. *Neuropsychological Rehabilitation, 19*(5), 625-644. doi:https://dx.doi.org/10.1080/09602010802613853

Boyle, G. J., Stern, Y., Stein, D. J., Golden, C. J., Sahakian, B. J., Lee, T. M.-C., & Chen, S.-H. A. (2023). *The SAGE Handbook of Clinical Neuropsychology: Clinical Neuropsychological Assessment and Diagnosis*: SAGE.

Brietzke, E., Gomes, F. A., Gerchman, F., & Freire, R. C. R. (2023). Should systematic reviews and meta-analyses include data from preprints? *Trends in psychiatry and psychotherapy, 45*. doi:10.47626/2237-6089-2021-0324

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*(6), 547-558.

Cerezo García, M., Plasencia, P. M., & Benito, Y. A. (2015). Alteration profile of executive functions in multiple sclerosis. *Acta Neurologica Scandinavica, 131*(5), 313-320. doi:10.1111/ane.12345

Chan, R. C. K., & Manly, T. O. M. (2002). The application of “dysexecutive syndrome” measures across cultures: Performance and checklist assessment in neurologically healthy and traumatically brain-injured Hong Kong Chinese volunteers. *Journal of the International Neuropsychological Society, 8*(6), 771-780. doi:10.1017/S1355617702860052

Channon, S., & Crawford, S. (1999). Problem-solving in real-life-type situations: the effects of anterior and posterior lesions on performance. *Neuropsychologia, 37*(7), 757-770. doi:10.1016/S0028-3932(98)00138-9

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychology review, 13*(4), 181-197. doi:10.1023/B:NERV.0000009483.91468.fb

Coetzer, R., Daisley, A., Newby, G., & Weatherhead, S. (2013). *Practical Neuropsychological Rehabilitation in Acquired Brain Injury : A Guide for Working Clinicians*. London, UNITED KINGDOM: Taylor & Francis Group.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). London;New York;: Psychology Press.

Downes, M. J., Brennan, M. L., Williams, H. C., & Dean, R. S. (2016). Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). *BMJ open, 6*(12), e011458-e011458. doi:10.1136/bmjopen-2016-011458

Elbaum, J. (2019). *Acquired Brain Injury: An Integrative Neuro-Rehabilitation Approach* (2nd 2019. ed.): Springer International Publishing.

Emmanouel, A., Mouza, E., Kessels, R. P. C., & Fasotti, L. (2014). Validity of the Dysexecutive Questionnaire (DEX). Ratings by patients with brain injury and their therapists. *Brain Injury, 28*(12), 1581-1589. doi:10.3109/02699052.2014.942371

Evans, J. J., Chua, S. E., McKenna, P. J., & Wilson, B. A. (1997). Assessment of the dysexecutive syndrome in schizophrenia. *Psychological Medicine, 27*(3), 635-646. doi:10.1017/S0033291797004790

Fukuta, D., & Mori, C. (2018). Executive dysfunctions and instrumental activities of daily living in mild neurocognitive disorder in Japan: A preliminary study. *International Medical Journal, 25(2)*, 86-89.

Grech, L. B., Kiropoulos, L. A., Kirby, K. M., Butler, E., Paine, M., & Hester, R. (2016). Coping mediates and moderates the relationship between executive functions and psychological adjustment in multiple sclerosis. *Neuropsychology, 30*(3), 361-376. doi:10.1037/neu0000256

Knight, C., Alderman, N., & Burgess, P. W. (2002). Development of a simplified version of the multiple errands test for use in hospital settings. *Neuropsychological Rehabilitation, 12*(3), 231-255.

Koerts, J., van Beilen, M., Leenders, K. L., Brouwer, W. H., Tucha, L., & Tucha, O. (2012). Complaints about impairments in executive functions in Parkinson's disease: The association with neuropsychological assessment. *Parkinsonism & related disorders, 18*(2), 194-197. doi:10.1016/j.parkreldis.2011.10.002

Marino, S. E., Meador, K. J., Loring, D. W., Okun, M. S., Fernandez, H. H., Fessler, A. J., . . . Werz, M. A. (2009). Subjective perception of cognition is related to mood and not performance. *Epilepsy & behavior, 14*(3), 459-464. doi:10.1016/j.yebeh.2008.12.007

Middleton, L. S., Denney, D. R., Lynch, S. G., & Parmenter, B. (2006). The relationship between perceived and objective cognitive functioning in multiple sclerosis. *Archives of Clinical Neuropsychology, 21*(5), 487-494. doi:10.1016/j.acn.2006.06.008

Mueller, J. A., & Dollaghan, C. (2013). A systematic review of assessments for identifying executive function impairment in adults with acquired brain injury. *Journal of speech, language, and hearing research, 56*(3), 1051-1064. doi:10.1044/1092-4388(2012/12-0147)

Norris, G., & Tate, R. L. (2000). The Behavioural Assessment of the Dysexecutive Syndrome (BADS): Ecological, Concurrent and Construct Validity. *Neuropsychological Rehabilitation, 10*(1), 33-45. doi:10.1080/096020100389282

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(1), 89-89. doi:10.1186/s13643-021-01626-4

Perry, S., & Coetzer, R. (2020). Self-awareness following acquired brain injury is linked with depression and executive function. *Cuadernos de Neuropsicología, 14*(3), 40-51.

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: a practical guide*. Oxford;Malden, Mass;: Blackwell Pub.

Preston, J., Hammersley, R., & Gallagher, H. (2013). The executive dysfunctions most commonly associated with multiple sclerosis and their impact on occupational performance. *The British Journal of Occupational Therapy, 76*(5), 225-233. doi:10.4276/030802213X13679275042726

Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: The Virtual Library Task. *Journal of the International Neuropsychological Society, 18*(3), 440-450. doi:10.1017/S1355617711001883

Weddell, R. A., & Wood, R. L. (2016). Exploration of correlates of self-reported personality change after moderate–severe traumatic brain injury. *Brain Injury, 30*(11), 1362-1371. doi:10.1080/02699052.2016.1195921

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome.* St. Edmunds, England: Thames Valley Test Company

Wood, R. L., & Liossi, C. (2006). The ecological validity of executive tests in a severely brain injured sample. *Archives of Clinical Neuropsychology, 21*(5), 429-437. doi:10.1016/j.acn.2005.06.014

Chapter 2

An investigation of current psychology practice in the clinical assessment of executive functions

Prepared in accordance with the author requirements for *The Clinical Neuropsychologist* journal.

https://www.tandfonline.com/action/authorSubmission?show=instructions&journalCode=ntcn20#word-limits

Plain Language Summary

Title

An investigation of current psychology practice in the clinical assessment of executive functions.

Background

Executive functions (EF) are mental skills that help us to plan, make decisions, multi-task and control our emotions (Barkley, 2012). While researchers and clinicians generally agree on this broad definition of EF, this consensus appears to break down when a more detailed and complete explanation of what these skills are and how they work is sought (Jurado & Rosselli, 2007). This disagreement has meant that there is a lack of clear guidelines on how EF difficulties should be assessed by clinicians.

Aims and Questions

This study aimed to explore how Clinical Psychologists assess EF, so that we can share their knowledge and practices widely. Additionally, it aimed to look at common challenges that they encounter when completing these assessments, and future recommendations for improvement.

Methods

A total of 106 Clinical Psychologists and Clinical Neuropsychologists (CPs) in the UK with at least one year of experience in adult neuropsychology took part in an online survey with closed and open questions. CPs were recruited using different platforms, including the British Psychological Society's Division of Neuropsychology group, neuropsychology special interest groups and social media. Open questions were analysed using reflexive thematic analysis, which allows patterns of meaning (themes) across all the participants' responses to be identified.

Results

The tests which are most commonly used by CPs to assess EF are: the D-KEFS Verbal Fluency, Trail Making (D-KEFS and other versions), and the Hayling Sentence Completion. Three important areas of EF assessments were identified. The first included having good knowledge and understanding of current theories of EF. The second involved assessing patients' ability to perform daily activities independently; currently this is difficult to do and the tasks used to assess this ability would benefit from being modernised and made to be more realistic using technology. The third included examining patterns across information

from different assessment sources to make sense of the findings. Challenges affecting the assessments of EF include: disagreements on how to define and understand the concept of EF; tests of EF being poor at showing how someone may perform in daily life and at measuring what they claim to measure; lack of information or statistics for patients from different cultures, ages and impairments that provide a basis for evaluating test results and making comparisons.

Conclusions

The findings of this study provided a better understanding of how EF are assessed, and this knowledge can now be shared widely with scientists and clinicians so they can improve their research and practice. More research is needed to improve EF theories and tests, so that we can develop guidelines that clinicians can follow.

References

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved* (1st ed.). The Guilford Press

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review, 17*(3), 213–233.

Abstract

Objective: This study investigated the practices and perspectives of Clinical Psychologists and Clinical Neuropsychologists (CPs) in the United Kingdom (UK) concerning the assessment of executive functions (EF).

Method: A total of 106 CPs with at least one year of experience in adult neuropsychology completed an online survey with a mix of closed and open questions. Descriptive statistics were used to present quantitative findings; qualitative responses were analysed using reflexive thematic analysis.

Results: The D-KEFS Verbal Fluency, Trail Making (D-KEFS and other versions), and the Hayling Sentence Completion were the most commonly used EF tests in this sample. Three areas key to EF assessments were identified: 1) sufficient knowledge and understanding of current EF frameworks; 2) assessment of functional ability in clinic and in real-life settings, which is currently limited and needs to be increased through the use of modern technology; 3) triangulation of information gathered from different data sources to formulate assessment findings. Barriers to the implementation of appropriate EF assessments were described, including inconsistencies in EF definitions and theoretical understanding, tests having limited ecological validity and psychometric rigour, and lack of culturally inclusive normative data.

Conclusions: This study provided empirical information on the practices, experiences, and challenges of UK-based CPs conducting EF assessments. Further research is warranted to improve current theoretical frameworks, psychometric validity and to develop evidence-based guidance to inform EF assessments.

Keywords: Executive functions; Neuropsychological assessment; Clinical Psychologists; Clinical Neuropsychologists; Reflexive thematic analysis.

Introduction

Executive functioning (EF) is a multidimensional construct which comprises a set of cognitive processes necessary for goal-directed behaviour and self-regulation of actions and emotions (Barkley, 2012), such as initiation, planning, problem solving, decision making, set shifting, inhibition, and cognitive flexibility (Jurado & Rosselli, 2007). EF plays a key role in many aspects of life, including work, relationships and managing finances (Diamond, 2013); as such, impairments in EF can have significant negative impacts on functioning. Although EF deficits have historically been linked to injuries to the brain, several other conditions are also characterised by EF difficulties, including neurodegenerative diseases such as dementia, multiple sclerosis and Parkinson's disease (Elliott, 2003). However, the accurate assessment of EF, which is crucial for the selection of appropriate interventions in any of these disorders, is not without its challenges (Chan, Shum, Toulopoulou, & Chen, 2008).

A multitude of definitions and models of EF have been proposed since the notion of the "executive brain" was first introduced in the 1970s (Pribram, 1973). Earlier models conceptualised EF as a unitary construct (Baddeley, 1986; Norman & Shallice, 1986) while later theoretical frameworks proposed more complex and multifaceted accounts (Miyake et al., 2000;Stuss & Alexander, 2000; see Chan et al., 2008 for a review of theoretical models of EF and associated instruments). A recent literature review found as many as 48 EF models that were studied or referenced across the 106 papers examined, with Miyake et al.'s (2000) unity/diversity framework being the most frequently cited (Baggetta & Alexander, 2016). Due to this lack of a universally agreed upon definition and explanatory model, the terminology used, and the number of individual processes included within the term 'EF' can vary widely. Another literature review found 68 different terms or processes that fell under the umbrella term of EF, which they subsequently reduced to 18 using latent semantic analysis, and 98 different tasks to assess them (Packwood, Hodgetts, & Tremblay, 2011). The lack of a clear definition and the variety of EF processes involved have undoubtedly played a role in the development of a substantial number of different tests to measure EF, which differ considerably in terms of number and type of skills assessed (Alvarez & Emory, 2006). It is also worth noting that some of the most widely used tests in current clinical practice were initially devised to detect deficits arising from frontal lobe damage (Jurado & Rosselli, 2007); however, it is now recognised that other brain regions are also required for intact EF (Alvarez & Emory, 2006), and that these traditional tests place greater focus on sensitivity to dysfunction than on specificity to distinct cognitive processes. Additionally, it has long been acknowledged that the very nature of EF can make it difficult to measure in clinical settings, given the structured format of tasks and the relatively controlled

environments they are conducted in (Banich, 2009). The predictive validity of EF tests is therefore limited, and test results are often not indicative of how individuals may perform in real-life scenarios (Shallice & Burgess, 1991).

This issue led to the development of ecologically valid psychometric tests, that is, tests that more closely resemble everyday situations (i.e., have verisimilitude) and that may in turn more accurately predict performance in everyday life (i.e., veridicality), as well as self-report and informant questionnaires which assess the presence of EF deficits and their impact on daily functioning (Burgess et al., 2006). While these more recent measures appear to better predict functional difficulties, there continues to be limited correspondence between test performance and functional deficits, as individuals with EF difficulties may perform within expected limits during assessments but can display significant impairments in specific aspects of everyday functioning (Burgess, Alderman, Evans, Emslie, & Wilson, 1998). Furthermore, a known sequela of executive dysfunction is reduced self-awareness of cognitive and behavioural impairments (Amanzio, Bartoli, Cipriani, & Palermo, 2020). For this reason, information on EF is usually obtained using a variety of additional measures, including clinical interviews, observations, and review of records (Suchy, Ziemnik, & Niermeyer, 2017), although this may be limited by the time constraints within which clinicians conduct assessments and by the measures and sources of information available to them.

The above complexities and ambiguities have resulted in the lack of a recommended or “gold standard” way of assessing EF, and little is known about how EF is conceptualised and assessed by individual clinicians in the United Kingdom (UK). Recent surveys of clinical neuropsychologists have mainly investigated general assessment practices for all cognitive domains, rather than EF specifically, and were conducted outside of the UK, including in France (Branco Lopes et al., 2021), Italy (Onida et al., 2019), Spain (Olabarrieta-Landa et al., 2016), Latin America (Arango-Lasprilla, Stevens, Morlett Paredes, Ardila, & Rivera, 2017), and the United States and Canada (Rabin, Paolillo, & Barr, 2016). Whilst Rabin et al.’s (2016) survey included some questions on the most commonly used instruments to assess EF, activities of daily living and subjective reports of cognitive difficulties, the findings relate to neuropsychology practices in North America and may not be generalisable to the UK. To date, and to our knowledge, no research has been conducted with Clinical Psychologists and Clinical Neuropsychologists (CPs) in the UK with a specific focus on the assessment of EF.

Aims and research questions

The primary aim of this study was to gather and formalise current "practice-based evidence" of UK-based CPs in the assessment of EF. To achieve this, the study's primary research questions were to identify the measures and sources of information typically used for assessing EF, and the rationales underpinning these practices. Another research question was to investigate the main challenges and barriers in the assessment and interpretation of EF results. A final research question set to explore CPs' perspectives and recommendations on how to improve EF assessments to inform future clinical practice.

Methods

Design

The study employed a mixed-methods, observational, cross-sectional design. It involved an online survey with a mix of closed and open-ended questions. While the survey design, which mostly generated quantitative data, was chosen to obtain as large and representative a sample as possible, the addition of qualitative data yielded richer insight into CPs' views and suggestions.

Participants

The study inclusion criteria were as follows:

A. be a qualified CP who holds registration with the Health and Care Professions Council (HCPC) as a Practitioner Psychologist;

B. have at least one year of clinical experience in adult neuropsychology;

C. live and work in the UK. If no longer practicing/retired, the participant should have worked as a CP in the UK at some point during their working life and should have obtained the experience stated in inclusion criterion B in the UK.

A total of 139 survey responses were submitted. Thirty-three responses were invalid, as determined by the time taken to complete the survey (< 200 seconds) and/or the provision of clearly unrelated or incompatible responses, a large proportion of which appeared to be generated by survey bots. A possible explanation for this may be that ineligible participants wished to obtain entries into the survey's prize draw. Three valid responses were excluded as they did not meet inclusion criterion A (n = 2) or B (n = 1). Responses from three further participants from the survey pilot (see 'Materials' section for details) were added to the dataset, leading to a total of 106 participants included in the final analysis.

Materials

The online survey (Appendix 2.1) was designed by the first author with input from the senior author. It was piloted with three University staff members who met the inclusion criteria, and with an additional staff member who did not meet the inclusion criteria but had experience in neuropsychology. All feedback received was incorporated in the final survey, but as the changes were editorial in nature, pilot data for the three eligible respondents were included in the final dataset. The survey was distributed electronically using the University of Glasgow online Qualtrics XM (Provo, UT) survey platform account. To maximise recruitment, participants who completed the survey were offered the opportunity to enter a prize draw to win one of five gift vouchers worth £20.

The final survey contained 53 questions across six main sections: 1) study eligibility (questions 1-3); 2) education and employment details (questions 4-13); 3) assessment of EF (questions 14-43); 4) interpretation of results of EF assessments (questions 44-47); 5) feedback of results of EF assessments (questions 48-50); 6) demographic details (questions 51-53). The survey included a range of quantitative and qualitative answer formats, including Likert scale, multiple-choice and free text boxes. In the interest of brevity, fourty-six questions were used to generate the analyses and results presented in this paper. The qualitative question included in this study asked participants “Do you have any suggestions to improve our way of assessing EF?”.

Procedure

The survey was open for a total of 13 weeks, from 30th November 2022 to 28th February 2023. Targeted and opportunistic sampling was used to reach the largest possible number of eligible CPs. Participants were recruited via interest- and profession-based e-mail lists and public forums, including the British Psychological Society (BPS) Division of Neuropsychology (DoN) newsletter, regional and special interest groups (see https://www.bps.org.uk/member-microsites/division-neuropsychology), directors of accredited postgraduate neuropsychology training courses, and social media posts on Twitter, Facebook and LinkedIn. A study advert and survey link (Appendix 2.2) were included in all requests for participation and circulation to other relevant contacts. Reminders were circulated every three to four weeks. The survey link directed participants to an online information sheet (Appendix 2.3) and consent form (Appendix 2.4). Consenting participants continued to the online survey.

Ethics, Governance and Data Protection

Ethical approval was obtained from the University of Glasgow Medical, Veterinary & Life Sciences College Ethics Committee on 24th October 2022 (application number: 200220005; Appendix 2.5). All participants provided informed consent prior to participation. Demographic data were stored in a separate file from participant responses. Each response was given a unique participant ID number to allow linking with demographic information. Participants were asked not to include personally identifiable information about themselves or people they work with in the responses (no such disclosures were made). Only fully anonymised data were used in the production of reports.

Data analysis

Descriptive statistics were used to analyse quantitative answers. Frequency distributions were generated by Qualtrics and exported into Microsoft Excel. The raw dataset was also exported into Microsoft Excel to obtain measures of central tendency and dispersion. Qualitative data included in the "Other" text boxes, which allowed participants to provide additional information on their quantitative responses, were coded and grouped into categories in Microsoft Excel following a content analysis approach.

Qualitative data obtained from the open-ended question were analysed manually on Microsoft Word using a six-step reflexive thematic analysis (TA) framework (Braun & Clarke, 2006). Reflexive TA was deemed to be an appropriate method for addressing the study aim related to suggestions for improving EF assessments, as it allowed for the explorations of participants' broad range of views and experiences of current challenges, limitations, and recommendations. A primarily inductive approach was used (i.e., codes and themes were developed from the data content). The analytic approach was situated within a critical realist framework (Terry, Hayfield, Clarke, & Braun, 2017), which allowed for a holistic understanding of participants' perceptions and experiences which are produced, and situated within, the broader socio-cultural contexts and healthcare system of the UK. The first author began by becoming thoroughly acquainted with the data, re-reading responses several times. Following this, she developed preliminary codes, which were reviewed by the senior author (JF). Initial themes were also reviewed by JF and refined until consensus was reached. The analysis was guided by Braun and Clarke's (2006) 15-point criteria checklist for conducting good thematic analysis.

Sample Size

Estimating the number of UK-based CPs that would meet the study's inclusion criteria was challenging as the professional organisation of most relevance to the target population, the BPS Division of Neuropsychology, is voluntary, and the only mandatory organisation to which they must belong, the HCPC, does not publish data on the specialisms registrants work within. A recent research project, which also surveyed UK CPs with experience in neuropsychology, obtained 78 responses (Baber, 2020). It was anticipated that a similar sample size could be achieved for this survey. As for the use of reflexive TA, Terry et al. (2017) recommend an indicative sample size of 30-100 responses for qualitative surveys conducted as part of professional Doctorate projects.

Reflexivity

Members of the research team included a trainee clinical psychologist and a clinical psychologist/neuropsychologist. Both have clinical and research experience in adult neuropsychology and a special interest in this study's topic. This prompted both authors to consider, and remain acutely aware of, possible ways in which data analyses could be influenced by their professional background, experiences, and prior assumptions. The practice of ongoing reflexivity ensured that both authors remained critical of their own stances and decisions in relation to this research throughout all its stages, from inception to completion.

Results

Demographic information

Key characteristics of the sample are summarised below, with full details presented in Appendix 2.6. An asterisk was used in tables where the number of responses was ≤5 to prevent identification of individuals.

Over 70% of participants were 30-49 years of age, with just over three quarters identifying as female (77.4%). The majority reported practicing in England (77.5%). See Table 2.1 for all demographic information.

Table 2.1

Characteristics of the sample (n = 106).

Demographic characteristic	Frequency	Percentage
Age range		
20-29 years	*	*
30-39 years	40	37.7%
40-49 years	36	34.0%
50-59 years	19	17.9%
60-69 years	6	5.7%
70+ years	*	*
Prefer not to say	*	*
Gender		
Female	82	77.4%
Male	24	22.6%
Non-binary / gender fluid	-	-
Prefer to self-describe:	-	-
Prefer not to say	-	-
Country/countries of practice[a]		
England	82	77.4%
Scotland	23	21.7%
Wales	*	*
Northern Ireland	*	*
Non-UK country - please specify:	*	*
Prefer not to say	-	-

Note. [a] Multiple responses were permitted, therefore responses do not total 100%. Percentages are calculated from the total respondents for this question, for each response option.

Qualifications

Most participants reported completing a Doctorate in Clinical Psychology or its equivalent (98.1%). Over half completed a PGDip/MSc in Clinical or Applied Neuropsychology (52.8%). Around a third indicated being on the Specialist Register of Clinical Neuropsychologists (SRCN; 35.8%) and 23.6% reported having the Qualification in Clinical Neuropsychology (QiCN).

Employment

Just under half of participants (46.2%) selected Clinical Psychologist as their current main role, and over a quarter (29.2%) reported working as Consultant Clinical Neuropsychologists. Most participants reported working in the NHS (88.7%) and over a quarter in independent or private practice (26.4%; see Table 2.2).

Table 2.2

Employment details of the sample (n = 106).

Employment information	Frequency	Percentage
Main role(s)[a]		
Clinical Psychologist	49	46.2%
Consultant Clinical Psychologist	12	11.3%
Clinical Neuropsychologist	16	15.1%
Consultant Clinical Neuropsychologist	31	29.2%
Researcher	4	3.8%
Other	*	*
Prefer not to say	-	-
Current workplace(s)[a]		
NHS	94	88.7%
Private healthcare provider	7	6.6%
Independent or private practice	28	26.4%
Charitable/third sector	*	*
Higher education institution	14	13.2%
N/A - no longer practicing/retired	-	-
Other	*	*
Prefer not to say	-	-

Note. [a] Multiple responses were permitted for this question.

Just under half of participants indicated working in community/outpatient adult neurorehabilitation (44.3%), followed by outpatient (40.6%) and acute (35.8%) neuropsychology settings, inpatient adult neurorehabilitation (33%) and memory clinic/dementia services (17%; Appendix 2.6.i) The clinical presentations participants reported working with most frequently included acquired brain injury (ABI; 77.4%), traumatic brain injury (TBI; moderate and severe: 68.9%; mild: 56.6%), stroke (63.2%), other neurological and neurosurgical presentations (53.8%) and functional neurological disorders (50.9%; Appendix 2.6.ii). Participants' years of post-qualification overall clinical experience ranged from less than one to 31+ years (median = 11). Of these, an average of 12 years was in adult neuropsychology (range = <1-31+ years; median = 11), which increased to 13.9 years when including pre-qualification experience (range = 1-31+ years; median = 12). Participants reported working an average of 36.8 hours a week including overtime and private work (range= 15-60) with 62.3% working full-time hours or more (≥37.5).

Over half of participants (53.8%) reported that 76-100% of their current work time required neuropsychology skills, while just under a quarter (23.6%) selected 51-75%. Over a quarter (27.4%) reported spending 80% of their neuropsychology work time on assessment and 20% on therapeutic intervention (see Table 2.3).

Table 2.3

Neuropsychology component of the participants' main role(s) (n = 106).

	Frequency	Percentage
Percentage of total work time characterised as requiring neuropsychology skills		
0%	-	-
1-25%	9	8.5%
26-50%	15	14.2%
51-75%	25	23.6%
76-100%	57	53.8%
Average balance of time spent between neuropsychological assessment and intervention		
100% assessment	7	6.6%
80% assessment / 20% therapeutic intervention	29	27.4%
60% assessment / 40% therapeutic intervention	21	19.8%
50% assessment / 50% therapeutic intervention	12	11.3%
40% assessment / 60% therapeutic intervention	15	14.2%
20% assessment / 80% therapeutic intervention	17	16.0%
100% therapeutic intervention	-	-

Self-reported competence

Participants' self-reported competence levels in EF assessment, formulation and feedback are illustrated in Table 2.4. The majority answered either "A lot" or "Somewhat" in all three domains.

Table 2.4

Self competence ratings in EF assessment, formulation, and feedback (n = 106).

Rating	Assessment n (%)	Formulation n (%)	Feedback n (%)
Not at all	-	-	-
A little	2 (1.9%)	1 (0.9%)	2 (1.9%)
Somewhat	36 (34.0%)	44 (41.5%)	38 (35.9%)
A lot	48 (45.3%)	43 (40.6%)	46 (43.4%)
A great deal	20 (18.9%)	18 (17.0%)	20 (18.9%)

Assessment of EF

Models of EF

Amongst the most frequently selected models of EF used were Norman and Shallice's (1980) supervisory attentional system model and its expansion by Shallice and Burgess

(1996), and Baddeley's (1986) working memory model (see Table 2.5). Additional models listed by participants included Zelazo and Carlson's (2012) hot and cool executive function in childhood and adolescence model (*n* = 1), Wilson and Betteridge's neurorehabilitation models (*n* = 1), Bunge et al.'s (2000) resource model of the neural basis of executive working memory (*n* = 1), and Crosson et al.'s (1989) pyramid model of self-awareness (*n* = 1). A proportion (*n* = 10) reported not using or finding any theoretical models helpful.

Table 2.5

Theoretical models of EF used by participants to guide their thinking.

EF model	Frequency	Percentage
Norman and Shallice's (1980) Supervisory Attentional System model and its expansion by Shallice and Burgess (1996)	79	74.5%
Baddeley's (1986) Working Memory model	78	73.6%
Stuss and Benson's (1986) tripartite model; Stuss and Alexander's (2007) model and its update (Stuss, 2011)	48	45.3%
Lezak's (1995) four-component model	22	20.8%
Ylvisaker's (1998) eight aspects of goal-directed task behaviour	20	18.9%
Duncan's (1996) Goal Neglect theory	19	17.9%
Miller and Cohen's (2001) integrative theory of prefrontal cortex function	18	17.0%
Damasio's (1996) Somatic Marker hypothesis	15	14.2%
Luria's (1966, 1973) Functional Unit model	8	9.4%
Miyake and Friedman's (2012) Unity/Diversity framework	7	7.5%

Note. Multiple responses were permitted for this question.

The most frequently used sources of information when assessing EF were the patient's current functioning, observations during interview and assessment, the client's self-report of difficulties, and the client's background history. All percentages of responses are presented in Table 2.6. The same information has also been presented using Figure 2.1 to aid data visualisation (see Appendix 2.7 for "other" responses).

Table 2.6

Most common sources of information gathered to assess EF.

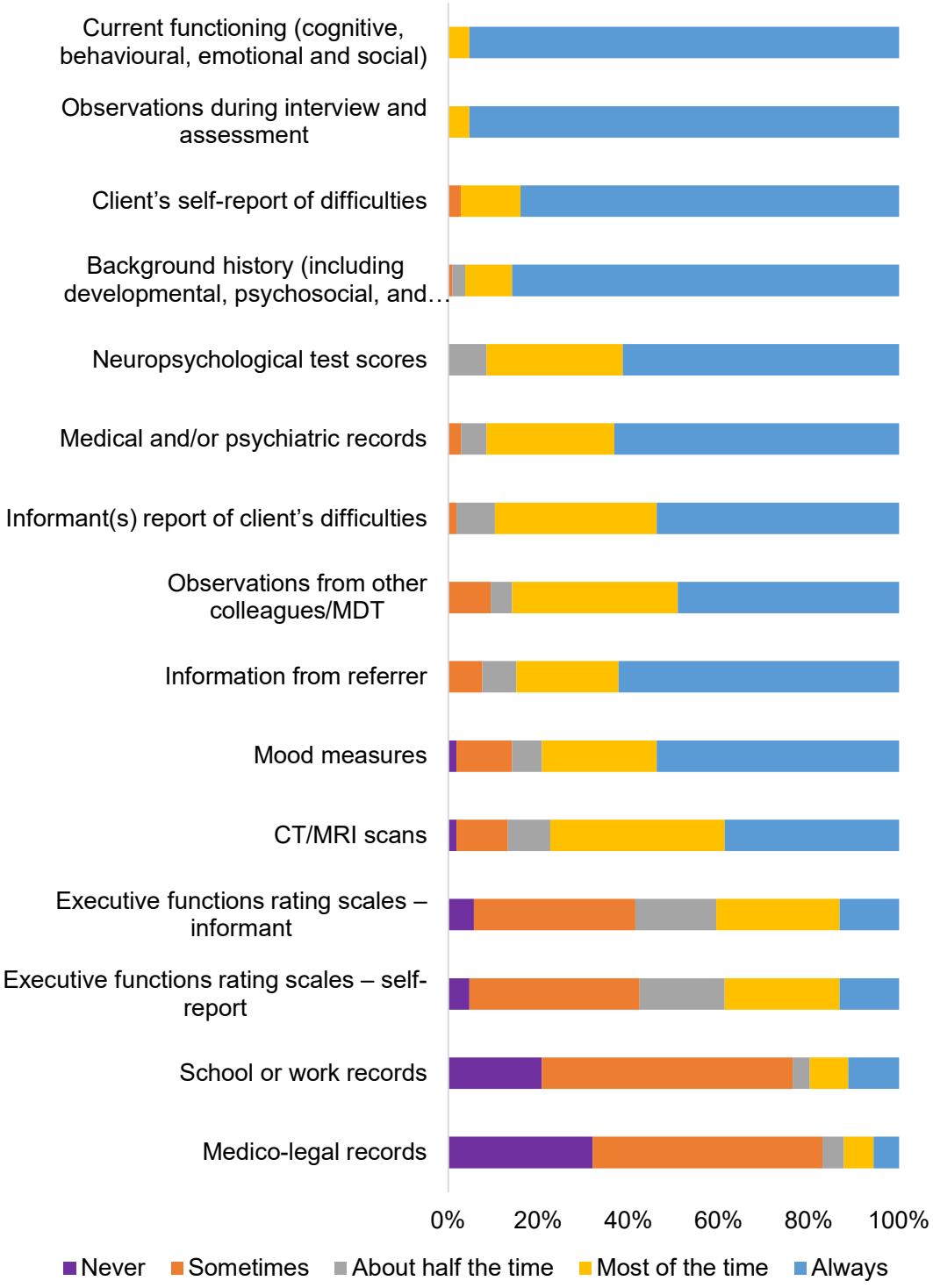
Rank	EF information source	Never	Sometimes	About half the time	Most of the time	Always
1	Current functioning (cognitive, behavioural, emotional and social)	-	-	-	4.7%	95.3%

2	Observations during interview and assessment	-	-	-	4.7%	95.3%
3	Client's self-report of difficulties	-	2.8%	-	13.2%	84.0%
4	Background history (including developmental, psychosocial, and medical)	-	0.9%	2.8%	10.4%	85.8%
5	Neuropsychological test scores	-	-	8.5%	30.2%	61.3%
6	Medical and/or psychiatric records	-	2.8%	5.7%	28.3%	63.2%
7	Informant(s) report of client's difficulties	-	1.9%	8.5%	35.8%	53.8%
8	Observations from other colleagues/MDT	-	9.4%	4.7%	36.8%	49.1%
9	Information from referrer	-	7.5%	7.5%	22.6%	62.3%
10	Mood measures	1.9%	12.3%	6.6%	25.5%	53.8%
11	CT/MRI scans	1.9%	11.3%	9.4%	38.7%	38.7%
12	Executive functions rating scales – informant	5.7%	35.8%	17.9%	27.4%	13.2%
13	Executive functions rating scales – self-report	4.7%	37.7%	18.9%	25.5%	13.2%
14	School or work records	20.8%	55.7%	3.8%	8.5%	11.3%
15	Medico-legal records	32.1%	50.9%	4.7%	6.6%	5.7%

Note. Multiple responses were permitted for this question. 'Always' and 'most of the time' ratings were combined and used to rank responses from highest to lowest combined %.

Figure 2.1

Most common sources of information gathered to assess EF.



Cognitive tests

Almost two-thirds of participants (60.4%) reported always including tests of EF in their neuropsychological assessments, whilst over a third reported including them most of the time (36.8%). A minority reported using them half of the time (1.9%) or just sometimes (0.9%), with no participant indicating they never use EF tests. Four participants provided example scenarios in which they may not include tests of EF in their assessments; these comprised lack of patient's engagement or consent, level and type of impairment limiting assessment, only using brief cognitive screens, time constraints, and using data from observations and day-to day function instead of EF tests. The top 15 stand-alone tests typically used by participants for assessing EF are presented in Table 2.7 (see Appendix 2.8 for all responses). The top 30 battery subtests used in EF assessments are reported in Table 2.8 (responses for each battery are presented in Appendix 2.9).

Table 2.7

Top 15 most typically used stand-alone tests of EF.

Stand-alone EF test	Frequency	Percentage
Trail-Making Test or Trails A & B	91	85.9%
Hayling Sentence Completion Test	88	83.0%
Brixton Spatial Anticipation Test	73	68.9%
Stroop Task	60	56.6%
Clock Drawing Test	53	50.0%
Controlled Oral Word Association Test	37	34.9%
Rey-Osterrieth Complex Figure Test	33	31.1%
Tower of London/Tower of Hanoi	31	29.3%
Symbol Digit Modalities Test	26	24.5%
Go/No Go Test	25	23.6%
Design Fluency Test	20	18.9%
Category Test	16	15.1%
Wisconsin Card Sorting Test	13	12.3%
Weigl Colour-Form Sort Test	11	10.4%
Cognitive Estimation Test	5	4.7%

Note. Multiple responses were permitted for this question.

Table 2.8

Top 30 most typically used battery subtests to assess EF.

Battery subtest	Frequency	Percentage
D-KEFS Verbal Fluency	100	94.3%
D-KEFS Trail Making	95	89.6%
BADS Zoo Map	82	77.4%
WAIS-IV Similarities	78	73.6%
D-KEFS Color-Word Interference	76	71.7%
BADS Key Search	74	69.8%

RBANS Semantic Fluency	73	68.9%
WAIS-IV Matrix Reasoning	73	68.9%
RBANS Digit Span	70	66.0%
RBANS Coding	69	65.1%
RBANS Figure Copy	68	64.2%
WAIS-IV Digit Span	68	64.2%
RBANS List Learning	65	61.3%
RBANS Story Memory - Immediate Recall	65	61.3%
RBANS List Recognition	65	61.3%
RBANS List Recall	64	60.4%
RBANS Figure Recall	64	60.4%
RBANS Line Orientation	63	59.4%
RBANS Picture Naming	63	59.4%
RBANS Story Memory - Delayed Recall	63	59.4%
WAIS-IV Block Design	63	59.4%
WAIS-IV Symbol Search	51	48.1%
WAIS-IV Coding	51	48.1%
WAIS-IV Vocabulary	49	46.2%
D-KEFS Tower	45	42.5%
WAIS-IV Arithmetic	44	41.5%
BADS Modified Six Elements	42	39.6%
WAIS-IV Information	34	32.1%
D-KEFS Design Fluency	32	30.2%
BADS Rule Shift Cards	30	28.3%

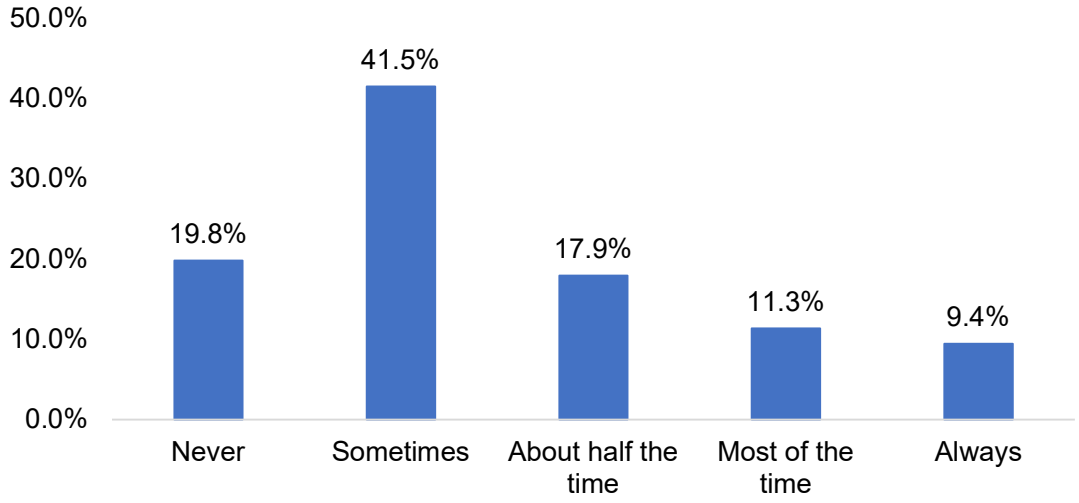
Note. Multiple responses were permitted for this question.

Functional tests

Over one-third of participants (41.5%) indicated they sometimes use information from functional tasks (either administered by them or a colleague) as part of their EF assessments, while only 9.4% reported always including them (Figure 2.2).

Figure 2.2

Frequency (%) of use of functional tasks (including if observed or administered by a colleague) to assess EF.



Sixty-two participants provided further information on the tasks they most often incorporate. The most frequently reported were a variety of functional tasks (*n* = 36), often carried out by Occupational Therapists (OTs), followed by Multiple Errands Tests (METs; *n* = 30), observations of clients in different settings (clinic, during testing, on other sessions; *n* = 19), self-reports, reports from relatives/carers as well as professionals from other disciplines (*n* = 5; see Appendix 2.10 for full summary of data).

Questionnaires and behaviour rating scales

The Dysexecutive Questionnaire (DEX) was the most frequently selected questionnaire used to assess EF, with three quarters of participants endorsing it (75.5%). This was followed by the Frontal Systems Behavior Scale (FrSBe; 26.4%) and the Behavior Rating Inventory of Executive Function-Adult (BRIEF-A; 20.8%). Eleven participants (10.4%) reported not using any questionnaires or behaviour rating scales to assess EF (all responses are presented in Appendix 2.11).

Average number of EF measures used

Over half of participants (53.8%) reported using three to five EF measures (including cognitive tests, functional tasks, and questionnaires or rating scales) on average. A third of participants (36.8%) reported that they use an average of six to eight. A smaller number selected one to two (4.7%) and nine to eleven (4.7%) as their responses.

Test selection for EF assessments

Approximately half of participants (55.7%) reported having some EF measures they always administer and others they add to their battery depending on the clinical presentation, while the remainder (44.3%) reported having a flexible battery and deciding which EF measures to administer after their clinical interview. Nobody selected the option "I have a fixed battery".

Important factors to the development of EF assessment approach

Over half of participants (53.8%) reported that supervision discussions were very important in the development of their approach to assessing EF, while a third (34.9%) rated it as extremely important. Specialist neuropsychology teaching was also rated as either very important (34.0%) or extremely important (35.8%) by over two-thirds of participants (see Table 2.9). Additional factors listed by participants included gaining clinical experience (*n* = 10) and availability of tests and resources in the service they worked in (*n* = 9; see Appendix 2.12 for all other responses).

Table 2.9

Importance of influential factors for the development of participants' EF approach.

Rank	Influential factor	Not at all	Slightly	Moderately	Very	Extremely
1	Supervision discussions	-	2.8%	8.5%	53.8%	34.9%
2	Specialist Clinical Neuropsychology teaching	2.8%	9.4%	17.9%	34.0%	35.8%
3	Feedback from colleagues	0.9%	10.4%	22.6%	52.8%	13.2%
4	Information from existing literature	0.9%	10.4%	34.0%	39.6%	15.1%
5	Adopting the same (or similar) approach as former supervisors	6.6%	25.5%	35.8%	24.5%	7.5%
6	Clinical Psychology Doctorate teaching	17.0%	32.1%	30.2%	17.0%	3.8%

Note. 'Very' and 'Extremely' ratings were combined and used to rank responses from highest to lowest.

Addressing cross-cultural challenges in the assessment of EF

The most common considerations adopted to address cross-cultural challenges during EF assessments are listed in Table 2.10. Other cross-cultural considerations listed by participants included gaining an understanding of the client's culture and social norms (n =

7) and relying on self and/or informant reports more than test data (n = 3). See Appendix 2.13 for all additional responses.

Table 2.10

Frequency of cross-cultural considerations use in EF assessments.

Rank	Cross-cultural consideration	Never	Sometimes	About half the time	Most of the time	Always
1	Administer tests that are as 'culture-fair' as possible (e.g., non-verbal tests)	1.9%	17.0%	9.4%	46.2%	25.5%
2	Adopt a more lenient/flexible approach to scoring / interpretation of scores	10.4%	27.4%	5.7%	32.1%	24.5%
3	Rely more on functional assessments	3.8%	26.4%	15.1%	32.1%	22.6%
4	Use interpreter services for interview and testing	8.5%	45.3%	4.7%	20.8%	20.8%
5	Ensure interpreters are trained in the joint administration of neuropsychological tests before proceeding	38.7%	24.5%	4.7%	11.3%	20.8%
6	Use tests and norms developed or adapted specifically for a given culturo-linguistic group	24.5%	45.3%	11.3%	15.1%	3.8%
7	Ask for input/help from other neuropsychologists that are fluent in the client's language	38.7%	36.8%	9.4%	8.5%	6.6%

Note. 'Always' and 'Most of the time' ratings were combined and used to rank responses from highest to lowest.

Top three main challenges with the selection of EF tests

Participants were asked to select up to three main challenges (from a list of options provided; see Table 2.11) associated with the selection of EF tests. Three participants reported tests not being available in their service and limited budget as an additional challenge under "Other" (Appendix 2.14 includes all "other" responses).

Table 2.11

Frequent challenges associated with the selection of EF assessment measures.

Challenge	Frequency	Percentage
Tests lack adequate ecological or predictive validity	77	72.6%

Tests lack adequate sensitivity to detect impairments	36	34.0%
Tests lack parallel or alternate forms	36	34.0%
Tests are culturally biased	34	32.1%
Tests lack adequate normative data	32	30.2%
Tests lack clear criteria for assessing change in an individual's function over time	25	23.6%
Tests lack adequate reliability	23	21.7%
Tests for a given patient population or executive functions domain lack consistent use by clinicians	11	10.4%
Tests are too expensive	8	7.5%

Note. Answer options for this question were adapted from Rabin et al.'s (2016) survey.

Formulation of EF assessments

Table 12 illustrates the main challenges associated with the interpretation of EF test results, from which participants could select up to three. Additional responses are included in Appendix 2.15.

Table 2.12

Frequent challenges associated with the interpretation of EF assessments.

Challenge	Frequency	Percentage
Limited or poor ecological validity of some tests	81	76.4%
Limited or poor psychometric data available for some tests	58	54.7%
Limitations in current theoretical knowledge	50	47.2%
Limited time available to complete a sufficiently thorough assessment	36	34.0%
Difficulty accessing information (e.g., medical records, scans, lack of informants, etc.)	30	28.3%

Feedback of EF assessment results

Feedback of EF assessments is most frequently provided to MDT members, followed by clients together with their significant others. See Table 2.13 for all participants' responses.

Table 2.13

Frequency of feedback provision of EF assessment results to relevant individuals.

Individual(s) receiving feedback	Never	Sometimes	About half the time	Most of the time	Always
Multi-disciplinary team members (MDT)	0.9%	11.3%	9.4%	36.8%	41.5%
Clients together with relatives/carers	-	13.2%	17.0%	56.6%	13.2%

Other healthcare professionals involved in the client's care	0.9%	21.7%	17.0%	37.7%	22.6%
Clients alone	2.8%	34.0%	17.0%	25.5%	20.8%
Relatives/carers alone	24.5%	48.1%	11.3%	14.2%	1.9%

Table 2.14 displays results for how often CPs provide feedback based on feedback modality. Participants indicated that they mostly provide feedback verbally, or in both written and verbal formats.

Table 2.14

Frequency of feedback by delivery modality.

Modality	Never	Sometimes	About half the time	Most of the time	Always
Verbal feedback	-	7.5%	1.9%	29.2%	61.3%
Both written and verbal feedback	-	9.4%	5.7%	37.7%	47.2%
Written feedback	-	10.4%	5.7%	25.5%	58.5%

Improving current assessment practices

Sixty-seven participants responded to the question "Do you have any suggestions to improve our way of assessing EF?". Ten participants reported having no suggestions and three responded with "not sure". The three main themes identified from the remainder of the responses (*n* = 54) are presented below (a summary of the themes and associated participants' quotes is included in Appendix 2.16).

Overarching theme: "Breaking free from tradition by driving and embracing change, innovation, and inclusivity"

Participants identified many limitations in the current approaches to the assessment of EF and provided a range of recommendations to address them. Overall, there was a strong sense that current knowledge and practices are outdated, including EF assessment methods, theoretical frameworks and their application to clinical practice, and psychometric data. CPs should increasingly move away from "blind loyalty" to more traditional practices and tests towards embracing new approaches and technological advances.

1) A call for improved ecological validity

1a. Standardisation and inclusion of functional assessments

The most prominent theme encapsulated the need for tests that have better ecological validity, represent everyday functions, and incorporate real-life situations. Participants

articulated that more standardised tasks of daily living and functional abilities should be developed and used, both in clinic and real-life settings, to improve EF assessments. Some of the participants highlighted the practical challenges of using more ecologically valid tests in clinical practice, including time efficiency and ease of administration:

> P28 *"We need efficient tasks with good norms and ecological validity, that can be carried out easily in clinical settings - not an easy task!"*

> P54 *"I've always liked some of the more behavioural tasks from the BADS (zoo / 6 elements) but slow to administer."*

1b. Modernising EF tests

Many participants suggested an increased use of digital technology, including computer-based tests and virtual reality (VR), to develop and incorporate more ecologically valid tests in the assessment process. A participant highlighted the need of making these affordable, while another noted that practical implementation of these digital tests in clinical practice will likely take time:

> P67 *"I think advances in technology with computerised testing are promising (e.g., incorporating item response theory and the Boston process approach) but probably a long way off application in clinical practice"*

1c. Integrating multiple data sources to enhance ecological validity of the overall assessment and its findings

Here, participants described the importance of including a wide range of measures in addition to formal cognitive tests when assessing EFs. Particular emphasis was placed on the clinical interview, which was mentioned across all comments in this sub-theme. Other measures included clinical observations, questionnaires, and medical data, which provide information that clinicians can triangulate to better capture everyday EF skills and functioning.

2) A need for improved EF knowledge and theory-practice links

2a. Improving the theoretical understanding of EF

Participants emphasised the need for an improved and more comprehensive theoretical understanding of EF to promote consistency across clinicians in the way EF is conceptualised and assessed. One participant commented on the lack of a consensus definition for EF, which is crucial to guide assessment; another highlighted the potential for

misinterpretation of assessment results due to perceived drawbacks in the current way EF skills are grouped under a unifying concept:

> P2 "*Reconsider the concept of EF. It's a broad term for many systems/functions and collating them into a whole reduces our understanding of brain and disease functioning...also probably leads to a misunderstanding/misinterpretation of assessment in some cases.*"

2b. Updating links between theory and practice

Participants reported a lot of the tests currently used are outdated and not specifically assessing EF in the way it is conceptualised according to more recent models. As a result, there is a need to bridge the current theory to practice gap:

> P84 *"I find that the theoretical/model based discussions in the literature often don't extend to clinical practice or consider the relationship with specific tests. There is a disconnect between theory and practice. My practice would benefit from more consistent links between the two. (…)"*

Suggestions included devising more comprehensive and updated tests; however, one participant queried the feasibility of this undertaking:

> P50 "*(...) I wonder whether we should start over again and design new tests. But logistically and practically how would this be possible?"*

2c. Ensuring clinicians have appropriate training and guidance

To avoid pitfalls in the assessment and interpretation of results, participants expressed the need for clinicians and trainees to have a good understanding of relevant theoretical frameworks. Two participants suggested creating best practice guidelines for EF assessments:

> P84 *"(…) In the same way that guidance has been developed in relation to best practice in assessing performance validity within neuropsychological tests, an attempt to do this with assessment of executive function would be incredibly helpful."*

3) Stronger psychometric rigour for all populations and impairment levels

Participants called for improved psychometric validity for EF tests, highlighting the need for more comprehensive and inclusive psychometric data; examples included developing better

norms for functional tests, older adults, non-native English speakers, and those with more severe impairments. These cultural considerations in EF test development and standardisation would ensure assessments are accessible to a wide range of individuals.

Discussion

This study explored current practices, challenges and recommendations regarding the clinical assessment of EF in the UK. A sample of 106 CPs with an average of 12 years of post-qualification experience in adult neuropsychology was surveyed. The average participant was female, between the age of 30-49 years, practiced in NHS England in a Clinical or Consultant Psychologist/Neuropsychologist role that required the use of neuropsychology skills at least 50% of the time. The most common work settings included neurorehabilitation, outpatient and acute settings, with ABI being the most common clinical presentation CPs worked with. Over half of the sample completed a post-graduate degree in Clinical or Applied Neuropsychology, and around a third indicated being on the SRCN. Most participants felt between 'somewhat' and 'a lot' competent in the assessment of EF, including in the formulation and feedback of findings.

Summary of findings within the context of the wider literature

The top three most used models of EF were Norman and Shallice's (1980) supervisory attentional system (SAS) model, Baddeley's (1986) working memory model, and Stuss and Benson's tripartite model (1986), along with their updates. This is not consistent with findings of a recent literature review which indicated that Miyake et al.'s (2000) model was the most cited out of all existing EF theoretical frameworks, with Baddeley's model being included in only three papers, and no paper citing the SAS or tripartite models (Baggetta & Alexander, 2016). It is worth noting that 43 of the 106 studies included in the review were reported to have put forward their own model of EF. This finding may point to a disparity between EF models used clinically in the UK and models used by researchers. Notably, the top two selected models (SAS and the working memory), both of which were endorsed by around three quarters of participants, were developed by UK-based researchers. It is possible that these models may also be more amenable to application in clinical practice, including test selection. This was broadly the case according to these findings. The top three selected EF options across both lists of stand-alone tests and batteries subtests were the D-KEFS Verbal Fluency (94.3%) and different versions of the Trail Making tests (D-KEFS = 89.6%; other versions, including Trails A & B = 85.9%). Both verbal fluency and trail making tasks tap

executive attention processes based on the SAS model (Chan et al., 2008). Tests that assess conflict processing and resolution, another aspect of supervisory attentional function, were also selected by many participants, with the D-KEFS Color-Word Interference (CWI) being more commonly used (71.7%) than the other various versions of the Stroop task (56.6%). Interestingly, the Stroop task was reported to be more commonly used in other European countries such as France, Italy, Spain, as well as in Latin America, where it ranked first in the list of top 20 instruments used by neuropsychologists (Arango-Lasprilla et al., 2017; Branco Lopes et al., 2021; Olabarrieta-Landa et al., 2016; Onida et al., 2019). Findings from the USA and Canada survey, which asked a more specific question about EF instruments, differ slightly from the present findings, as the Wisconsin Card Sorting Test (WCST) was ranked as the most used EF test, followed by Trail Making tests, the D-KEFS battery and the Stroop test (including D-KEFS CWI as well as other Stroop versions), which was only selected by 22.6% participants (Rabin et al., 2016). Lower usage rates of the WCST in the UK (12.3%) may be attributed to questions about its predictive validity in clinical populations put forward by UK-based researchers (Burgess et al., 2006), and/or due to widely known issues with its scoring and interpretation (Miles et al., 2021).

Some of the tests that were specifically designed to assess different components of the SAS are also typically used by the sample of CPs surveyed; these include the Hayling Sentence Completion test and the Brixton Spatial Anticipation Test, which ranked 2nd (83%) and 3rd (68.9%) in the most used stand-alone tests list. The Modified Six Element (MSE) test from the Behavioural Assessment of Dysexecutive Syndrome (BADS; Wilson et al., 1996) was also developed based on the SAS model; however, it was reported to be used less frequently (39.6%). A possible explanation may lie in the administration requirements of the MSE, as highlighted by one of the survey's participants; while the Hayling and the Brixton take approximately five and five to ten minutes to administer, respectively, the MSE may take longer as patients are given ten minutes to complete the task, and a few additional minutes are required to set up test materials and provide instructions. Regarding working memory models of EF, tests such as the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) and the WAIS-IV Digit Span were typically used by 66% and 64.2% participants, respectively. The WAIS-IV Letter-Number Sequencing was less commonly used (19.8%). An interesting, and perhaps unexpected, finding was the frequency with which most RBANS subtests were selected. Whilst some tap EF-related processes, including working memory and verbal fluency, other tests are clearly designed to assess separate cognitive domains, such as memory and language. It is possible that many CPs selected all subtests in this battery as it may be frequently included in their assessments as a fixed battery to identify and characterise cognitive deficits, including in EF.

The top three most common sources of information used when assessing EF were the patient's current functioning, observations during interview and assessment, and the client's self-report of difficulties. Informant reports, however, do not appear to be gathered as often (ranking at 7th place). The use of EF questionnaires to collect informant or self-report data (amongst which the DEX was the most used) does not appear to be as frequent, ranking at 12th and 13th place, respectively. This is an interesting finding considering that reduced self-awareness is a common issue in patients with EF deficits. Functional tests are also not as frequently used by CPs in their EF assessments as indicated by their responses. Overall, over half of the sample reported using between three to five EF measures on average in their assessments, followed by around a third who indicated using an average of six to eight.

Discussions in supervision, specialist neuropsychology teaching and feedback from colleagues were the top three most influential factors for the development of CPs' approach to assessing EF. Of note, teaching on the Clinical Psychology doctorate course was the least influential factor of those provided in the list. Interestingly, availability of tests and resources in services was provided as a factor by a few CPs, highlighting possible systemic factors which may influence, and potentially limit, CPs approaches to EF assessments in the UK.

This survey also explored approaches to address possible cross-cultural challenges of EF assessments. It should be acknowledged that responses to this question may not be specific to the assessment of EF, but likely refer to neuropsychological assessments in general. It is also worth noting that only a limited number (≤5) of CPs from either Wales or Northern Ireland took part in this survey; as a result, cross-cultural practices that may be common in these countries, where some communities are likely to be first-language Welsh or Gaelic, may not have been captured. The most endorsed approach by CPs who answered this question was to administer tests that are as culturally fair as possible, for example non-verbal tests. However, this relies on the assumption that clinicians have the necessary knowledge of cultural differences and what may represent a more culturally appropriate assessment (see Franzen et al., 2022 for a summary of key barriers to cross-cultural neuropsychological assessments in Europe). Adopting a more flexible or lenient approach to scoring and score interpretation, and relying more on functional assessments, were also amongst the most selected options. Almost half (45.3%) reported only sometimes using interpreter services for interview and testing. A potential explanation for this may lie in the limited availability of resources, limited time to plan and book an interpreter, and/or known challenges with interpreter-mediated neuropsychological testing, including measurement precision and reliability issues (Casas et al., 2012).

Almost three quarters of participants reported lack of adequate ecological or predictive validity as a key challenge in relation to the selection of EF tests. This finding echoes the results of Rabin et al.'s survey (2016), where this issue was also the most endorsed in a similar question about all neuropsychological instruments (not specific to EF). Other main challenges identified included lack of adequate sensitivity to detect impairments, lack of parallel or alternate forms and tests being culturally biased. Limited ecological validity was also endorsed by 76.4% of CPs as one of the major challenges to interpretation of EF assessment results. Limited or poor psychometric data available was listed as the second most frequent challenge by 54.7% of participants; it is worth noting that fewer CPs (30.2%) endorsed this option as one of their "top three" challenges for test selection. This possibly indicates that these tests might be selected despite limited psychometric validity, and interpretation of scores may be more heavily influenced by qualitative observations and the overall assessment. The third most selected challenge involved limitations in current theoretical knowledge, which was endorsed by approximately half of all participants. This is a well-known pitfall in the field of EF assessments, as previously discussed.

Last but not least, feedback of EF assessments is mainly provided verbally; around 78% of CPs reported feeding back the results of their EF assessments to MDT members most of the time or always. Feedback is also often provided to clients together with the relatives or carers and other professionals involved in the client's care. Feedback to clients alone or relatives/carers alone is less frequently provided.

Suggestions for improvement

Participants' qualitative responses provided a more in-depth insight into this survey's findings. Three key themes were identified using TA. It was evident from CPs' views that a substantial change is needed in the way EF is conceptualised and assessed, by updating current theories and practices. The first and more prevalent theme of improving ecological validity of current assessments was one that resonated with, and expanded upon, quantitative responses in the present survey as well as data from the available literature (Burgess et al., 1998). While most participants reported only sometimes including functional tasks in their practice, they called for an increased standardisation and use of these tasks to improve EF assessments. These findings highlighted potential barriers that may limit the use of these tasks currently, including limited psychometric validity, ease of administration, time and financial constraints. Participants also suggested an increased use of technology to improve EF assessments, with virtual reality (VR) being mentioned by many. Incorporating VR can improve verisimilitude of EF tests with a more immersive, three-dimensional

simulation of real-life situations (Renison, Ponsford, Testa, Richardson, & Brownfield, 2012), although costs and limited psychometric validation may hinder its use and utility in clinical practice. The use of gamification, VR, and wearable technology, such as smartwatches and other digital devices, may also offer promising and more ecologically valid methods to assess EF, especially in younger "digital" generations who have grown up using technology; however, the feasibility, acceptability and validity of these assessment tools should be carefully assessed before they are employed in clinical practice. In line with this, the third theme identified the need for improved psychometric validity and availability of normative data for different populations and impairment types and levels to improve overall ecological validity of assessments. Overall, theme 1 and 3 identified a need for both empirical instruments and functional tasks to be improved and included in assessments. While it has been recognised that clinic-based EF tests can provide useful information particularly to inform neurorehabilitation plans, more naturalistic assessments of functional competence (such as the MET) are needed in conjunction with data from observations and informants for the correct identification of EF deficits (Manchester, Priestley, & Jackson, 2004). Integration and triangulation of information across multiple data sources was indeed recommended by participants to improve ecological validity of the overall assessment, which is in line with recommendations from the literature (Suchy et al., 2017). However, caution should be taken to carefully select the number of EF measures. CPs should balance the need for a comprehensive and ecologically valid assessment without over-testing to reduce the likelihood of detecting impairments simply due to random variation rather than true executive dysfunction (Russell, Russell, & Hill, 2005).

The theme of improved EF knowledge and theory-practice links also resonated with previous literature, highlighting this as one of the current major barriers to EF assessments (Chan et al., 2008). Participants spoke about EF being poorly defined, with concerns about the suitability of currently available models. It is possible that more recently developed models of EF may be too broad or complex, which may hinder their integration in clinical settings. A need for education and training for CPs was also highlighted, with some suggesting the development of best practice guidelines for the assessment of EF. As the Doctoral training in Clinical Psychology was reported to be the least influential factor to the development of CP's approaches to EF assessment, some improvement and standardisation of neuropsychology teaching, particularly on complex areas like EF assessment, may be needed to improve training.

Implications and recommendations for future research

Findings from the current study may be of relevance to both clinicians and researchers to inform test selection and consider current issues in EF assessments. They emphasise the need of further research, with collaborations between clinicians and researchers being of particular value for the development and/or update of EF frameworks. The identification of clinically relevant brain biomarkers of EF may also help inform theoretical work on this topic as well as the development of computerised EF tests, although limited research funding in this area may constrain opportunities to apply these advances in clinical settings. Large normative studies are needed to improve inadequate and outdated norms currently available for some tests; however, efforts of this type and scale are difficult and expensive, with limited support available from funding agencies (Casaletto & Heaton, 2017).

Additional research in this area should include interviews with CPs as well as other professionals that regularly assess EF, such as OTs, to capture their practices and perspectives on this complex topic. The creation of empirical guidelines on EF assessment based on this and future work may prove incredibly helpful. Clear guidelines may aid CPs with more complex assessments, including mental capacity assessments, where detecting EF deficits and their impact on mental capacity can be especially challenging (George & Gilbert, 2018).

Strengths and limitations

The current study is, to our knowledge, the first to comprehensively explore practices and opinions related to EF assessments in a sample of CPs in the UK. Another strength of this study lies in its sample size, which is larger than other existing studies in the UK (Baber, 2020). While other neuropsychology surveys obtained a higher number of responses, it is important to note that most of them only gathered quantitative data on generic assessment practices. This study's exploration of CP's views and opinions on EF practices and challenges through collection of qualitative data has significantly added to the current limited understanding of this complex area of neuropsychological assessment.

Limitations include challenges estimating the sample frame for this study's target population, which means we cannot be certain about the representativeness of the study sample. The online format of the survey may have advantaged CPs who are more comfortable with using technology; on the other hand, CPs who are not part of professional or special interest groups, those who have limited time, and/or are less comfortable with partaking in research may be under-represented. Qualitative responses to the open-ended question were only provided by around half of the total sample, potentially increasing bias. It should also be

noted that the briefer and more constrained nature of free-text box responses can limit participants' ability to describe their views in more detail, and the authors' opportunities to clarify and contextualise responses.

Conclusions

EF is a complex construct which is conceptualised and assessed with significant variance in both research and clinical practice, leading to difficulties in the identification and rehabilitation of EF deficits. In this study, CPs provided information about their current practices in EF assessment; they also described some of the common challenges and potential solutions to address them. Future research would benefit from collaboration between professionals and between academics and clinicians to explore how theoretical and clinical understanding and assessment of EF could be improved and further advanced.

The following recommendations may go some way towards reducing issues pertinent to EF assessments in clinical settings:

- Review and development of theoretical frameworks of EF to improve integration in clinical practice.
- Modernisation and improved psychometric validity, including ecological validity, of EF tests for all populations and impairment levels, which would require considerable investment from funding bodies and test publishers.
- Assessment of functional competence to be used in conjunction with more conventional EF tests.
- Creation of best practice guidelines, with the aim to: 1) improve theory-practice links; 2) enhance CPs' understanding of current models and their limitations; 3) emphasise the use and triangulation of multiple sources of information; 4) reduce variability and misinterpretation of EF assessments.

References

Alvarez, J. A., & Emory, E. (2006). Executive function and the frontal lobes: a meta-analytic review. *Neuropsychology review, 16*(1), 17-42. doi:10.1007/s11065-006-9002-x

Amanzio, M., Bartoli, M., Cipriani, G. E., & Palermo, S. (2020). Executive Dysfunction and Reduced Self-Awareness in Patients With Neurological Disorders. A Mini-Review. *Frontiers in psychology, 11*, 1697-1697. doi:10.3389/fpsyg.2020.01697

Arango-Lasprilla, J. C., Stevens, L., Morlett Paredes, A., Ardila, A., & Rivera, D. (2017). Profession of neuropsychology in Latin America. *Applied neuropsychology. Adult, 24*(4), 318-330. doi:10.1080/23279095.2016.1185423

Baber, Z. (2020). *An exploration of the perspectives of neuropsychologists working with clients from ethnically, culturally and linguistically diverse backgrounds.* (Doctoral dissertation, University of East London).

Baddeley, A. D. (1986). *Working memory* (Vol. 11;11.;). Oxford: Clarendon.

Baggetta, P., & Alexander, P. A. (2016). Conceptualization and Operationalization of Executive Function. *Mind, brain and education, 10*(1), 10-33. doi:10.1111/mbe.12100

Banich, M. T. (2009). Executive Function: The Search for an Integrated Account. *Current directions in psychological science : a journal of the American Psychological Society, 18*(2), 89-94. doi:10.1111/j.1467-8721.2009.01615.x

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved.* (1st ed.): Guilford Press.

Branco Lopes, A., Leal, G., Malvy, L., Wauquiez, G., Ponchel, A., Rivera, D., & Arango-Lasprilla, J. C. (2021). Neuropsychology in France. *Applied neuropsychology. Adult, 28*(3), 328-339. doi:10.1080/23279095.2019.1633329

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. doi:10.1191/1478088706qp063oa

Burgess, P. W., Alderman, N., Evans, J. O. N., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*(6), 547-558. doi:10.1017/S1355617798466037

Burgess, P. W., Alderman, N., Forbes, C., Costello, A., M-A.Coates, L., Dawson, D. R., . . . Channon, S. (2006). The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society, 12*(2), 194-209. doi:10.1017/S1355617706060310

Casaletto, K. B., & Heaton, R. K. (2017). Neuropsychological Assessment: Past and Future. *Journal of the International Neuropsychological Society, 23*(9-10), 778-790. doi:10.1017/S1355617717001060

Casas, R., Guzmán-Vélez, E., Cardona-Rodriguez, J., Rodriguez, N., Quiñones, G., Izaguirre, B., & Tranel, D. (2012). Interpreter-Mediated Neuropsychological Testing of Monolingual Spanish Speakers. *Clinical neuropsychologist, 26*(1), 88-101. doi:10.1080/13854046.2011.640641

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology, 23*(2), 201-216. doi:10.1016/j.acn.2007.08.010

Diamond, A. (2013). Executive Functions. *Annual review of psychology, 64*(1), 135-168. doi:10.1146/annurev-psych-113011-143750

Elliott, R. (2003). Executive functions and their disorders: Imaging in clinical neuroscience. *British medical bulletin, 65*(1), 49-59.

Franzen, S., Watermeyer, T. J., Pomati, S., Papma, J. M., Nielsen, T. R., Narme, P., . . . European Consortium on Cross-Cultural, N. (2022). Cross-cultural neuropsychological assessment in Europe: Position statement of the European Consortium on Cross-Cultural Neuropsychology (ECCroN). *Clinical neuropsychologist, 36*(3), 546-557. doi:10.1080/13854046.2021.1981456

George, M., & Gilbert, S. (2018). Mental Capacity Act (2005) assessments: Why everyone needs to know about the frontal lobe paradox. *The neuropsychologist, 5*(1), 59-66.
Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: a review of our current understanding. *Neuropsychology review, 17*(3), 213-233. doi:10.1007/s11065-007-9040-z
Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: coming out of the office. *Brain injury, 18*(11), 1067-1081. doi:10.1080/02699050410001672387
Miles, S., Howlett, C. A., Berryman, C., Nedeljkovic, M., Moseley, G. L., & Phillipou, A. (2021). Considerations for using the Wisconsin Card Sorting Test to assess cognitive flexibility. *Behavior Research Methods, 53*(5), 2083-2091. doi:10.3758/s13428-021-01551-3
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.
Norman, D. A., & Shallice, T. (1986). Attention to Action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and Self-Regulation: Advances in Research and Theory Volume 4* (pp. 1-18). Boston, MA: Springer US.
Olabarrieta-Landa, L., Caracuel, A., Pérez-García, M., Panyavin, I., Morlett-Paredes, A., & Arango-Lasprilla, J. C. (2016). The profession of neuropsychology in Spain: results of a national survey. *Clinical neuropsychologist, 30*(8), 1335-1355. doi:10.1080/13854046.2016.1183049
Onida, A., Di Vita, A., Bianchini, F., Rivera, D., Morlett-Paredes, A., Guariglia, C., & Arango-Lasprilla, J. C. (2019). Neuropsychology as a profession in Italy. *Applied neuropsychology. Adult, 26*(6), 543-557. doi:10.1080/23279095.2018.1466782
Packwood, S., Hodgetts, H. M., & Tremblay, S. (2011). A multiperspective approach to the conceptualization of executive functions. *Journal of clinical and experimental neuropsychology, 33*(4), 456-470. doi:10.1080/13803395.2010.533157
Pribram, K. H. (1973). The primate frontal cortex–executive of the brain. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes.* (pp. 293-314): Elsevier.
Rabin, L. A., Paolillo, E., & Barr, W. B. (2016). Stability in Test-Usage Practices of Clinical Neuropsychologists in the United States and Canada Over a 10-Year Period: A Follow-Up Survey of INS and NAN Members. *Archives of clinical neuropsychology, 31*(3), 206-230. doi:10.1093/arclin/acw007
Renison, B., Ponsford, J., Testa, R., Richardson, B., & Brownfield, K. (2012). The ecological and construct validity of a newly developed measure of executive function: the virtual library task. *Journal of the International Neuropsychological Society, 18*(3), 440-450.
Russell, E. W., Russell, S. L., & Hill, B. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of clinical neuropsychology, 20*(6), 785-794.
Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain, 114*(2), 727-741. doi:10.1093/brain/114.2.727
Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological research, 63*(3-4), 289-298. doi:10.1007/s004269900007
Stuss, D. T., & Benson, D. F. (1986). The frontal lobes. New York: Raven Press.
Suchy, Y., Ziemnik, R. E., & Niermeyer, M. A. (2017). Assessment of executive functions in clinical settings. In E. Goldberg (Ed.), *Executive functions in health and disease* (pp. 551-569): Elsevier.
Terry, G., Hayfield, N., Clarke, V., & Braun, V. (2017). Thematic analysis. In C. Willig & W. Stainton Rogers (Eds.), *The Sage handbook of qualitative research in psychology.* (2nd ed., pp. 17-36). Thousand Oaks, CA: SAGE Inc.

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural Assessment of the Dysexecutive Syndrome.* St. Edmunds, England: Thames Valley Test Company

Appendix 1.1. PRISMA Checklist

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	6
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	7
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	9
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	9-10
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	10-11
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	11
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	11
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	11-12
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	12
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	12
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	12
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	12-13
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	12
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	13
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	12-13
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	13
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	13
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	13
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	N/A
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	12-13
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	12-13

PRISMA 2020 Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	14
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	14
Study characteristics	17	Cite each included study and present its characteristics.	14-15
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	15
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	15
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	15, 36-41
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	36-41
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	40-41
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	40-41
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	41-44
	23b	Discuss any limitations of the evidence included in the review.	44
	23c	Discuss any limitations of the review processes used.	44-45
	23d	Discuss implications of the results for practice, policy, and future research.	45-46
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	10
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	10
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	10
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	N/A
Competing interests	26	Declare any competing interests of review authors.	N/A
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	N/A

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71
For more information, visit: http://www.prisma-statement.org/

Appendix 1.2 Search Strategies

MEDLINE(R) ALL

1 (("behavio* assessment" adj3 "dysexecutive syndrome") or BADS).mp.

2 DEX.mp.

3 "dysexecutive questionnaire".mp.

4 1 and 2

5 1 or 3

6 4 or 5

7 limit 6 to yr="1996 -Current"

8 limit 7 to english language

Embase

1 (("behavio* assessment" adj3 "dysexecutive syndrome") or BADS).mp.

2 DEX.mp.

3 "dysexecutive questionnaire".mp.

4 1 and 2

5 1 or 3

6 4 or 5

7 limit 6 to yr="1996 -Current"

8 limit 7 to english language

PsycINFO

S7 S4 OR S5 Limiters - Publication Year: 1996-2023; English

S6 S4 OR S5 Expanders - Apply equivalent subjects

S5 S1 OR S3 Expanders - Apply equivalent subjects

S4 S1 AND S2 Expanders - Apply equivalent subjects

S3 TX "dysexecutive questionnaire" Expanders - Apply equivalent subjects

S2 TX "DEX" Expanders - Apply equivalent subjects

S1 TX ((("behavio* assessment") N2 ("dysexecutive syndrome")) or "BADS") Expanders - Apply equivalent subjects

CINAHL

S7 S4 OR S5 Limiters - English Language

S6 S4 OR S5 Expanders - Apply equivalent subjects

S5 S1 OR S3 Expanders - Apply equivalent subjects

S4 S1 AND S2 Expanders - Apply equivalent subjects

S3 TX "dysexecutive questionnaire" Expanders - Apply equivalent subjects

S2 TX "DEX" Expanders - Apply equivalent subjects

S1 TX ((("behavio* assessment") N2 ("dysexecutive syndrome")) or "BADS") Expanders - Apply equivalent subjects

Web of Science – Core Collection

1: TS=((((("behavio* assessment") near/2 ("dysexecutive syndrome")) or "BADS"))

2: TS=("DEX")

3: TS=("dysexecutive questionnaire")

4: #1 AND #2

5: #1 OR #3

6: #4 OR #5

7: #4 OR #5 Timespan: 1996-01-01 to 2023-12-31

Appendix 1.3 Screening Tool

https://osf.io/6z3rb

Appendix 1.4 AXIS Tool Quality Ratings

Question	Boelen et al. (2009)	Burgess et al. (1998)	Cerezo García et al. (2015)	Chan & Manly (2002)	Channon & Crawford (1999)	Emmanouel et al. (2014)	Evans et al. (1997)	Fukuta & Mori (2018)	Grech et al. (2016)	Knight et al. (2002)	Koerts et al. (2012)	Norris & Tate (2000)	Preston et al. (2013)	Weddell & Wood (2016)	Wilson et al. (1996)
INTRODUCTION															
1. Were the aims/objectives of the study clear?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
METHODS															
2. Was the study design appropriate for the stated aim(s)?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
3. Was the sample size justified?	N	N	N	N	N	N	N	N	N	N	N	N	Y	Y	N
4. Was the target/reference population clearly defined? (Is it clear who the research was about?)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
5. Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	N	N	DK	N	DK	N	DK	DK	Y	N	N	N	Y	N	N
6. Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
7. Were measures undertaken to address and categorise non-responders?	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N
8. Were the risk factor and **outcome variables** measured appropriate to the aims of the study?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
9. Were the risk factor and **outcome variables** measured correctly using instruments/measurements that had been trialled, piloted or published previously?	Y	Y	DK	Y	Y	N	Y	DK	Y	Y	DK	Y	Y	Y	Y
10. Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. p-values, confidence intervals)	N*	N*	Y	N*	N*	N*	N*	Y	Y	N*	Y	N*	Y	Y	N*
11. Were the methods (including statistical methods) sufficiently described to enable them to be repeated?	Y	N	Y	N	N	Y	N	Y	Y	N	Y	Y	Y	Y	N
RESULTS															
12. Were the basic data adequately described?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
13. Does the response rate raise concerns about non-response bias? (reversed item)	DK	DK	DK	DK	DK	DK	DK	DK	DK	DK	DK	DK	DK	DK	DK
14. If appropriate, was information about non-responders described?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N	N	N/A
15. Were the results internally consistent?	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
16. Were the results presented for all the analyses described in the methods?	Y	N	Y	N	Y	Y	DK	Y	Y	N	Y	Y	Y	Y	DK
DISCUSSION															
17. Were the authors' discussions and conclusions justified by the results?	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
18. Were the limitations of the study discussed?	Y	N	N	Y	Y	N	N	Y	Y	Y	Y	N	Y	Y	N
OTHER															
19. Were there any funding sources or conflicts of interest that may affect the authors' interpretation of the results? (reversed item)	N	Y	N	N	N	N	DK	DK	DK	Y	DK	DK	N	N	Y
20. Was ethical approval or consent of participants attained?	DK	DK	Y	Y	Y	DK	DK	Y	Y	DK	Y	DK	Y	Y	DK

Note. * indicates that the p-value threshold for analyses was provided in the Results section. Quality of reporting domain (items 1, 4, 10, 11, 12, 16 and 18); Study design quality domain (items 2, 3, 5, 8, 17, 19 and 20); Introduction of bias domain (items 6, 7, 9, 13, 14 and 15).

Appendix 2.1 Online Survey

https://osf.io/z9jqd

Appendix 2.2 Study Advert

https://osf.io/snpuk

Appendix 2.3 Information Sheet

https://osf.io/37z95

Appendix 2.4 Consent Form

https://osf.io/2vuks

Appendix 2.5 Ethics Approval Letter

University of Glasgow

Dr Jessica Fish

MVLS College Ethics Committee

An investigation of current psychology practice in the clinical assessment of executive functions 200220005

The College Ethics Committee has reviewed your application and has agreed that there is no objection on ethical grounds to the proposed study. We are happy therefore to approve the project, subject to the following conditions

- Project end date as stipulated in original application.
- The data should be held securely for a period of ten years after the completion of the research project, or for longer if specified by the research funder or sponsor, in accordance with the University's Code of Good Practice in Research: (http://www.gla.ac.uk/media/media_227599_en.pdf)
- The research should be carried out only on the sites, and/or groups defined in the application.
- Any proposed changes in the protocol should be submitted for reassessment, except when it is necessary to change the protocol to eliminate hazard to the subjects or where the change involves only the administrative aspects of the project. The Ethics Committee should be informed of any such changes.
- For projects requiring the use of an online questionnaire, the University has an Online Surveys account for research. To request access, see the University's application procedure at https://www.gla.ac.uk/research/strategy/ourpolicies/useofonlinesurveystoolforresearch/.
- You should submit a short end of study report within 3 months of completion.

Yours sincerely

[redacted]

Dr Terry Quinn

Terry Quinn
FWSO, FESO, MD, FRCP, BSc (hons), MBChB (hons)
Reader / Honorary Consultant

College of Medicine, Veterinary & Life Sciences
School of Cardiovascular and Metabolic Health
[redacted]

The University of Glasgow, charity number SC004401

Appendix 2.6 Demographic Information

2.6.i

Participants' work settings.

Work setting	Frequency	Percentage
Community/Outpatient Adult Neurorehabilitation	47	44.3%
Neuropsychology - Outpatients	43	40.6%
Neuropsychology - Acute	38	35.8%
Inpatient Adult Neurorehabilitation	35	33.0%
Memory Clinic or Dementia services	18	17.0%
Community Older Adult Mental Health	8	7.5%
Clinical Health	7	6.6%
Inpatient Older Adult Mental Health	*	*
Community Adult Mental Health	*	*
Adult Forensic	*	*
Adult Learning Disability	-	-
Inpatient Adult Mental Health	-	-
Other	6	5.7%

Note. Multiple response options were available for this question.

2.6.ii

Clinical presentation(s) participants work with most frequently.

Clinical presentation(s)	Frequency	Percentage
Acquired brain injury	82	77.4%
Traumatic brain injury – moderate and severe	73	68.9%
Stroke	67	63.2%
Traumatic brain injury – mild	60	56.6%
Other neurological and neurosurgical presentations	57	53.8%
Functional neurological disorders	54	50.9%
Multiple sclerosis	52	49.1%
Mild cognitive impairment and dementias	51	48.1%
Epilepsy	42	39.6%
Movement disorders	30	28.3%
Clinical health conditions	30	28.3%
Psychiatric disorders	24	22.6%
Addictions	7	6.6%
Autism Spectrum Disorders (ASDs)	6	5.7%
ADHD	*	*
Learning disabilities	*	*
Other	*	*

Note. Multiple response options were available for this question.

Appendix 2.7 Sources of information to assess EF – Other

EF information source	Sometimes	About half the time	Most of the time	Always
MET and functional assessments (e.g., OT tasks, JEF, ADLs, etc.)	3.8%	1.9%	1.9%	1.9%
Observations outside formal assessment (e.g., on the ward, in group/social settings, ADLs)	0.9%	0.9%	-	1.9%
Other psychometric measures	-	-	0.9%	-
Behaviour charts	0.9%	-	-	-
Diary sheets	0.9%	-	-	-
Witness statements	-	-	0.9%	-
Capacity / risk assessments	-	-	0.9%	-
Support workers / homecare	-	-	-	0.9%

Appendix 2.8 Most typically used stand-alone tests of EF

Stand-alone EF test	Frequency	Percentage
Trail-Making Test (TMT) or Trails A & B	91	85.9%
Hayling Sentence Completion Test	88	83.0%
Brixton Spatial Anticipation Test	73	68.9%
Stroop Task	60	56.6%
Clock Drawing Test	53	50.0%
Controlled Oral Word Association Test (COWAT)	37	34.9%
Rey-Osterrieth Complex Figure Test (ROCFT)	33	31.1%
Tower of London/Tower of Hanoi	31	29.3%
Symbol Digit Modalities Test (SDMT)	26	24.5%
Go/No Go Test	25	23.6%
Design Fluency Test	20	18.9%
Category Test	16	15.1%
Wisconsin Card Sorting Test (WCST)	13	12.3%
Weigl Colour-Form Sort Test (WCFS)	11	10.4%
Cognitive Estimation Test (CET)	5	4.7%
Continuous Performance Task (CPT)	3	2.8%
Concept Generation Test (CGT)	2	1.9%
Iowa Gambling Task	2	1.9%
Measure of Everyday Planning (MEP)	2	1.9%
Halstead Category Test or Booklet Category Test	1	0.9%
Jansari assessment of Executive Functions (JEF)	1	0.9%
Cambridge Executive Functioning Assessment – Executive Functions (CEFA-EF)	-	-
Contingency Naming Test (CNT)	-	-
Executive Interview (EXIT-25)	-	-
Functional Assessment of Verbal Reasoning and Executive Strategies (FAVRES)	-	-
Five-Point Test	-	-
Paced Auditory Serial Addition Test (PASAT)	-	-
Ruff Figural Fluency Test (RFFT)	-	-
Self-Ordered Pointing Test (SOPT)	-	-

Stand-alone EF test reported by participants under "Other" responses.

Other - stand-alone test	Frequency	Percentage
Colour Trails	1	0.9%
Institute of Cognitive Neurology (INECO) Frontal Screening (IFS)	1	0.9%
FRONTIER Executive Screen (FES)	1	0.9%
Backward Digit Span	1	0.9%

Appendix 2.9 Most typically used battery subtests to assess EF

Behavioural Assessment of Dysexecutive Syndrome (BADS)

Subtest	Frequency	Percentage
Zoo Map	82	77.4%
Key Search	74	69.8%
Modified Six Elements	42	39.6%
Rule Shift Cards	30	28.3%
Action Program	24	22.6%
Temporal Judgment	12	11.3%

Note. 88 respondents selected at least one BADS subtest.

Cambridge Neuropsychological Automated Battery (CANTAB)

Subtest	Frequency	Percentage
Motor Screening	1	0.9%
Spatial Recognition Memory	1	0.9%
Spatial Span	1	0.9%
Spatial Working Memory	1	0.9%
Big Little Circle	1	0.9%
Intra-Extra Dimensional Shift	1	0.9%
Reaction Time	1	0.9%
Stockings of Cambridge	1	0.9%
Matching to Sample	-	-
Delayed Matching to Sample	-	-
Pattern Recognition Memory	-	-
Paired Associate Learning	-	-
Rapid Visual Information Processing	-	-

Note. 3 respondents selected at least one CANTAB subtest.

Delis-Kaplan Executive Function System (D-KEFS)

Subtest	Frequency	Percentage
Verbal Fluency	100	94.3%
Trail Making	95	89.6%
Color-Word Interference	76	71.7%
Tower	45	42.5%
Design Fluency	32	30.2%
Sorting	27	25.5%
Twenty Questions	22	20.8%
Proverb	13	12.3%
Word Context	6	5.7%

Note. 102 participants selected at least one D-KEFS subtest.

Frontal Assessment Battery (FAB)

Subtest	Frequency	Percentage
Similarities	28	26.4%
Go-No Go	27	25.5%
Motor Series (“Luria”)	26	24.5%
Lexical Fluency	19	17.9%
Conflicting Instructions	17	16.0%
Prehension Behaviour	15	14.2%

Note. 35 respondents selected at least one FAB subtest.

Kaplan Baycrest Neurocognitive Assessment (KBNA)

Subtest	Frequency	Percentage
Practical Problem Solving	10	9.4%
Orientation	7	6.6%
Word Lists 1 (immediate recall)	6	5.7%
Picture Naming	6	5.7%
Verbal Fluency	6	5.7%
Praxis	6	5.7%
Sequences	5	4.7%
Complex Figure 1 (copy)	5	4.7%
Word Lists 2 (free and cued recall and recognition)	5	4.7%
Clocks	4	3.8%
Complex Figure 2 (recall and recognition)	4	3.8%
Spatial Location	4	3.8%
Conceptual Shifting	4	3.8%
Numbers	3	2.8%
Motor Programming	3	2.8%
Reading Single Words	3	2.8%
Picture Recognition	3	2.8%
Expression of Emotion	3	2.8%
Picture Description (Oral)	3	2.8%
Symbol Cancellation	2	1.9%
Sentence Reading – Arithmetic	2	1.9%
Auditory Comprehension	2	1.9%
Repetition	2	1.9%
Picture Description (Written)	2	1.9%
Auditory Signal Detection	1	0.9%

Note. 18 respondents selected at least one KBNA subtest.

Neuropsychological Assessment Battery (NAB) – Executive and Daily Living Tests

Subtest	Frequency	Percentage
Mazes	10	9.4%
Categories	7	6.6%
Judgment	6	5.7%
Daily Living Memory	4	3.8%
Map Reading	4	3.8%
Word Generation	3	2.8%
Driving Scenes	3	2.8%
Bill Payment	2	1.9%

Note. 12 respondents selected at least one NAB subtest.

Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)

Subtest	Frequency	Percentage
Semantic Fluency	73	68.9%
Digit Span	70	66.0%
Coding	69	65.1%
Figure Copy	68	64.2%
List Learning	65	61.3%
Story Memory - Immediate Recall	65	61.3%
List Recognition	65	61.3%
List Recall	64	60.4%
Figure Recall	64	60.4%
Line Orientation	63	59.4%
Picture Naming	63	59.4%
Story Memory - Delayed Recall	63	59.4%

Note. 78 respondents selected at least one RBANS subtest.

Wechsler Adult Intelligence Scale 4th Edition (WAIS-IV)

Subtest	Frequency	Percentage
Similarities	78	73.6%
Matrix Reasoning	73	68.9%
Digit Span	68	64.2%
Block Design	63	59.4%
Symbol Search	51	48.1%
Coding	51	48.1%
Vocabulary	49	46.2%
Arithmetic	44	41.5%
Information	34	32.1%
Visual Puzzles	29	27.4%
Letter-Number Sequencing	21	19.8%
Comprehension	13	12.3%
Cancellation	9	8.5%
Figure Weights	5	4.7%
Picture Completion	5	4.7%

Note. 91 respondents selected at least one WAIS-IV subtest.

Battery subtests reported by participants under "Other" (n=13).

Battery subtest	Frequency	Percentage
WMS-IV Symbol Span	3	2.8%
BMIPB-II Speed of information processing task	2	1.9%
WASI	1	0.9%
TOPF	1	0.9%
Short Parallel Assessments of Neuropsychological Status (SPANS)	1	0.9%
Test of Everyday Attention (TEA) - Map Search, Elevator Counting, Elevator counting with distraction, Telephone search, Telephone search dual task	1	0.9%
BMIPB-II Figure Copy	1	0.9%
CVLT-II	1	0.9%
Single word cues for dynamic aphasia / poor initiation in FTD	1	0.9%
Oxford cognitive Screen (OCS) Task switching	1	0.9%
Cognitive Linguistic Quick Test (CLQT)	1	0.9%
Neuropsychological Assessment Battery screening module (NAB-S)	1	0.9%

Appendix 2.10 Functional tasks used – Other

Functional tasks used to assess EF as reported by participants under "Other" (n=62)

Code	Frequency	Percentage
Functional tasks	36	34.0%
METs (including OxMET, JEF, etc.)	30	28.3%
Observations	19	17.9%
Reports from colleagues / MDT	5	4.7%
Cognitive tests	4	3.8%
Reports from client and relative or carer	3	2.8%
Playing games	1	0.9%
Mental capacity assessments	1	0.9%

Appendix 2.11 Most frequently used EF questionnaires or rating scales

Questionnaire or behaviour rating scale	Frequency	Percentage
Dysexecutive Questionnaire (DEX)	80	75.5%
Frontal Systems Behavior Scale (FrSBe)	28	26.4%
Behavior Rating Inventory of Executive Function-Adult (BRIEF-A)	22	20.8%
I don't use any questionnaires and/or behaviour rating scales to assess executive functions	11	10.4%
Cognitive Failures Questionnaire (CFQ)	10	9.4%
Neuropsychiatric Inventory (NPI)	8	7.5%
Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE)	5	4.7%
Frontal Behavior Inventory (FBI)	4	3.8%
Behavioral Dyscontrol Scale (BDS)	2	1.9%
Iowa Rating Scales of Personality Change (IRSPC)	1	0.9%
Barkley Deficits in Executive Functioning Scales (BDEFS)	-	-
Comprehensive Assessment for Dementia in People with Down Syndrome and Others with Intellectual Disabilities (CAMDEX-DS) – Informant Interview	-	-
Comprehensive Executive Function Inventory (CEFI)	-	-
Frontal Lobe Personality Scale (FLOPS)	-	-
Problem Solving Inventory (PSI)	-	-

EF questionnaires or behaviour rating scales reported by participants under "Other"

Questionnaire or behaviour rating scale	Frequency	Percentage
Edinburgh Cognitive and Behavioural ALS Screen (ECAS) Behaviour Screen – Carer Interview	3	2.8%
Swansea Neurobehavioural Outcome Scale (SASNOS)	2	1.9%
Adult executive functioning inventory (ADEXI)	1	0.9%
Everyday Memory Questionnaire - Revised (EMQR)	1	0.9%
Minnesota Multiphasic Personality Inventory—2—Restructured Form (MMPI–2–RF)	1	0.9%
Lille Apathy Rating Scale (LARS)	1	0.9%
BADS	1	0.9%

Appendix 2.12 Important factors to the development of current approach to EF assessment – Other

	Very important		Extremely important	
Code	**Frequency**	**Percentage**	**Frequency**	**Percentage**
Clinical experience (including acute, outpatient, rehab and diagnostic settings)	7	6.6%	3	2.8%
Availability of tests and/or resources in the service	2	1.9%	7	6.6%
Experience of test use	-	-	3	2.8%
Learning / feedback from colleagues through supervision, MDT etc.	3	2.8%	-	-
Normative data (availability, quality)	2	1.9%	1	0.9%
Academic degrees, involvement in vivas for QiCN/MSc and conferences	3	2.8%	-	-
Feedback from clients and family	2	1.9%	-	-
Cognitive / EF models	2	1.9%	-	-
Medicolegal case review	2	1.9%	-	-
Time limits	1	0.9%	1	0.9%
Awareness of which tests are used in similar services to allow for repeat assessment / what tests referrers may be familiar with	-	-	2	1.9%
Observations	1	0.9%	-	-
Reflection on own clinical practice	1	0.9%	-	-

Appendix 2.13 Cross-cultural considerations for EF assessments – Other

Code	Never (n)	Sometimes (n)	About half the time (n)	Most of the time (n)	Always (n)
Gain an understanding of the client's culture and social norms		1	1	1	4
Rely on self/informant reports more than test data				2	1
Ask for input/help from other *disciplines* that are fluent in the client's language (this is because this may be accessible, whilst NPs would not be)		1		1	
Consider the level of impairment within cultural factors					1
Carefully assess educational history					1
Formulation					1
Try to borrow tests from other services due to limited availability of culturally sensitive tests in my service.					1
Suggest a repeat assessment to monitor change (e.g., dementia assessments)			1		
Use of models such as RACE-L (Boakye & Mwale)				1	
Have a few short verbal and written words and phrases		1			
Call the DoN or BPS for advice			1		
Include a clear statement in my notes and report					1
Always check the client has understood the task before completing					1
Supervision					1
Seek new tests e.g., Colour Trails Test*					
Not needed often but currently reviewing why in terms of barriers to referrals*					
Discuss what the cultural specific issues are*					
Same interpreters are not available for multiple sessions*					

Note. * indicates that no frequency rating was provided for this item.

Appendix 2.14 Challenges with the selection of EF measures – Other

Code	Frequency	Percentage
Tests are not available in the service / limited budget	3	2.8%
Tests lack close correspondence with theoretical models	1	0.9%
Tests rely on English language or English language features	1	0.9%
Tests can be limited if the client has motor or language difficulties	1	0.9%
Subtests from commonly used batteries (e.g., BADS) aren't designed to be used independently and 'profile scores' from this are only of qual utility	1	0.9%
Tests are not age or IQ normed	1	0.9%
Tests only assess a few aspects of EF, tests necessarily rely on too many other cognitive skills	1	0.9%
Difficulty translating scores on tests into meaningful information about what they person struggles with and useful rehabilitation goals.	1	0.9%
Tests, although necessary, should not supplant good collateral information on day-to-day planning, problems solving and organisation skills.	1	0.9%

Appendix 2.15 Challenges with the formulation of EF assessments – Other

Code	Frequency	Percentage
Limited or poor construct/criterion validity	1	0.9%
Lack of validity in general	1	0.9%
Issues associated with the "frontal lobe paradox"	1	0.9%
Difficulties in estimating pre-morbid ability	1	0.9%
Difficulties separating the role of motivation/personality/values from more purely neurological deficit	1	0.9%
Difficulties assessing change from premorbid ability (no clear consensus regarding correlations between different EFs and IQ)	1	0.9%
Test scoring for the Hayling and Brixton needs improving	1	0.9%
Difficulties interpreting functional data reliably	1	0.9%
Critical areas missing from assessment, mainly social cognition and theory of mind.	1	0.9%
Impact of co-morbidity factors	1	0.9%
Interplay of EF skills with other cognitive domains	1	0.9%
I don't feel limited	1	0.9%

Appendix 2.16 Summary of themes and participants' quotes

Theme	Subordinate theme	Participant's quote
1. A call for improved ecological validity	**1a. Standardisation and inclusion of functional assessments**	P6 "Tests which have a functional element incorporated"
		P24 "Further developments of standardised functional assessments such as hospital or home versions of the Multiple Errands test."
		P28 "We need efficient tasks with good norms and ecological validity, that can be carried out easily in clinical settings - not an easy task!"
		P39 "More assessments based on real life situations and function making the assessment outcome more 'meaningful'."
		P47 "More ecologically valid (MET-type) activities that could be used in the clinic room"
		P48 "We need better tests, I'd like to see ecologically valid tests for each aspect of EF so you can break down where someone is struggling and develop detailed rehab plans."
		P51 "I think more standardisation of functional tasks would be useful."
		P54 "I've always liked some of the more behavioural tasks from the BADS (zoo / 6 elements) but slow to administer."
		P63 "I think more practical and everyday assessments are important, i.e. demonstrate how you would use this object, explain / show me how you would make a cup of tea. I know there are tests out there that cover this, including occupational therapy assessments, but more everyday assessments in real life settings would be more helpful."
		P73 "Standardised scoring, more functional assessments that can be completed in clinic settings"
		P76 "Better norms for more functional tests"

		P87 “More development of functional assessments and less reliance and verbal and pen-and-paper assessment - or preferably a battery that includes both”
		P92 “More assessment involving daily living tasks”
		P106 “new batteries assessing more real-life functions e.g. ability to switch between functional tasks”
	1b. Modernising EF tests	P1 “I think we need more functional assessments and computer based tests. E.g. using zoo map test with a 18 year old who is rarely using pen and paper”
		P4 “More modern and digital testing. VR assessments (akin to OCS virtual MET)”
		P13 “A good computer version of the Iowa task. (…) Computer / VR version of the marketplace test.”
		P64 “Develop digital versions and versions using virtual reality”
		P67 “I think advances in technology with computerised testing are promising (e.g., incorporating item response theory and the Boston process approach) but probably a long way off application in clinical practice”
		P94 “We need to utilise digital platforms and ideas/concepts from virtual and gaming domains”
		P104 “Use of VR and making this affordable”
		P106 “perhaps the use of virtual reality/ technology would make assessments somewhat more ecologically valid”
	1c. Integrating multiple data sources to enhance ecological validity of the overall assessment and its findings	P7 “Clinical interview with client and families is also under-rated”
		P18 “Always try to assess day to day functional competence in addition to using formal interviews, rating scales, and cognitive testing”

		P33 “Test selection should be guided by the clinical interview and observations as they will often tell you what the nature of the problem is”
		P38 “I think more test batteries should emphasise the role of questionnaires and structured interviews, and make clear how these link to the 'tests' included within the battery, so that it encourages clinicians (particularly those less experienced) to triangulate between data sources when considering executive difficulties.”
		P91 “Always use a combination of formal tests, questionnaires, observations, observations of significant others, and medical data”
2. A need for improved EF knowledge and theory-to-practice links	**2a. Improving the theoretical understanding of EF**	P2 “Reconsider the concept of EF. It's a broad term for many systems/functions and collating them into a whole reduces our understanding of brain and disease functioning...also probably leads to a misunderstanding/misinterpretation of assessment in some cases. Improving researcher/test developer and clinician understanding of criterion and construct validity of tests would also be helpful!”
		P95 “If we could agree on what executive function was, that would be a start.”
		P87 “Greater understanding of executive functioning non-neuro populations, and non-psychiatric populations - how "normal" are different types of difficulties”
	2b. Updating links between theory and practice	P32 “-Many of the assessments are out-of-date, very costly and lacking in ecological validity. -Increased link to up-to-date cognitive models/theory”
		P39 “More research using factor analysis to identify more options of as close to ‘pure’ testing of executive function as possible.”
		P48 “I’d also like to see better social cognition and frontal paradox measures, what we have is outdated and not related to current understanding of emotions.”
		P50 “Tests which have been designed to identify frontal lobe damage are used to assess executive functions. I wonder whether we should start over again and design new tests. But logistically and practically how would this be possible?”

		P84 “I find that the theoretical/model based discussions in the literature often don’t extend to clinical practice or consider the relationship with specific tests. There is a disconnect between theory and practice. My practice would benefit from more consistent links between the two.”
	2c. Ensuring clinicians have appropriate training and guidance	P7 “This is a huge question! I think a good theoretical understanding of EF is required to know which tests tap into which areas of EF in order to get a good assessment.”
		P30 “Ensuring trainees have good supervisors who understand executive functioning. In my view many qualified CP's don't but feel the have the competencies to assess it.”
		P33 “It’s an area in the assessment that seems to create anxiety for juniors in terms of administration and interpretation”
		P71 “Collaborative development and work by neuropsychology with OTs (and other professionals), families and clients to create a ‘best practice guide’ or clinical consensus of how best to assess exec function. Perhaps to include considerations at different stages of the rehab pathway, being mindful of limitations of assessments in certain environment - e.g ward environment often reducing the opportunity for exec function difficulties to be highlighted. Also to include some ‘best practice’ examples and examples of how to collate existing assessment tools based on different theoretical models. This could also be linked to assessment of capacity, insight and awareness …..”
		P84 “In the same way that guidance has been developed in relation to best practice in assessing performance validity within neuropsychological tests, an attempt to do this with assessment of executive function would be incredibly helpful.”
3. Stronger psychometric rigour for all tests, populations and impairment levels		P13 “Better norms for the BADS”

		P25 “Yes! please develop exec tests which are IQ normed, Age normed and easy to interpret. Anything available in other languages would be great (particularly for informants/family to complete as well as the client).”
		P26 “More challenging normed functional tasks”
		P53 “Having some more options that are less reliant on visual skills”
		P54 “It's always going to be tricky to marry formal psychometrics to the unstructured world we live in.”
		P58 “Having more options for people with aphasia, particularly receptive aphasia, would be helpful. Also more options for people with visual difficulties.”
		P74 “Better norms in older adult samples e.g. trails A+B”
		P77 “More standardisation across teams. Simpler tests for those with more severe impairments”
		P97 “norming on wider population”
		P100 “It would be helpful to have more culturally relevant normative data”
		P103 “Better norms, particularly for non-native English speakers, and for impairment types.”

Appendix 2.17 MRP Proposal

https://osf.io/5xqhg

Appendix 2.18 APA Style JARS Mixed Methods Research (JARS–Mixed) Reporting Guidelines

Based on Journal Article Reporting Standards - Mixed Methods Design - Table 1 (apa.org)
See also Journal Article Reporting Standards - Quantitative Design - Table 1 (apa.org) and Journal Article Reporting Standards - Qualitative Design - Table 1 (apa.org)

JARS-Mixed		
1) Title Page		
a) Title	Refrain from using words that are either qualitative (e.g., "explore," "understand") or quantitative (e.g., "determinants," "correlates"), because mixed methods stands in the middle between qualitative and quantitative research.	p. 50
	Reference the mixed methods, qualitative methods, and quantitative methods used	N/A
2) Abstract		
	Specify the type of mixed methods design used. See the note on types of designs in the Research Design Overview section of this table.	p. 53
	Consider using one keyword that describes the type of mixed methods design and one that describes the problem addressed.	p. 53
	Describe your approach(es) to inquiry and, if relevant, how intersecting approaches to inquiry are combined when this description will facilitate the review process and intelligibility of your paper. If your work is not grounded in a specific approach(es) to inquiry or your approach would be too complicated to explain in the allotted word count, however, it would not be advisable to provide explication on this point in the abstract	p. 53
3) Introduction		
a) Description of Research Problems/Questions	This section may convey barriers in the literature that suggest a need for both qualitative and quantitative data.	p. 54-55
b) Study Objectives/Aims/Research Goals	State three types of research objectives/aims/goals: qualitative, quantitative, and mixed methods. Order these goals to reflect the type of mixed methods design used	p. 56
	Describe the ways approaches to inquiry were combined, as it illuminates the objectives and mixed methods rationale (e.g., descriptive, interpretive, feminist, psychoanalytic, postpositivist, critical, postmodern, constructivist, or pragmatic approaches).	p. 56

4)	**Method**		
a)	Research Design Overview		
i)	General overview	Explain why mixed methods research is appropriate as a methodology given the paper's goals.	p. 56
		Identify the type of mixed methods design used and define it.	p. 56
		Indicate the qualitative approach to inquiry and the quantitative design used within the mixed methods design type (e.g., ethnography, randomized experiment).	p. 56
		If multiple approaches to inquiry were combined, describe how this was done and provide a rationale (e.g., descriptive, interpretive, feminist, psychoanalytic, postpositivist, critical, postmodern, constructivist, or pragmatic approaches), as it is illuminating for the mixed method in use.	p. 56, 58
		Provide a rationale or justification for the need to collect both qualitative and quantitative data and the added value of integrating the results (findings) from the two databases.	p. 56
ii)	Participants or Other Data Sources	When data are collected from multiple sources, clearly identify the sources of qualitative and quantitative data (e.g., participants, text), their characteristics, and the relationship between the data sets, if there is one (e.g., an embedded design).	p. 57
		State the data sources in the order of procedures used in the design type (e.g., qualitative sources first in an exploratory sequential design followed by quantitative sources), if a sequenced design is used in the mixed methods study.	p. 57
iii)	Researcher Description	Because mixed methods research includes qualitative research, and reflexivity is often included in qualitative research, we recommend statements as to how the researchers' backgrounds influence the research.	p. 59
b)	Participant Recruitment		
i)	Participant Sampling or Selection	Describe the qualitative and the quantitative sampling in separate sections.	p. 58-59
		Relate the order of the sections to the procedures used in the mixed methods design type.	p. 58
ii)	Participant Recruitment	Discuss the recruitment strategy for qualitative and quantitative research separately.	p. 57
c)	Data Collection		

i) Data Collection/Identification Procedures	See the JARS–Qual and JARS–Quant Standards.	p. 57-58
ii) Recording and Transforming the Data	See the JARS–Qual Standards	p. 58
d) Data Analysis	Devote separate sections to the qualitative data analysis, the quantitative data analysis, and the mixed methods analysis. This mixed methods analysis consists of ways that the quantitative and qualitative results will be "mixed" or integrated according to the type of mixed methods design used.	p. 58
e) Validity, Reliability, and Methodological Integrity	Indicate methodological integrity, quantitative validity and reliability, and mixed methods validity or legitimacy. Further assessments of mixed methods integrity are also indicated to show the quality of the research process and the inferences drawn from the intersection of the quantitative and qualitative data.	p. 58-59
5) Findings/Results		
a) Findings/Results Subsections	Indicate how the qualitative and quantitative results were "mixed" or integrated (e.g., discussion; tables of joint displays; graphs; data transformation in which one form of data is transformed to the other, such as qualitative text, codes, themes are transformed into quantitative counts or variables).	Discussion
	In mixed methods research, the Findings section typically includes sections on qualitative findings, quantitative results, and mixed methods results. This section should mirror the type of mixed methods design in terms of sequence (i.e., whether quantitative strand or qualitative strand comes first; if both are gathered at the same time, either qualitative findings or quantitative results could be presented first).	p. 59-75
6. Discussion		
i) Discussion Subsections	Typically, the Discussion section, like the Method and Findings/Results, mirrors in sequence the procedures used in the type of mixed methods design. It also reflects on the implications of the integrated findings from across the two methods	p. 75-81