An exploration of the measurement and models of frontal functions in clinical and non-clinical populations

Hannah Catherine Wakely

Thesis submitted in partial fulfilment of the degree of Doctor of Clinical Psychology

University of East Anglia

Faculty of Medicine and Health Sciences

Submission date: 16th June 2020

Candidate registration number: 100225183

Thesis portfolio word count (excluding appendices): 32,991

Overall abstract for thesis portfolio

Objective: The thesis portfolio aimed to assess the psychometric properties and conceptual structure of rating scale measures of frontal functions.

Methods: A systematic review of the literature collected data on the validity and reliability of executive function rating scales with various clinical and non-clinical groups. Alongside this, a validation study explored the psychometric properties of the revised dysexecutive questionnaire (DEX-R) in a non-clinical population. In total, 140 participants took part, some completing the DEX-R at two different timepoints and another validated measure, the FrSBe. Factor analysis and Rasch analysis were used to explore underlying subconstructs. Correlations of mood and demographic variables were also conducted.

Results: There were 24 studies which met criteria for the systematic review. Papers used a variety of EF rating scales across different clinical and non-clinical groups. The DEX was the most widely used measure. Quality varied, many papers would have benefited from the use of a reference standard. In the empirical paper, the DEX-R was found to be a valid and reliable measure of dysexecutive problems in a non-clinical sample. It was determined to be multidimensional and a factor analysis resulted in three factors. Responses correlated with age and brief measures of anxiety and depression.

Conclusions: Rating scale measures supplement neuropsychological testing well in their ecological validity and in capturing the wide-ranging difficulties individuals may face. Understanding individual differences has clinical benefits for interpreting assessments, particularly in variation of responses influenced somewhat by age and mood. Establishing robust sub-scales that map onto models may have useful clinical applications to understand specific areas of strength and limitations relevant for rehabilitation or adapting psychological therapies.

Table of Contents

Acknowledgements

I would like to dedicate this thesis to my father, who has been a constant cheerleader throughout my education and career. His passion for helping others is what inspired me to join this profession. Without his support it would not have been possible to begin this journey, and his unrelenting belief in my abilities motivated me to keep going. Thank you also to my mother, brother, partner and friends for their encouragement. I am hugely appreciative to all who gave up their time to take part in this research.

The support from my research supervisors has been invaluable, Fergus Gracey and Ratko Radakovic – I am very grateful for all your energy and encouragement throughout the process and for igniting my passion in research. Thank you for joining me on a journey delving into the wonders of the frontal functions of the brain.

Introduction to the thesis portfolio

This thesis portfolio includes two papers, a systematic review and an empirical research study. These both investigate psychometric properties of rating scales measuring frontal functions of the brain.

What are frontal functions?

The term 'frontal functions' encompasses cognitive, behavioural and emotional processes thought to originate in the frontal areas of the brain. These include flexible thinking, planning, monitoring, social behaviour, decision making, initiation, inhibition and emotional regulation (Lezak, 1995). The frontal areas of the brain have been implicated in neurological, developmental and mental health difficulties (Bombois et al., 2007; Cullen, 2016; Hill, 2004; Morice & Delahunty, 1996; Swanberg, Tractenberg, Mohs, Thal & Cummings, 2004; Zinn, Bosworth, Hoenig & Swartzwelder, 2007). There can be profound functional difficulties which people can face as a result which vary from person to person, impacting upon activities of daily living, occupational and educational activities (Drakopoulos, Sparding, Clements, Pålsson, & Landén, 2020; Goel, Grafman, Tajik, Gana, & Danto, 1997; Grant & Adams, 2009; Laakso et al., 2019; Ponsford, Draper & Schönberger, 2008; Vaughan & Giovanello, 2010). Different theoretical models have been proposed to account for the kinds of problems observed in people with acquired damage to the frontal areas of the brain. There have been several synonymous terms used interchangeably in attempting to define these as a psychological construct, including frontal functions, executive function and dysexecutive problems.

Theories and models of frontal functions

Hierarchical models of frontal functions

Hierarchical views formed the early understanding of the functions of the frontal areas of the brain, proposing these control and regulate other lower-level cognitive processes (Lezak, 1982; Luria, 1995). Initial theoretical frameworks suggested a unitary process involved in attentional control of these processes, known as a single central executive (Baddeley & Hitch, 1974; Grafman, 1989; Norman & Shallice, 1986; Pribram, 1960). Baddeley and Wilson (1988) went beyond just concentrating on the cognitive elements of frontal functions, additionally incorporating the behavioural and emotional components. This included behaviours observable by others, such as decision making and impulsivity. For instance, the observable aspects of decision making would include struggling with complex or conflicting demands. They use the term dysexecutive syndrome to understand these observable aspects of frontal dysfunction. It has been argued that the use of the word syndrome may be problematic in the conceptualisation of these as there is variation in the type of symptoms experienced following a brain injury (Damasio, Tranel & Damasio, 1991). For example, some may experience more cognitive impairments, but the behavioural and emotional process remain intact, or vice versa (Damasio et al., 1991; Stuss, 2007).

Executive function models

The term 'executive function' was defined by Lezak (1995) as ''those capacities that enable a person to engage successfully in independent, purposive, self-serving behaviour'' (p42). Executive function is thought by some to account for the cognitive aspects of frontal lobe function. Diamond (2013) distinguishes between 'core' and 'higher-order' EF components. Components of 'core' EF included working memory, inhibitory control, and cognitive flexibility, whereas components of 'higher-order' EF are considered to be reasoning, problem-solving and planning. Some frameworks considered these processes to originate solely within the frontal lobes. However, impairment of EF has been

evident in those with injury to differing areas of the brain (Stuss, 2011). Burgess (2004) also argued that terms including 'frontal lobe disorder' had limitations in not accounting for functions linked to other areas of the brain which could co-ordinate these processes.

The Stuss model

Stuss's model (2011) is based on mapping underlying differences in frontal brain structures with lesion and imaging research to dissociate specific functions. He identifies four different processes: (1) Executive Cognitive functions in dorsolateral prefrontal cortex (PFC) regions, (2) Energisation in superior medial areas involved in initiation, (3) Behavioural and Emotional Self-Regulation in ventromedial areas and (4) Metacognition in anterior medial regions of the PFC which coordinate all the aforementioned processes. He also addresses how the connections from the prefrontal cortex to other areas of the brain can lead to the similar dysexecutive presentations typically seen in those thought to have 'pure' frontal brain injuries (Stuss, 2011).

Measurement of frontal functions

Issues with the measurement of frontal functions

Inconsistency in operationally defining frontal functions makes them difficult to measure, with clinicians completing a range of different tasks to tap into the different processes thought to underlie these (Cicerone, Levin, Malec, Stuss & Whyte, 2006). Clinicians and researchers aim to utilise assessments which can assist with predicting problems in day-to-day life, providing diagnosis based on identifying impaired and spared functions, and planning and evaluating rehabilitation interventions. Additionally, assessment may assist in identifying the mechanisms of frontal or other cognitive processes in non-clinical and clinical populations and relating these to underlying brain structures or everyday behaviours. Currently, measurement of such difficulties associated with frontal functions includes traditional standardised neuropsychological tests, ecological

neuropsychological measures and rating scale measures. Due to the task impurity problem with the measurement of EF's, clinicians often rely on completing numerous different neuropsychological tasks in order to tap into the range of EF associated difficulties which may be faced by individuals (Miyake & Friedman, 2012). This is due to the large amount of variance which can influence upon performance in these tasks, which can be attributed to non-EF specific factors, such as visual processing (Miyake & Friedman, 2012). Additionally, such neuropsychological measures mainly tap into the cognitive processes of EF (Chan, Shum, Toulopoulou, & Chen, 2008). Advantages of rating scale measurement over these other tools are in their ability to gather a personalised account from the person on functional difficulties faced in their everyday life. They therefore may be able to capture behavioural, cognitive, social and emotional difficulties making them a potentially ecologically valid means for measurement within the constraints of a clinical setting. Their measurement has important implications in identifying the most suitable interventions to support people with the functional difficulties of which they report. Evidence based practice in neurorehabilitation typically involves a mixture of approaches aiming to restore functions alongside those aiming to compensate for the impact of deficits in everyday life (Turner-Stokes, 2003). Neurorehabilitation additionally aims to improve self-management; however, EF difficulties can impair the skills required to self-manage effectively in everyday life. Due to the challenge in its measurement and the tendency for those available to primarily highlight the 'core' cognitive EF's, dysexecutive problems specific to emotional, social and behavioural domains can be missed during assessment (Chan et al., 2008). This may result in individuals continuing to experience difficulties following their rehabilitation (Cicerone et al., 2006). It is also recognised that there can be individual differences in non-clinical groups, specifically for EF's related to updating, shifting and inhibition (Miyake & Friedman, 2012). Miyake and Friedman (2012) found that some EF

components show unity and diversity, at a latent level show some heritability and that such individual differences have some consistency through development. They also found that there are links with self-regulatory behaviours and specific EF individual differences.

Rating scale measurement

Rating scale measures may offer an ecologically valid means to capture both patient and informant views on challenges associated with dysexecutive problems (Burgess, Alderman, Evans, Emslie & Wilson, 1998). Further development of these rating scale measures may have useful clinical applications in being able to highlight specific areas of strength and difficulties which can become a focus for a person's individual rehabilitation plan (Cicerone et al., 2006). Cicerone et al. (2006) detail how rating scales can supplement traditional neuropsychological testing to highlight the specific challenges to target as part of neurorehabilitation. For example, based on Stuss's (2011) theoretical model, whether problem-solving, prompting, goal management or emotional regulation strategies would be most beneficial for a person depending on whether their difficulties are due to either activation, awareness, cognitive and/or emotional processes. These can be captured within subscales in these rating scales, such as with the development of the Dysexecutive Questionnaire Revised (DEX-R) where one study aimed to refine the DEX to align it to the Stuss model with an ABI sample (Simblett, Ring & Bateman, 2017).

Clinical implications and scope of thesis portfolio

A systematic review of measures would enable clinicians and researchers to identify the most psychometrically robust rating scale measures of frontal functions. The inconsistency of the conceptual understanding means there is variability in what these rating scales claim to measure, for example the tendency to focus on cognitive processes and omit social and emotional prefrontal processes in measures such as the Adult Executive Functioning Inventory (ADEXI: Thorell & Nyberg, 2008). There has

traditionally been a large effort in the measurement of cognitive processes associated with EF, however, for clinicians, there may be challenges in finding a measure which also includes the social, emotional and behavioural processes known to affect those with frontal function impairments. Using additional neuropsychological testing to capture these is likely to be time consuming and limited by the constraints of a clinical setting. As part of the thesis, a systematic review was carried out focussing on addressing these challenges, taking the Lezak (1995, p42) definition of executive function to allow for the inclusion of a range of not just cognitive, but also of emotional, social and behavioural processes. What constitutes as robust psychometric properties lie in their ability to demonstrate sound validity and reliability (Messick, 1989). As well as the review establishing which rating scales measuring executive function show the best psychometric properties it will also identify whether subconstructs/factors are evident and if these subconstructs map onto conceptual models of executive function. Factor analysis and Rasch analysis can be used to explore the subconstructs of measures, and to measure construct validity driven by theoretical understandings (Boone, 2016; Browne & Cutik, 1993; Brown, 2015; Coffman, 2014; Wright, 1996). Rasch analysis enables clinicians to understand whether the measurement tool they are using is unidimensional or multidimensional. Factor analysis groups items into related factors based on responses given. As dysexecutive problems encompass behavioural, cognitive and emotional difficulties being able to establish subconstructs allows the measurement of the different ways people may be impacted by such difficulties (Damasio et al., 1991; Stuss, 2007). In addition, Rasch analysis confirms whether the measure is indeed interval, as opposed to making assumptions of what is typically classed as an ordinal measure. The review sought to examine whether these rating scales apply better for specific clinical or non-clinical groups to support clinicians in identifying which may be most appropriate within their setting.

The empirical paper will then extend on the systematic review by taking a conceptually based rating scale of frontal function to further explore its psychometric properties with a non-clinical population. This paper will assess the validity, reliability and factor structure of the revised version of the Dysexecutive Questionnaire (DEX-R) and whether it is multidimensional in a non-clinical sample and if so whether the factor structure maps onto theoretical conceptualisations of frontal lobe functioning. The DEX-R has preliminarily been demonstrated to be a valid and reliable measure of dysexecutive problems in an ABI and a healthy ageing sample (Dimitriadou, Michaelides, Bateman & Constantinidou, 2018; Simblett et al., 2017). The use of a non-clinical sample may highlight similar individual differences as found in previous research (Miyake & Friedman, 2012). An extended discussion chapter will bring together the conclusions from both papers.

Chapter 1

Systematic review

Prepared for submission to *Neuropsychological Rehabilitation*

The Validity and Reliability of Executive Function Rating Scales: A Systematic Review

Hannah Wakely[1], Fergus Gracey[1], Pernille Myhre[1] and Ratko Radakovic[2].

[1]Department of Clinical Psychology and Psychological Therapies, University of East Anglia, United Kingdom;

[2]School of Health Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, United Kingdom. Norfolk and Norwich University Hospital, Norwich, United Kingdom.

The Euan MacDonald Centre for Motor Neurone Disease, University of Edinburgh, Edinburgh, United Kingdom. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, United Kingdom. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom.

Correspondence should be addressed to Fergus Gracey, Department of Clinical Psychology and Psychological Therapies, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ.

E-mail: f.gracey@uea.ac.uk

Total word count: 13025

Word count excluding abstract, table and references: 8467

Abstract

Background: Rating scales are used to measure executive function in addition to traditional cognitive assessments in research and clinical practice. A recent review of the literature on their psychometric properties has not been conducted. The main objective was therefore to review the psychometric properties of rating scales measuring executive function.

Methods: Searches were performed up to April 2020 in EMBASE, PsycINFO, CINAHL and MEDLINE. Included papers used an EF rating scale, reporting both a reliability and validity statistic. Quality assessment was completed using a modified version of the QUADAS-2. Data extraction and a narrative synthesis of the data followed.

Results: 24 papers were included in the review with 8449 participants. The EF rating scales included the DEX, BRIEF, FrSBe/FLOPS, BASC, ADEXI, FBI, ECQ and BDEFS. A range of clinical and non-clinical groups were included, and the factor structures varied within and between the rating scales. Most had at least adequate validity and reliability. The quality of papers was mixed; many did not include an adequate reference standard.

Conclusions: The DEX, FrSBe and BRIEF-A were the most widely used rating scales with adequate to excellent reliability and validity across clinical and non-clinical groups. Papers were limited in utilising test-retest reliability and concurrent validity to compare to existing EF rating scales. Future research using discriminative analysis could further enhance the use of these measures.

Keywords:

Executive Function; Rating Scales; Validity; Reliability

Prospero registration:

www.crd.york.ac.uk/prospero/;CRD42019139013

Introduction

The term executive function (EF) has been defined as “those capacities that enable a person to engage successfully in independent, purposive, self-serving behaviour’’ (Lezak, 1995, p42). EF is “important to just about every aspect of life” (Diamond, 2013, p. 137) therefore the recognition of such difficulties in clinical settings are important in being able to provide relevant interventions. Failure to capture EF difficulties may result in individuals continuing to experience problems in their everyday functioning following rehabilitation (Cicerone, Levin, Malec, Stuss, & Whyte, 2006).

There have been challenges in the measurement of executive function, particularly due to constraints on testing within a clinical environment. Clinicians can use performance based tests and rating scales to measure EF. Performance based tests include tasks thought to tap into certain processes of EF and may be based on a particular cognitive model or aim to capture everyday function (or both). Various traditional tasks of EF such as the stroop task and the trail making task have had issues in sensitivity to detect frontal deficits despite reported or observed difficulties in everyday function (Shallice & Burgess, 1991). Therefore, a limitation is the lack of ecological validity by failing to highlight challenges in functional difficulties observed or reported by individuals (Holst & Thorell, 2018; Manchester, Priestley, & Jackson, 2004; Shallice & Burgess, 1991; Stuss et al., 1983). In addition, problem solving abilities require novelty of tasks and so the ability for repeated measurement is a shortcoming of these tests (Holst & Thorell, 2018).

There have been attempts to address these issues of ecological validity of testing in a clinical environment by either comparing existing tests to everyday behaviours (veridicality) or designing new tests which resemble an everyday task (verisimilitude) (Chaytor & Schmitter-Edgecombe, 2003). The Multiple Errands Test and the Six Elements Test were designed to provide a less structured assessment to increase demands on EF

components, particularly those associated with higher-order EF's (Diamond, 2013; Malloy & Grace, 2005; Shallice & Burgess, 1991). However, there are still some structure and implicit prompts present through clinical administration which may not mirror difficulties present in everyday life. In addition, these can be time-consuming to administer and may not measure all the underlying components (including behavioural, social and emotional components) thought to underpin EF (Burgess, Alderman, Evans, Emslie, & Wilson, 1998; Chan, Shum, Toulopoulou, & Chen, 2008; Duggan, Garcia-Barrera & Müller, 2018).

There is therefore a gap between our conceptual understanding of EF, and the cognitive, behavioural, emotional and social difficulties that can be seen in clinical settings and experienced in everyday life. To overcome this, rating scale measures have been developed to capture reports of these functional challenges which may be experienced (Duggan et al., 2018; Isquith Roth, & Gioia, 2013; Toplak, West & Stanovich, 2013). The use of self-ratings to measure EF has shown to better capture such processes when compared to EF neuropsychological tests (Barkley & Fischer, 2011; Barkley & Murphy, 2010; Holst & Thorell, 2018; Toplak et al., 2013). However, the poor correlations in different types of measurement may be explained by these measuring differing sub-constructs of EF, or frontal functions (Burgess et al., 1998; McAuley, Chen, Goos, Schachar, & Crosbie, 2010; Toplak et al., 2013). Due to these differences in measurement, it has been recommended neither method be used standalone and instead self-ratings can complement traditional testing well in enhancing overall EF assessment (Toplak et al., 2013). Furthermore, challenges and different views on operationalising and conceptualising EF may explain the inconsistency of EF rating scales and the subconstructs and factors they aim to measure (Duggan et al., 2018; Garcia-Barrera, Karr, & Kamphaus, 2013). A number of different EF scales have been developed and applied across different clinical populations and translated into different languages. These are available for children

and adults. EF rating scales also offer informant versions further enhancing clinical assessments which is of particular use for people who are experiencing reduced self-awareness of their difficulties. The multiple factors thought to underpin EF has been another challenge in its measurement, behaviour rating scales have been developed with subscales to capture these different components.

The psychometric properties of EF rating scales continue to be explored to improve their measurement. Reliability is determined by how replicable the measure is (Clark-Carter, 2009; Messick, 1989). Internal reliability measures the consistency of a measure to establish if the questions relate to each other. Test-retest reliability measures whether the measure is consistent over time, such as by comparing scores on the measure at two different time points which enables the accountability of day-to-day variability. Inter-rater reliability compares the degree of agreement in scores between more than one rater (Clark-Carter, 2009). Those with impairments to EF can have difficulties with self-awareness, and therefore the development of measures with sound inter-rater reliability is important. Validity refers to whether the measure actually measures what it sets out to, such as the construct of EF and its underlying sub-constructs (Messick, 1989). There are different types of validity, including concurrent, construct, face, content and criterion validity (Clark-Carter, 2009; Messick, 1989). Both reliability and validity are important in establishing the psychometric properties of measures as the presence of reliability correlates with increased validity of measures (Litwin, 1995). Recommendations on the interpretation of such psychometric properties is available in the literature (Hermans, van der Pas, & Evenhuis, 2011). Factor analysis and Rasch analysis have been used as a measure of construct validity. They provide a basis to understand whether scales are measuring multiple components and how specific items group together (Boone, 2016;

Brown, 2015; Browne & Cutik, 1993; Coffman, 2014; Wright, 1996). Using subscales can allow for the detection of where tailored interventions can be used.

When identifying which psychometrics are of most importance when reviewing rating scales, a number of considerations can be incorporated. In order to understand what a useful clinical measure is we need to know whether the content is useful and relevant to the patient and the clinician, established by content validity. This also involves the measure relating to the neuropsychological construct being assessed via construct validity. In addition, when yielding a score, we need to know whether this score is interpretable (Fermanian, 2005). In terms of classical test theory, a more reliable measure would increase the confidence of knowing where the true score lies. Sound reliability of a measure demonstrates internal consistency of the construct being measured and its stability over time. When looking at the psychometrics of clinical measures which incorporate different domains, factor analysis and item response theory can provide further important information to check factor structure and the reliability and validity of subscales (Wu, Tam & Jen, 2016). Furthermore, factor analysis can contribute to determining the conceptual nature of a measure. This has clinical advantages in being able to profile strengths and difficulties in a particular domain. Concurrent validity involves comparing responses to a 'gold standard' measure to identify whether something meaningful is being measured. It is important to be sure EF is being measured in a way comparable to a gold standard. For example, identifying the presence of frontal damage or of EF impairments. Therefore, the psychometric properties of particular interest in establishing the robustness of a clinical measure include content, construct and concurrent validity, and reliability. Important but less high priority for a review include how the measure subscales align with neuropsychological models, and the extent to which individual items are sensitive to a

range of abilities. The stability of such psychometric properties can be guided by whether these are evident across diverse samples or contexts on a measure by measure basis.

Malloy and Grace (2005) completed a review which reported the psychometric properties of rating scale measures of frontal functions. This included the Behaviour Rating Inventory of Executive Functions (BRIEF; Gioia, Isquith, Guy & Kenworthy, 2000), the Dysexecutive Questionnaire (DEX; Burgess et al., 1998), the Frontal Systems Behaviour Scale (FrSBe; Grace & Malloy, 2001), and the Iowa Rating Scales of Personality Change (IRSPC; Barrash, Tranel, & Anderson, 2000). The review also included the Frontal Behaviour Inventory (FBI; Kertesz, Davidson, & Fox, 1997) and the Neuropsychiatric Inventory (NPI; Cummings et al., 1994). Whilst these were developed to capture neuropsychological features of dementia, their inclusion in the review was due to their potential to measure deficits associated with frontal functions. The psychometric properties of the FrSBe were the most robust, evidencing both reliability and validity across different clinical groups. The FrSBe and BRIEF provided comprehensive norms, not provided by the other rating scales. The BRIEF also demonstrated reliability, however the review reports upon the original child version. Since publication of the review, the adult version has become widely available (Roth, Isquith & Gioia, 2005). Only the FrSBe and IRSPC evidenced classification between frontal and non-frontal brain injured groups. Although, the FBI and NPI were sensitive to behaviour changes such as disinhibition but these were largely based on dementia research. The FBI was considered to be highly reliable and valid. There was no reliability or norms reported for the DEX, and validity was evidenced by a factor analysis. Brain injury groups only differed to control groups based on informant versions of the DEX. Research presented in the review considered how this may relate to differences in awareness.

There has however since been a surge in research in this area as well as on operationalising EF and frontal functions. There are several EF rating scales available but there has not been a recent review on how these compare to each other in terms of psychometric quality. In addition, it is unclear whether different rating scales are more suited for particular clinical groups and whether their psychometric quality is stable across diverse contexts. Diverse application across clinical and non-clinical populations and contexts would allow a greater comparison on how each rating scale performs. For example, whether the DEX performs well in differing clinical groups and across different countries. If these have been applied broadly, they may highlight whether a rating scale is particularly sensitive to EF impairments associated with a specific condition. A lack of systematic reviews in this area poses challenges for clinicians and researchers in identifying the most robust rating scales to measure specific components of executive function. Therefore, the aim of the current systematic review was to focus on the psychometric properties across different EF rating scales to address this gap in the literature. This will contribute by guiding clinicians in their decision making on selecting an EF rating scale based on robustness and/or conceptual structure.

Review questions:

1. Which EF rating scale measures are the most robust for clinical application across patients with neuropsychological deficits?
    a. Which rating scales measuring executive function show the best psychometric properties?
    b. Do these apply better for specific clinical or non-clinical populations?
    c. What subconstructs / factors are evident and do these map onto models of executive function?

Methodology

The PRISMA guidelines were used as a guide in completing the systematic review (Moher et al., 2009). A protocol for this review was registered with the PROSPERO systematic review protocol registry (www.crd.york.ac.uk/prospero/;CRD42019139013).

Search Strategy:

Searches of the literature were conducted between 1985 and April 2020 in the following four electronic databases: MEDLINE, PsycINFO, EMBASE, and Cumulative Index to Nursing and Allied Health Literature (CINAHL). Backward citation searching of included papers was conducted, as well as reviewing google scholar for commonly cited papers not identified from the main search. Two additional papers were identified in this way. The search terms included syntax and Boolean operators adapted for each database (i.e. AND, OR). Search terms were grouped to include an executive function term, a scale term, a validity term and a reliability term. For executive function, the search terms included: "dysexecutive syndrome" OR "exec* func*" OR "dysexec*" OR "Frontal Function*" OR "Frontal System*" OR "Exec* Dysfunct*" OR "Executive Impairment*" OR "Frontal Lobe Syndrome*" OR “Metacognition” OR “Supervisory Attention” OR “Higher Cognition”. For rating scale, the search terms included: "Rat* Scale" OR "Summed Rat* Scale*" OR "Psych* Rat* Scale*" OR "Psychiatric Status Rating Scales" OR "Questionnaire*" OR "Measurement*" OR "Self-report Measure*" OR "Outcome Measure*" OR "Psychomet*". For validity, the search terms included: "Valid*" OR "Test Valid*" OR "Statistical Valid*". For reliability, the search terms included: "Reliab*" OR "Test Reliab*" OR "Interrater Reliab*" OR "Statistical Reliab*". Limits applied included, being published since 1985, being published in English and using human subjects. Only those papers published prior to the search date are included in the review.

Eligibility Criteria:

The population under study included participants with a neurological condition or illness, neurodevelopmental disorder, or those with a diagnosed mental health condition. The review did not limit to specific clinical populations, and those reporting findings from research with non-clinical populations were also included. The review focussed solely on adult populations.

Studies were required to fulfil the following criteria:

- Published since 1985.
- Used rating scale methods for assessing executive function.
- Report psychometric properties (both reliability and validity).
- Published in peer-reviewed journals.

Studies were however excluded, if:

- They were not published in English.
- The measurement was not an EF rating scale.
- If either only reliability or validity were reported.
- The sample only included those aged under 18 years of age.

Data extraction

Papers were screened and selected by one reviewer (H.W.) based on the above eligibility criteria. Duplicate papers were then removed using the software Endnote. Papers were screened for eligibility based on the title and abstract. A second reviewer (P.M.) examined a randomly selected 20% of these papers. A third reviewer was available for adjudicating should this have been required. The full-text versions of these papers were then evaluated for inclusion. Relevant data was extracted including: Author and year of publication, the number of participants, the population under study, the EF rating scale used, the type of rating (self-report or informant), the psychometric properties and their

corresponding statistics (reliability, validity and factor structure). Conceptual models were commented on where reported.

Risk of bias (quality) assessment

Each article was subjected to a quality assessment using a modified version of the Quality Assessment of Diagnostic Accuracy Studies version 2 (QUADAS-2: Whiting et al., 2011) checklist. This assesses four domains: patient selection, index test, reference standard, and flow and timing. The authors recommend modifying the quality assessment to suit the review question by adding or omitting signalling questions. As the current review is not of diagnostic studies the following modifications were made: one signalling question was removed from Domain one and Domain three, two were removed from Domain two, and an additional signalling question more specific to a psychometric review was added to both Domain one and two derived from the original version of the QUADAS-2. All four signalling questions for Domain four were retained. The modified version of the QUADAS-2 can be found in Appendix D.

The patient selection was broad to include clinical and non-clinical populations, the target condition was EF, the index test was the method by which EF was assessed, and the reference standard was the method used to classify the presence of EF. If papers did not include a reference standard, then this was rated as not applicable. This was to allow for being able to discriminate between papers that had high risk of bias when a reference standard was used.

Analysis

Narrative synthesis involved integration and comparison of extracted data as follows:

1. Looking at psychometric properties across all measures
2. Comparing psychometric properties across participant groups
3. Comparing factor structure across measures and participant groups

The interpretations of psychometric properties were guided by the literature (Cheung & Wang, 2017; Hermans et al., 2011). The quality assessment ratings regarding the risk of bias and applicability for the four domains of the QUADAS-2 were presented in a table. Risk of bias and concerns regarding applicability were rated as either 'low', 'high', 'unclear', or 'not applicable'. A graph summarises the proportion of papers with bias across signalling questions to allow recommendations to improve quality to be made.

Results

Study Selection

Figure 1 details the process for which the papers were retrieved from the database searches. A total of 2512 papers were retrieved, with two additional papers identified from manual searches. Following the screening and review stages there were 24 papers included in the current systematic review. The study characteristics are detailed in table 1 and include author, sample, EF rating scale measure, type of validity and reliability, and for those using factor analysis, the number and labels of factors reported. The table also presents the quality of the papers included in the review based on the modified version of the QUADRAS-2, additionally figure 2 shows the proportion of papers meeting each signalling question of the QUADRAS-2.

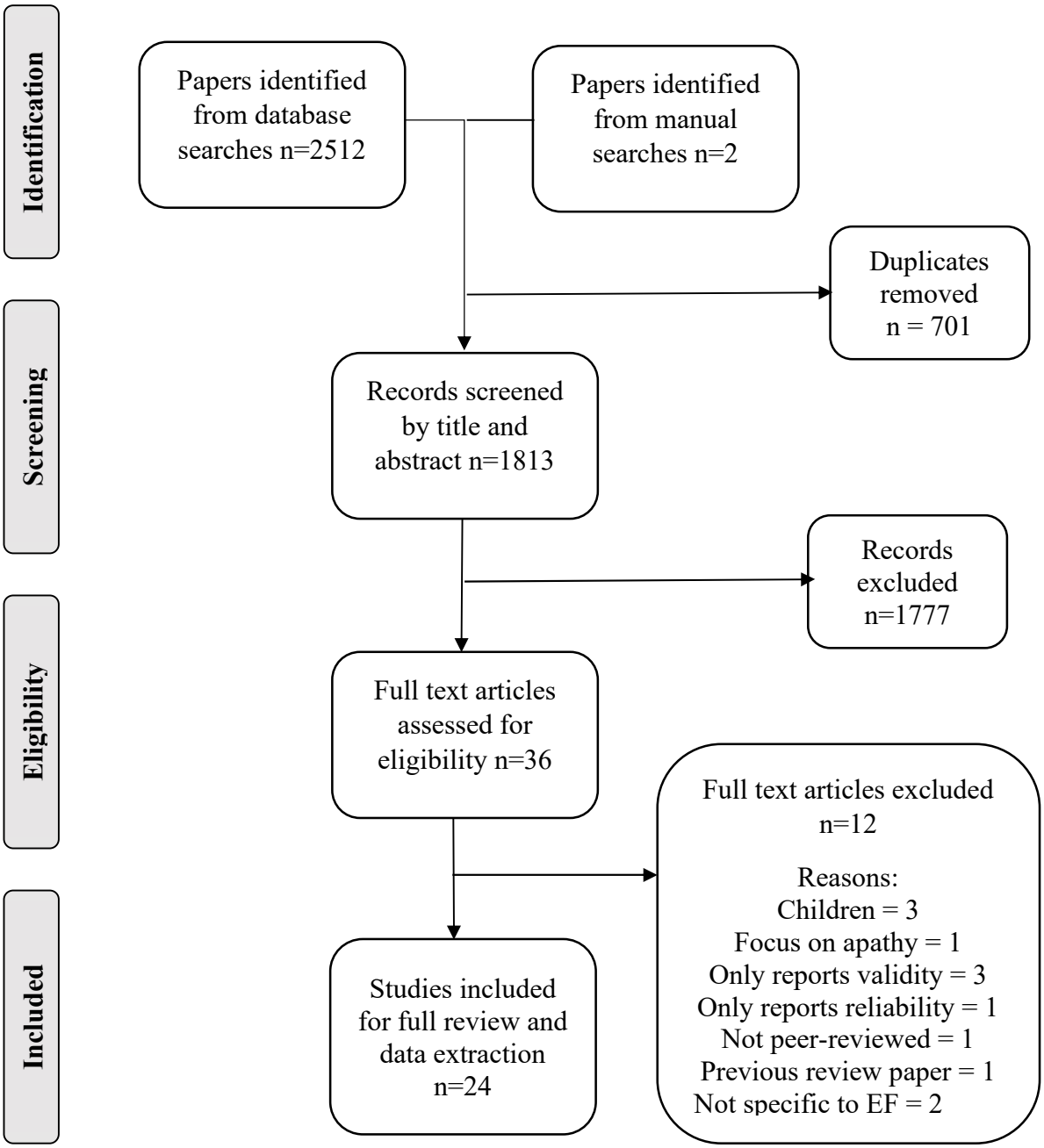


Figure 1. PRISMA Flowchart detailing review process

Participant Characteristics

A total of 8449 participants took part in the studies detailed in the review. Of these, 2250 were from clinical groups (1183 brain injury, 194 dementia, 741 mental health and 132 ADHD), 4922 were from non-clinical groups, 253 on healthy ageing, and 1042 informants. Their ages ranged from 18 – 90 years. The types of ratings included eight self-report, six were informant based (i.e. family member, carer, staff or researcher), and seven including both self and informant ratings. Ten papers explored psychometric properties with either solely or including those with an acquired or traumatic brain injury, two with neurodegenerative conditions, five mental health, two neurodevelopmental which focused on ADHD, and six included non-clinical groups. A number included healthy controls which was the case in an additional six papers. Of the non-clinical papers, one focussed explicitly on older adults (Dimitriadou, Michaelides, Bateman & Constantinidou, 2018).

Study Characteristics

The majority of papers reported the psychometric properties of already standardised EF rating scale measures, with one reporting on the development of a new measure (Coolidge & Griego, 1995). There were seven which assessed the DEX or modifications of the DEX, five with the BRIEF-A (Roth et al., 2005), four which assessed the FrSBe, two for the Barkley Deficits in Executive Functioning Scale (BDEFS: Barkley, 2011) and one paper each for the FBI, Frontal Lobe Personality Scale (FLOPs: Grace & Malloy, 1992), the Behavioural Assessment System for Children (BASC: Reynolds & Kamphaus, 1992), the ADEXI, and the Executive-Complaints Questionnaire (ECQ: Mías, Ruiz, Causse, & Verónica, 2017). Most rating scales assessed were the English versions, two included translated versions in Spanish (Caracuel et al., 2012; Vélez-Pastrana et al., 2016), one in Persian (Mani et al., 2018), one in Italian (Milan et al., 2008), one in Dutch (Beerten-

Duijkers, Vissers, Rinck, Barkley & Egger, 2019), and one in Japanese (Shinagawa et al., 2007). All translated versions were also then back translated.

Only papers which included both a type of validity and reliability statistic were included in the review. Type of validity statistic included: Construct Validity (often reported as being established by Factor Analysis or Rasch Analysis), Convergent validity, Discriminant validity, Content validity, Concurrent validity and Criterion Validity. A limited number of papers compared responses to another rating scale measure of EF. The reliability statistics included internal consistency, split-half reliability, test-retest reliability and interrater reliability. The timeframe for test-retest reliability varied between one week and three months. Of those where a factor structure was reported, these ranged from between two and five factors as being present in the rating scale.

Risk of Bias and Quality Assessment

All papers were assessed for risk of bias and applicability using a modified version of the QUADAS-2. The scores for the domain's patient selection, index test, reference standard, and flow and timing can be found in Appendix E. Figure 2 details aspects which consistently may have introduced bias. Only 15 studies included a reference standard, those where a reference standard was not a measurement of EF were rated as high, which was the case for five studies. No papers scored high across all domains; however, the majority of papers were unclear in at least one domain. Typically, the index test was appropriate for the review question, however eight papers did not detail the administration clear enough to know whether bias may have occurred. Seven papers scored high in patient selection due to not using a consecutive or random sample and included either payment or were part of a treatment study which may have led to different motivations to take part to those who did not. This could bias the measurement of EF as motivation is thought to be an underlying component of it (Pessoa, 2009). All but three papers did recruit a sample which

was applicable to their review question. The flow and timing domain was often difficult to establish due to limited information provided in the papers. The paper by Coolidge and Griego (1995) was particularly difficult to interpret due to the limited information in their paper. Shinagawa et al. (2007) was the only paper to score low across all domains in both bias and applicability. Beerten-Duijkers et al. (2019) and Milan et al. (2008) were both particularly strong in their quality, both only would have benefited from a clear description of the patient selection. Caracuel et al. (2012) excluded those who were not self-reliant in ADL's, and therefore could have limited applicability due to EF measures being used to understand such challenges.

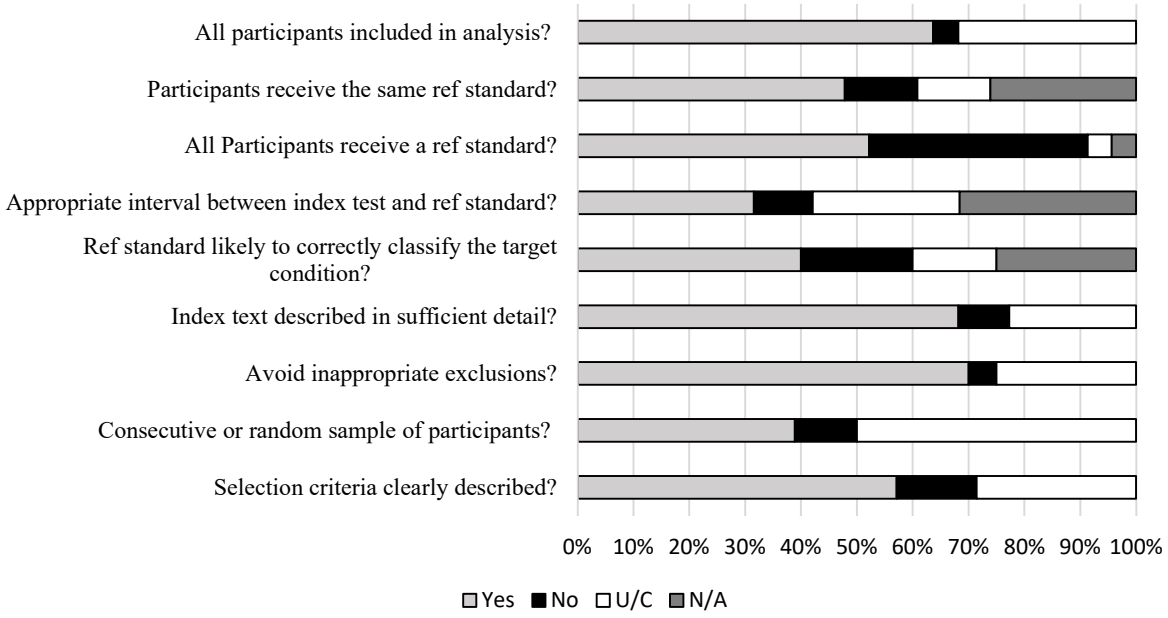


Figure 2.

Proportion of ratings to signalling questions on the QUADAS-2.

Table 1.

Data Extraction Table

Author (Year)	Reliability Type and Statistic	Validity Type and Statistic	Quality*	Type of Rating	Factor Structure	Participants (n, clinical group)
Papers assessing the DEX (n=4)						
Bodenburg & Dopslaff (2008)	Internal consistency (Cronbach's alpha) **α =**.85	Construct validity: via EFA	- -	Self-report	4 factors: - Initiate and sustain - Impulse control and sequencing - Excitability - Regard for social standards)	ABI 191 ABI
Hellebrekers et al. (2017)	Internal consistency (Cronbach's alpha) ABI **α=**.89 Informants **α=** .89 Test-retest (5 weeks, Spearman): ABI: r =.88 Informants: r =.60	Construct validity: via EFA	- - - +	Self-report and informant report	2 factors (Self-rating version): - Initiating & sustaining actions - Impulse control & sequencing of heard information 3 factors (Informant version): - Initiating, sustaining actions & regard for social situations - Impulse control & sequencing of heard information - Planning & decision-making	ABI 105 ABI 105 Informants
Simblett et al. (2012)	Internal consistency (PSI) Multiple analysis: .68 – .81	Internal construct validity: reported as examined using Rasch Analysis	-	Self-report and informant	3 factors: - Executive Cognition - Metacognition - Behavioural-Emotional Self-Regulation	ABI 271 (181 TBI, 84 non-traumatic BI, 6 other)
Shinagawa et al. (2007)	Internal consistency (Cronbach's alpha) **α =**.93 Test–retest (n=44, one month, ICC) Total: r =.95	Concurrent validity: FAB vs DEX factors: Apathy: r = .45 p<.01 Planning and monitoring: r = .65 p<.01 Hyperactivity: r = .31 p<.05 NPI Apathy and DEX Apathy: r = .37 p<.01	- - - -	Informant report	3 factors: - Apathy - Planning and monitoring - Hyperactivity	Neurodegenerative 122 (Caregiver for a person with Alzheimer's Disease)

	Factor 1: r = .93 Factor 2: r =.97 Factor 3: r = .95	Construct validity: via EFA				
Papers assessing modifications of the DEX (n=3)						
Shaw et al. (2015)	Internal consistency (Cronbach's alpha) Total: **α**=.85 Subscales: Factor 1: **α**=.80 Factor 2: **α**=.75 Factor 3: **α**=.60 Samples: Community: **α**=.90 Psychiatric: **α**=.91 Neurological: **α**=.91	Concurrent Validity (Total score) SDS: r = .54, *p*<.01 BAI: r = .62, *p*<.01 DASS-S: r = .58, *p*<.01 GSES: r = -.66, *p*<0.01 SWLS: r = -.47, *p*<.01 Criterion-Related Validity (Discriminant Function Analysis): Total Score correctly classified 68.6% of cases (λ = 0.90, χ2 [2] = 102.51, *F* [2.990] = 54.00, *p*<.001). Criterion Validity (Total Score): A one-way between-groups ANOVA, statistically significant difference among the groups, *F*(3, 992) = 36.38, *p*< .001. The post-hoc comparisons (Games-Howell test) community group reported significantly fewer levels of dysexecutive syndrome than the depressed and anxious (psychiatric) groups. Construct Validity: via CFA and EFA	- - + +	Self-report	3 factors - Inhibition - Volition - Social Regulation	Mix clinical and non-clinical Total 997 Community (n = 663) Psychiatric (depressed [n = 92] and anxious [n = 122]), "Neurologically impaired" (n = 120)
Simblett et al. (2017)	Rasch Analysis Internal consistency (PSI): Modified Metacognition: .82 Modified Executive Cognition: .92 Modified Behavioural-	Internal construct validity: reported as examined using Rasch Analysis	-	Self-report and informant report	4 factors: - Metacognition - Executive Cognition - Behavioural-Emotional Self-Regulation - Activation	ABI 136 (ABI) 71 (Family members)

	Emotional Self-Regulation: .76 Activation: .88					
Dimitriadou et al. (2018)	Internal consistency (Cronbach's alpha): Total: Self **α=** .88 Informant **α=** .91 Scale ranges: Self **α=** .71-.80 Informant **α=** .77-.85 Inter-rater: Non-significant differences between self and informant	Factorial Validity: via EFA/CFA Convergent validity: "The significant positive correlations between the three symptom factors" Self: r = .65 – .77 Informant: r = .59 – .81	- - -	Self-report and informant report	3 factors: - Motivation and Attention - Flexibility, Fluency and Working Memory - Social Self-Regulation	Healthy ageing / non-clinical Older adults (n = 235) Informants (n = 187)
Papers assessing the BRIEF-A (n=5)						
Ciszewski et al. (2014)	Internal consistency (Cronbach's alpha): BRI/MCI/GEC: **α=**.93/.94/.96 Nine clinical scales (**α**=.69 to .91)	Construct validity: EFA and CFA Convergent validity: (Self-report vs informant-report, Pearson's) r = .85, *p*<.01	- -	Self-report and informant report	EFA 2 factor structure - BRI - MCI CFA (did not fit well)	Mental health 252 (Eating Disorder) 31 informants
Rouel et al. (2016)	Internal consistency (Cronbach's alpha): BRI **α=** .93 MCI **α=** .95 GEC **α=** .97 Test-retest (2 months; n = 30, ICC) BRI: r = .96	Convergent validity: WCST perseveration r= .03 NS EF Composite r= .14 NS TMT-B derived r= .02 NS RCFT r=.17 NS Digit Span backwards r= −.01 NS Content validity: via EFA	- - +	Self-report	2 factors: - BRI - MCI	Mental health 98 obese participants (BED binge eating disorder)

	MCI: r = .93 GEC: r = .94					
Hauser et al. (2013)	Internal consistency (Cronbach's alpha): BRI/MCI/GEC: Deaf non-ADHD: **α**=.91/.95/.95 Hearing non-ADHD: **α**=.90/.93/.95 Deaf ADHD: **α**=.91/.95/.96 Hearing ADHD: **α**=.84/.93/.93	Convergent/Predictive Validity (CAARS-S:L): Deaf Group ADHD Index = .83 Hearing Group ADHD Index = .81 Deaf Group DSM-IV Symptoms Total = .92 Hearing Group DSM-IV Symptoms Total = .93 Correlation to manual (type of correlation not reported): Deaf ADHD r = .53 *p*= .07 Hearing ADHD r = .92, *p*<.01	+ + +	Self-report	-	Neurodevelopmental / Non-clinical 360 college students: (151 Deaf non-ADHD; 128 Hearing non-ADHD; 25 Deaf ADHD; 56 Hearing ADHD)
Mani et al. (2018)	Internal consistency (Cronbach's alpha): Subscales: Inhibit **α**=.69 Shift **α**=.77 Emotional control **α**=.84 Self-monitor **α**=.70 Initiate **α**=.72 Working memory **α**=.78 Plan/organize **α**=.80 Task monitor **α**=.65 Organization of material **α**=.78 Test-retest reliability (n= 60, one month, Pearson):	Content validity: The content of the scale was confirmed by researchers using the manual and EF theory. Face validity: 5 psychiatrists checked the final version (not the researchers). Construct Validity: via Factor analysis	- - + +	Self-report	CFA – number not reported but states confirms the original structure of the BRIEF-A	Non-clinical 318 Students/ employees

	r = .78, $p < .001$					
Waid-Ebbs et al. (2012)	Internal consistency (Cronbach's alpha): BRI **α=** .94 MCI **α=** .96 Item/person reliability and separation: BRI =.85/.93 MCI =.86/.94	Construct Validity: via CFA and Rasch Analysis	-	Informant report	2 factors: - BRI - MCI	ABI 90 TBI and 89 informants
Papers assessing the FrSBe (n=4)						
Caracuel et al. (2012)	Rasch Analysis Internal consistency (PSI): Sample A (ABI) Apathy = .70 - .74 Disinhibition =.73 Executive dysfunction = .71 - .74 Sample B (Relatives): Apathy = .87 Disinhibition = .79 - .83 Executive dysfunction = .86 Sample C (Control): Apathy = 0.71-.72 Disinhibition = .72 -.73 Executive dysfunction = .71-.75	Construct Validity: reported as examined using Rasch Analysis	+ +	Self-report and informant report	3 factors: - Apathy - Disinhibition - Executive Dysfunction	Mixed sample Total 245 Spanish subjects: Sample A: 65 TBI / stroke (45 TBI, 20 Stroke) Sample B: Family-rating of the same 65 participants Sample C: 115 healthy individuals

Carvalho et al. (2013)	Internal consistency: (Cronbach's alpha) Original Model Total **α**= .95 Apathy **α**= .88 Disinhibition **α**= .84 Executive Dysfunction **α**= .91 Revised model: Total **α**=.93 Apathy **α**= .81 Disinhibition **α**= .82 Executive Dysfunction **α**= .92	Construct Validity: via CFA	-	Informant report (family version)	3 factors: - Apathy - Disinhibition - Executive Dysfunction	Neurological / Neurodegenerative 494 informants "various neurological conditions" (both dementia and ABI)
Niemeier et al. (2013)	Internal consistency: (Cronbach's alpha) Self-T1 **α**=.89 Self-T2 **α**=.89 Family-T1 = .92 Family-T2 = .93 Test-retest (7 days, reported only as *r*): Self: r = .54 Family: r = .72	Convergent validity with PCRS (only reported as *r)*: Self-T1: r = -.50 Self-T2: r = -.53 Family-T1: r = -.65 Family-T2: r = -.71 Construct Validity: via CFA and EFA	- +	Self-report and informant report	CFA: 1 & 3 factor models did not fit the data EFA: 4 factors (no factor labels) (four separate EFA produced separate factor loadings each time, all retained four factors)	ABI 101 (TBI) 38 individuals (37.6%) with moderate TBI and 63 (62.4%) with severe TBI
Velligan et al. (2002)	Internal consistency: (Cronbach's alpha) Total: **α**=.94 Apathy **α**=.88 Disinhibition **α**=.86 Executive Dysfunction **α**=.91 Test-retest: (3 months, Pearson's) Total r = .78	Criterion Validity (Comparison of patients to controls): Total: $F_{3,177}$=51.86, p<.0001 Apathy*:* $F_{1,179}$=156.82, p<.0001; Disinhibition: $F_{1,179}$=16.24, p<.0001; Executive Dysfunction: $F_{1,179}$=58.35, p<.0001 Convergent validity (Spearman's rho)	- - + +	Informant report (Staff version, researcher)	-	Mental health 131 (Schizophrenia)

	Apathy r = .68, *p*<.01 Disinhibition r = .65, *p*<.01 Executive Dysfunction r = .65, *p*<.01	Verbal Fluency: (Apathy r= -.47, *p*<.01; Disinhibition r =.16 NS; Executive function - r =.43 *p*<.01) Trails B Errors: (Apathy r = -.17, NS; Disinhibition r =.42 *p*<.01; Executive function r =-.38 *p*<.01) Trails B Time: (Apathy r =-.30, *p*<.01; Disinhibition r =.33 *p*<.01; Executive function r =-.48 *p*<.01) Continuous performance test false alarms (Apathy r =.11 NS; Disinhibition r =.22 NS; Executive Function r =-.26 *p*<.01).				
Papers assessing the FLOPS (n=1)						
Grace et al. (1999)	Internal consistency (Cronbach's alpha) **α**=.96 Split-half = .93	Construct validity: Pre-Post comparison by family, not significantly correlated for frontal ABI (r = .30, *p*=.16). Significant difference in pre-post scores for frontal ABI (t = -6.21, *p* <.001), no significant correlation or difference for non-frontal ABI group (t = -1.69, *p* = .11).	- +	Informant report (Family)	-	Mixed 87 (24 frontal ABI, 15 non-frontal ABI, 48 healthy controls)
Papers assessing the BDEFS (n=2)						
Beerten-Duijkers et al. (2019)	Internal consistency (Cronbach's alpha) Total **α**=.94 Subscales: Self-management to time **α**=.86 Self-organization **α**=.92 Self-regulation of emotion **α**=.90 Self-restraint **α**=.81 Self-motivation **α**=.75	Concurrent Validity DEX: r = .92, *p*<.05 BIS r =.43 *p*<.05 92% clinical agreement between English and Dutch versions of the BDEFS	- - -	Self-report	-	Non-clinical 85 Dutch Adults

Vélez-Pastrana et al. (2016)	Internal consistency (Cronbach's alpha) Self-management to time **α**=.93 Self-organisation/ problem solving **α**=.94 Self-regulation of emotion **α**=.92 Self-restraint **α**=.89 Self-motivation **α**=.86	Construct Validity via EFA and CFA	- + +	Self-report	5 factors: - Self-Organization/ Problem Solving - Self-Management to Time - Self-Regulation of Emotion - Self-Restraint - Self-Motivation	Non-clinical 452 Latino community adults
Papers assessing the FBI (n=1)						
Milan et al. (2008)	Internal consistency (Cronbach's alpha) Total **α**=.93 Inter-rater (Cohen k coefficient): k = .92, *p*<.0001 Test–retest (2 weeks, Cohen k coefficient): k = .90, *p*<.0001	Concurrent validity: (NPI-P: Frontal items) r = .45; *p*<0.01 (FAB) r = .31; *p*< .01 Factorial Validity via EFA Discriminant validity: 24 FBI sub-scores correctly classified 100% fv-FTD, 90.9% AD and 73.3% with VaD (Wilks k= 0.0945; F = 4.317; *P* < .0001).	- - -	Informant report	5 factors	Neurodegenerative 72 (dementia; 35 FTD, 22 AD, 15 VaD)
Papers assessing the BASC (n=1)						
Duggan et al. (2018)	Internal consistency (Cronbach's alpha) Total **α**=.84 Problem Solving **α**=.70 Attentional Control **α**=.80 Behavioural Control **α**=.66 Emotional Control **α**=.70	Convergent Validity (BRIEF-A): BASC Problem Solving (BRIEF Plan/Organize r = .65, *p*<.01) BASC Attentional Control (BRIEF Working Memory r = .62, *p*<.01) BASC Behavioural Control (BRIEF Inhibit r = .49, *p*<.01)	-	Self-report	4 Factors: - Problem Solving - Attentional Control - Behavioural Control - Emotional Control	Non-clinical Study 1: 765 (US College – non-clinical) Study 2: 197 (University Students – non-clinical)

		CFA				
Papers assessing the ADEXI (n=1)						
Holst & Thorell (2018)	Internal consistency (Cronbach's alpha) Full scale **α**=.91, inhibition **α**=.77, working memory **α**=0.90 Test-retest reliability (2–3 weeks, bivariate & ICC): r =.68 and .72 for bivariate correlations and between r =.62 and .72 for ICC Interrater reliability: r = .53 for bivariate and r =.49 for the ICC	Convergent validity (ADEXI vs BDEFS r= .48-.72, and correlations between ADEXI scores and scores from 'laboratory' measures of EF r <.30) Discriminant validity, analyses of variance (ANOVAs) were used to study group difference for the ADEXI: classified 85% of the participants in the correct category with a sensitivity of 86% and a specificity of 84% Factorial validity via Factor Analysis	- +	Self-report	2 factors: - Working memory - Inhibition	Mixed sample 202 (adults with ADHD n = 51, adults diagnosed with other psychiatric disorders n = 46, and a non-clinical sample of university students n = 105)
Papers assessing the ECQ (n=1)						
Miranda et al. (2019)	Internal consistency (Cronbach's alpha) Total: **α**=.90 Subscales: Executive Attention **α**=.84 Behavioural Flexibility **α**=.81 Inhibitory Control **α**=.58	Construct Validity via EFA and CFA Convergent Validity: Average Variance Extracted: Executive Attention (.49), Behavioural Flexibility (.38) and Inhibitory Control (.30) Divergent Validity ($p < .001$)	-	Self-report	3 factors - Executive Attention - Behavioural Flexibility - Inhibitory Control	Non-clinical 672 Spanish speaking Argentinians
Papers assessing the Development of a self-report measure of EF from the 200-item Coolidge Axis II Inventory (n=1)						
Coolidge & Griego (1995)	Internal consistency (Cronbach's alpha) Total **α**=.72 Subscales: Decision-making difficulties **α**=.77 Poor planning **α**=.63 Task completion **α**=.66	Construct Validity: via FA	0	Self-report	3 factors: - Decision-making difficulties - Poor planning - Task incompletion	Mixed sample 1,223 non-clinical 17 closed head-injured patients

Note. Abbreviations: ABI = acquired brain injury, AD = Alzheimer's dementia, FTD = frontotemporal dementia, VaD = vascular dementia, EFA = exploratory factor analysis, CFA = confirmatory factor analysis, ICC = intraclass correlation coefficient, PSI = Person Separation Index (equivalent to Cronbach's alpha), NS = not significant, ADEXI = Adult executive functioning inventory, BASC = Behaviour Assessment System for Children, BDEFS = Barkley Deficits in Executive Functioning Scale, DEX = dysexecutive questionnaire, ECQ = Executive Complaints Questionnaire, FAB = Frontal Assessment Battery, BRIEF = Behaviour Rating Inventory of Executive Functions, FBI= Frontal Behaviour Inventory, FrSBe = Frontal Systems Behaviour Scale, GEC = Global Executive Composite (BRIEF), BRI = Behavioural Regulation Index (BRIEF), BIS = Barratt Impulsiveness Scale. MCI = Metacognition Index (BRIEF), NPI-P = Neuropsychiatric Inventory, SDS = Self-Rating Depression Scale, BAI = Beck Anxiety Inventory, DASS-S = Depression Anxiety Stress Scales, GSES = General Self-Efficacy Scale, SWLS = Satisfaction with Life Scale, WCST = Wisconsin Card Sorting Test, TMT-B = Trail Making Test Part B, CAARS-S:L = Conner's Adult ADHD Rating Scales-Self-Report: Long Version, PCRS = Patient Competency Rating Scale, RCFT = Rey Osterrieth Complex Figure Test, T1 = time one, T2 = time two, DSM-IV = Diagnostic and Statistical Manual of Mental Disorders, fourth edition. *Quality ratings based on the four domains of the QUADAS-2: **-** donates a 'low' rating in a domain; **+** indicates a 'high' rating in a domain; 0 indicates unclear or n/a across all four domains and therefore uninterpretable.

Synthesis of results

Psychometric properties across all measures

The psychometric properties of the DEX were explored in seven papers, these included those where adaptations to the original DEX was made. All but two papers used the DEX with neurological samples, with one other using a healthy ageing group and the other a mixed sample. The DEX displayed good levels of reliability across studies for the self and informant versions. The subscales in purely ABI samples also had high internal consistency when mapped against the Stuss model of frontal functions. However, it should be noted that this was a modified version of the DEX, known as the DEX-R where additional items and rewording of the original DEX have been made (Simblett et al., 2017). The test-retest reliability of the DEX was stronger for self-ratings than for informant versions in one paper when these were completed five weeks apart (Hellebrekers et al., 2017). However, a translated version of the DEX demonstrated high test-retest reliability for informant versions in a separate paper completed one month apart (Shinagawa et al., 2007). The two papers used contrasting aetiology samples, one with an ABI sample and the other with an Alzheimer's dementia sample. Both papers scored relatively equally in their quality ratings. The DEX had low to moderate correlations when compared against a separate rating scale, and low correlation with the FAB, a neuropsychological test (Shinagawa et al., 2007). However, only the apathy domain was compared with the rating scale, the NPI apathy subscale, which itself is not an EF rating measure. In addition, the NPI apathy subscale has not itself been robustly subjected to validity testing (Cummings et al., 1994). Only the planning and monitoring subscale of the DEX had a moderate correlation. Moderate correlations were found between the DEX and depression, anxiety and self-efficacy rating scales, however these do not form adequate validity testing as they are not existing valid EF rating scales to be compared to (Shaw, Oei & Sawang, 2015).

Shaw et al. (2015) recognise this limitation recommending concurrent validation against a separate EF rating scale be applied in future research. Both shorter and longer versions of the DEX had good reliability (Shaw et al., 2015; Simblett et al., 2017).

The BRIEF-A was assessed in five papers across different clinical and non-clinical groups. The internal consistency was strong for the total and index scores, however, was more varied across the nine subscales. This was found in both papers which explored the subscale reliability (Ciszewski et al., 2014; Mani et al., 2018). The test-retest reliability was high after one month and very high after two months although some methodological quality issues were raised in both papers (Mani et al., 2018; Rouel et al., 2016). In terms of the validity of the BRIEF-A, the self and informant versions correlated well, but there was little or no correlation when compared against neuropsychological tests (Rouel et al., 2016). Interestingly, when scores were compared again the BRIEF-A manual norms, of those with ADHD only those without hearing impairments correlated well (Hauser et al., 2013). However, this paper had many methodological flaws.

The FrSBe was assessed by four papers, this included different clinical and non-clinical groups. The internal consistency of the subscales varied and was robust for the self and informant total scores. In one paper, the clinical and control samples had similar internal consistency scores across subscales, but the informant version had increased internal consistency scores (Caracuel et al., 2012). However, one of the methodological flaws raised with this paper was that all participants were required to be independent in activities of daily living, and it is questionable whether this limits the generalisability to other ABI groups when assessing EF. Particularly as one of the benefits of such rating scales are in identifying functional difficulties which may not be captured by EF tests. Test-retest reliability was high for total scores and moderate for the subscales, in addition, the total scores correlated well for informant ratings and moderate for self-ratings. These

were for retests completed both one week and three months apart. However, one study produced test-retest statistics following treatment, which could impact on the consistency across time (Velligan et al., 2013). The authors report validity by comparing to a patient competency rating scale and neuropsychological tests (Niemeier et al., 2013; Velligan et al., 2002). Although the correlations were moderate and high with the PCRS, it is not an EF rating scale and therefore does not demonstrate construct validity by comparing to another EF rating scale. It is however a measure of self-awareness of deficits which are more distinct in EF deficits. Verbal fluency was the only test which had moderate correlation with the subscales, but this was just for the apathy and executive function subscales (Velligan et al., 2002). The FLOPS had impressive psychometric properties; however, this has subsequently been revised and developed into the FrSBe.

The FBI and ADEXI were both only assessed by one paper each. They had excellent internal consistency for total scores. The test-retest and interrater reliability was more robust for the FBI. Both had little or no correlation when compared to neuropsychological tests (Holst et al., 2017; Milan et al., 2008). The ADEXI had some correlation with the BDEFS, the BDEFS was validated against models of ADHD and EF (Barkley, 2011). The psychometric properties of the BDEFs were explored by two papers. The Spanish version of the BDEFS had good to excellent internal consistency, its factor structure mapped well onto the original version, although some items loaded onto differing factors to those in the manual (Vélez-Pastrana et al., 2016). A non-clinical sample were recruited via convenience sample and the reference standard was an ADHD rating scale, whilst this may support its validity based on the Barkley model, a more robust validity measure would be comparing it to existing EF rating scales. The Barkley model was based on capturing both the “hot” and “cold” EF components. The Dutch version of the BDEFS found it correlated well with the DEX, and less so with the BIS-II. The DEX and BDEFS

both include subscales viewing EF as multidimensional. Although, the self-regulation subscale showed poor correlation with the 'emotion' items on the DEX (Beerten-Duijkers et al., 2019). The one paper assessing the BASC did use an existing EF measure as their reference standard by comparing scores to the BRIEF-A, which yielded low and moderate positive correlations (Duggan et al., 2018). The psychometric properties for a shortened version of the Coolidge Axis II Inventory were reported in one paper and were less robust, with questionable and acceptable internal consistency reported (Coolidge & Griego, 1995). The applicability of this paper raised concerns due to only including 17 participants with a head injury, despite a large sample size using a non-clinical sample. The psychometric properties of the ECQ development and validation focused on Spanish speaking countries and was limited to non-clinical groups. Its convergent validity was assessed using average variance extracted, which was below the agreement level required across subscales (Cheung & Wang, 2017).

Psychometric properties across participant groups

ABI samples were explored by eleven papers, seven of which were the sole clinical group in the study. Five of these papers examined the original and modified versions of the DEX (Bodenburg et al., 2008; Hellebrekers et al., 2017; Shaw et al., 2015; Simblett et al., 2012; Simblett et al., 2017). Test-retest reliability was greater for self-reports than informants on the original DEX, but this type of reliability was only performed by one paper (Hellebrekers et al., 2017). Two papers by the same authors addressed issues of measurement found within the DEX for ABI samples (Simblett et al., 2012; Simblett et al., 2017). These aimed to develop a more psychometrically robust revised version, and indeed demonstrates internal consistency ranging between adequate and excellent for its revised subscales. Its validity was supported by mapping onto a theoretical model of frontal functions (Stuss, 2011). The quality could have been improved by providing details of a

reference standard. However, all papers assessing the psychometric properties of the DEX with ABI samples would have benefited from either using or providing more details of a reference standard using an existing EF rating scale to establish external validity. Only Shaw et al. (2015) included a validity statistic that went beyond examining construct validity by factor or Rasch analysis, however the clinical groups (ABI and mental health) were combined together in the analysis. ABI samples were also examined in papers using the FrSBe (Caracuel et al., 2012; Carvalho et al., 2013; Niemeier et al., 2013), BRIEF-A (Waid-Ebbs et al., 2012), the FLOPS (Grace et al., 1999) and part of a new EF measure development (Coolidge & Griego, 1995). However, Coolidge and Griego (1995) only included 17 participants as part of their ABI sample finding they scored significantly higher than a control sample. Papers using the FrSBe had issues with bias, and one paper combined mixed neurological aetiologies. The BRIEF-A had good to excellent reliability, however as only one paper was retrieved, there was limited reports of validity statistics applied to ABI. Overall, the DEX and its variants appear to be more robust with ABI samples.

Both papers exploring psychometric properties using a neurodegenerative sample displayed excellent reliability but were less able to evidence validity (Milan et al., 2008; Shinagawa et al., 2007). They showed limited bias and were applicable to similar samples and used different EF rating scales, the FBI and the DEX. Milan et al. (2008) reported that the FBI was able to correctly classify by type of dementia, with the following being correctly classified: 100% fv-FTD, 90.9% AD and 73.3% with VaD (Wilks $k= 0.0945$; $F = 4.317$; $P < 0.0001$). However, this was based on small sample sizes (35 carers for the FTD group) and was not compared against a healthy control group. Additionally, they found that the FBI misclassified 26.7% of those from the VaD group into the FTD group.

Therefore, this 100% statistic may not be informative, and the results presented here should be interpreted with caution.

The two papers examining the BRIEF-A with eating disorders sample had excellent reliability for total scores and the indices, but less so for the nine subscales (Ciszewski et al., 2014; Rouel et al., 2016). Limited validity was demonstrated when comparing to neuropsychological tests, however self and informant versions correlated well with each other. The FrSBe was used with a sample of people diagnosed with schizophrenia, with good reliability and validity (Velligan et al., 2002). The DEX also showed excellent reliability when used with a sample who had mental health conditions (Shaw et al., 2015). Shaw et al. (2015) additionally found that the DEX correctly classified 68.6% of participants in a mixed sample, a quarter of these included participants with a mental health condition, compared to most of the non-clinical participants being correctly classified. They also included neurological samples in this analysis but do not specify the proportion correctly classified.

Two papers which focused on ADHD were retrieved. These assessed the ADEXI and the BRIEF-A, both had excellent reliability. However, the responses on the BRIEF-A by the ADHD sample only correlated well and significantly with the manual for those without hearing impairments compared to those with hearing impairments (Hauser et al., 2013). Holst et al. (2017) combined an ADHD and mental health sample and the ADEXI correctly classified 85% of participants, with sensitivity of 86% and specificity of 84%.

Six papers assessed the psychometric properties of EF rating scales with non-clinical populations. The most robust in its reliability was the DEX-R with a healthy ageing sample for both the self and informant versions (Dimitriadou et al., 2018). The BASC did not correlate well with the BRIEF-A (Duggan et al., 2018). However, in a

separate paper the inhibition subscale of the BRIEF-A was the only one to be valid with a non-clinical sample (Shaw et al., 2015). Test-retest reliability was only explored by one paper, with the BRIEF-A which was highly correlated one month apart (Mani et al., 2018). The BDEFS also had excellent internal consistency with a non-clinical sample, but this was only explored by one paper (Beerten-Duijkers et al., 2019). Miranda et al. (2019) found excellent internal consistency with an Argentinian population, despite not meeting criteria for convergent validity the three factors were confirmed via a CFA (Cheung & Wang, 2017).

Factor structure across measures

Twenty papers report a factor structure using either factor or Rasch analysis which varied across and within the same EF rating scale and participant groups. This included eleven using clinical samples, six using non-clinical samples and three applying factor analysis with mixed samples. These ranged from between two and five factors. Only the two papers using eating disorder samples with the BRIEF-A agreed on their factor structure, which was in line with the manuals two indices (Ciszewski et al., 2014; Rouel et al., 2016). However, a two-factor model did not fit well when a CFA was then applied (Ciszewski et al., 2014). Carvalho et al. (2013) found the original version of the FrSBe to be a good fit, however, they produced an alternative model with a slightly improved fit for their mixed neurological aetiology sample, using a reduced model where eight of the original items were removed. Niemeier et al. (2013) attempted numerous EFA and CFA and did not find these fit the proposed subscales of the FrSBe, although they suggest this may be due to their smaller restricted sample size only including those with moderate and severe TBI. Although five factors were reported in the FBI as accounting for 65% of the variance, over 40% was explained by one factor with only three of the items not loading onto it (excessive jocularity, incontinence and alien hand) (Milan et al., 2008). Vélez-

Pastrana et al. (2016) found the Spanish version of the BDEFS had the same factor structure as the original English version.

For the DEX and its modified versions, there was one paper which reports a two-factor model (Hellebrekers et al., 2017), three which report a three-factor model (Dimitriadou et al., 2018; Shaw et al., 2015; Shinagawa et al., 2007) and one paper reports a four-factor model (Bodenburg et al., 2008). One paper found different factor structures for self-versus informant versions (Hellebrekers et al., 2017). For the BRIEF-A, three papers found a two-factor model (Ciszewski et al., 2014; Rouel et al., 2016; Waid-Ebbs et al., 2012) and one paper confirmed fit with the original subscales (Mani et al., 2018). The FrSBe showed both three factors (Caracuel et al., 2012; Carvalho et al., 2013), and four factors (Niemeier et al., 2013). There were five factors in the FBI (Milan et al., 2008), two factors for the ADEXI (Holst et al., 2017), four for the BASC (Duggan et al., 2018), three factors on the development of a new EF rating scale measure (Coolidge & Griego, 1995) and three for the ECQ (Miranda et al., 2019).

Across clinical and non-clinical groups, the most widely reported factors included initiation, apathy or motivation (seven papers), inhibition (eight papers) and self-regulation (14 papers). The initiation, apathy or motivation, and self-regulatory factors were represented equally across groups, whereas the inhibition factor only appeared in one of the six non-clinical papers in comparison to seven of the eleven clinical papers. The three papers using mixed clinical samples all found an inhibition factor, these papers all had an issue with bias. Although self-regulation was present across studies, in non-clinical papers these were reported as discrete emotional, social or behavioural self-regulation factors, whereas in clinical papers these were typically combined as a behavioural and emotional self-regulation factor. Other factors appearing at least three times across groups were executive cognition, decision making and problem solving, and metacognition. Executive

cognitive factors were generally equally presented across groups. The metacognition factor appeared more so in the papers with clinical groups, and factors relating to decision making and problem solving were more present in the non-clinical papers. The DEX and its variants captured a broad range of factors, both the BRIEF-A and FrSBe included factors in line with their indices/subscales. Other rating scales only included one paper where a factor structure was reported.

Discussion

Summary of evidence

The aim of this systematic review was to evaluate which EF rating measures are the most robust for clinical application across patients with neuropsychological deficits. Twenty-four studies met inclusion for the final review and included the evaluation of the psychometric properties of a range of EF rating scales across different participant groups. Many papers focused on ABI samples, additionally, papers included mental health, neurodegenerative, and ADHD samples. Additionally, the psychometric properties using non-clinical samples were investigated in a number of papers. The rating scales included the DEX, BRIEF-A, FrSBe/FLOPS, FBI, BASC, BDEFS, ECQ and the adapted Coolidge Axis II Inventory.

The majority of papers used the DEX (or its modifications), the BRIEF-A, or the FrSBe. Overall, these measures had acceptable to excellent reliability and validity for both their total and subscale scores. Informant-rating forms appeared to perform better than self-rated versions, particularly for the DEX and FrSBe. There was greater variation in identified subscales and in the psychometric properties across clinical groups, and although psychometric analyses were conducted across measures, only three studies assessed sensitivity / specificity of classification or identification of clinical conditions, with only

one in neurodegenerative conditions and two comparing mental health populations with either ADHD or non-clinical groups (Holst et al., 2017; Milan et al., 2008; Shaw et al., 2015). Internal reliability was the most widely reported type of reliability reported. There were promising results for stability and consistency over different time points, but this approach was used less frequently. Validity was assessed in different ways, some comparing to other EF rating scales, non-EF rating scales and neuropsychological tests. This led to issues in an applicable reference standard being used to establish the external validity of rating scales, as many would have benefited from comparing to an existing EF rating scale measure. Most papers analysed subscale factors, typically finding two to four factors in both clinical and non-clinical groups, which commonly covered initiation, apathy or motivation, inhibition, or self-regulation in general or broken down into cognitive, behavioural and emotional factors. Other factors included metacognition, working memory, and decision making or problem solving.

The variability within the factors thought to underpin the DEX may be due to the different attempts to modify the original version (Bodenburg et al., 2008; Dimitriadou et al., 2018; Shaw et al., 2015; Simblett et al., 2017). Indeed Shaw et al. (2015) compare the original DEX with a revised version incorporating 15 of the 20 items. They completed numerous CFA against four existing factor models reported in the literature (Burgess et al., 1998; Chan, 2001; Mooney et al., 2006; Wilson et al., 1996), finding that these models did not fit the data. A revised version of the DEX with modifications to improve fit incorporated an additional 14 items with four subscales, following analysis revealing the DEX to not be a unidimensional measure (Simblett et al., 2017). However, when the same measure was used with a healthy ageing sample, three subscales were revealed (Dimitriadou et al., 2018).

Returning to the review conducted by Malloy and Grace (2005), the current systematic review supports the robustness of the FrSBe in its reliability and validity across different clinical and non-clinical groups. This review expands on the evidence of the psychometric properties of the BRIEF, as it includes the adult version which had not be devised prior to their review (Roth et al., 2005). The BRIEF-A maintains consistent reliability in the total and index scores, however, was less consistent across the nine subscales. The FBI was again considered to be highly reliable, however this was based on one paper using a dementia sample. Therefore, this review does not extend on its use with different clinical groups. The DEX had limited psychometric properties reported in the review by Malloy and Grace (2005), the current review significantly expands on this. Three papers report on the classification of groups. The FBI was reported to correctly classify based on different types of dementia, discriminating with 100% accuracy frontal-temporal dementia. But this was based on small samples, without a healthy control group and misclassified 26.7% of those with vascular dementia into the frontal-temporal dementia group. Additionally, the ADEXI was reported to correctly classify 85% of participants, with a sensitivity of 86% and a specificity of 84% (Holst & Thorell, 2018). An abbreviated version of the DEX was able to correctly classify 68.6% of cases from clinical groups (Shaw et al., 2015). However, the clinical groups included neurological and mental health samples, and only those in the mental health group scored significantly higher than controls. It was not clear whether those in the neurological group experienced frontal deficits, and therefore may have been more diverse than those in the mental health group in terms of the commonality of cognitive functioning. As found previously, there was again limited reporting of the sensitivity and specificity of EF rating scales used with clinical groups, and therefore remains a gap requiring further research.

Strengths and Limitations

Limitations of this review relate to the exclusion of papers based on language and the requirement for these to be published in peer reviewed journals. Additional papers may have had important contributions towards understanding the psychometric properties of EF rating scales, particularly in regard to their application across diverse contexts. The overarching goal of the review was to identify which EF rating scales were the most robust for clinical application across patients with neuropsychological deficits. This was dependent on a range of these measures being applied broadly across these contexts, which they were not. One research question related to how EF rating scales compare against each other. As each measure was not compared across similar populations and countries/contexts this heterogeneity became problematic to synthesis and compare the stability of the psychometrics, reducing the robustness of the review findings. A further issue with the heterogeneity of EF constructs across studies and rating scale measures provided a further barrier to the synthesis of results. This posed a challenge in interpreting how these apply to EF theory, as each measure should capture the different constructs of EF so that they can distinguish between clinical groups with their differing profiles of strengths and difficulties. Comparing less diverse samples across papers would allow for enhanced understanding of the psychometric quality of rating scales but would be limited in how generalisable these would be in different clinical contexts and conditions. Therefore, this review was not able to easily conclude on the most robust measures due to these limitations with heterogeneity. Robust psychometric statistics required both validity and reliability to be assessed by papers, additional papers reporting on either reliability, validity or factor structure are likely to be more widely available and may have strengthened the ability to understand how these apply to different clinical groups. Furthermore, this requirement limited the retrieval of papers which may have exclusively focussed on the content, concurrent or construct validity of individual rating scales. These

validity statistics were outlined as being useful determinants of their psychometric quality. A strength of the review is the consistent approach taken in the interpretation of reported statistics based on a previous systematic review (Hermans et al., 2011). There currently is not a quality check available explicitly for use with psychometric studies. The QUADAS-2 is available for diagnostic studies and the authors recommend modifying it to suit specific review questions. It has previously been applied to systematic reviews of psychometric studies. Due to the constraints of the thesis, there was no reviewer for the final full-text review and quality checks which were only completed by one researcher which could lead to bias. However, the high level of agreement found between reviewers during the screening phase provides a degree of confidence (Appendix F). The challenges in conceptualising EF meant broader search terms were included.

In terms of the key limitations of the reviewed studies, one challenge was the lack of a reference standard in many of the papers, which usually is available in diagnosable conditions. However, EF already has issues in conceptualisation and measurement, and therefore it is difficult to ascertain what the best reference standard would be. There was no universal reference standard used, varied approaches were taken such as brain imaging, clinical interviews, EF rating scales and neuropsychological tests. Some papers used non-EF rating scales which have limited applicability when establishing whether the index test is a valid measure of EF. Shaw et al. (2015) recommend further research with the DEX assesses its concurrent validity by comparing to another EF rating scale measure. There were varied correlations against neuropsychological tests, however this again could be due to differing components of EF being measured. One study showed strong correlation between a verbal fluency task and the apathy subscale of the FrSBe (Velligan et al., 2002). This is in keeping with the Stuss model in which poor verbal fluency performance may be observed in those reporting poor energisation abilities. Therefore, EF rating measures need

to be more specific in relating tests to specific questionnaire domains. The factors reported in the papers reflect the 'higher-order' EF's more so than 'core' EF processes commonly measured by neuropsychological tests. This supports the use of rating scale measures being used alongside neuropsychological tests to enhance assessment, as opposed to one replacing the other. The recognition of different methods capturing different processes was what initially led to the development of the DEX (Burgess et al., 1998). Furthermore, a huge challenge in the ability to report on the quality of papers with the limited information provided by some of the authors. This meant it could not be determined whether they had met the specified criteria, whilst other papers were more transparent in their reporting of the methodological process and limitations of these. Therefore, this represents a confound in the reporting of quality within the current review. Papers using neurodegenerative samples were rated as being at the lowest risk of bias and most applicable, whereas papers using ADHD, mental health or mixed samples had a high level of bias in at least one domain. The DEX and BRIEF-A studies had lower risks of bias in all but one paper each, whilst papers using the FrSBe scored high in at least one domain. Returning to the extent that measures looked at content, concurrent, construct validity and reliability as outlined in the introduction, different approaches were used to achieve this. Reports of the content validity of rating scales were limited. Where construct validity was reported this was often via a factor analysis. Concurrent validity was reported in some of the papers by comparing to an EF rating scale which has already been validated. However, it was difficult to ascertain what would be a 'gold standard' measure to compare to. Many papers assessed validity through discriminative or convergent validity which have the added value of comparing to neuropsychological tests or looking at the sensitivity and specificity of measures. Whilst these validity types were not reported as being of most importance, they may be useful for clinicians in determining how a particular rating scale overlaps with such

tests or how well they would identify an EF impairment. All papers were required to report at least one reliability statistic which mostly included internal consistency with some additionally reporting test-retest reliability. The secondary purpose of utilising factor analysis or item response theory was applied in a number of papers.

Implications for future research

The review of specific measures was limited due to the restricted application of these measures across different clinical groups and contexts. The review by clinical group was again limited due to insufficient numbers of studies using one measure across different groups or multiple measures across matched groups. This led to the issue with heterogeneity and could be addressed in two ways. Firstly, by conducting further research using more systematic data collection applying each measure across different clinical groups and contexts. Secondly, by a comparison of different measures with single clinical or non-clinical groups. If this research had been done, then this would have allowed us to systematically compare and contrast between different measures for the same clinical group, as well as examining the performance of a specific measure across multiple contexts and populations. Furthermore, it would enhance our understanding of the EF constructs reported and how these vary across rating scale measures and clinical groups.

Conclusions

To conclude, the psychometric properties of a range of EF rating scale measures have been studied. The DEX, BRIEF-A and FrSBe were the most widely used across clinical and non-clinical groups with robust reliability and validity statistics reported. Papers would have benefited from assessing concurrent validity by comparing against existing EF rating scales. Additionally, further evidence to demonstrate their consistency over time would improve robustness. The challenges of the conceptualisation of EF and its underlying sub-constructs has perhaps led to there being no universally agreed factor

structure present across measures. Interestingly, non-clinical groups also show multiple underlying factors. Attempts have been made to adapt existing measures, such as the DEX, to measure conceptual theoretical frameworks such as those proposed by Stuss. Further research in the development of a quality assessment suited to psychometric studies would be beneficial for future reviews of this nature.

References

Barkley, R. A. (2011). *Barkley Deficits in Executive Functioning Scale (BDEFS).* New York: Guilford Press.

Barkley, R. A., & Fischer, M. (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, *36*(2), 137-161. doi:10.1080/87565641.2010.549877

Barkley, R. A., & Murphy, K. R. (2010). Impairment in occupational functioning and adult ADHD: the predictive utility of executive function (EF) ratings versus EF tests. *Archives of Clinical Neuropsychology*, *25*(3), 157-173. doi:10.1093/arclin/acq014

Barrash, J., Tranel, D., & Anderson, S. W. (2000). Acquired personality disturbances associated with bilateral damage to the ventromedial prefrontal region. *Developmental Neuropsychology*, *18*(3), 355-381. doi:10.1207/S1532694205Barrash

Beerten-Duijkers, J. C., Vissers, C. T., Rinck, M., Barkley, R. A., & Egger, J. I. (2019). Dutch translation and adaptation of the Barkley Deficits in Executive Functioning Scale (BDEFS). *Archives of Clinical Psychiatry (São Paulo)*, *46*(4), 89-96. doi:10.1590/0101-60830000000202

Bodenburg, S., & Dopslaff, N. (2008). The Dysexecutive Questionnaire advanced: Item and test score characteristics, 4-factor solution, and severity classification. *The Journal of Nervous and Mental Disease*, *196*(1), 75-78. doi:10.1097/NMD.0b013e31815faa2b

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, *15*(4), rm4. doi:10.1187/cbe.16-04-0148

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136-136.

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. doi:10.1017/S1355617798466037

Caracuel, A., Verdejo-García, A., Fernández-Serrano, M. J., Moreno-López, L., Santago-Ramajo, S., Salinas-Sánchez, I., & Pérez-García, M. (2012). Preliminary validation of the Spanish version of the Frontal Systems Behavior Scale (FrSBe) using Rasch analysis. *Brain Injury*, *26*(6), 844-852. doi:10.3109/02699052.2012.655365

Carvalho, J. O., Ready, R. E., Malloy, P., & Grace, J. (2013). Confirmatory factor analysis of the frontal systems behavior scale (FrSBe). *Assessment*, *20*(5), 632-641. doi:10.1177/1073191113492845

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

Cheung, G. W., & Wang, C. (2017). Current approaches for assessing convergent and discriminant validity with SEM: Issues and solutions. In *Academy of Management Proceedings* (Vol. 2017, No. 1, p. 12706). Briarcliff Manor, NY 10510: Academy of Management.

Cicerone, K., Levin, H., Malec, J., Stuss, D., & Whyte, J. (2006). Cognitive rehabilitation interventions for executive function: moving from bench to bedside in patients with traumatic brain injury. *Journal of Cognitive Neuroscience*, *18*(7), 1212-1222. doi:10.1162/jocn.2006.18.7.1212

Ciszewski, S., Francis, K., Mendella, P., Bissada, H., & Tasca, G. A. (2014). Validity and reliability of the Behavior Rating Inventory of Executive Function—Adult Version in a clinical sample with eating disorders. *Eating Behaviors*, *15*(2), 175-181. doi:10.1016/j.eatbeh.2014.01.004

Clark-Carter, D. (2009). *Quantitative psychological research: The complete student's companion*. Psychology press.

Coffman, T. P. (2014). Psychometric properties of the Barkley Deficits in Executive Functioning Scale (BDEFS) in a college student population. Florida State University. Retrieved from: https://diginole.lib.fsu.edu/islandora/object/fsu%3A185371

Coolidge, F. L., & Griego, J. A. (1995). Executive functions of the frontal lobes: Psychometric properties of a self-rating scale. *Psychological Reports*, *77*(1), 24-26. doi:10.2466/pr0.1995.77.1.24

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, *44*(12), 2308-2308. doi:10.1212/WNL.44.12.2308

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System.* San Antonio: Psychological Corporation.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168. doi:10.1146/annurev-psych-113011-143750

Dimitriadou, M., Michaelides, M. P., Bateman, A., & Constantinidou, F. (2018). Measurement of everyday dysexecutive symptoms in normal aging with the Greek version of the dysexecutive questionnaire-revised. *Neuropsychological Rehabilitation*, 1-20. doi:10.1080/09602011.2018.1543127

Dubois, B., Slachevsky, A., Litvan, I., & Pillon, B. F. A. B. (2000). The FAB: A frontal assessment battery at bedside. *Neurology*, *55*(11), 1621-1626. doi:10.1212/WNL.55.11.1621

Duggan, E. C., Garcia-Barrera, M. A., & Müller, U. (2018). Derivation, replication, and validity analyses of a screener for the behavioral assessment of executive functions in young adults. *Assessment*, *25*(7), 867-884. doi:10.1177/1073191116670511

Fermanian, J. (2005). Validation of assessment scales in physical medicine and rehabilitation: how are psychometric properties determined? *Ann Readapt Med Phys, 48*(6), 281–287. doi:10.1016/j.annrmp.2005.04.004

Garcia-Barrera, M. A., Karr, J. E., & Kamphaus, R. W. (2013). Longitudinal applications of a behavioral screener of executive functioning: Assessing factorial invariance

and exploring latent growth. *Psychological Assessment*, *25*(4), 1300. doi.org/10.1037/a0034046

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238. doi:10.1076/chin.6.3.235.3152

Grace, J., & Malloy, P. F. (2001). *Frontal Systems Behavior Scale (FrSBe): Professional Manual.* Lutz, FL: Psychological Assessment Resources, Inc.

Grace, J., Stout, J. C., & Malloy, P. F. (1999). Assessing frontal lobe behavioral syndromes with the frontal lobe personality scale. *Assessment*, *6*(3), 269-284. doi:10.1177/107319119900600307

Hauser, P. C., Lukomski, J., & Samar, V. (2013). Reliability and validity of the BRIEF-A for assessing deaf college students' executive function. *Journal of Psychoeducational Assessment*, *31*(4), 363-374. doi:10.1177/0734282912464466

Hellebrekers, D., Winkens, I., Kruiper, S., & Van Heugten, C. (2017). Psychometric properties of the awareness questionnaire, patient competency rating scale and Dysexecutive Questionnaire in patients with acquired brain injury. *Brain Injury*, *31*(11), 1469-1478. doi:10.1080/02699052.2017.1377350

Hermans, H., van der Pas, F. H., & Evenhuis, H. M. (2011). Instruments assessing anxiety in adults with intellectual disabilities: A systematic review. *Research in Developmental Disabilities*, *32*(3), 861-870. doi:10.1016/j.ridd.2011.01.034

Holst, Y., & Thorell, L. B. (2018). Adult executive functioning inventory (ADEXI): Validity, reliability, and relations to ADHD. *International Journal of Methods in Psychiatric Research*, *27*(1), e1567. doi:10.1002/mpr.1567

Isquith, P. K., Roth, R. M., & Gioia, G. (2013). Contribution of rating scales to the assessment of executive functions. *Applied Neuropsychology: Child*, *2*(2), 125-132. doi.org/10.1080/21622965.2013.748389

Kertesz, A., Davidson, W., & Fox, H. (1997). Frontal behavioral inventory: Diagnostic criteria for frontal lobe dementi. *Canadian Journal of Neurological Sciences*, *24*(1), 29-36. doi:10.1017/S0317167100021053

Litwin, M. S. (1995). *How to measure survey reliability and validity* (Vol. 7). Sage.

Malloy, P., & Grace, J. (2005). A review of rating scales for measuring behavior change due to frontal systems damage. *Cognitive and Behavioral Neurology*, *18*(1), 18-27. doi: 10.1097/01.wnn.0000152232.47901.88

Manchester, D., Priestley, N., & Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury*, *18*(11), 1067-1081. doi.org/10.1080/02699050410001672387

Mani, A., Ghelijkhani, S., Haghighat, R., Ahmadzadeh, L., Chohedri, E., & Heydari, S. T. (2018). Validity and reliability of the Persian version of the self-report form of Behavior Rating Inventory of Executive Function-Adult version (BRIEF-A). *Shiraz E-Medical Journal*, *19*(2). doi:10.5812/semj.14295

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function?. *Journal of the International Neuropsychological Society*, *16*(3), 495-505. doi:10.1017/S1355617710000093

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. doi:10.3102/0013189X018002005

Mías, C. D., Ruiz, A. R., Causse, M. B., & Verónica, M. (2017). Detección de Deterioro Cognitivo y factores demográficos asociados en adultos mayores de 50 años de la Región del Alto Valle: Neuquén y Río Negro [Cognitive Impairment detection and demographic factors associated in adults over 50 years of the Upper Valley Region: Neuquén and Rio Negro]. *Revista Argentina de Neuropsicología, 30,* 34–49. https://www.researchgate.net/profile/Carlos_Mias/publication/319942673_Deteccion_de_Deterioro_Cognitivo_y_factores_demograficos_asociados_en_adultos_mayores_de_50_anos_de_la_Region_del_Alto_Valle_Neuquen_y_Rio_Negro/links/59c2a8a7458515af3060c0d4/Deteccion-de-Deterioro-Cognitivo-y-factores-demograficos-asociados-en-adultos-mayores-de-50-anos-de-la-Region-del-Alto-Valle-Neuquen-y-Rio-Negro.pdf

Milan, G., Lamenza, F., Iavarone, A., Galeone, F., Lore, E., De Falco, C., ... & Postiglione, A. (2008). Frontal Behavioural Inventory in the differential diagnosis of dementia. *Acta Neurologica Scandinavica*, *117*(4), 260-265. doi:10.1111/j.1600-0404.2007.00934.x

Miranda, A. R., Rivadero, L., Serra, S. V., & Soria, E. A. (2019). Multidomain self-report assessment of fronto-executive complaints in Spanish-speaking adults. *Psychology & Neuroscience*. doi:10.1037/pne0000187

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, *151*(4), 264-269. doi:10.1371/journal.pmed1000097

Niemeier, J. P., Perrin, P. B., Holcomb, M. G., Nersessova, K. S., & Rolston, C. D. (2013). Factor structure, reliability, and validity of the Frontal Systems Behavior Scale (FrSBe) in an acute traumatic brain injury population. *Rehabilitation Psychology*, *58*(1), 51. doi:10.1037/a0031612

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology*, *51*(6), 768-774. doi:10.1002/1097-4679(199511)51:6<768::AID-JCLP2270510607>3.0.CO;2-1

Pessoa, L. (2009). How do emotion and motivation direct executive control?. *Trends in Cognitive Sciences*, *13*(4), 160-166. doi:10.1016/j.tics.2009.01.006

Prigatano, G. P., Fordyce, D., Zeiner, H., Rouechel, J., Pepping, M., & Wood, B. (1986). *Neuropsychological rehabilitation after brain injury*. Baltimore, MD: Johns Hopkins University Press.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior assessment system for children–Third edition (BASC-3).* Bloomington, MN: Pearson.

Roth R, Isquith P, Gioia G. (2005). *Behavior Rating Inventory of Executive Function-Adult Version.* Lutz, FL: PAR

Rouel, M., Raman, J., Hay, P., & Smith, E. (2016). Validation of the Behaviour Rating Inventory of Executive Function–Adult Version (BRIEF-A) in the obese with and without binge eating disorder. *Eating Behaviors*, *23*, 58-65. doi:10.1016/j.eatbeh.2016.07.010

Shallice, T. I. M., & 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

Shaw, S., Oei, T. P., & Sawang, S. (2015). Psychometric validation of the Dysexecutive Questionnaire (DEX). *Psychological Assessment*, *27*(1), 138. doi:10.1037/a0038195

Shinagawa, Y., Nakaaki, S., Hongo, J., Murata, Y., Sato, J., Matsui, T., ... & Furukawa, T. A. (2007). Reliability and validity of the Japanese version of the Dysexecutive Questionnaire (DEX) in Alzheimer's disease: Validation of a behavioral rating scale to assess dysexecutive symptoms in Japanese patients with Alzheimer's disease. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, *22*(10), 951-956. doi:10.1002/gps.1768

Simblett, S. K., Badham, R., Greening, K., Adlam, A., Ring, H., & Bateman, A. (2012). Validating independent ratings of executive functioning following acquired brain injury using Rasch analysis. *Neuropsychological Rehabilitation*, *22*(6), 874-889. doi:10.1080/09602011.2012.703956

Simblett, S. K., Ring, H., & Bateman, A. (2017). The Dysexecutive Questionnaire Revised (DEX-R): An extended measure of everyday dysexecutive problems after acquired brain injury. *Neuropsychological Rehabilitation, 27*(8), 1124-1141. doi.org/10.1080/09602011.2015.1121880

Stuss, D. T. (2011). Functions of the frontal lobes: Relation to executive functions. *Journal of the International Neuropsychological Society*, *17*(5), 759-765. doi:10.1017/S1355617711000695

Stuss, D. T., Benson, D. F., Weir, W. S., Naeser, M. A., Lieberman, I., & Ferrill, D. (1983). The involvement of orbitofrontal cerebrum in cognitive tasks. *Neuropsychologia, 21*(3), 235-248. doi:10.1016/0028-3932(83)90040-4

Thorell, L. B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology*, *33*(4), 536-552. doi:10.1080/87565640802101516

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct?. *Journal of Child Psychology and Psychiatry*, *54*(2), 131-143. doi.org/10.1111/jcpp.12001

Vélez-Pastrana, M. C., González, R. A., Rodríguez Cardona, J., Purcell Baerga, P., Alicea Rodríguez, Á., & Levin, F. R. (2016). Psychometric properties of the Barkley Deficits in Executive Functioning Scale: A Spanish-language version in a community sample of puerto rican adults. *Psychological Assessment*, *28*(5), 483. doi:10.1037/pas0000171

Velligan, D. I., Ritch, J. L., Sui, D., DiCocco, M., & Huntzinger, C. D. (2002). Frontal Systems Behavior Scale in schizophrenia: relationships with psychiatric symptomatology, cognition and adaptive function. *Psychiatry Research*, *113*(3), 227-236. doi:10.1016/S0165-1781(02)00264-0

Waid-Ebbs, J. K., Wen, P. S., Heaton, S. C., Donovan, N. J., & Velozo, C. (2012). The item level psychometrics of the Behaviour Rating Inventory of Executive Function-adult (BRIEF-A) in a TBI sample. *Brain Injury*, *26*(13-14), 1646-1657. doi:10.3109/02699052.2012.700087

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., ... & Bossuyt, P. M. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529-536. doi:10.7326/0003-4819-155-8-201110180-00009

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome: BADS*. London: Pearson.

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(1), 3-24.

Wu, M., Tam, H. P., & Jen, T. H. (2016) Classical Test Theory. In: *Educational Measurement for Applied Researchers* (pp. 73-90). Springer, Singapore. doi:10.1007/978-981-10-3302-5_5

Chapter 2

Bridging chapter

Bridging Chapter

The systematic review established a range of valid and reliable rating scales used in the measurement of executive function. These could provide an ecologically valid alternative or supplement to traditional neuropsychological testing. The most widely used measure was the DEX including those where modifications to the DEX were made. Most of the reviewed papers assessed psychometric properties of EF rating scales using different clinical groups. Papers also used non-clinical populations to assess psychometric properties with individual variations being found. The review highlighted inconsistency in the underlying subconstructs in the measurement of EF, and differences in the rating measures subscales to reflect this. The Stuss model fractionates frontal functions into four distinct components. This has benefits over unitary models to capture the wide-ranging difficulties that can be experienced.

The Dysexecutive Questionnaire (DEX: Burgess, Alderman, Evans, Emslie & Wilson, 1998) was subjected to modifications based on psychometric investigations to produce a conceptually sound measure of dysexecutive problems (Simblett & Bateman, 2011; Simblett et al., 2017). This revised version of the DEX includes subscales mirroring the Stuss model. Two papers were retrieved in the review assessing the DEX-R, demonstrating it to be a valid and reliable measure with a brain-injured and healthy ageing population (Dimitriadou, Michaelides, Bateman, & Constantinidou, 2018; Simblett et al., 2017). Papers assessing the DEX-R would have benefited from the use of a reference standard to further improve quality. Indeed, this was a recommendation by Shaw, Oei and Sawang (2015) who suggested future research on the DEX should assess concurrent validation by comparing to another validated EF rating scale. Also, test-retest reliability may further enhance the psychometric properties of the DEX-R.

What is known so far about the DEX-R is based on clinical populations. This provides useful information for clinicians working with these client groups however tells us less about how it is applied to non-clinical populations. There is known individual variation of dysexecutive problems in non-clinical samples (Chan, 2001; Miyake & Friedman, 2012). Any construct is going to vary in the population and share some overlap with clinical groups. Whether certain items on the DEX-R are more likely to be endorsed by clinical or non-clinical groups is not yet known. Understanding the extent that these behaviours are present in a non-clinical population would support clinicians to determine what scores are significantly unlikely and could indicate a clinical impairment. Further analysis with non-clinical populations may assist in understanding the impact of individual differences which may contribute to variability in the factor structure of dysexecutive problems. A key clinical challenge is whether a person being assessed may have had some of the behavioural characteristics associated with these difficulties prior to their injury. Therefore, understanding the most commonly reported characteristics in the general population could contribute to what is classed as clinical. This could support clinicians and researchers to understand what is being measured in addition to any frontal deficits.

The empirical paper will therefore assess the measurement properties of the DEX-R further with a non-clinical population by assessing its test-retest and internal reliability, and validity. Further exploration of the factor structure of the DEX-R will be conducted using factor analysis and Rasch analysis to investigate whether the sub-constructs on which it was based with the Stuss model continue to be found in a non-clinical population. Secondary analysis will establish influences of mood and demographic factors such as ageing.

Chapter 3

Empirical paper

Prepared for submission to *Neuropsychological Rehabilitation*

Psychometric Properties of The Revised Dysexecutive Questionnaire (DEX-R) In A Non-Clinical Population

Hannah Wakely[1], Ratko Radakovic[2], Andrew Bateman[3], Sara Simblett[4], Jessica Fish[5] and Fergus Gracey[1]

[1]Department of Clinical Psychology, University of East Anglia, United Kingdom; [2]School of Health Sciences, Faculty of Medicine and Health Sciences, University of East Anglia, United Kingdom. Norfolk and Norwich University Hospital, Norwich, United Kingdom. The Euan MacDonald Centre for Motor Neurone Disease, University of Edinburgh, Edinburgh, United Kingdom. Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, United Kingdom. Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, United Kingdom; [3]School of Health and Social Care, University of Essex; [4]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London; [5]Department of Clinical Neuropsychology and Clinical Health Psychology, The University of Glasgow.

Correspondence should be addressed to Fergus Gracey, Department of Clinical Psychology, Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ.
E-mail: f.gracey@uea.ac.uk

Total word count: 10604

Word count excluding abstract, table and references: 7485

Abstract

Aims: The aim of this study was to assess the psychometric properties of the revised version of the Dysexecutive Questionnaire (DEX-R) with a non-clinical sample.

Methods: The study was hosted online, with 140 participants completing the DEX-R, GAD-2 and PHQ-2. A proportion also completed the FrSBe, with some additionally completing the DEX-R again three weeks later. Correlations of demographic factors and symptoms of anxiety and depression were conducted. Rasch and factor analysis were also used to explore underlying subconstructs.

Results: The measures did not display normal distributions, and so transformations and non-parametric statistics were applied. The DEX-R correlated highly with the FrSBe, indicating sound concurrent validity. Internal consistency, split-half reliability and test-retest reliability were excellent. Age and symptoms of depression and anxiety correlated with DEX-R scores, with older age associated with less dysexecutive problems. The Rasch analysis confirmed the multidimensionality of the rating scale, and a three-factor structure was found relating to activation-self-regulatory, cognitive and social-emotional processes. Frequencies of responses on DEX-R items varied, many were not fully endorsed.

Conclusion: Interpretations of dysexecutive problems should also consider mood and individual variation. Comparison to clinical groups could identify what constitutes as clinical levels of dysexecutive symptoms.

Keywords: Dysexecutive problems; Rating Scales; Validity; Reliability

Introduction

According to the charity Headway, there were an estimated 348,453 hospital admissions due to acquired brain injury (ABI) in 2016 - 2017 in the United Kingdom (Headway, 2018). An ABI is an acute injury to the brain that happens after birth. It is 'acquired' in the essence that the person did not have this neurological injury prior to the event. It, therefore, does not include those of a progressive nature nor those which have a genetic predisposition. The damage can be classified as either focal (localised) or non-focal (diffuse) and includes both traumatic and non-traumatic brain injuries (Teasell, 2007; Turner-Stokes, 2003). Turner-Stokes (2003) explains how ABI can arise through "trauma, vascular accident (e.g. stroke), cerebral anoxia, other toxic or metabolic insult (e.g. hypoglycemia), infection (e.g. encephalitis) or other inflammation (e.g. vasculitis)" (p14). These conditions can result in physical, cognitive, communication and emotional difficulties (Wilson, Gracey, Evans & Bateman, 2009). Those who have sustained a traumatic brain injury often present with difficulties associated with frontal lobe function (McDonald, Flashman, & Saykin, 2002), this is due to the size, structure and location making the frontal lobes particularly vulnerable following road traffic accidents and assaults (Cicerone, Levin, Malec, Stuss, & Whyte, 2006; Levin, et al., 1987). Research on the structure and function of the frontal areas of the brain highlights several roles in cognitive, behavioural and emotional processes. These include flexible thinking, planning, monitoring, social behaviour, decision making, initiation, inhibition and emotional regulation (Lezak, 1995). The term dysexecutive problems has been used to describe difficulties with these functions, which can have a profound impact on a person's level of independence resulting in challenges in day-to-day life (Hanks, Rapport, Millis, & Deshpande, 1999). They are thought to affect around 40% of people who have a stroke (Hoffmann & Schmitt, 2006; Pohjasvaara et al., 2002).

Different theoretical models have been proposed to account for the kinds of problems observed in people with acquired damage to the frontal lobes. Early models best understood these as being a unitary process (Baddeley & Hitch, 1974; Grafman, 1989; Norman & Shallice, 1986; Pribram, 1960). However, evidence from the clinical and research literature began to consider how various frontal functions were, in fact, dissociable due to differences in presenting functional difficulties reported by individuals (Damasio, Tranel & Damasio, 1991; Stuss, 2007). Also, the focus on such processes being executed solely in the frontal lobes shifted to the recognition of the involvement of wider neural networks and circuits (Burgess, 2004; Fuster, 2008; Stuss, 2011). Despite these shifts in understanding, there continue to be discrepancies in operationally defining these processes. Although, most models recognise some overlaps being present in their unity and diversity (Miyake & Friedman, 2012).

There have also been challenges in the measurement of dysexecutive problems, with various different approaches to their assessment. Neuropsychological tests can be time-consuming requiring complex interpretations. Additionally, they have been criticised for lacking ecological validity due to testing in a well-structured environment with cues being provided (Burgess, Alderman, Evans, Emslie & Wilson, 1997; Damasio et al., 1991; Eslinger & Damasio, 1985). As a result, these tests often fail to highlight difficulties in this area despite reports of challenges in day-to-day life (Shallice & Burgess, 1991; Stuss & Benson, 1983). To overcome this, self-report measures have been developed to capture challenges faced in everyday life. A limitation of using these with people with dysexecutive problems is the issue of reduced self-awareness, meaning they may be more likely to underreport such difficulties (Simblett, Ring & Bateman, 2017). Informant versions are available for those close to the person to be able to corroborate or assess the discrepancy compared to the self-report version. Different rating scales have been

developed and are available for clinical use, however, there are issues in the standardisation and interpretation of scores. This is because the nature of these difficulties, such as decision making, perseveration and flexibility, could lead to issues with the reliability of item responses.

The Dysexecutive Questionnaire (DEX; Burgess et al., 1998) forms part of the Behavioural Assessment of the Dysexecutive Syndrome (BADS: Wilson, Alderman, Burgess, Emslie & Evans, 1996). The DEX is a self-report measure of dysexecutive problems, designed to predict everyday difficulties. There are 20 items measuring behavioural, cognitive, motivational and emotional changes from pre-morbid functioning generating in a single score.

Simblett and Bateman (2011) assessed the psychometric properties of the DEX using item response theory by deploying Rasch analysis techniques. Their analysis suggested the DEX not to be a unidimensional measure of dysexecutive problems, instead capturing underlying sub-constructs thought to underpin these difficulties. Therefore, a total score on self-report measures may not best capture these challenges. In further research, Simblett et al. (2017) made amendments to the wording of some of the items in the DEX as well as including an additional 14 items to expand its measurement to incorporate Stuss (2011) proposed categories of dysexecutive problems. After applying Rasch techniques, data from a clinical sample suggested the revised version of the DEX mapped onto the Stuss model capturing four separate sub-constructs of executive cognitive functions, metacognition, activation and, behavioural and emotional self-regulation. Such development has useful clinical applications to highlight specific areas of strength and difficulties which can assist in diagnosis, neuropsychological formulation or become a focus for a person's individual rehabilitation plan. For example, based on Stuss's (2011) theoretical model, whether goal management or emotional regulation strategies would be

most beneficial. The DEX-R has some additional evidence of psychometric quality when applied to healthy ageing and mental health samples (Dimitriadou et al., 2018; Loschiavo-Alvares et al., 2013). Although neither factor analysis with these groups yielded factors that aligned with Stuss's model.

Neurological disorders are known to contribute to reports of dysexecutive problems, however, additional understandings of how individual variation may manifest is useful for clinicians to understand. Research has highlighted individual variations in reported levels of dysexecutive problems in non-clinical populations as measured by the DEX questionnaire (Chan, 2001). One variable relates to age. Normal ageing processes have been associated with a decline in various cognitive functions associated with prefrontal areas (Van Petten et al., 2004; West, 1996). This may be more prominent in cognitive changes related to the dorsolateral regions, with less change from normal ageing being found in ventromedial areas thought to underpin the emotional processing aspects of dysexecutive problems (MacPherson, Phillips, & Della Sala, 2002). The relationship between normal ageing and dysexecutive problems may have implications in the management of activities of daily living. Literature also suggests the prefrontal regions of the brain mature later than more posterior areas, with development continuing throughout adolescence and towards a person's early 30's (Barkley, 2012; Coffman, 2014). The role of the ventromedial prefrontal cortex (VMPFC) has been particularly implicated in social and emotional processing (Burnett, Bird, Moll, Frith & Blakemore, 2009; Pfeifer et al., 2011; Sebastian et al., 2011). This could have implications in the social and emotional functions associated with the VMPFC being less developed in younger people. Negative affect has been found to mediate the increased reports of dysexecutive problems in younger people (Gerstorf, Siedlecki, Tucker-Drob, & Salthouse, 2008). Correlations have been found between dysexecutive and anxiety and depression symptoms, which may relate to

cognitive variability in mood (Shaw, Oei & Sawang, 2015). In addition, it has been proposed that the development of working memory abilities arise later than processes relating to set-shifting/cognitive flexibility between adolescence and young adulthood (Huizinga, Dolan, & van der Molen, 2006). Another factor contributing to variation in reports of dysexecutive problems includes education level (Faria, Alves, & Charchat-Fichman, 2015; Foss et al., 2013).

There is currently a gap in the literature relating to the psychometric properties of the DEX-R with a non-clinical population. It would be useful to establish variations within the DEX-R, due to reports of dysexecutive problems in non-clinical populations with potential individual differences contributing. Additional benefits in collating this data would be used in normative data to allow comparison in how an individual may be expected to perform at a given age if they had no prior condition or injury.

Research Questions

The proposed study aimed to assess the psychometric properties of the DEX-R further. The primary research questions were:

1. What are the measurement properties of the DEX-R within a non-clinical population?
   1a. Is the DEX-R a reliable measure of dysexecutive problems?
   1b. Is the DEX-R a valid measure of dysexecutive problems when compared to an existing valid self-report measure?
2. Does the DEX-R perform as an interval level measure as established by item response theory?
3. What is the factor structure of the DEX-R in a non-clinical population?
   3a. Does the factor structure align with the Stuss model?

In addition, there were secondary research questions:

4. What are the effects of demographic and mood variables on DEX-R and DEX-R subscale performance?

   4a. What are the effects of age on DEX-R and DEX-R subscale performance?

Method

Design

Quantitative methods were used to test the research questions utilising a within-subjects cross-sectional design. Participants took part in all aspects of the study by completing all questionnaire measures as well as the option to complete the test-retest phase at a second timepoint.

Participants

The study aimed to capture a broad sample of the population; participants were only required to be aged 18 years of age or over. As the study investigated whether there are changes based on age, there was no upper age limit. Questions relating to health were included as part of the study to allow for monitoring whether clinical factors explained variance in the data, should this have arisen. Participants were recruited into the study online via a snowball sampling recruitment method whereby information about the study was distributed online through the research team's networks, including social media. Separate sample size estimates were calculated for each question, the largest requirement being for correlation analyses, requiring at least 109 participants. This was calculated using G* Power 3.1.9 (Faul & Erdfelder, 1992), with power set at 0.9 to detect a medium effect size and probability was set at .05. This calculation was repeated for the multiple regression analysis with the addition of there being four variables included in the model,

which indicated 82 participants were required for this analysis. The literature was consulted for the required sample size for other analyses.

Measures

Dysexecutive Questionnaire-Revised

The DEX-R is a 37-item questionnaire measuring dysexecutive problems which were developed following research conducted by Simblett and Bateman (2011) on the DEX Questionnaire (Burgess et al., 1998). A Rasch analysis found the DEX Questionnaire to not be a unidimensional measure of dysexecutive problems, suggesting it measures more than one construct. An additional 14 items were included and rewording of the DEX was made to improve fit to the Stuss model, resulting in the DEX-R. The psychometric properties of the DEX-R have been explored with a clinical sample (Simblett et al., 2017). This showed good internal reliability. The measurement of underlying sub-constructs of dysexecutive problems appeared to map well onto the Stuss model, namely, activation regulatory functions, behavioural-emotional self-regulatory functions, metacognitive functions and executive cognitive functions. These four terms are also used to name the differing subscales within the DEX-R. It is measured using a Likert scale, with response options within the DEX-R including: “Very often”, “Fairly often”, “Sometimes”, “Occasionally”, and “Never”. These responses are coded as 0 (*never*) to 4 (*very often*), with higher scores indicating greater reports of dysexecutive problems.

Frontal Systems Behaviour Scale (FrSBe)

The FrSBe is an already validated measure of the self-report of dysexecutive problems which has been normed against non-clinical samples, responses therefore formed part of the concurrent validity testing phase (Grace & Malloy, 2001). Other measures were considered however these were either not as well validated or the subscales did not align as

well with the DEX-R. Responses are coded as 1 (a*lmost never*) to 5 (*almost always*), however, reverse scoring is applied to a selection of items. Higher scores indicate more reported dysexecutive problems.

The FrSBe includes 46-items generating a total score or split across three sub-systems: executive dysfunction, apathy and disinhibition. It is for use with those aged between 18 – 95 years old. It takes approximately 10 minutes to administer and 15 minutes for scoring. Research has demonstrated high internal consistency for total score and subscale scores ranging from 0.78 – 0.94 in neurological, mental health, non-clinical samples (Grace & Malloy, 2001; Stout, Ready & Grace, 2003; Velligan, Ritch, Sui, DiCocco & Huntzinger, 2002). Construct validity and factor analysis supports the three factors of apathy, executive function and disinhibition thought to underpin the measure in various samples (Carvalho, Ready, Malloy, & Grace, 2013; Grace, Stout, Malloy, 1999; Stout et al., 2003). Coefficients are reported from a normative sample, 0.92 for the total score, 0.78 for apathy, 0.80 for disinhibition and 0.87 for executive subscales (Malloy & Grace, 2005).

The Generalized Anxiety Disorder Scale

The Generalized Anxiety Disorder scale (GAD-2) is a short, self-report measure of anxiety with a sensitivity of 65% and specificity of 88% for any anxiety disorder (Skapinakis, 2007). It asks participants how often in the past two weeks have two criteria occurred: “Feeling nervous, anxious or on edge” and “Not being able to stop or control worrying”. It is measured using a Likert scale, with response options “Not at all”, “Several days”, “More than half the days” and “Nearly every day”. These are coded as 0 (*not at all*) to 3 (*nearly every day*). Higher scores suggest an increased presence of anxiety symptoms, with a clinical cut-off equal to or above three points (Skapinakis, 2007).

The Patient Health Questionnaire

The Patient Health Questionnaire (PHQ-2) is a short, self-report measure with a sensitivity of 79% and specificity of 86% for detecting symptoms of depression (Löwe, Kroenke & Gräfe, 2005). The PHQ-2 is formed of two questions asking how often individuals have experienced the following in the past two weeks: “Little interest or pleasure in doing things” and “Feeling down, depressed and hopeless”. It is also measured using a Likert scale, with response options “Not at all”, “Several days”, “More than half the days” and “Nearly every day”. These are coded as 0 (*not at all*) to 3 (*nearly every day*). Higher scores suggest an increased presence of depressive symptoms with a clinical cut-off with a score of three or above (Löwe et al., 2005).

Demographic Questions

Demographic questions included questions on age, gender, highest education level, years of education and ethnicity.

Health Questions

Additional questions regarding health included: “Have you ever been formally diagnosed or hospitalised for the following conditions?”, and included neurodegenerative conditions (e.g. dementia, Parkinson’s disease, Huntington’s disease, Multiple sclerosis), neurodevelopmental conditions (e.g. autism spectrum disorder, attention deficit disorder (ADHD), learning disability), acquired brain injury, stroke, and mental health conditions (e.g. Bipolar disorder, Schizophrenia or Psychotic Illness). An “other” or “prefer not to answer” option was also available. These were included to monitor whether clinical factors explained variance in the data.

Procedure

The study was made available online via Qualtrics survey software (Snow & Mann, 2013). It was circulated online by the research team's network. Participants were first directed to the participant information sheet and could opt-in by providing an email address. A link to the study with a password for access was sent, this enabled access to a consent form. Participants were then directed to the DEX-R questionnaire, the FrSBe, PHQ-2, GAD-2, demographic questions and questions regarding their health. Participant were assigned with an ID number and they could opt in to be sent another link three weeks later to complete the DEX-R for a second time as part of the test-retest phase.

Ethical Considerations

The study was approved by the Faculty of Medicine and Health Sciences Research Ethics Committee at the University of East Anglia (UEA), reference number 201819 – 032 (Appendix G). Participants gave informed consent and they were made aware of their right to withdraw by closing the survey.

Analysis

Data cleaning was completed to ensure responses in the spreadsheet were in accordance with instructions for the questionnaire, such as the removal of incomplete datasets. Parametric assumptions were checked using histograms and Shapiro-Wilk tests. Where item responses were not normally distributed transformations were attempted, otherwise non-parametric alternatives were used. Homogeneity of variance was checked using Levene's test of equality of variance for t-tests. The study used these techniques as the removal of outliers would limit interpretations of non-clinical responses on the DEX-R, which could make comparisons with clinical groups difficult. If limited data relating to clinical factors were retrieved to provide analysis on variance, then these responses were removed from the analysis. These were retained for future analysis due to consent being

gained for these purposes. To provide consistency in reporting, interpretations of psychometric properties were derived from the literature (Hermans, van der Pas & Evenhuis, 2011).

Data were analysed using the Statistical Package for the Social Sciences (SPSS), R, and RUMM 2020/2030.

1. What are the measurement properties of the DEX-R within a non-clinical population?

Measurement properties of the DEX-R were explored by assessing its reliability and validity. Internal reliability of the DEX-R was assessed using Cronbach's standardised **α** and split-half reliability which measure the consistency of the questionnaire to establish whether the questions relate to each other (Cronbach, 1951; Messick, 1989). A criticism of these are their lack of accountability for day-to-day variability. Therefore, test-retest reliability was also assessed using Intra Class Correlation (ICC) to measure whether the questionnaire is consistent over time. A three-week interval for the test-retest phase was chosen in line with previous research (Cummings et al., 1994; Gioia, Isquith, Guy & Kenworthy, 2000; Holst & Thorell, 2018). A score of 0.7 or above is recommended to establish adequate reliability (Hermans et al., 2011).

Validity refers to whether the questionnaire actually measures what it sets out to, in this instance, whether the DEX-R measures dysexecutive problems (Messick, 1989). This was assessed through concurrent validation, by establishing if there was any correlation between the DEX-R and another validated measure of dysexecutive problems, the FrSBe. This was completed using Pearson Product Moment Correlation Coefficient.

2. Does the DEX-R perform as an interval level measure as established by item response theory?

Rasch analysis was completed using the software RUMM2030 (Andrich, Sheridan & Luo, 2009). It is underpinned by item response theory which aims to calibrate both the difficulty of items as well as an individual's ability. It establishes whether a questionnaire can be classed as an interval level measurement, as opposed to ordinal. Whether the DEX-R performs as a unidimensional measure was also explored with Rasch analysis because this identifies whether it is formed of subscales. If the chi-square value is not significant this confirms there is a misfit with the Rasch model, and therefore infers it is a unidimensional measure. If the data does differ significantly, this implies that the DEX-R is not a unidimensional measure, and therefore measuring multiple subconstructs. Multidimensionality was explored further by the factor analysis detailed below.

3. What is the factor structure of the DEX-R in a non-clinical population?

Exploratory factor analysis (EFA) was used to establish whether the underlying structure and latent constructs match onto the Stuss model and whether this supports previous research with the DEX-R with a clinical population (Simblett et al., 2017). The decision on the number of factors to extract was determined by a parallel analysis using R software (Horn, 1965). It is recommended that an oblique rotation is first applied, and if the factor correlations are above .32 then this rotation is maintained (Pedhazur & Schmelkin, 1991; Tabachnick, Fidell & Ullman, 2007). SPSS was used to run the Principal Axis Factoring. It is recommended that factor loadings below 0.3 are suppressed (Field, 2013).

4. **What are the effects of demographic and mood variables on DEX-R and DEX-R subscale performance?**

In order to determine whether demographic or mood variables are associated with variation in scores on the DEX-R, the study also compared subgroups (e.g. gender) and

correlations with continuous variables (e.g. age) with total DEX-R scores and each DEX-R subscale. This used Pearson Product Moment Correlation Coefficient. These variables included age, gender, years of education, anxiety and depression scores. The GAD-2 and PHQ-2 were to be analysed as a continuous measure unless a high proportion scored above the established cut off, in which case the groups were to be compared between those scoring above and below three. Symptoms associated with depression and anxiety can be related to cognitive variability, such as with problem-solving, worry and flexibility. As multiple correlations were being used, the Bonferroni correction was applied with the alpha level set at 0.01 Regression analysis was used to identify which factors predict dysexecutive domains or total score. Gender was the only category variable and was converted into a binary variable. Non-parametric tests were used as the DEX-R, GAD-2, and PHQ-2 total scores were not normally distributed, and the latter were unable to reach normality via transformation. Descriptive statistics were used to explore the frequency of participant responses on the items of the DEX-R.

Results

Recruitment took place from the 5th July 2019 to the 16th December 2019 where 140 people participated, of whom 99 completed the test-retest phase, and 60 the validity phase. Fifteen participants reported being diagnosed or formally hospitalised for at least one of the predefined health conditions. Due to the small number reporting such conditions, group comparisons could not be made and they were therefore excluded from the analysis (see Appendix M). There were 125 participants included in the analysis (80% female) aged between 19 to 69 years (M = 37.7, SD = 12.6), 82% were educated to at least degree level and 77% were White British. See table 1 below for further demographic details.

Table 1.

Demographic Information (n = 125)

		n (%)	*Mean*	SD
Gender				
	Male	25 (20)		
	Female	100 (80)		
Age			37.7	12.6
	18-21	2 (2)		
	22-28	28 (22)		
	29-38	48 (38)		
	39-55	33 (26)		
	56-76	14 (11)		
	76+	0 (0)		
Years of education*			17.5	3.7
	Up to 11 years	6 (5)		
	12 – 14 years	21 (17)		
	15 – 16 years	16 (13)		
	17+ years	79 (65)		
Highest level of education				
	Degree or equivalent	103 (82)		
	Higher Education	4 (3)		
	A Level or equivalent	7 (6)		
	GCSE grades A*-C or equivalent	2 (2)		
	Other qualification	9 (7)		
	No qualification	0 (0)		
	Other / Prefer not to say	0 (0)		
Ethnicity				
	White (British)	97 (77)		
	Other white background	15 (12)		
	White (Irish)	4 (3)		
	Indian	2 (2)		
	Other Asian background	2 (2)		
	Mixed White and Asian	2 (2)		
	Other mixed background	1 (1)		
	Black African	1 (1)		
	Mixed White and Black Caribbean	1 (1)		

Note. *3 missing data for years of education. Source for categorisations, Office for National Statistics

1a. Is the DEX-R a reliable measure of dysexecutive problems?

The DEX-R had excellent internal consistency with Cronbach's **α** at .93 for time one and .94 for time two (Hermans et al., 2011). Cronbach's **α** were also conducted to establish the level of consistency if each item were removed (see table 2). Removal of specific items did not yield significant changes to the DEX-R reliability, with **α** ranging from .93 and .94. Split-half reliability was .93 for time one, and .94 for time two. A high degree of reliability was found between DEX-R scores on two-time points. The average measure ICC was .92 with a 95% confidence interval from .88 to .95 ($F(88,88)= 12.4$, $p<.001$). The median interval between the two phases were 23 days (interquartile range: 21 – 28 days). Table 2 displays scores given for the first phase of completion of the DEX-R.

Table 2.

Descriptive statistics and internal consistency for DEX-R items and total score

Item	Mean (SD)	Item-total correlation	Cronbach's α if item deleted
1. Impulsivity	1.06 (0.81)	r = .38	.93
2. Prospective memory	1.13 (0.96)	r = .60	.93
3. Apathy	1.14 (0.93)	r = .46	.93
4. Initiation	1.40 (1.10)	r = .68	.93
5. Planning	0.90 (1.02)	r = .50	.93
6. Social disinhibition	1.14 (0.90)	r = .56	.93
7. Intention	1.38 (1.13)	r = .65	.93
8. Verbal aggression	1.98 (0.89)	r = -.02	.94
9. Verbal fluency	1.38 (0.92)	r = .54	.93
10. Anger	1.00 (0.86)	r = .43	.93
11. Perseveration	0.56 (0.85)	r = .60	.93
12. Performance monitoring	0.63 (0.64)	r = .43	.93
13. Abstract thinking	0.62 (0.79)	r = .57	.93
14. Metaworry	1.07 (1.08)	r = .60	.93
15. Lack of concern	0.94 (1.05)	r = .17	.93
16. Blunted affect 1	0.77 (0.91)	r = .41	.93

17. Working memory	1.03 (0.96)	r = .63	.93
18. Lack of social composure	1.03 (1.02)	r = .51	.93
19. Insight	0.46 (0.81)	r = .65	.93
20. Inertia	0.98 (0.98)	r = .47	.93
21. Temporal sequencing	0.50 (0.79)	r = .57	.93
22. Cognitive control	1.18 (1.04)	r = .65	.93
23. Variable motivation	0.61 (0.85)	r = .57	.93
24. Physical aggression	0.26 (0.66)	r = .38	.93
25. Organisational ability	1.10 (1.09)	r = .58	.93
26. Inability to inhibit responses	0.77 (0.85)	r = .59	.93
27. Confabulation	0.14 (0.35)	r = .23	.93
28. Emotional lability	0.37 (0.67)	r = .52	.93
29. Distraction	1.33 (0.97)	r = .65	.93
30. Restlessness	0.79 (0.81)	r = .47	.93
31. Cognitive confidence	0.73 (0.76)	r = .42	.93
32. Knowing doing dissociation	0.56 (0.77)	r = .57	.93
33. Blunted affect 2	0.75 (0.85)	r = .57	.93
34. Information processing	0.62 (0.90)	r = .61	.93
35. No concern for social rules	0.59 (.874)	r = .23	.93
36. Complex attention	0.89 (0.84)	r = .50	.93
37. Decision making	1.27 (1.09)	r = .64	.93

1b. Is the DEX-R a valid measure of dysexecutive problems when compared to an existing valid self-report measure?

The DEX-R had good concurrent validity when compared to responses given on another validated measure of dysexecutive problems, the FrSBe (Grace & Malloy, 2001). Both the total scores on the DEX-R and FrSBe were first transformed to achieve adequate normality, the correlation between DEX-R and FrSBe was r = .83, $p < .01$.

2. Does the DEX-R perform as an interval level measure as established by item response theory?

The responses on the DEX-R did not show fit to the Rasch model which suggests it measures more than one subconstruct ($\chi 2(74, N = 140) = 205.54, p < .001$). This is in keeping with the aim of the development of the DEX-R to measure underlying subconstructs of dysexecutive problems. Differential item functioning using Bonferroni correction only yielded differences by gender for one item relating to the expression of emotion. Many of the questions showed disordered thresholds, and scale responses were not all endorsed on items, therefore it was not possible to confirm the interval nature of the scale as these have not yet arrived at a stable solution. Due to the small number in the clinical groups, we could not compare the level of endorsement or differential item functioning. Upon inspection of the frequency of responses, those at the higher end of the scale where dysexecutive problems would be scored as occurring "fairly often" or "very often" were rarely rated for many of the items. Twelve of the items were not fully endorsed, an additional 10 were only fully endorsed based on the response of one person rating "very often" on those items. Confabulation had almost no endorsement, with 100% of participants responding as 'never' or only 'occasionally'. Also, performance monitoring was not rated by participants as occurring fairly or very often, and 82% of participants rated item 24, physical aggression, as 'never' occurring. This is consistent with the DEX-R being applied to a non-clinical sample. However, despite this being a sample who did not report health conditions associated with disruption to frontal functioning, a small number of participants did endorse some items as occurring often or very often. Whilst this study aimed to recruit a non-clinical sample, there is potential for those amongst an apparently 'healthy' population to endorse behaviours and items that are often attributed to frontal impairments. This may in part be due to individual variation, where some participants show a lower level of ability, or it may reflect measurement error where the true score lies within a wide range of impaired to not impaired. Items which were endorsed as occurring

more often included items corresponding to prospective memory, initiation, planning, intention, verbal aggression, metaworry, distraction and decision making (Appendix N).

3. What is the factor structure of the DEX-R in a non-clinical population? Does the factor structure align with the Stuss model?

The parallel analysis retained three factors. The principal axis factoring using the oblique rotation resulted in appropriate correlations for this rotation method to be applied (Pedhazur & Schmelkin, 1991; Tabachnick et al., 2007). The factor analysis had adequate sampling (Kaiser-Meyer-Olkin: .83) and correlation (Bartlett's Test of Sphericity: < 0.01). The factor loading matrix is presented in table 3. The three factors accounted for 42.2% of the total variance, with the first factor accounting for 31.1%. Items with factor loadings below 0.3 were excluded. The internal consistency of factor one was excellent, good for factor two and questionable for the third factor (Hermans et al., 2011). As there were only three factors, the results, therefore, do not align with the Stuss model of frontal functions. The development of the DEX-R found some items did not map onto any of the subscales, these were retained due to their clinical utility (Simblett et al., 2017). Therefore, despite not all items achieving factor loadings above 0.4, no attempts were made to purify the model to preserve this utility.

Seven of the items cross-loaded onto more than one factor. The 19 items loading onto factor one related to those processes associated with the medial/dorsal domain. These items spanned across the proposed Stuss subscales, although mainly encompassed those from the activation subscale. These factors commonly share themes of initiation, maintenance and responsiveness, such as the ability to activate or inhibit a behaviour or thought and was therefore labelled as ‘activation-self regulation’. The 17 items loading onto factor two appeared to relate to dorsolateral domains, typically these items appear to

represent cognitive dysexecutive symptoms such as planning, decision-making, abstract thinking, memory and attention. Although it was recognised that both blunted affect items additionally loaded onto this factor, and all seven cross-loadings involved this factor. The higher factor loadings mainly included those in the proposed executive cognition Stuss subscale. This factor was therefore labelled as 'cognition'. Items loading onto factor three related to processes associated with the orbitofrontal areas, these also shared the blunted affect items. Additionally, more items corresponded to the Stuss behavioural-emotional self-regulation subscale. However, the items also appear to relate to social- self regulatory dysexecutive symptoms, therefore this factor was labelled 'social-emotional'.

Table 3.

DEX-R items according to Stuss subscales and EFA with Promax rotation

Item Number	**Stuss subscale according to Simblett et al. (2017)**	**Item Description**	**Factor**		
			1	**2**	**3**
3.	Activation	Apathy	**.79**	-.26	-.04
26.	Metacognition	Inability to inhibit responses	**.68**	-.21	.29
30.	Activation	Restlessness	**.65**	-.08	-.08
11.	Activation	Perseveration	**.64**	-.15	.27
32.	Behavioural-Emotional Self-Regulation	Knowing doing dissociation	**.64**	-.11	.18
22.	Previously metacognition, later removed	Cognitive control	**.62**	.23	-.19
28.	Behavioural-Emotional Self-Regulation	Emotional lability	**.60**	.07	-.15
4.	Activation	Initiation	**.59**	.25	-.13
19.	Behavioural-Emotional Self-Regulation (previously Activation)	Insight	**.57**	-.01	.28
10.	Metacognition	Anger	**.53**	-.01	-.09
25.	Executive Cognition	Organisational ability	**.51**	.17	-.02
9.	Executive Cognition	Verbal fluency	**.47**	**.30**	-.20
24.	Behavioural-Emotional Self-Regulation	Physical aggression	**.46**	.00	-.08
23.	Activation	Variable motivation	**.45**	**.32**	-.18
7.	Activation	Intention	**.44**	**.34**	-.08
14.	Metacognition	Metaworry	**.40**	.29	-.01

37.	Activation	Decision making	**.40**	**.38**	-.08
6.	Metacognition	Social disinhibition	**.34**	.23	.10
21.	Executive Cognition	Temporal sequencing	**.32**	.15	.25
27.	Behavioural-Emotional Self-Regulation	Confabulation	.28	.06	-.12
13.	Executive Cognition	Abstract thinking	-.10	**.84**	-.10
5.	Executive Cognition	Planning	-.07	**.76**	-.16
36.	Executive Cognition	Complex attention	-.08	**.64**	.05
31.	Metacognition	Cognitive confidence	-.07	**.61**	-.07
17.	Executive Cognition	Working memory	.12	**.56**	.07
29.	Executive Cognition	Distraction	.19	**.54**	.01
20.	Behavioural-Emotional Self-Regulation	Inertia	-.09	**.52**	.16
33.	No subscale	Blunted affect 2	-.10	**.49**	**.41**
2.	Previously Executive Cognition, later removed	Prospective memory	.15	**.48**	.08
12.	Activation	Performance monitoring	-.07	**.36**	**.33**
1.	Metacognition	Impulsivity	.02	**.33**	.14
18.	Metacognition	Lack of social composure	.11	**.30**	.29
35.	Metacognition	No concern for social rules	-.20	.01	**.69**
15.	Behavioural-Emotional Self-Regulation	Lack of concern	.04	-.11	**.39**
16.	Behavioural-Emotional Self-Regulation	Blunted affect 1	-.06	**.30**	**.37**
8.	Behavioural-Emotional Self-Regulation	Verbal aggression	-.21	-.05	**.37**
34.	Executive Cognition	Information processing	.19	.26	**.37**
Eigenvalues			10.93	1.48	1.38
% of variance			31.12	5.69	5.14
Cronbach's **α**			.92	.88	.68

Note. Factor loadings > .30 are shown in bold

4. **What are the effects of demographic and mood variables on DEX-R and DEX-R subscale performance? In particular, what are the effects of age on DEX-R and DEX-R subscale performance?**

Demographic details of the participants are detailed in table 1. Pearson Product Moment Correlation Coefficient analysis was conducted to establish any influence on participants reports of dysexecutive problems. Gender (r = -.02, p = .836) and years of

education (r = -.52, p = .551) were not significantly correlated to DEX-R total and factor scores (p>.05). No significant differences were found between males and females t(123) = -.208, p = .84. A negative correlation was found between responses on the first DEX-R administration and age (r = -.27, p = .002). Age significantly correlated with factor one, r = -.34, p = <.01 and factor two r = -.24, p = <.01, but not with factor three, r = .03, p = .76. A cut off of three as specified by the literature for the GAD-2 and PHQ-2 indicated 18% of participants scoring above the anxiety threshold, and 7% scoring above the depression threshold. Due to the uneven group sizes correlation analysis was used to preserve validity. Spearman Rho correlations was applied when analysing the PHQ-2 and GAD-2. Anxiety scores correlated with dysexecutive problems, r = .45, p = <.01. Depression scores were moderately correlated to dysexecutive problems, r = .58, p = <.01. The scores on the GAD-2 significantly correlated with factor one, r = .51, p = <.01 and factor two r = .47, p = <.01, but not with factor three, r = .16, p = .16. Whereas, the scores on the PHQ-2 significantly correlated with all the factors, factor one, r = .57, p = <.01, factor two r = .45, p = <.01, and factor three, r = .25, p <.01.

The effect of background variables on DEX-R responses was analysed using a multiple regression, the DEX-R total score was the dependent variable, and age, gender, GAD-2 and PHQ-2 scores were the independent variables. In model one, age and gender were kept constant and explained 9.5% of the variance whereas in model two the additional inclusion of the GAD-2 and PHQ-2 scores explained 31.9%, F(4,120) = 14.05, p < .001. The results found gender not to significantly predict DEX-R scores (p >.05*)*. In the first model, DEX-R scores decreased by .44 for every year older a participant was, however when the model also accounted for mood, this decreased to a reduction of .17 for every year older but was no longer a significant effect. Controlling for age, gender, anxiety and depression scores, the regression coefficient (B = 2.51, 95% CI (0.21, 4.81) p < .05)

for the GAD-2 indicates that for each increased score on the GAD-2, the total DEX-R score will increase by 2.51. Furthermore, within the same model, the regression coefficient (B = 5.28, 95% CI (2.67, 7.88) $p < .05$) for the PHQ-2 indicates that for each increased score on the PHQ-2, the total DEX-R score will increase by 5.28.

Discussion

The research aimed to assess the psychometric properties of the DEX-R with a non-clinical sample. The DEX-R was demonstrated to be a valid measure when compared to an already validated measure of dysexecutive problems, the FrSBe. Also, the DEX-R was reliable, evidenced by its high internal, test-retest and split-half reliability.

The individual variability of dysexecutive problems in a non-clinical population supports previous research (Chan, 2001). The influence of demographic factors and mood were also explored, with age and mood found to significantly correlate with the DEX-R as has previously been found in the literature with the DEX and the DEX-R (Dimitriadou et al., 2018; Shaw et al., 2015). Age negatively correlated with the DEX-R, indicating that older participants reported less dysexecutive problems. This is inconsistent with wider literature that cognitive functions associated with prefrontal areas decline with normal ageing (Van Petten et al., 2004; West, 1996). This is likely to be due to the sample demographics, with the oldest participant being 69 years old and only 11% were aged over 56 years of age. The effect of age seemed to be removed when incorporating mood into the model, linking somewhat to previous research where negative affect mediated responses of dysexecutive problems reported by younger people (Gerstorf et al., 2008). These findings could be due to the younger age of the sample and explained by theories of brain maturation, with prefrontal areas developing into people's early 30's (Barkley, 2012; Coffman, 2014). The ventromedial areas of the prefrontal cortex are known to mature later

than other regions and are implicated in social and emotional functions, and may therefore be less developed in this sample (Burnett et al., 2009; Gerstorf et al., 2008; Pfeifer et al., 2011; Sebastian et al., 2011). Furthermore, the demographics of the sample which are predominately younger, white and female are known to have increased prevalence of anxiety levels (Jenkins, Ducker, Gooding, James & Rutter-Eley, 2020). This might contribute to the correlations with symptoms of anxiety and depression. Although the current study's reported prevalence is lower than that reported in the literature, with 18% scoring above the anxiety threshold and 7% above the depression threshold.

The Rasch analysis evidenced the DEX-R as being multidimensional, supporting its development to capture underlying subconstructs of dysexecutive problems (Simblett et al., 2017). A factor analysis found three factors representing activation- self regulatory, cognitive and social-emotional functions. When compared to the proposed Stuss subscales, the factors appeared to share some overlap with the 'activation', 'executive cognition' and 'behaviour-emotional self-regulation' subscales, most notably activation with the first factor leading to the shared label 'activation', however also encompassing self-regulatory items. Those items corresponding with the 'metacognition' subscale appeared to be equally distributed across the three factors, which may relate to the application with a non-clinical sample. Despite the results not aligning fully with the Stuss model, the three factors do appear to have some overlap with theoretical conceptualisations of frontal lobe functioning. All the subscales correlated with symptoms of depression, albeit only a weaker correlation for the social-emotional factor. Only the activation-self regulatory and cognition factors correlated with symptoms of anxiety. The social-emotional factor may therefore represent reduced emotional reactivity or neutrality, such as those items loading onto it including both blunted affect items, a lack of concern and no concern for social rules. Furthermore, the cognitive factor correlations may be driven by those executive

cognitive factors associated with depression and anxiety, such as difficulties with problem-solving and decision making, and working memory. The activation-self regulatory factor may correlate with symptoms of depression and anxiety as they account for components of apathy and a lack of motivation. Therefore, individual differences found in the study, particularly the different components forming factor one, may reflect an overlay with the cognitive symptoms of depression such as apathy. The issue in determining overlap and directionality of mood and dysexecutive problems is further complicated by the lack of diversity in the sample. The demographics of the sample are reported to experience a higher incidence of anxiety (Jenkins et al., 2020). It may be that particular items on the DEX-R are more likely to be rated as occurring more frequently in those experiencing symptoms of anxiety or depression, such as working memory, problem solving and decision making as previously noted. This might account for why DEX-R scores increased by two for every GAD-2 score, and by five for every PHQ-2 score. In terms of age, the later maturation of the prefrontal areas associated with social and emotional variations in those under the age of 30 has already been discussed. It is recognised that any construct will show some individual variation, and the distributions of scores will share some overlap between clinical and non-clinical groups.

The DEX-R was initially developed to map on to the Stuss model. Both the self and informant report versions have demonstrated validity and reliability in both acquired brain injury and healthy ageing samples (Dimitriadou et al., 2018; Simblett et al., 2017). The current study adds to the literature regarding the robustness of the psychometric properties of the DEX-R by evidencing its stability and consistency over time. Additionally, it extends upon the previous literature by comparing to a dysexecutive measure validated in non-clinical and clinical groups, evidencing concurrent validity of the DEX-R.

Furthermore, its application with a non-clinical population provides some consideration for clinicians on how individual differences and mood may contribute to responses.

The DEX-R mapped on to the Stuss model when applied to a brain injury group, with four subscales supporting the theoretical domains of metacognition, activation, executive-cognition, and behavioural-emotional regulation (Simblett et al., 2017; Stuss, 2007). Its application in a healthy ageing and a bipolar sample instead yielded three factors (Social Self-Regulation, Motivation and Attention, and Flexibility, Fluency and Working Memory). Both authors report these factors were in line with Fuster's (2008) theory, which outlines the role of three prefrontal circuits (orbitofrontal, anterior cingulate and dorsolateral) in cognitive and emotion frontal processes. The current study supported previous research on the DEX-R being multidimensional, and factor analysis yielded three factors more in line with this research. However, a confirmatory factor analysis would be required to confirm fit to the model.

The differences in the underlying subscales and subconstructs found within the DEX and DEX-R may relate in part to the clinical population of which they are applied to, as it is well understood that such difficulties are noted in neurological, neurodevelopmental and mental health conditions. The frequency of responses shows two-thirds received a full-range of responses, however, this reduced to 40% where more than one individual rated each response. The items likely have different meanings to people, which in part may relate to the presence of clinical factors. A comparison of the endorsement of items against clinical groups could provide insights into those more highly rated by clinical groups. Those who score highly on the less endorsed items may more likely indicate a problem beyond individual variation. Attempts were not made to purify the model using Rasch techniques in order to retain the clinical utility of the DEX-R.

Strengths and Limitations

A limitation of the current study is the highly educated and largely white and female sample, this limits its generalisability more widely in terms of how well the measure performs across diverse groups. This skew in the sample means how well the DEX-R performs with less well educated, non-white, male populations cannot be commented on here. This poses limitations for clinicians working with these groups as the psychometrics of the measure reported are based on the homogeneous sample, restricting its application. Given the scope of the research was to measure dysexecutive problems, a limitation is that those with such difficulties might be less likely to enrol in taking part in the online study. This limits the representativeness of the sample as the nature of these difficulties such as, for example, 'energisation' abilities which involve the activation and initiation of behaviour may mean they are less motivated to volunteer in online research (Stuss, 2011). Furthermore, the study therefore cannot contribute to previous research on individual variation of reported dysexecutive problems relating to education level (Faria et al., 2015; Foss et al., 2013). The secondary research question relating to healthy ageing was challenging to measure due to the small number of participants over the age of 60, as research on healthy ageing typically indicates less change prior to this age. The study would have benefited from including additional questions on demographic variables, such as occupation and lifestyle (e.g. alcohol consumption and smoking status) as these can contribute to variability in neuropsychological abilities (Fisher et al., 2014; Glass et al., 2009).

The use of a survey to administer the rating scales may be influenced by survey bias due to the self-reporting of behaviour, cognition and mood. To overcome this, participants were not required to give their personal details and by doing it online it was hoped this would reduce such bias. A forced-response option was applied to the survey to

ensure there was no missing data. This was because missing data would have meant participants data could not be used. Nevertheless, the use of forced responses in survey research can increase reactance (Stieger, Reips & Voracek, 2007). Suggestions to mitigate reactance include removal for the requirement to include personal information and by providing a 'degree of autonomy'. This autonomy can be applied via the requirement to manually take part in the survey, rather than a freely accessible one. A strength of the study is the use of such a manual login, which has been additionally found to lead to increased quality of data (Heerwegh & Loosveldt, 2002).

The current study measured constructs which incorporate factors which could be impeded by dysexecutive factors, such as motivation and apathy. An email reminder was used to prompt and remind participants of the study. The use of a survey was beneficial in applying the test-retest phase, as this could be distributed via a second survey, rather than individuals being required to follow-up themselves. A potential confound could have been that those who are more well planned and organised would be more likely to complete this phase, therefore the control of this confound is a strength of the study. There is not a universal standard on the timeframe for test-retest reliability. The requirement is for it to be long enough for memory to not confound results, and not too long that other factors account for responses (e.g. rehabilitation, neurodegeneration). A three-week interval has been applied in similar research (Cummings et al., 1994; Gioia et al., 2000; Holst & Thorell, 2018) and was suitable for the constraints of a doctoral thesis. The correlation between the results were high to indicate reliability, as they were not perfectly correlated this indicates it is less likely memory dictated responses.

Questions relating to health were included as part of the study to allow for monitoring whether clinical factors explained variance in the data, should this have arisen. Unfortunately, only limited data from clinical groups were received and they were

therefore excluded from the analysis. However, consent was gained to retain their data for future analysis, which could be used in studies directly recruiting such clinical groups.

Clinical implications

The use of rating scales enhances the assessment of dysexecutive problems in capturing everyday experiences of individuals, which are not necessarily measured by neuropsychological tests alone (Malloy & Grace, 2005). For example, the BADS was found to correlate with the behaviour and cognition factors, but not the emotion factor on the DEX. The factors in the current study appear to capture emotional processes well and may particularly relate to emotional regulatory models (Salas, Gross & Turnbull, 2019). Furthermore, the correlations with depression scores may indicate the measurement of apathy. The psychometric properties of the DEX-R have previously been assessed with ABI, bipolar disorder and healthy ageing samples. The current study used a non-clinical sample which highlighted how age and mood can influence responses on the DEX-R. In regard to the decrease in reports of dysexecutive problems with increased age in a normally distributed sample, this has implications for clinicians to consider theories of prefrontal brain maturation into their assessments and interpretations. Such as considering how social and emotional functions might be less developed in younger people and adapting neurorehabilitation strategies with this in mind. A further finding was the presence of behaviours typically attributed to dysexecutive problems in clinical settings also occurring in a non-clinical sample. Certain items were more highly rated than others indicating that individual variation is found not just in clinical groups. This individual variation has implications for clinicians in their interpretation of assessment scores, such as an awareness that this natural variation in scores on these items may have been present prior to clinical diagnosis. Therefore, clinicians would need to be mindful in not immediately concluding that someone has a dysexecutive impairment if a score on an item might be

common in the general populations. A further clinical implication is the inclusion of mood scales in their assessments and to consider health factors in such interpretations.

Future research

The current study focused on reports of dysexecutive problems in a non-clinical population, and there was a lack of endorsement for some of the items. Future research could consider a comparison to clinical groups on how responses differ between clinical and non-clinical groups via differential item functioning. Additional investigation via a Rasch analysis could establish subscales, which may highlight differences between groups such as whether awareness and metacognitive processes are more applicable to those with prefrontal brain injuries. Finally, engaging with participants from different ethnic groups and a range of educational backgrounds could be made to improve the generalisability of findings.

Conclusion

The current study supports the psychometric properties of the DEX-R as found in previous research, being both a valid and reliable measure of dysexecutive problems. The number of factors had similarities with two previous studies. There was individual variation in responses, influenced somewhat by mood and age. This may have implications for clinicians in their interpretation of DEX-R scores. Future research could consider comparing the responses of clinical and non-clinical groups.

References

Andrich, D., Sheridan, B., & Luo, G. (2009). *RUMM 2030 (Beta version for Windows)* Perth. Western Australia: RUMM Laboratory Pty Ltd.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

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

Burgess, P. W. (2004). Theory and methodology in executive function research. In *Methodology of Frontal and Executive Function* (pp. 87-121). Routledge.

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. doi:10.1017/S1355617798466037

Burnett, S., Bird, G., Moll, J., Frith, C., & Blakemore, S. J. (2009). Development during adolescence of the neural processing of social emotion. *Journal of Cognitive Neuroscience*, *21*(9), 1736-1750. doi.org/10.1162/jocn.2009.21121

Busch, R. M., McBride, A., Curtiss, G., & Vanderploeg, R. D. (2005). The components of executive functioning in traumatic brain injury. *Journal of Clinical and Experimental Neuropsychology, 27*(8), 1022-1032. doi.org/10.1080/13803390490919263

Callahan, C. D. (2001). The assessment and rehabilitation of executive function disorders. *Rehabilitation of neuropsychological disorders: A practical guide for rehabilitation professionals,* 75-106.

Carvalho, J. O., Ready, R. E., Malloy, P., & Grace, J. (2013). Confirmatory factor analysis of the frontal systems behavior scale (FrSBe). *Assessment*, *20*(5), 632-641. doi:10.1177/1073191113492845

Chan, R. C. (2001). Dysexecutive symptoms among a non-clinical sample: a study with the use of the Dysexecutive questionnaire. *British Journal of Psychology, 92*(3), 551-565. doi.org/10.1348/000712601162338

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (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

Cicerone, K., Levin, H., Malec, J., Stuss, D., & Whyte, J. (2006). Cognitive rehabilitation interventions for executive function: Moving from bench to bedside in patients with traumatic brain injury. *Journal of Cognitive Neuroscience*, *18*(7), 1212-1222. doi.org/10.1162/jocn.2006.18.7.1212

Coffman, T. P. (2014). *Psychometric properties of the Barkley Deficits in Executive Functioning Scale (BDEFS) in a college student population.* Florida State University. Retrieved from: https://diginole.lib.fsu.edu/islandora/object/fsu%3A185371

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, *44*(12), 2308-2308. doi:10.1212/WNL.44.12.2308

Damasio, A.R., Tranel, D., & Damasio, H. (1991). Somatic markers and the guidance of behavior. In H. Levin, H. Eisenberg, & A. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 217-228). New York: Oxford University Press.

Dimitriadou, M., Michaelides, M. P., Bateman, A., & Constantinidou, F. (2018). Measurement of everyday dysexecutive symptoms in normal aging with the Greek version of the dysexecutive questionnaire-revised. *Neuropsychological Rehabilitation*, 1-20. doi:10.1080/09602011.2018.1543127

Eslinger, P. J., & Damasio, A. R. (1985). Severe disturbance of higher cognition after bilateral frontal lobe ablation patient EVR. *Neurology, 35*(12), 1731-1731. doi.org/10.1212/WNL.35.12.1731

Faria, C. D. A., Alves, H. V. D., & Charchat-Fichman, H. (2015). The most frequently used tests for assessing executive functions in aging. *Dementia & Neuropsychologia*, *9*(2), 149-155. doi:10.1590/1980-57642015DN92000009

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Fisher, G. G., Infurna, F. J., Grosch, J., Stachowski, A., Faul, J. D., & Tetrick, L. E. (2014). Mental work demands, retirement, and longitudinal trajectories of cognitive functioning. *J. Occup. Health Psychol.* 19, 231–242. doi: 10.1037/a0035724

Foss, M. P., Carvalho, V. A. D., Machado, T. H., Reis, G. C. D., Tumas, V., Caramelli, P., ... & Porto, C. S. (2013). Mattis Dementia Rating Scale (DRS): Normative data for the Brazilian middle-age and elderly populations. *Dementia & Neuropsychologia*, *7*(4), 374-379. doi:10.1590/S1980-57642013DN74000004

Fuster, J. M. (2008). *The prefrontal cortex*. (4th ed.). China: Elsevier.

Gerstorf, D., Siedlecki, K. L., Tucker-Drob, E. M., & Salthouse, T. A. (2008). Executive dysfunctions across adulthood: Measurement properties and correlates of the DEX self-report questionnaire. *Aging, Neuropsychology, and Cognition*, *15*(4), 424-445. doi:10.1080/13825580701640374

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). Test review behavior rating inventory of executive function. *Child Neuropsychology*, *6*(3), 235-238. doi:10.1076/chin.6.3.235.3152

Glass, J. M., Buu, A., Adams, K. M., Nigg, J. T., Puttler, L. I., Jester, J. M., & Zucker, R. A. (2009). Effects of alcoholism severity and smoking on executive neurocognitive function. *Addiction*, *104*(1), 38-48. doi: 10.1111/j.1360-0443.2008.02415.x

Grace, J., & Malloy, P. F. (2001). *Frontal Systems Behavior Scale (FrSBe): Professional Manual.* Lutz, FL: Psychological Assessment Resources, Inc.

Grace, J., Stout, J. C., & Malloy, P. F. (1999). Assessing frontal lobe behavioral syndromes with the frontal lobe personality scale. *Assessment*, *6*(3), 269-284. doi:10.1177/107319119900600307

Grafman, J. (1989). Plans, actions, and mental sets: Managerial knowledge units in the frontal lobes. In E. Perecman (Eds.), *Integrating theory and practice in clinical neuropsychology* (pp. 93-138). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Hanks, R. A., Rapport, L. J., Millis, S. R., & Deshpande, S. A. (1999). Measures of executive functioning as predictors of functional ability and social integration in a rehabilitation sample. *Archives of Physical Medicine and Rehabilitation, 80*(9), 1030-1037. doi.org/10.1016/S0003-9993(99)90056-4

Headway (2018). In *Statistics.* Retrieved from: https://www.headway.org.uk/about-brain-injury/further-information/statistics/statistics-resources/

Heerwegh, D., & Loosveldt, G. (2002). Web surveys: The effect of controlling survey access using PIN numbers. *Social Science Computer Review*, *20*(1), 10-21. doi: 10.1177/089443930202000102

Hermans, H., van der Pas, F. H., & Evenhuis, H. M. (2011). Instruments assessing anxiety in adults with intellectual disabilities: A systematic review. *Research in Developmental Disabilities*, *32*(3), 861-870. doi:10.1016/j.ridd.2011.01.034

Hoffmann, M., & Schmitt, F. (2006). Metacognition in stroke: bedside assessment and relation to location, size, and stroke severity. *Cognitive and Behavioral Neurology, 19*(2), 85-94. doi:10.1097/01.wnn.0000213903.59827.50

Holst, Y., & Thorell, L. B. (2018). Adult executive functioning inventory (ADEXI): Validity, reliability, and relations to ADHD. *International Journal of Methods in Psychiatric Research*, *27*(1), e1567. doi:10.1002/mpr.1567

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179-185. doi:10.1007/BF02289447

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. *Neuropsychologia*, *44*(11), 2017-2036. doi:10.1016/j.neuropsychologia.2006.01.010

Jenkins, P. E., Ducker, I., Gooding, R., James, M., & Rutter-Eley, E. (2020). Anxiety and depression in a sample of UK college students: a study of prevalence, comorbidity, and quality of life. *Journal of American College Health*, 1-7. doi:10.1080/07448481.2019.1709474

Levin, H. S., Amparo, E., Eisenberg, H. M., Williams, D. H., High Jr, W. M., McArdle, C. B., & Weiner, R. L. (1987). Magnetic resonance imaging and computerized tomography in relation to the neurobehavioral sequelae of mild and moderate head injuries. *Journal of Neurosurgery*, *66*(5), 706-713. doi:10.3171/jns.1987.66.5.0706

Lezak, M. D. (1995). *Neuropsychological Assessment*. New York: Oxford University Press

Loschiavo-Alvares, F. Q., Sediyama, C. Y. N., Vasconcelos, A. G., Neves, F., Correa, H., Malloy-Diniz, L. F., & Bateman, A. (2013). Clinical application of DEX-R for

patients with bipolar disorder type I and II. *Clinical Neuropsychiatry*, *10*(2), 86-95. https://www.researchgate.net/publication/248381422_Clinical_Application_of_DEX-R_for_patients_with_Bipolar_Disorder_type_I_and_II

Löwe, B., Kroenke, K., & Gräfe, K. (2005). Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *Journal of Psychosomatic Research*, *58*(2), 163-171. doi:10.1016/j.jpsychores.2004.09.006

MacPherson, S. E., Phillips, L. H., & Della Sala, S. (2002). Age, executive function and social decision making: A dorsolateral prefrontal theory of cognitive aging. *Psychology and Aging*, *17*(4), 598. doi:10.1037/0882-7974.17.4.598

Malloy, P., & Grace, J. (2005). A review of rating scales for measuring behavior change due to frontal systems damage. *Cognitive and Behavioral Neurology*, *18*(1), 18-27. doi: 10.1097/01.wnn.0000152232.47901.88

McDonald, B. C., Flashman, L. A., & Saykin, A. J. (2002). Executive dysfunction following traumatic brain injury: neural substrates and treatment strategies. *NeuroRehabilitation, 17*(4), 333-344. doi:10.3233/NRE-2002-17407

McGuire, B. E., Morrison, T. G., Barker, L. A., Morton, N., McBrinn, J., Caldwell, S., ... & Walsh, J. (2014). Impaired self-awareness after traumatic brain injury: inter-rater reliability and factor structure of the Dysexecutive Questionnaire (DEX) in patients, significant others and clinicians. *Frontiers in Behavioral Neuroscience, 8*, 352. doi.org/10.3389/fnbeh.2014.00352

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. doi:10.3102/0013189X018002005

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8-14. doi:10.1177/0963721411429458

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* (Vol. 4, pp. 1–18). New York: Plenum

Pedhazur, E. J. & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, N]: Erlbaum.

Pfeifer, J. H., Masten, C. L., Moore III, W. E., Oswald, T. M., Mazziotta, J. C., Iacoboni, M., & Dapretto, M. (2011). Entering adolescence: Resistance to peer influence, risky behavior, and neural changes in emotion reactivity. *Neuron, 69*(5), 1029-1036. doi.org/10.1016/j.neuron.2011.02.019

Pohjasvaara, T., Leskelä, M., Vataja, R., Kalska, H., Ylikoski, R., Hietanen, M., ... & Erkinjuntti, T. (2002). Post-stroke depression, executive dysfunction and functional outcome. *European Journal of Neurology, 9*(3), 269-275. doi.org/10.1046/j.1468-1331.2002. 00396.x

Pribram, K.H. (1960), 'The intrinsic systems of the forebrain', in *Handbook on Physiology,* Neurophysiology II, ed. J. Field, H.W. Magoun & V.E. Hall. Washington, DC: American Physiological Society.

Salas, C. E., Gross, J. J., & Turnbull, O. H. (2019). Using the process model to understand emotion regulation changes after brain injury. *Psychology & Neuroscience*, *12*(4), 430. doi:10.1037/pne0000174

Sebastian, C. L., Fontaine, N. M., Bird, G., Blakemore, S. J., De Brito, S. A., McCrory, E. J., & Viding, E. (2011). Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, *7*(1), 53-63. doi.org/10.1093/scan/nsr023

Shallice, T. I. M., & 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

Shaw, S., Oei, T. P., & Sawang, S. (2015). Psychometric validation of the Dysexecutive Questionnaire (DEX). *Psychological Assessment*, *27*(1), 138. doi:10.1037/a0038195

Simblett, S. K., & Bateman, A. (2011). Dimensions of the Dysexecutive Questionnaire (DEX) examined using Rasch analysis. *Neuropsychological Rehabilitation, 21*(1), 1–25. http://doi.org/10.1080/09602011.2010.531216

Simblett, S. K., Ring, H., & Bateman, A. (2017). The Dysexecutive Questionnaire Revised (DEX-R): An extended measure of everyday dysexecutive problems after acquired brain injury. *Neuropsychological Rehabilitation, 27*(8), 1124-1141. doi.org/10.1080/09602011.2015.1121880

Skapinakis, P. (2007). The 2-item Generalized Anxiety Disorder scale had high sensitivity and specificity for detecting GAD in primary care. *Evidence-Based Medicine*, *12*(5), 149-149. doi:10.1136/ebm.12.5.149

Snow, J., & Mann, M. (2013). *Qualtrics survey software: handbook for research professionals*. Qualtrics Labs, Inc.

Stieger, S., Reips, U. D., & Voracek, M. (2007). Forced-response in online surveys: Bias from reactance and an increase in sex-specific dropout. *Journal of the American*

Society for Information Science and Technology, *58*(11), 1653-1660. doi:10.1002/asi.20651

Stout, J. C., Ready, R. E., Grace, J., Malloy, P. F., & Paulsen, J. S. (2003). Factor analysis of the Frontal Systems Behavior Scale (FrSBe). *Assessment*, *10*(1), 79-85. doi:10.1177/1073191102250339

Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary.* American Chemical Society.

Stuss, D. T. (2007). New approaches to prefrontal lobe testing. *The Human Frontal Lobes: Functions and Disorders, 2*, 292-305. doi:10.1017/S1355617711000695

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 901-915. doi:10.1098/rstb.2007.2096

Stuss, D. T. (2011). Functions of the frontal lobes: relation to executive functions. *Journal of the International Neuropsychological Society*, *17*(5), 759-765. doi:10.1017/S1355617711000695

Stuss, D. T., Benson, D. F., Weir, W. S., Naeser, M. A., Lieberman, I., & Ferrill, D. (1983). The involvement of orbitofrontal cerebrum in cognitive tasks. *Neuropsychologia, 21*(3), 235-248. doi:10.1016/0028-3932(83)90040-4

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5). Boston, MA: Pearson.

Teasell, R., Bayona, N., Marshall, S., Cullen, N., Bayley, M., Chundamala, J., ... & Lippert, C. (2007). A systematic review of the rehabilitation of moderate to severe

acquired brain injuries. *Brain Injury*, *21*(2), 107-112. doi:10.1080/02699050701201524

Turner-Stokes L. *Rehabilitation following Acquired Brain Injury: National Clinical Guidelines*. London: Royal College of Physicians/ British Society of Rehabilitation Medicine, 2003.

Van Petten, C., Plante, E., Davidson, P. S., Kuo, T. Y., Bajuscak, L., & Glisky, E. L. (2004). Memory and executive function in older adults: Relationships with temporal and prefrontal gray matter volumes and white matter hyperintensities. *Neuropsychologia*, *42*(10), 1313-1335. doi:10.1016/j.neuropsychologia.2004.02.009

Velligan, D. I., Ritch, J. L., Sui, D., DiCocco, M., & Huntzinger, C. D. (2002). Frontal Systems Behavior Scale in schizophrenia: relationships with psychiatric symptomatology, cognition and adaptive function. *Psychiatry Research*, *113*(3), 227-236. doi:10.1016/S0165-1781(02)00264-0

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin, 120*(2), 272. https://www.researchgate.net/profile/Robert_West12/publication/14374675_An_application_of_prefrontal_cortex_theory_to_cognitive_aging/links/0a85e531cae9774487000000.pdf

Wilson, B. A., Alderman, N., Burgess, P. W., Emslie, H., & Evans, J. J. (1996). *Behavioural assessment of the dysexecutive syndrome: BADS*. London: Pearson.

Wilson, B.A., Gracey, F., Evans, J.E., & Bateman, A. (2009). *Neuropsychological rehabilitation: Theories, models, therapy and outcome.* Cambridge, UK: Cambridge University Press.

Chapter 4

Extended Methodology

Extended Methodology

This extended methodology chapter provides supplementary information regarding the empirical paper including, sample size and power calculations, measures, procedure, ethical considerations and analysis.

Sample size and power calculations

Two approaches were used to estimate the required sample size for each analysis. This included reviewing the research literature and calculations conducted using the computer package G* Power 3.1.9 (Faul & Erdfelder, 1992).

Calculations using G*Power:

Questions relating to validity, test-retest reliability, mood and demographic factors were answered using correlational analysis. G* Power analysis determined 109 participants were required with sufficient power of 0.9, to detect a medium effect size, or 64 participants if power were set to 0.8. For multiple regression analysis, G* Power output indicated 82 participants were required, again for a medium effect size to be detected with power set at 0.9 with four variables included in the model. Probability for both of these were set at .05.

Recommendations from the literature:

Internal reliability: A sample size of at least 30 participants are recommended when using a single Cronbach's alpha coefficient (Bujang, Omar & Baharum, 2018).

Rasch analysis: For a Rasch analysis, a sample size of 243 is acceptable based on consultation of the literature suggesting parameters for sample size based on 99% confidence (Linacre, 1994). Pilot Rasch analysis has been conducted with samples sizes of

at least 100 participants (Simblett et al., 2017). It is recommended that at least 100 participants are required to avoid disordering parameters (Chen et al., 2014).

Factor analysis: The literature recommends a sample size of at least 50 for exploratory factor analysis (de Winter, Dodou, & Wieringa, 2009). SPSS calculates the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy which indicates whether factor analysis is appropriate with the amount of data, and therefore whether an increased sample size is required. A KMO above 0.7 is classed as 'good', above 0.8 is 'great', and those above 0.9 are 'superb' (Field, 2013; Hutcheson & Sofroniou, 1999).

Measures

Full copies of the DEX-R and FrSBe are not provided in the thesis portfolio as they are subject to copyright. The additional 14 questions included in the DEX-R are reported in the literature (Simblett et al., 2017). Permission was required to use the FrSBe and DEX-R for research purposes, license agreements between authors and Pearson and PAR can be found in the appendices (Appendix K and L).

Procedure

Qualtrics survey software hosted the research, a three-year license was granted to the researcher by the University of East Anglia. Due to the included measures being subject to copyright, the license for use required the online survey be password protected, and for participants to be emailed the link with the password. A forced response option was applied to the questionnaire in order to limit missing data or errors. Where participants partially completed the research, this data was recorded by Qualtrics. Data was first downloaded on to Excel where these incomplete datasets were removed. Data was then paired for participants who took part in the test-retest reliability phase, however errors in the ID code meant that two of these could not be used.

Ethical Considerations

Informed Consent

Detailed information about the study was provided on the participant information sheet which detailed the aims and purpose of the study to allow for informed consent.

Risks and benefits from taking part

The study took no longer than 30 minutes to complete and was, therefore, hoped that no psychological distress would arise from the study. The participant information sheet detailed that should participants have concerns following the nature of the questionnaire, such as memory or emotional difficulties, they could discuss this with their General Practitioner. Participants were informed that dysexecutive problems are reported in non-clinical populations to allow normalising their responses to questions.

Right to withdraw

Participants were made aware that they could withdraw from the study by exiting the survey. The participant information sheet highlighted that due to the data being anonymised, once they have completed the survey their data would not be able to be removed. They were made aware they could contact the researcher should they wish to withdraw from the second phase of the study as the researcher would stop the reminder email from being sent out through Qualtrics.

Confidentiality

Participants were assigned with a participant code to ensure confidentiality was upheld. The only personal detail collected was an email address to send the link to the study online and for the test-retest phase which did not form part of the analysis. Email

addresses were stored separately from the anonymised research data and were password protected.

Data Protection

Guidance from the Data Protection Act and The General Data Protection Regulation (GDPR) (EU) 2016/679 were followed. In line with UEA data protection policy, anonymised study data will be archived by UEA for 10 years after the study ends, after which it will be destroyed.

Analysis

Test-Retest reliability

Psychometric research traditionally uses Pearson's correlation to determine consistency over time. However, limitations have been raised due to its theoretical standpoint of measuring relationships between different variables. Therefore, in the case of test-retest reliability it was determined as not appropriate to use due to this method measuring the relationship between the same variable (Yen & Lo, 2002). Additionally, when applied in these cases Pearson's correlation lacks the ability to detect systematic errors. Yen and Lo (2002) suggest using intra-class correlations to limit these issues and was therefore the measure used for test-retest reliability in the current study.

Factor Analysis

Different methods were considered to decide on the approach to take on determining the number of factors to extract. The default option in SPSS is the mostly commonly reported method where factors with eigenvalues greater than one are retained (Kaiser, 1960). However, it is argued that the retention of factors above eigenvalues of one is arbitrary, and that this method can overestimate the numbers to retain due to reporting an

upper bound, despite interpretations of this method being on reporting exact numbers to retain (Hayton, Allen & Scarpello, 2004). Another method considered was using a scree plot of eigenvalues to identify the point of inflexion, extracting those above this point (Cattell, 1966). The literature specifies that this method is more reliable in samples above 200 participants, and therefore was not appropriate for the current study due to the samples size being below this (Stevens, 2002). Parallel analysis was the final factor retention method considered (Horn, 1965). This compares two sets of data, the eigenvalues of randomly generated data with the actual data eigenvalues. Those factors where eigenvalues are above the 95th percentile of the randomly generated data are extracted (Çokluk & Koçak, 2016). Hayton, Allen and Scarpello (2004) describe it as the most precise method for factor retention, and it was deemed the most appropriate method for the current study.

Additional analysis

Further exploratory analyses were conducted which were secondary to the research questions outlined in chapter 3. These included analyses using the test-retest reliability data, the FrSBe subscales and comparison of groups. The results of which can be found in chapter 5.

Chapter 5

Extended Results

Extended Results

Additional exploratory analyses were conducted beyond the scope of the empirical paper, and therefore were not included in chapter 3. These are instead reported here and include analyses using the DEX-R responses collected from the test-retest reliability phase, the FrSBe subscales, and the comparison of groups.

Additional demographic and mood correlations

Correlations calculated between the first administration of the DEX-R and demographic factors are reported in chapter 3. These were repeated for the DEX-R responses collected during the test-retest phase. Table 1 compares these, and the relationships found from the analysis using the first administration of the DEX-R were replicated.

Table 1.

DEX-R relationship with Age, Education and Gender

		Age	Years of education	Gender	GAD-2	PHQ-2
DEX-R, time 1	r	-.27	-.05	-.02	0.45	0.58
(n = 125)	p	.002	.551	.836	<0.01	<0.01
DEX-R, time 2	r	-.29	-.03	-.16	0.41	0.49
(n = 89)	p	.006	.770	.138	<0.01	<0.01

Additional validity analysis

The FrSBe subscales were correlated against the three DEX-R factors extracted, all correlations were significant and are detailed in table 2.

Table 2.

Comparison between FrSBe subscale scores and the three DEX-R factors generated (n = 53)

FrSBe Subscale		EFA Factors		
		Activation-self-regulation	Cognition	Social-emotional
FrSBe Dysexecutive	r	.65	.75	.41
	p	.000	.000	.002
FrSBe Apathy	r	.68	.60	.35
	p	.000	.000	.011
FrSBe Disinhibition	r	.68	.64	.66
	p	.000	.000	.000

Clinical Variables

Only 15 participants reported a health condition which could potentially influence responses on the DEX-R. Due to the small sample, discriminant analysis could not be computed. However, when comparing scores between the two groups, those with a reported health condition scored significantly higher on the DEX-R (Mdn = 42) than those without a health condition (Mdn = 30) *(U*=635.500, *p*=.042). There was also a significant difference in the DEX-R scores for those scoring above (Mdn = 47) and below (Mdn = 26) the cut off on the GAD-2 (*U* = 682.500, *p* = .000), and above (Mdn = 52) and below (Mdn = 29) the cut off on the PHQ-2 (*U* = 359.000, *p* = .049). Further small-scaled analysis was conducted using the data retrieved from those reporting a health condition. These participants were matched with a non-clinical counterpart based on age, gender, ethnicity and education level. This matching was only achievable for 13 of these participants, however, this analysis found no significant differences in DEX-R scores between those

reporting a health condition (Mdn = 35) compared to the matched controls who did not report a health condition (Mdn = 21) (p = .113). Therefore, once demographic factors were controlled for these apparent differences between the clinical group and the rest of the sample were no longer found. These differences in scores may be attributable to demographics, however, all of these analyses would benefit from larger sample sizes to compare groups.

Chapter 6

Overall discussion and critical review

Overall discussion and critical review

This final chapter of the thesis portfolio will summarise and appraise the findings from the systematic review and the main study. Wider clinical implications and considerations for future research will then be explored. Finally, the chapter will close with final conclusions.

Main findings

The thesis portfolio intended to explore the measurement of frontal functions in clinical and non-clinical populations. The systematic review identified various rating scales of executive function used across differing clinical and non-clinical groups. The DEX was the most widely used rating scale and additional papers were retrieved where modifications to the DEX had been made. These had good psychometric properties across different groups but there were inconsistencies in the factors reported. The DEX-R was found to be a multidimensional measure with both an ABI and healthy ageing sample. Previous measurement of the validity of the DEX-R were limited in not comparing it to an already validated measure and assessing consistency over time could strengthen its reliability. Individual variation with non-clinical groups has been found previously (Chan, 2001).

The empirical paper therefore sought to address these gaps by establishing the psychometric properties of the DEX-R in a non-clinical population. The DEX-R was determined to be a valid and reliable measure of dysexecutive symptoms. The Rasch analysis confirmed the multidimensionality of the DEX-R. A three-factor structure was found, in line with two previous studies on the DEX-R. This included an activation-self-regulatory factor, cognitive factor and social-emotional factor. Factors relating to initiation, cognition, and self-regulation were widely reported in the systematic review with non-clinical samples. The DEX-R when used with a non-clinical population shared some overlay, but did not fully align with, the Stuss model.

There were individual variations found in reports of dysexecutive problems in the non-clinical population, in line with previous research (Chan, 2001). Mood, notably symptoms of depression, correlated more so with higher DEX-R scores. Additionally, younger age was associated with increased reports of dysexecutive problems, again consistent with the literature (Gerstorf et al., 2008). These associations were replicated when the correlations were repeated using the DEX-R responses from the test-retest phase. Weaker correlations of symptoms of depression and anxiety were found for the social-emotional factor and appeared to relate more so with neutrality of emotions, such as blunted affect. Papers included in the systematic review were selected based on their reports of psychometric properties, some additionally included analysis relating to mood. Shaw, Oei and Sawang (2015) reported similar correlation co-efficients as the empirical paper for depression and the DEX, although they found greater associations with anxiety measures. Additionally, they found those with a diagnosed depressive or anxiety condition scored significantly higher on the DEX compared to a non-clinical sample. Five papers in the systematic review included samples with mental health conditions, where the number of factors ranged from two to three. Across papers the rating scales capture behavioural and emotional processes, going beyond the traditional focus of measuring mainly cognitive difficulties reported (Chan et al., 2008; Damasio, Tranel & Damasio, 1991; Stuss, 2007). The systematic review highlighted how decision making and initiation factors were more present in non-clinical samples, in the empirical paper these were two of the most endorsed items of the DEX-R with a non-clinical sample. When considered with the correlations of symptoms of depression, these factors may represent the cognitive aspects of these symptoms. Furthermore, the systematic review consistently found that inhibition was not widely reported in non-clinical samples and did not form part of the factor descriptions in the empirical paper when a factor analysis was applied with a non-clinical sample.

Individual differences therefore perhaps contribute to the variance influencing performance in neuropsychological tests (Miyake & Friedman, 2012). The conceptualisation of these as a "syndrome" is also further challenged due to the varied responses on the DEX-R in both papers (Damasio et al., 1991).

Strength and limitation of the thesis portfolio

Both the systematic review and empirical paper contribute to the literature on the measurement of frontal functions. They both extend on the understanding that the variation in symptoms goes beyond clinical groups, with such individual variability being found in non-clinical groups. Both papers included a consistent and standardised interpretation of the psychometric properties (Hermans et al., 2011). A limitation of the empirical paper and that of those in the systematic review using non-clinical groups, were that these often-included individuals with a higher level of educational attainment, were mostly white and female. This limits the generalisation of findings to a more diverse population as it is important to consider how well measures perform across diverse groups.

It had been hoped that this thesis would contribute further on the conceptual nature of frontal functions. However, this was difficult due to the complexities of analysing psychometrics and differences in the methods used, the varied labels of the factors, and the variation found across samples. The Stuss model was predominately considered, particularly with its relevance to the DEX and DEX-R. There is a usefulness in considering the unity and diversity of frontal functions further, such as predictions when using measures with clinical groups with particular patterns of neurological injury.

Constraints relating to the completion of such a study as part of a doctoral training programme meant that funding limited the amount of FrSBes which could be purchased for use, and the time frame for test-retest reliability was reduced (albeit still acceptable given similar literature including those in the systematic review). The quality of the empirical

paper would have been improved if all participants completed the FrSBe, as outlined by the QUADAS-2. The power analysis indicated a minimum of 64 participants for correlation analysis. This was reached for the test-retest reliability analysis, mood and demographic correlation analyses, however not for the validity analysis where 54 participants were included. The sample size for the multiple regression and pilot Rasch were achieved. The factor analysis sample size was classed as 'great' based on the KMO (Field, 2013; Hutcheson & Sofroniou, 1999). Unfortunately, there were delays experienced during the ethics process and whilst obtaining the budget. This reduced the expected data collection time which could have focussed on recruiting a more balanced and representative sample.

A strength was conducting the concurrent validation analysis as recommended by papers in the systematic review (Shaw et al., 2015). Additionally, a strength is that the systematic review supported the use of the FrSBe as a validated alternative to be used as part of assessing the DEX-Rs validity. However, one of the papers found that the disinhibition scale on the FrSBe was the only one valid in a non-clinical population (Caracuel et al., 2012). This was because the apathy and executive dysfunction scales were not classed as unidimensional, and therefore in non-clinical groups may capture different processes. Unfortunately, the empirical paper did not receive a large enough clinical sample to explore these differences further. Those participants who regrettably were excluded from the analysis on this basis did provide their consent for their data to be used as part of future research, the scope of which could specifically address clinical comparisons. Although age and mood showed interesting correlations, these were not explored in greater detail as the focus of the thesis portfolio was in regard to psychometric properties. A strength based on the findings from both papers, is the applicability of rating scales measuring frontal functions beyond cognitive factors. The factor analyses across

papers highlight how underlying subconstructs capture a broad picture of challenges reported, additionally encompassing social, emotional and behavioural processes. Furthermore, the cross-loadings of seven items across factors show some association between factors, which may reflect the unity and diversity previously reported (Miyake & Friedman, 2012). These findings demonstrate the utility of rating scale measures as part of neuropsychological testing, with clinical benefits to capture specific difficulties and being able to tailor interventions most appropriate for a person. The clinical use of both standardised neuropsychological tests and rating scales together can enhance assessment, and address the gap previously identified in capturing social-emotional process to improve ecological validity (Chan et al., 2008). Additionally, in the current context whereby social distancing measures are in place, the methods explored in the thesis may contribute to identifying valid and reliable methods which could be conducted remotely.

The systematic review demonstrated how informant versions of the rating scales further enhances assessment and has additional benefits where individuals may have reduced awareness of their difficulties. The empirical paper did not explore this further; informant reports were not included due to the scope of the research being conducted as part of a doctoral thesis online. Hosting the study online produced its own challenges and opportunities. The hope was that this would enable a more representative sample to take part as the study could reach more widely. The use of an ID code provided anonymity; however, some responses could not be used as they did not match at the retest stage.

Many psychological measures apply statistical analyses appropriate to interval levels of measurement, despite many of these actually being ordinal. Rasch analysis was applied in papers using the DEX-R retrieved in the systematic review which indicated it performed as an internal level of measurement. However, the empirical paper was unable to replicate these findings due to the limited endorsement of item responses, again the

implication being the inclusion of clinical groups to achieve such endorsement. Rasch analysis is also concerned with the relationship between item difficulty and person ability (Bond, 2015). Therefore, further analysis using item response theory could explore the extent to which items may perform differently between clinical and non-clinical groups to further understand the variation in likelihood of rating at certain levels across items as found in the frequency of responses in the empirical paper.

Papers differed as to whether an oblique rotation or orthogonal rotation was applied in a factor analysis, depending on whether the authors felt the factors should, or should not be correlated with each other. Furthermore, papers contrasted in whether correlated factors evidenced construct validity. One paper reports the correlation found did therefore demonstrate construct validity (Dimitriadou et al., 2018), whereas a separate paper used the lack of correlation to also indicate construct validity (Shaw et al., 2015). It would be expected that the factors would have some correlation as they are measuring the same underlying construct. Guidance from the literature was sought in order to base the decision on best standards. The empirical paper did find the three factors correlated, and therefore the rotation was applied.

Clinical implications

Both papers report upon the psychometric properties of different rating scales, and how these translate across different clinical groups. The systematic review presents the differences by clinical group, which can support clinicians in identifying the measure most relevant to their clinical setting. In addition, these papers highlight the individual variability in non-clinical groups that can arise due to age or mood, for example. The implications for clinicians are holding these variations in mind in their interpretation of scores. The papers retrieved in the systematic review mostly appeared to view it as a multidimensional construct, with different rating scales captured these underlying

constructs through the use of subscales. The empirical paper supported the idea of multidimensionality, and contributed to the three-factors as reported in other papers of the DEX-R. The DEX-R was developed to capture the four processes in the Stuss model for those with ABI. Perhaps the different factors reported in this paper and others are due to the different groups these have been applied with. The questions likely have different meanings to different people, and the level at which they impact on everyday life. This was reflected in the lack of endorsement of many items in the empirical paper. Different items on the DEX-R correspond to different subconstructs, as individual variability is known in both clinical and non-clinical groups the use of subscales can guide clinicians to individualised interventions, for example, prompting, goal-management, and emotional regulation strategies. Furthermore, assessment may identify a presence of a deficit compared to what would be expected in a non-clinical group which may contribute to diagnosis. Also, in guiding rehabilitation in formulating why a person is having difficulty in everyday life tasks to improve performance, as well as uses as an outcome evaluation.

Future research

The systematic review included papers where responses on rating scales were compared between clinical and non-clinical groups. The empirical paper added to the current literature on the DEX-R by exploring its psychometric properties and factor structure with a non-clinical sample. Previous research utilising the DEX-R has included samples of people with ABI, bi-polar disorder and an exploration of healthy ageing. Future research could compare clinical and non-clinical groups, to further understand variation in factor scores in accordance with underlying neurological networks of specific clinical disorders. This would further support clinicians to be able to select the most robust measure applicable to their setting. The systematic review highlighted only three papers assessing classification of clinical conditions. One paper attempted a discriminant analysis

correctly classifying a non-clinical and psychiatric group, but not the neurological group . Shaw et al. (2015) does not outline whether this neurological group includes those with or without frontal deficits (or both). The mental health group may therefore be less diverse representing a common difference in cognitive functioning compared to the non-clinical group. Shaw et al. (2015) recommended further research assess the discriminative ability of the DEX. Since revisions to the DEX have since been made, leading to the development of the DEX-R, this may mean improvements are more sensitive to classifying different clinical groups. The empirical paper did not achieve an adequate sample size to conduct such an analysis, therefore recommendations are made for future research to consider this further.

Conclusion

The different models and measurement of frontal functions has led to the use of rating scale measures to enhance clinical assessment, by their ability to gain everyday experiences of people and the challenges which they may experience. The systematic review mostly reported on the robustness of psychometric properties in the DEX and its variants, the BRIEF-A and the FrSBe. The DEX-R was developed to map on to conceptual frameworks and the empirical paper applied it with a non-clinical sample expanding on its reported reliability, validity and factor structure. Further development of the DEX-R could compare clinical and non-clinical groups to understand differences in endorsement of response, further adding to the robustness of the measure. This could consider whether those behaviours typically understood to indicate frontal dysfunction may not only arise due to frontal brain injury. This has important implications when interpreting responses on rating scale measures in clinical and research settings as indicating such difficulties. And finally, the continuum of dysexecutive symptoms found in non-clinical groups, as well as

inconsistencies found in clinical groups, may challenge the concept of there being a dysexecutive ‘syndrome’.

Additional references

Baddeley, A. D., & Hitch, G. (1974). Working memory. In *Psychology of learning and motivation* (Vol. 8, pp. 47-89). Academic press.

Baddeley, A., & Wilson, B. (1988). Frontal amnesia and the dysexecutive syndrome. *Brain and Cognition*, *7*(2), 212-230. doi.org/10.1016/0278-2626(88)90031-0

Bps.org.uk. (2014). Code of Human Research Ethics. [online] Available at: http://www.bps.org.uk/system/files/Public%20files/code_of_human_research_ethics_dec_2014_inf180_web.pdf

Bombois, S., Debette, S., Delbeuck, X., Bruandet, A., Lepoittevin, S., Delmaire, C., ... & Pasquier, F. (2007). Prevalence of subcortical vascular lesions and association with executive function in mild cognitive impairment subtypes. *Stroke*, *38*(9), 2595-2597. doi:10.1161/STROKEAHA.107.486407

Bond, T. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Boone, W. J. (2016). Rasch analysis for instrument development: why, when, and how?. *CBE—Life Sciences Education*, *15*(4), rm4. doi:10.1187/cbe.16-04-0148

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136-136.

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for Cronbach's alpha test: a simple guide for researchers. *The*

Malaysian Journal of Medical Sciences: MJMS, *25*(6), 85. doi:10.21315/mjms2018.25.6.9

Burgess, P. W. (2004). Theory and methodology in executive function research. In *Methodology of Frontal and Executive Function* (pp. 87-121). Routledge.

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. doi:10.1017/S1355617798466037

Caracuel, A., Verdejo-García, A., Fernández-Serrano, M. J., Moreno-López, L., Santago-Ramajo, S., Salinas-Sánchez, I., & Pérez-García, M. (2012). Preliminary validation of the Spanish version of the Frontal Systems Behavior Scale (FrSBe) using Rasch analysis. *Brain Injury*, *26*(6), 844-852. doi:10.3109/02699052.2012.655365

Cattell, R. B. (*1966*). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245 – 276.

Chan, R. C. (2001). Dysexecutive symptoms among a non-clinical sample: a study with the use of the Dysexecutive questionnaire. *British Journal of Psychology, 92*(3), 551-565. doi.org/10.1348/000712601162338

Chan, R. C., & Manly, T. (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

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (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

Chen, W. H., Lenderking, W., Jin, Y., Wyrwich, K. W., Gelhorn, H., & Revicki, D. A. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research*, *23*(2), 485-493. doi:10.1007/s11136-013-0487-5

Child, D. (1990). *The essentials of factor analysis* (2nd ed.). London: Cassel Educational Limited.

Cicerone, K., Levin, H., Malec, J., Stuss, D., & Whyte, J. (2006). Cognitive rehabilitation interventions for executive function: moving from bench to bedside in patients with traumatic brain injury. *Journal of Cognitive Neuroscience*, *18*(7), 1212-1222. doi.org/10.1162/jocn.2006.18.7.1212

Ciszewski, S., Francis, K., Mendella, P., Bissada, H., & Tasca, G. A. (2014). Validity and reliability of the Behavior Rating Inventory of Executive Function—Adult Version in a clinical sample with eating disorders. *Eating Behaviors*, *15*(2), 175-181. doi:10.1016/j.eatbeh.2014.01.004

Coffman, T. P. (2014). Psychometric properties of the Barkley Deficits in Executive Functioning Scale (BDEFS) in a college student population. Florida State University. Retrieved from: https://diginole.lib.fsu.edu/islandora/object/fsu%3A185371

Çokluk, Ö., & Koçak, D. (2016). Using Horn's parallel analysis method in exploratory factor analysis for determining the number of factors. *Educational Sciences: Theory and Practice*, *16*(2), 537-551. doi:10.12738/estp.2016.2.0328

Cullen, K. R., LaRiviere, L. L., Vizueta, N., Thomas, K. M., Hunt, R. H., Miller, M. J., ... & Schulz, S. C. (2016). Brain activation in response to overt and covert fear and

happy faces in women with borderline personality disorder. *Brain Imaging and Behavior*, *10*(2), 319-331. doi:10.1007/s11682-015-9406-4

Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, *44*(12), 2308-2308. doi:10.1212/WNL.44.12.2308

Damasio, A.R., Tranel, D., & Damasio, H. (1991). Somatic markers and the guidance of behavior. In H. Levin, H. Eisenberg, & A. Benton (Eds.), *Frontal lobe function and dysfunction* (pp. 217-228). New York: Oxford University Press.

de Winter, J. D., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147-181. doi:10.1080/00273170902794206

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168. doi:10.1146/annurev-psych-113011-143750

Dimitriadou, M., Michaelides, M. P., Bateman, A., & Constantinidou, F. (2018). Measurement of everyday dysexecutive symptoms in normal aging with the Greek version of the dysexecutive questionnaire-revised. *Neuropsychological Rehabilitation*, 1-20. doi:10.1080/09602011.2018.1543127

Drakopoulos, J., Sparding, T., Clements, C., Pålsson, E., & Landén, M. (2020). Executive functioning but not IQ or illness severity predicts occupational status in bipolar disorder. *International Journal of Bipolar Disorders*, *8*(1), 7. doi:10.1186/s40345-019-0168-6

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [Computer program]. *Bonn, FRG: Bonn University, Department of Psychology*.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (3rd ed.). Sage.

Gerstorf, D., Siedlecki, K. L., Tucker-Drob, E. M., & Salthouse, T. A. (2008). Executive dysfunctions across adulthood: measurement properties and correlates of the DEX self-report questionnaire. *Aging, Neuropsychology, and Cognition*, *15*(4), 424-445. doi:10.1080/13825580701640374

Goel, V., Grafman, J., Tajik, J., Gana, S., & Danto, D. (1997). A study of the performance of patients with frontal lobe lesions in a financial planning task. *Brain: A Journal of Neurology*, *120*(10), 1805-1822. doi:10.1093/brain/120.10.1805

Grafman, J. (1989). Plans, actions, and mental sets: Managerial knowledge units in the frontal lobes. In E. Perecman (Eds.), *Integrating theory and practice in clinical neuropsychology* (pp. 93-138). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Grant, I., & Adams, K. (2009). *Neuropsychological assessment of neuropsychiatric and neuromedical disorders*. Oxford University Press.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor Retention Decisions in Exploratory Factor Analysis: A Tutorial on Parallel Analysis. *Organizational Research Methods, 7*(2), 191-205. doi:10.1177/1094428104263675

Hermans, H., van der Pas, F. H., & Evenhuis, H. M. (2011). Instruments assessing anxiety in adults with intellectual disabilities: A systematic review. *Research in Developmental Disabilities*, *32*(3), 861-870. doi:10.1016/j.ridd.2011.01.034

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26-32. doi:10.1016/j.tics.2003.11.003

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Hutcheson, G. D., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Sage.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Laakso, H. M., Hietanen, M., Melkas, S., Sibolt, G., Curtze, S., Virta, M., ... & Jokinen, H. (2019). Executive function subdomains are associated with post-stroke functional outcome and permanent institutionalization. *European Journal of Neurology*, *26*(3), 546-552. doi:10.1111/ene.13854

Lezak, M. D. (1995). *Neuropsychological Assessment*. New York: Oxford University Press

Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, *7*(4), 328.

Loschiavo-Alvares, F. Q., Sediyama, C. Y. N., Vasconcelos, A. G., Neves, F., Correa, H., Malloy-Diniz, L. F., & Bateman, A. (2013). Clinical application of DEX-R for patients with bipolar disorder type I and II. *Clinical Neuropsychiatry*, *10*(2), 86-95. https://www.researchgate.net/publication/248381422_Clinical_Application_of_DEX-R_for_patients_with_Bipolar_Disorder_type_I_and_II

Luria, A. R. (2012). *Higher cortical functions in man*. Springer Science & Business Media.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, *18*(2), 5-11. doi:10.3102/0013189X018002005

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8-14. doi:10.1177/0963721411429458

Morice, R., & Delahunty, A. (1996). Frontal/executive impairments in schizophrenia. *Schizophrenia Bulletin*, *22*(1), 125-137. doi:10.1093/schbul/22.1.125

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* (Vol. 4, pp. 1–18). New York: Plenum

Ponsford, J., Draper, K., & Schönberger, M. (2008). Functional outcome 10 years after traumatic brain injury: Its relationship with demographic, injury severity, and cognitive and emotional status. *Journal of the International Neuropsychological Society*, *14*(2), 233-242. doi:10.1037/a0017729

Pribram, K. H., (1960). The intrinsic systems of the forebrain, in *Handbook on Physiology,* Neurophysiology II, ed. J. Field, H.W. Magoun & V.E. Hall. Washington, DC: American Physiological Society. http://karlpribram.com/wp-content/uploads/pdf/theory/T-006.pdf

Shaw, S., Oei, T. P., & Sawang, S. (2015). Psychometric validation of the Dysexecutive Questionnaire (DEX). *Psychological Assessment*, *27*(1), 138. doi:10.1037/a0038195

Shinagawa, Y., Nakaaki, S., Hongo, J., Murata, Y., Sato, J., Matsui, T., ... & Furukawa, T. A. (2007). Reliability and validity of the Japanese version of the Dysexecutive Questionnaire (DEX) in Alzheimer's disease: Validation of a behavioral rating scale to assess dysexecutive symptoms in Japanese patients with Alzheimer's disease. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences*, *22*(10), 951-956. doi:10.1002/gps.1768

Simblett, S. K., Badham, R., Greening, K., Adlam, A., Ring, H., & Bateman, A. (2012). Validating independent ratings of executive functioning following acquired brain injury using Rasch analysis. *Neuropsychological Rehabilitation*, *22*(6), 874-889. doi:10.1080/09602011.2012.703956

Simblett, S. K., Ring, H., & Bateman, A. (2017). The Dysexecutive Questionnaire Revised (DEX-R): An extended measure of everyday dysexecutive problems after acquired brain injury. *Neuropsychological Rehabilitation, 27*(8), 1124-1141. doi.org/10.1080/09602011.2015.1121880

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences (4th ed.)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Stout, J. C., Ready, R. E., Grace, J., Malloy, P. F., & Paulsen, J. S. (2003). Factor analysis of the Frontal Systems Behavior Scale (FrSBe). *Assessment*, *10*(1), 79-85. doi:10.1177/1073191102250339

Stuss, D. T. (2007). New approaches to prefrontal lobe testing. *The Human Frontal Lobes: Functions and Disorders, 2*, 292-305.

Stuss, D. T. (2011). Functions of the frontal lobes: Relation to executive functions. *Journal of the International Neuropsychological Society*, *17*(5), 759-765. doi:10.1017/S1355617711000695

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 901-915. doi:10.1098/rstb.2007.2096

Swanberg, M. M., Tractenberg, R. E., Mohs, R., Thal, L. J., & Cummings, J. L. (2004). Executive dysfunction in Alzheimer disease. *Archives of Neurology*, *61*(4), 556-560. doi:10.1001/archneur.61.4.556

Thorell, L. B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology*, *33*(4), 536-552. doi:10.1080/87565640802101516

Vaughan, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, *25*(2), 343. doi:10.1037/a0017729

Wright, B. D. (1996). Comparing Rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(1), 3-24.

Yen, M., & Lo, L. H. (2002). Examining test-retest reliability: an intra-class correlation approach. *Nursing Research*, *51*(1), 59-62.

Zinn, S., Bosworth, H. B., Hoenig, H. M., & Swartzwelder, H. S. (2007). Executive function deficits in acute stroke. *Archives of Physical Medicine and Rehabilitation*, *88*(2), 173-180. doi:10.1016/j.apmr.2006.11.015

Appendices

Appendix A: Author Guidelines for Neuropsychological Rehabilitation

Instructions for authors

COVID-19 impact on peer review
As a result of the significant disruption that is being caused by the COVID-19 pandemic we understand that many authors and peer reviewers will be making adjustments to their professional and personal lives. As a result they may have difficulty in meeting the timelines associated with our peer review process. Please let the journal editorial office know if you need additional time. Our systems will continue to remind you of the original timelines but we intend to be flexible.

Thank you for choosing to submit your paper to us. These instructions will ensure we have everything required so your paper can move through peer review, production and publication smoothly. Please take the time to read and follow them as closely as possible, as doing so will ensure your paper matches the journal's requirements.

AUTHOR SERVICES
Supporting Taylor & Francis authors

For general guidance on every stage of the publication process, please visit our Author Services website.

EDITING SERVICES
Supporting Taylor & Francis authors

For editing support, including translation and language polishing, explore our Editing Services website

SCHOLARONE MANUSCRIPTS™

This journal uses ScholarOne Manuscripts (previously Manuscript Central) to peer review manuscript submissions. Please read the guide for ScholarOne authors before making a submission. Complete guidelines for preparing and submitting your manuscript to this journal are provided below.

This title utilises format-free submission. Authors may submit their paper in any scholarly format or layout. References can be in any style or format, so long as a consistent scholarly citation format is applied. For more detail see the format-free submission section below.

About the Journal

Neuropsychological Rehabilitation is an international, peer-reviewed journal publishing high-quality, original research. Please see the journal's Aims & Scope for information about its focus and peer-review policy.

Please note that this journal only publishes manuscripts in English.

Neuropsychological Rehabilitation accepts the following types of article: original articles, scholarly reviews, book reviews.

Open Access

You have the option to publish open access in this journal via our Open Select publishing program. Publishing open access means that your article will be free to access online immediately on publication, increasing the visibility, readership and impact of your research. Articles published Open Select with Taylor & Francis typically receive 32% more citations* and over 6 times as many downloads** compared to those that are not published Open Select.

Your research funder or your institution may require you to publish your article open access. Visit our Author Services website to find out more about open access policies and how you can comply with these.

You will be asked to pay an article publishing charge (APC) to make your article open access and this cost can often be covered by your institution or funder. Use our APC finder to view the APC for this journal.

Please visit our Author Services website or contact openaccess@tandf.co.uk if you would like more information about our Open Select Program.

*Citations received up to Jan 31st 2020 for articles published in 2015-2019 in journals listed in Web of Science®.
**Usage in 2017-2019 for articles published in 2015-2019.

Peer Review and Ethics

Taylor & Francis is committed to peer-review integrity and upholding the highest standards of review. Once your paper has been assessed for suitability by the editor, it will then be single blind peer reviewed by independent, anonymous expert referees. Find out more about what to expect during peer review and read our guidance on publishing ethics.

Preparing Your Paper

All authors submitting to medicine, biomedicine, health sciences, allied and public health journals should conform to the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, prepared by the International Committee of Medical Journal Editors (ICMJE).

Clinical trials: must conform to the Consort guidelines http://www.consort-statement.org. Submitted papers should include a checklist confirming that all of the Consort requirements have been met, together with the corresponding page number of the manuscript where the information is located. In addition, trials must be pre-registered on a site such as clinicaltrials.gov or equivalent, and the manuscript should include the reference number to the relevant pre-registration.

Systematic reviews: submitted papers should follow PRISMA http://www.prisma-statement.org/ guidelines and submission should also be accompanied by a completed PRISMA checklist, together with the corresponding page number of the manuscript where the information is located.

Single-case studies: submitted papers should follow SCRIBE guidelines (http://psycnet.apa.org/fulltext/2016-17384-001.html) and include a completed SCRIBE checklist together with the corresponding page number of the manuscript where the information is located.

Observational studies: submitted papers should follow the STROBE guidelines (https://www.strobe-statement.org/index.php?id=strobe-home) and also include a completed checklist of compliance, together with the corresponding page number of the manuscript where the information is located.

Qualitative studies: should follow the COREQ guidelines (http://www.equator-network.org/reporting-guidelines/coreq/) and be accompanied by a completed COREQ checklist of compliance, together with the corresponding page number of the manuscript where the information is located.

The EQUATOR Network (Enhancing the Quality and Transparency of Health Research) website provides further information on available guidelines.

Structure

Your paper should be compiled in the following order: title page; abstract; keywords; main text introduction, materials and methods, results, discussion; acknowledgments; declaration of interest statement; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figures; figure captions (as a list).

Word Limits

Please include a word count for your paper. There are no word limits for papers in this journal.

Format-Free Submission

Authors may submit their paper in any scholarly format or layout. Manuscripts may be supplied as single or multiple files. These can be Word, rich text format (rtf), open document format (odt), or PDF files. Figures and tables can be placed within the text or submitted as separate documents. Figures should be of sufficient resolution to enable refereeing.

- There are no strict formatting requirements, but all manuscripts must contain the essential elements needed to evaluate a manuscript: abstract, author affiliation, figures, tables, funder information, and references. Further details may be requested upon acceptance.
- References can be in any style or format, so long as a consistent scholarly citation format is applied. Author name(s), journal or book title, article or chapter title, year of publication, volume and issue (where appropriate) and page numbers are essential. All bibliographic

entries must contain a corresponding in-text citation. The addition of DOI (Digital Object Identifier) numbers is recommended but not essential.
- The journal reference style will be applied to the paper post-acceptance by Taylor & Francis.
- Spelling can be US or UK English so long as usage is consistent.

Note that, regardless of the file format of the original submission, an editable version of the article must be supplied at the revision stage.

Taylor & Francis Editing Services

To help you improve your manuscript and prepare it for submission, Taylor & Francis provides a range of editing services. Choose from options such as English Language Editing, which will ensure that your article is free of spelling and grammar errors, Translation, and Artwork Preparation. For more information, including pricing, visit this website.

Checklist: What to Include

1. **Author details.** Please ensure everyone meeting the International Committee of Medical Journal Editors (ICMJE) requirements for authorship is included as an author of your paper. All authors of a manuscript should include their full name and affiliation on the cover page of the manuscript. Where available, please also include ORCiDs and social media handles (Facebook, Twitter or LinkedIn). One author will need to be identified as the corresponding author, with their email address normally displayed in the article PDF (depending on the journal) and the online article. Authors' affiliations are the affiliations where the research was conducted. If any of the named co-authors moves affiliation during the peer-review process, the new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after your paper is accepted. Read more on authorship.
2. Should contain an unstructured abstract of 200 words.
3. You can opt to include a **video abstract** with your article. Find out how these can help your work reach a wider audience, and what to think about when filming.
4. Between 5 and 5 **keywords**. Read making your article more discoverable, including information on choosing a title and search engine optimization.
5. **Funding details.** Please supply all details required by your funding and grant-awarding bodies as follows:
   For single agency grants
   This work was supported by the [Funding Agency] under Grant [number xxxx].
   For multiple agency grants
   This work was supported by the [Funding Agency #1] under Grant [number xxxx]; [Funding Agency #2] under Grant [number xxxx]; and [Funding Agency #3] under Grant [number xxxx].
6. **Disclosure statement.** This is to acknowledge any financial interest or benefit that has arisen from the direct applications of your research. Further guidance on what is a conflict of interest and how to disclose it.
7. **Data availability statement.** If there is a data set associated with the paper, please provide information about where the data supporting the results or analyses presented in the paper can be found. Where applicable, this should include the hyperlink, DOI or other persistent identifier associated with the data set(s). Templates are also available to support authors.
8. **Data deposition.** If you choose to share or make the data underlying the study open, please deposit your data in a recognized data repository prior to or at the time of submission. You will be asked to provide the DOI, pre-reserved DOI, or other persistent identifier for the data set.
9. **Geolocation information.** Submitting a geolocation information section, as a separate paragraph before your acknowledgements, means we can index your paper's study area accurately in

JournalMap's geographic literature database and make your article more discoverable to others. More information.

10. **Supplemental online material.** Supplemental material can be a video, dataset, fileset, sound file or anything which supports (and is pertinent to) your paper. We publish supplemental material online via Figshare. Find out more about supplemental material and how to submit it with your article.
11. **Figures.** Figures should be high quality (1200 dpi for line art, 600 dpi for grayscale and 300 dpi for colour, at the correct size). Figures should be supplied in one of our preferred file formats: EPS, PS, JPEG, TIFF, or Microsoft Word (DOC or DOCX) files are acceptable for figures that have been drawn in Word. For information relating to other file types, please consult our Submission of electronic artwork document.
12. **Tables.** Tables should present new information rather than duplicating what is in the text. Readers should be able to interpret the table without reference to the text. Please supply editable files.
13. **Equations.** If you are submitting your manuscript as a Word document, please ensure that equations are editable. More information about mathematical symbols and equations.
14. **Units.** Please use SI units (non-italicized).

Using Third-Party Material in your Paper

You must obtain the necessary permission to reuse third-party material in your article. The use of short extracts of text and some other types of material is usually permitted, on a limited basis, for the purposes of criticism and review without securing formal permission. If you wish to include any material in your paper for which you do not hold copyright, and which is not covered by this informal agreement, you will need to obtain written permission from the copyright owner prior to submission. More information on requesting permission to reproduce work(s) under copyright.

Disclosure Statement

Please include a disclosure statement, using the subheading "Disclosure of interest." If you have no interests to declare, please state this (suggested wording: *The authors report no conflict of interest*). For all NIH/Wellcome-funded papers, the grant number(s) must be included in the declaration of interest statement. Read more on declaring conflicts of interest.

Clinical Trials Registry

In order to be published in a Taylor & Francis journal, all clinical trials must have been registered in a public repository at the beginning of the research process (prior to patient enrolment). Trial registration numbers should be included in the abstract, with full details in the methods section. The registry should be publicly accessible (at no charge), open to all prospective registrants, and managed by a not-for-profit organization. For a list of registries that meet these requirements, please visit the WHO International Clinical Trials Registry Platform (ICTRP). The registration of all clinical trials facilitates the sharing of information among clinicians, researchers, and patients, enhances public confidence in research, and is in accordance with the ICMJE guidelines.

Complying With Ethics of Experimentation

Please ensure that all research reported in submitted papers has been conducted in an ethical and responsible manner, and is in full compliance with all relevant codes of experimentation and legislation. All papers which report in vivo experiments or clinical

trials on humans or animals must include a written statement in the Methods section. This should explain that all work was conducted with the formal approval of the local human subject or animal care committees (institutional and national), and that clinical trials have been registered as legislation requires. Authors who do not have formal ethics review committees should include a statement that their study follows the principles of the Declaration of Helsinki.

Consent

All authors are required to follow the ICMJE requirements on privacy and informed consent from patients and study participants. Please confirm that any patient, service user, or participant (or that person's parent or legal guardian) in any research, experiment, or clinical trial described in your paper has given written consent to the inclusion of material pertaining to themselves, that they acknowledge that they cannot be identified via the paper; and that you have fully anonymized them. Where someone is deceased, please ensure you have written consent from the family or estate. Authors may use this Patient Consent Form, which should be completed, saved, and sent to the journal if requested.

Health and Safety

Please confirm that all mandatory laboratory health and safety procedures have been complied with in the course of conducting any experimental work reported in your paper. Please ensure your paper contains all appropriate warnings on any hazards that may be involved in carrying out the experiments or procedures you have described, or that may be involved in instructions, materials, or formulae.

Please include all relevant safety precautions; and cite any accepted standard or code of practice. Authors working in animal science may find it useful to consult the International Association of Veterinary Editors' Consensus Author Guidelines on Animal Ethics and Welfare and Guidelines for the Treatment of Animals in Behavioural Research and Teaching. When a product has not yet been approved by an appropriate regulatory body for the use described in your paper, please specify this, or that the product is still investigational.

Submitting Your Paper

This journal uses ScholarOne Manuscripts to manage the peer-review process. If you haven't submitted a paper to this journal before, you will need to create an account in ScholarOne. Please read the guidelines above and then submit your paper in the relevant Author Centre, where you will find user guides and a helpdesk.

Please note that *Neuropsychological Rehabilitation* uses Crossref™ to screen papers for unoriginal material. By submitting your paper to *Neuropsychological Rehabilitation* you are agreeing to originality checks during the peer-review and production processes.

On acceptance, we recommend that you keep a copy of your Accepted Manuscript. Find out more about sharing your work.

Data Sharing Policy

This journal applies the Taylor & Francis Basic Data Sharing Policy. Authors are encouraged to share or make open the data supporting the results or analyses presented in their paper where this does not violate the protection of human subjects or other valid privacy or security concerns.

Authors are encouraged to deposit the dataset(s) in a recognized data repository that can mint a persistent digital identifier, preferably a digital object identifier (DOI) and recognizes a long-term preservation plan. If you are uncertain about where to deposit your data, please see this information regarding repositories.

Authors are further encouraged to cite any data sets referenced in the article and provide a Data Availability Statement.

At the point of submission, you will be asked if there is a data set associated with the paper. If you reply yes, you will be asked to provide the DOI, pre-registered DOI, hyperlink, or other persistent identifier associated with the data set(s). If you have selected to provide a pre-registered DOI, please be prepared to share the reviewer URL associated with your data deposit, upon request by reviewers.

Where one or multiple data sets are associated with a manuscript, these are not formally peer reviewed as a part of the journal submission process. It is the author's responsibility to ensure the soundness of data. Any errors in the data rest solely with the producers of the data set(s).

Publication Charges

There are no submission fees, publication fees or page charges for this journal.

Colour figures will be reproduced in colour in your online article free of charge. If it is necessary for the figures to be reproduced in colour in the print version, a charge will apply.

Charges for colour figures in print are £300 per figure ($400 US Dollars; $500 Australian Dollars; €350). For more than 4 colour figures, figures 5 and above will be charged at £50 per figure ($75 US Dollars; $100 Australian Dollars; €65). Depending on your location, these charges may be subject to local taxes.

Copyright Options

Copyright allows you to protect your original material, and stop others from using your work without your permission. Taylor & Francis offers a number of different license and reuse options, including Creative Commons licenses when publishing open access. Read more on publishing agreements.

Complying with Funding Agencies

We will deposit all National Institutes of Health or Wellcome Trust-funded papers into PubMedCentral on behalf of authors, meeting the requirements of their respective open access policies. If this applies to you, please tell our production team when you receive

your article proofs, so we can do this for you. Check funders' open access policy mandates here. Find out more about sharing your work.

My Authored Works

On publication, you will be able to view, download and check your article's metrics (downloads, citations and Altmetric data) via My Authored Works on Taylor & Francis Online. This is where you can access every article you have published with us, as well as your free eprints link, so you can quickly and easily share your work with friends and colleagues.

We are committed to promoting and increasing the visibility of your article. Here are some tips and ideas on how you can work with us to promote your research.

Article Reprints

You will be sent a link to order article reprints via your account in our production system. For enquiries about reprints, please contact the Taylor & Francis Author Services team at reprints@tandf.co.uk. You can also order print copies of the journal issue in which your article appears.

Queries

Should you have any queries, please visit our Author Services website or contact us here.

Updated 04-5-2020

Appendix B: PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	15
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	16
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	17 to 22
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	22
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	23
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	24
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	23
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	23
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	22 to 24

Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	23 to 24
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	24
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	25
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	24
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	25
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	25 to 26
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	n/a
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	26
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	28
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	29
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	31 to 40
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	41 to 49
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	30
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	n/a
DISCUSSION			

Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	49 to 51
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	51 to 55
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	55 to 56
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	n/a

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

Appendix C: Interpretation of psychometric properties (adapted based on Hermans et al. (2011))

Psychometric Statistic	Interpretation
Internal Consistency	
<0.50	Unacceptable
0.50-0.59	Poor
0.60-0.69	Questionable
0.70-0.79	Acceptable
0.80-0.89	Good
≥0.90	Excellent
Correlation Coefficients (Pearson's product-moment and Spearman rank)	
<0.29	Little or no correlation
0.30-0.49	Low correlation
0.50-0.69	Moderate correlation
0.70-0.89	High correlation
≥0.90	Very high correlation
Intra-class Correlation Coefficient	
<0.40	Poor
0.40-0.59	Fair
0.60-0.74	Good
≥0.75	Excellent
Kappa Statistic	
<0.00	Poor
0.00-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

Appendix D: Modified QUADAS-2

DOMAIN 1: PARTICIPANT SELECTION
A. Risk of Bias

- Were selection criteria clearly described? *(Yes/No/Unclear)*
- Was a consecutive or random sample of participants enrolled? *(Yes/No/Unclear)*
- Did the study avoid inappropriate exclusions? *(Yes/No/Unclear)*

Could the selection of participants have introduced bias? RISK: LOW /HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the included participants do not match the review question? CONCERN: LOW /HIGH/UNCLEAR

DOMAIN 2: INDEX TEST(S)
A. Risk of Bias

- Was the execution of the index text described in sufficient detail to permit replication of the test? (including details of it being appropriately translated where applicable) *(Yes/No/Unclear)*

Could the conduct or interpretation of the index test have introduced bias? RISK: LOW /HIGH/UNCLEAR

B. Concerns regarding applicability

Is there concern that the index test, its conduct, or interpretation differ from the review question? CONCERN: LOW /HIGH/UNCLEAR

DOMAIN 3: REFERENCE STANDARD
A. Risk of Bias

- Is the reference standard likely to correctly classify the target condition? *(Yes/No/Unclear/N/A)*

Could the reference standard, its conduct, or its interpretation have introduced bias? RISK: LOW /HIGH/UNCLEAR

B. Concerns regarding applicability
Is there concern that the target condition as defined by the reference standard does not match the review question? **CONCERN: LOW /HIGH/UNCLEAR**

DOMAIN 4: FLOW AND TIMING

A. Risk of Bias

- Was there an appropriate interval between index test(s) and reference standard? *(Yes/No/Unclear/N/A)*
- Did all participants receive a reference standard? *(Yes/No/Unclear/N/A)*
- Did participants receive the same reference standard? *(Yes/No/Unclear/N/A)*
- Were all participants included in the analysis? (were withdrawals from the study explained?) *(Yes/No/Unclear)*

Could the participant flow have introduced bias? RISK: LOW /HIGH/UNCLEAR

Appendix E: Quality Assessment using the QUADAS-2

Table 1.

Risk of bias and applicability ratings across domains for all included papers.

Study	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Beerten-Duijkers et al. (2019)	U/C	LOW	LOW	LOW	LOW	LOW	LOW
Bodenburg et al. (2008)	LOW	U/C	N/A	LOW	LOW	LOW	N/A
Caracuel et al. (2012)	HIGH	U/C	N/A	HIGH	HIGH	LOW	N/A
Carvalho et al. (2013)	LOW	U/C	U/C	U/C	LOW	LOW	U/C
Ciszewski et al. (2014)	LOW	LOW	N/A	U/C	HIGH	LOW	LOW
Coolidge et al. (1995)	U/C	U/C	N/A	U/C	HIGH	LOW	U/C
Dimitriadou et al. (2018)	LOW	LOW	N/A	LOW	LOW	LOW	N/A
Duggan et al. (2018)	U/C	U/C	LOW	U/C	LOW	LOW	LOW
Grace et al. (1999)	LOW	LOW	LOW	HIGH	LOW	LOW	LOW
Hauser et al. (2013)	HIGH	U/C	HIGH	HIGH	LOW	LOW	HIGH
Hellebrekers et al. (2017)	LOW	LOW	LOW	HIGH	LOW	LOW	LOW
Holst et al. (2017)	HIGH	LOW	LOW	U/C	LOW	LOW	LOW
Mani et al. (2018)	LOW	LOW	HIGH	U/C	LOW	LOW	HIGH
Milan et al. (2008)	U/C	LOW	LOW	LOW	LOW	LOW	LOW

Miranda et al. (2019)	U/C	U/C	N/A	LOW	LOW	LOW	N/A
Niemeier et al. (2013)	LOW	U/C	HIGH	U/C	LOW	LOW	HIGH
Rouel et al. (2016)	HIGH	LOW	LOW	LOW	LOW	LOW	LOW
Shaw et al. (2015)	HIGH	LOW	HIGH	LOW	LOW	LOW	HIGH
Shinagawa et al. (2007)	LOW	LOW	LOW	LOW	LOW	LOW	LOW
Simblett et al. (2012)	LOW	U/C	N/A	U/C	LOW	LOW	N/A
Simblett et al. (2017)	U/C	LOW	N/A	U/C	LOW	LOW	N/A
Vélez-Pastrana et al. (2016)	HIGH	LOW	HIGH	U/C	LOW	LOW	HIGH
Velligan et al. (2002)	HIGH	LOW	LOW	HIGH	LOW	LOW	LOW
Waid-Ebbs et al. (2012)	U/C	LOW	N/A	U/C	LOW	LOW	N/A

Note. N/A = not applicable, U/C = unclear

Appendix F: Reviewer agreement

Table 2.

A sample of the papers reviewed by a second reviewer in the systematic review

	Author (H.W)		Reviewer (P.M)	
Author & Title	Include/Exclude	Reason for exclusion	Include/Exclude	Reason for exclusion
Badrkhahan et al. (2019)	Exclude	Not an EF measure	Exclude	Not on EF
Dilandro (2008)	Exclude	Not an EF measure	Exclude	Not on EF
Duggan et al. (2018)	**Include**		**Include**	
Dukart et al. (2015)	Exclude	Not an EF measure	Exclude	Not on EF
Fegyveres et al. (2008)	Exclude	Not an EF measure	Exclude	Not on EF
Hauser et al. (2013)	**Include**		**Include**	
Julayanont et al. (2015)	Exclude	Not an EF measure	Exclude	Not on EF
Mani et al. (2018)	**Include**		**Include**	
Manivannan et al. (2019)	Exclude	Not a rating scale measure	Exclude	No rating scale
Niemeier et al. (2013)	**Include**		**Include**	
Park et al. (2012)	Exclude	Only reports validity	Exclude	No reliability measure
Rand et al. (2009)	Exclude	Not a rating scale measure	Exclude	Not a rating scale
Simões (2012)	Exclude	Does not assess psychometrics of EF rating scales	Exclude	Review - not including EF rating scale
van Beilen et al. (2005)	Exclude	Not a rating scale measure	Exclude	Test - Not a rating scale
Velligan et al. (2002)	**Include**		**Include**	
Waldon et al. (2016)	Exclude	Child sample	Exclude	Children
Wang et al. (2017)	Exclude	Not an EF measure	Exclude	Not an EF rating scale

Viklund et al. (2019)	**Include**		**Include**	
Winkens et al. (2009)	Exclude	Not an EF measure	Exclude	Not on EF
Withrington et al. (2014)	Exclude	Only reports validity	Exclude	Not mentioning reliability

Appendix G: Research Ethics Committee approval

Faculty of Medicine and Health Sciences Research Ethics Committee

University of East Anglia

Hannah Wakely
MED

Research & Innovation Services
Floor 1, The Registry
University of East Anglia
Norwich Research Park
Norwich, NR4 7TJ

Email: fmh.ethics@uea.ac.uk

Web: www.uea.ac.uk/researchandenterprise

18 February 2019

Dear Hannah

Title: Psychometric Properties of The Revised Dysexecutive Questionnaire (DEX-R) In A Non-Clinical Population.

Reference: 201819 - 032

Thank you for your response to the recommendations from the FMH Ethics Committee to your proposal. I have considered your amendments and can now confirm that your proposal has been approved.

Please can you ensure that any further amendments to either the protocol or documents submitted are notified to us in advance, and also that any adverse events which occur during your project are reported to the Committee.

Approval by the FMH Research Committee should not be taken as evidence that your study is compliant with GDPR and the Data Protection Act 2018. If you need guidance on how to make your study GDPR compliant, please contact your institution's Data Protection Officer.

Please can you also arrange to send us a report once your project is completed.

Yours sincerely

Professor M J Wilkinson
Chair, FMH Research Ethics Committee

Appendix H: Participant Information Sheet

Participant Information Sheet: January 2019 Version 2

Study Title: "Psychometric properties of the revised Dysexecutive Questionnaire (DEX-R) in a non-clinical population"

Thank you for your interest in this study which forms part of a Doctorate in Clinical Psychology thesis. Before you decide whether you would like to take part, please read the following information carefully. If you would like more information about the study, please do not hesitate to email me any questions.

What is this research looking at?

This research is looking at how we measure a particular set of mental abilities known as 'executive' functions in the general population. These relate to thinking skills such as our memory, concentration, emotional regulation and motivation. There are variations between people in the general population, and it is this difference that the research is interested in. We already understand that difficulties in these mental abilities can be more pronounced for those who have a neurological illness or brain injury.

Do I have to take part?

If you want to take part, you will need to provide informed consent. This is your agreement that you have been given enough information about the study to help you decide whether or not to take part. You do not have to take part, and your participation is voluntary.

What will happen if I agree to take part?

First you must confirm that you are over 18 and that you speak English. You will then be asked to complete a consent form. You will then be asked to complete two questionnaires made up of 83 questions in total, and answer a few questions about you, all of which should take no longer than 30 minutes. We want to repeat one of the questionnaires a second time. After three weeks you will be emailed again to complete this. You will be given an online debrief which will address the aims of the study. You will also be provided with contact information for the researcher and supervisor should you want further information.

Are there any problems with taking part?

We do not think taking part will cause any difficulties or problems, other than requiring about half an hour of your time. Some people might have concerns about their thinking skills. These questionnaires cannot tell us if you have a particular health condition that

might affect thinking skills. Although the questions relate to challenges found in the typical population, if taking part does raise any questions or concerns, you should discuss this with your general practitioner.

Will it help me if I take part?

No, this study will not help you if you take part. Taking part in the study will add to our understanding of these types of mental skills in the general population and in people with health conditions involving the brain.

How will you store the information that I give you?

The information generated from this study will be strictly confidential and stored in accordance with the law on keeping information safe the General Data Protection Act. All data will be anonymous, and each person's responses assigned a unique code to help identify it. Where your email is kept for communication purposes it will not be linked to your data in any way. All data will be kept in the custody of the research team at UEA. You will be assigned a unique code to maintain anonymity, this will ensure no data can be traced back to your name. Data will be archived for 10 years after the end of the study, after this period it will be securely destroyed. All electronic data will be stored on an encrypted USB and password protected computer. Data will only be accessed by the lead researcher and research team. We will only require an email address for distributing the project, no other personal identifiable information will be collected.

How will the data be used?

Data will be analysed and used for a trainee clinical psychologist as part of their thesis for completion of the Doctorate in Clinical Psychology (DClinPsy). It may be used for further analyses by members of the research team for example to compare scores against patients with brain injuries.

What happens if I agree to take part, but change my mind later?

Your participation is voluntary, and you are under no obligation to take part. You can withdraw at any time during completion of the survey, but after that, as the data is anonymous, we won't be able to withdraw your data. If you ask to withdraw after completing the first set of questionnaires, then we will stop the email being sent reminding you to complete the second set of questionnaires.

How do I know that this research is safe for me to take part in?

The study has been reviewed by staff at UEA, the research team and has been approved by the University of East Anglia Research Ethics Committee.

If you have any questions or concerns about this research, please contact the researchers or department.

Questions regarding this research can be directed to the research team via:

Researcher Contact details: Hannah Wakely: h.wakely@uea.ac.uk

Supervisor: Dr Fergus Gracey: f.gracey@uea.ac.uk

Concerns or complains about the research can be made through the Head of Department:

Professor Niall Broomfield, Head of Department.

Norwich Medical School

Faculty of Medicine and Health Sciences

University of East Anglia

Norwich

NR4 7TJ

Appendix I: Consent Form

Consent Form: January 2019 Version 2

Study Title: "Psychometric properties of the revised Dysexecutive Questionnaire (DEX-R) in a non-clinical population"

Name of Researcher: Hannah Wakely

Thank you for your interest in this study exploring psychometric properties of a measure of dysexecutive problems. The term 'dysexecutive' refers to variations in thinking skills such as our memory, concentration, emotional regulation and motivation. The following questionnaire is part of a research project by Hannah Wakely for a thesis research project as part of the Doctorate in Clinical Psychology.

Your participation is voluntary, and you are free to withdraw at any time during completion of the survey without giving any reason and without it affecting you in any way.

Your email address will not be shared outside of the research team or published in the final report(s) from this study.

If you are interested in taking part, we would like you to confirm the following:

I have read the preceding information describing this study ☐

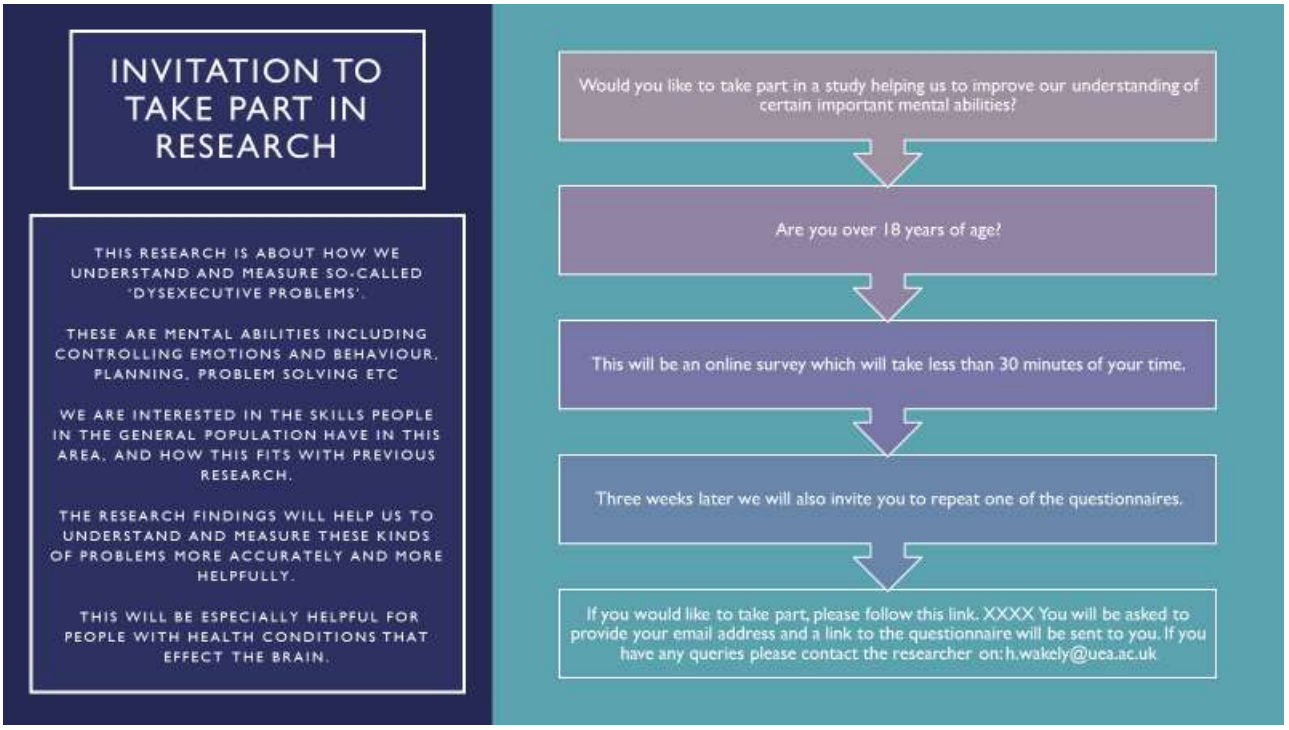
I understand I am free to withdraw at any point by closing the survey window, but not once I have completed the survey ☐

I understand my anonymous data may be used by the research team in future research studies ☐

I am 18 years of age or older ☐

I consent to taking part in the research study ☐

Appendix J: Advertising Materials



Appendix K: Pearson Permission Form (DEX-R)

Pearson
Clinical Assessment
80 Strand
London WC2R 0RL, UK
T +44 (0)20 7010 2860
E info@pearsonclinical.co.uk
W pearsonclinical.co.uk

Permission Agreement

This Permission Agreement (herein as "Agreement") entered into as 27 July 2018 between Pearson, Clinical Assessment, a division of Pearson Education Ltd, at 80 Strand, London WC2R 0RL (herein "Publisher") and

NAME: University of East Anglia

ADDRESS: Department of Clinical Psychology
Norwich Medical School
University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

(herein "Licensee") WITNESS: Dr Fergus Gracey

WHEREAS the Publisher is the owner of the DEX-R (herein the "Work(s)"); and

WHEREAS the Licensee wishes to use the Work in a computerised survey for research purposes only (herein the "Licensed Use").

NOW, THEREFORE, the Publisher and the Licensee agree as follows:

1. The Licensee may either produce, have produced, and/or distribute the Work(s) specified above, solely for the Licensed Use and subject to the terms and conditions of this Agreement.
2. The Work(s) shall be identified by title on any reproduction unless otherwise expressly provided in this Agreement.
3. The Licensed Use specifically excludes the right to print, copy or distribute in any form, or to translate, adapt, revise, or to exhibit, represent, record or reproduce any portion of the Work(s), either separately or as part of any other larger publication, except as otherwise expressly provided herein. The Licensee agrees to abide by copyright, patent, and trademark laws of such other laws as governed by and interpreted under the laws of the United Kingdom.
4. The Licensee will take all necessary precautions to safeguard all test materials by limiting access to only those individuals or agencies with a responsible, professional interest in the test's security.
5. The Licensee shall submit to the Publisher a copy of any publication which incorporates the Work(s), indicating contemplated design and format, and shall not produce or have produced such edition

Pearson
Clinical Assessment
80 Strand
London WC2R 0RL, UK
T +44 (0)20 7010 2860
E info@pearsonclinical.co.uk
W pearsonclinical.co.uk

without approval of content and design in writing from the Publisher. The Publisher agrees to review such copy and to indicate approval or make recommendations for changes within fifteen (15) days of its receipt.

6. It is understood and agreed that while the Work(s) may be used for the purposes of the research identified within the Licensed Use, no commercial use of the Works may be made. By way of example, the Licensee may not sell copies of the Work(s) to any third parties, nor charge for access to the Work(s). All unused copies of the Work prepared by the Licensee hereunder must be destroyed upon completion of the trial identified within the Licensed Use, or upon expiry or termination of this Agreement, whichever is the earlier

7. All rights in the Work(s) not granted to the Licensee by this Agreement are expressly reserved to the Publisher.

8. The Licence granted herein shall be for a period commencing September 2018 and terminating March 2020 whereupon the Licensed Use shall cease. Licensee must obtain written permission for any extension of this Agreement.

9. Any reproduction of any portion of the Work(s) or subsequent reference to the Work(s) in reports/articles resulting from the Licensed Use herein shall bear the following notices:

   Dex Questionnaire Revised (DEX-R). Copyright © 1996, 2013 by Barbara A Wilson, Nick Alderman, Paul W Burgess, Hazel Emslie and Jonathan J Evans. Published by Pearson Education Ltd. Adapted and reproduced with permission from Pearson Education Ltd. All rights reserved. FOR RESEARCH PURPOSES ONLY

10. In any published or unpublished reports of the Licensed Use authorized under this Agreement, a specific acknowledgement of the permission shall be made, including reference to the full title of the Work(s), the copyright notice, the author, and the Publisher.

11. Pearson Education Ltd will not charge a fee for up to 750 administrations. If additional reproductions above those authorized in this agreement become necessary, the Publisher must be contacted in advance.

12. The Licence granted herein is non-exclusive and non-transferable to any third party without written permission from the Publisher. The Licensee will remain liable for any breach of this Agreement by any such affiliate or collaborator.

13. The Licensee acknowledges and agrees that the rights granted under this Agreement by the Publisher do not extend to any third party material which the Publisher notifies the Licensee is included in the Work(s). The Licensee is fully responsible for obtaining any necessary permission from the copyright holder for any such third-party material.

Pearson
Clinical Assessment
80 Strand
London WC2R 0RL, UK
T +44 (0)20 7010 2860
E info@pearsonclinical.co.uk
W pearsonclinical.co.uk

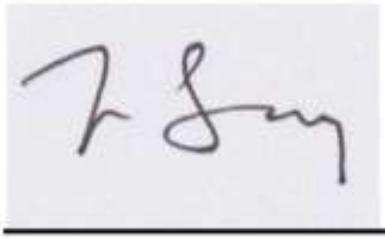
Pearson

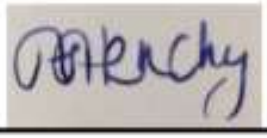
14. This Agreement shall become effective only if it is executed by the Licensee within thirty (30) days of the effective date shown above.

15. This instrument contains the entire Agreement between the parties and there are merged herein all prior and collateral understandings and agreements. No amendment or modification of this Agreement shall be valid unless in writing and signed by both parties.

16. Regardless of the place of its physical execution or performance, this Agreement shall be governed by English and Welsh laws and shall be subject to the exclusive jurisdiction of the courts of England and Wales.

Pearson, Clinical Assessment
A division of Pearson Education Ltd





31/7/2018

Dr Fergus Gracey

Senior Research Fellow and Clinical Neuropsychologist

Faye Henchy
30 July 2018
Pearson Assessment

80 Strand
London
WC2R 0RL

Appendix L: PAR license agreement (FrSBe)

Creating Connections. Changing Lives. PAR®

LICENSE AGREEMENT

THIS AGREEMENT, made this May 16, 2019, by and between Psychological Assessment Resources, Inc., a Florida Corporation, with its principal offices located at 16204 North Florida Avenue, Lutz, Florida 33549, hereinafter referred to as PAR, and Hannah C. Wakely, with her principal offices located at the University of East Anglia, Department of Clinical Psychology, Norwich Medical School, Norwich Research Park, Norwich NR4 7TJ, United Kingdom, hereinafter referred to as Licensee.

1) RECITALS

PAR has developed and holds all copyrights and distribution rights to certain psychological tests and related materials as listed in Schedule A, hereinafter called "Test". The Test consists of PAR's items, scoring keys, scales, profiles, standard-score conversion tables, norms tables, interpretive information, and related materials created, prepared, devised, and combined by PAR for the administration, scoring, reporting, and analysis of the Test, and includes the words, symbols, numbers, and letters used to represent the Test. Licensee desires to develop automated procedures for the secure and encrypted administration of the Test through Licensee's secure internet assessment website utilizing Qualtrics. The access to Licensee's website will be by invitation only in connection with Licensee's research titled, *Further validation of the DEX-Revised Questionnaire on a healthy population* and to subjects for this research purpose only (the "Limited Purpose(s)"). Unless permitted to do so by a separate license agreement, Licensee only has the right to use the Test for the Limited Purpose described above.

In consideration of the mutual covenants and promises expressed herein and other good and valuable considerations, it is agreed as follows:

2) LICENSE

PAR hereby grants to Licensee, subject to the terms of this Agreement, a non-transferable, non-refundable, non-exclusive license to place the Test on Licensee's Website for the Limited Purpose described in Section 1 above. Licensee agrees to hold secure and treat as proprietary all information transferred to it from PAR. Licensee shall carefully control the use of the Test for the Limited Purpose described in this Agreement. Licensee's use of the Test will be under the supervision or in consultation with a qualified psychologist or other qualified individual and consistent with the then current edition of the Standards for Educational and Psychological Testing published by the American Psychological Association.

16204 N. Florida Ave. | Lutz, FL 33549 | 813.968.3003 | parinc.com

3) TERMS AND TERMINATION

The initial term of this Agreement shall extend from May 16, 2019 through December 31, 2019, and may be extended only by mutual agreement of the parties. Notwithstanding any other provision of this Agreement, this Agreement may be terminated if any of the following events occur:

(a) Termination is mutually agreed to by the parties.

(b) Licensee defaults in the performance of any of its duties hereunder.

On the effective date of expiration or termination of this Agreement pursuant to subsections (a) and (b) above, all rights in this Agreement revert to PAR. Computer software programs written by or for Licensee remain the property of Licensee. Licensee warrants that upon expiration or termination of this Agreement under subsections (a) and (b) above, and except as set forth in any separate license agreement relating thereto, all portions of the Test licensed hereunder shall be removed from Licensee's Website. Failure to cease all uses of the Test shall constitute copyright infringement.

4) TERMINATION RIGHTS

In the event of termination pursuant to paragraph 3 above for any reason, PAR shall not be liable to Licensee for compensation, reimbursement or damages for any purpose, on account of any expenditures, investments, leases or commitments made or for any other reason whatsoever based upon or growing out of this Agreement.

5) CONDITIONS OF USE

PAR shall have the right to review, test, and approve that portion of Licensee's Website which includes the Test. Following PAR's approval of that portion of Licensee's Website containing the Test, the manner in which the Test appears on such Website shall not be changed in any material way without prior approval of PAR.

The computer programs developed by Licensee and used in any phase of administration and scoring of the Test shall be fully tested by Licensee and shall be encrypted and reasonably protected from access, intrusion and changes by persons who are not authorized agents of Licensee. In addition to the foregoing, Licensee shall exert all reasonable commercial efforts to prevent the Programs, and any accompanying code for the administration of the Test from being accessed, viewed or copied by others. Licensee warrants the accuracy of such scoring and reporting.

6) PROPRIETARY RIGHTS

PAR is the owner of all right, title and interest in the Test. Licensee shall acquire no right or interest in the Test, by virtue of this Agreement or by virtue of the use of the Test, except the right to use the Test in accordance with the provisions of this Agreement. Licensee shall not modify or revise the Test in any manner without written approval by PAR. All uses of the Test by Licensee shall inure to the benefit of PAR. Licensee agrees not to challenge or otherwise interfere with the validity of the Test or PAR's ownership of them.

7) ROYALTIES

Licensee agrees to pay PAR a royalty fee for use of the Test and copyrighted materials contained therein, at the rate of $5.00 USD per each test administration of the Test. Licensee will also provide PAR with an itemized accounting of all administrations of each Test administered by Licensee during the term of this agreement. Licensee shall pay to PAR Three Hundred and Ninety US Dollars ($390.00 USD) as an initial license fee ($5.00 USD per administration for 78 administrations), which is due and payable upon the signing of this License Agreement. Licensee shall also pay PAR $5.00 USD per each test administered for any tests administered above 78 by January 15, 2020.

Licensing fees paid to PAR will be payable in US Dollars drawn on a US bank. Any taxes levied on fees by Licensee's government, or fees deducted by Licensee's bank (originating or intermediary) and/or financial institution, shall be paid by Licensee and shall not reduce the amount due to PAR.

For the purposes of this Agreement, an administration of the Test includes any instance where the Test is completed wholly or in part by a subject.

8) ACCOUNTING

Licensee shall develop secure computerized accounting methods acceptable to PAR. Such accounting methods must include an electronic counting mechanism which will accurately record the number of administrations of each Test used. Licensee will keep accurate financial records of all transactions relating to the use of the Test, and PAR shall have the right to examine the software and records of Licensee pertaining to the use of the Test. Licensee will make such software and records accessible to PAR or its nominee during normal working hours upon not less than five (5) business days' prior written notice. Licensee shall retain such software and records for at least one year from the date this Agreement expires or the effective termination date.

The Website shall contain the following copyright notice:

> "Adapted and reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc. (PAR), 16204 North Florida Avenue, Lutz, Florida 33549, from the Frontal Systems Behavior Scale by Janet Grace, PhD and Paul F. Malloy, PhD, Copyright 1992, 2000, 2001 by PAR. Further reproduction is prohibited without permission of PAR."

9) INDEMNITY

Licensee agrees to indemnify PAR and hold PAR harmless against any claim or demand or against any recovery in any suit (including taxes of any kind, reasonable attorney's fees, litigation costs, and other related expenses) that may be:

(a) brought by or against PAR, arising or alleged to have arisen out of the use of the Test by Licensee;

(b) sustained or incurred by PAR, arising or alleged to have arisen in any way from the breach of any of Licensee's obligations hereunder; or

(c) incurred by PAR in any litigation to enforce this Agreement, including litigation against Licensee.

10) ASSIGNMENT

Licensee shall not assign this Agreement or any license, power, privilege, right, or immunity, or delegate any duty, responsibility, or obligation hereunder, without the prior written consent of PAR. Any assignment by PAR of its rights in the Test shall be made subject to this Agreement.

11) GOVERNING LAW

This Agreement shall be construed according to the laws of the State of Florida of the United States of America. Venue for any legal action relative to this Agreement shall be in the appropriate state court in Hillsborough County, Florida, or in the United States District Court for the Middle District of Florida, Tampa division. Licensee agrees that, in any action relating to this Agreement, the Circuit Court in Hillsborough County, Florida or the United States District Court for the Middle District of Florida, Tampa Division, has personal jurisdiction over Licensee, and that Licensee waives any argument it may otherwise have against the exercise of those courts' personal jurisdiction over Licensee.

12) SEVERABILITY

If any provision of this Agreement shall, to any extent, be invalid and unenforceable such provision shall be deemed not to be part of this Agreement, and the parties agree to remain bound by all remaining provisions.

13) EQUITABLE RELIEF

Licensee acknowledges that irreparable damage would result from unauthorized use of the Test and further agrees that PAR would have no adequate remedy at law to redress such a breach. Therefore, Licensee agrees that, in the event of such a breach, specific performance and/or injunctive relief, without the necessity of a bond, shall be awarded by a Court of competent jurisdiction.

14) ENTIRE AGREEMENT OF THE PARTIES

This instrument embodies the whole Agreement of the parties. There are no promises, terms, conditions, or obligations for the Test licensed hereunder other than those contained herein; and this Agreement shall supersede all previous communications, representations, or agreements, either written or verbal, between the parties hereto, with the exception of any prior agreements that have not previously been terminated by written consent of both parties or by one party if the terms of the agreement allow. This Agreement may be changed only by an agreement in writing signed by both parties.

15) NOTICES AND MODIFICATIONS

Any notice required or permitted to be given under this Agreement shall be sufficient if in writing and if sent by certified or registered mail postage prepaid to the addresses first herein above written or to such addresses as either party may from time to time amend in writing. No letter, telegram, or communication passing between the parties hereto covering any matter during this contract, or periods thereafter, shall be deemed a part of this Agreement unless it is distinctly stated in such letter, telegram, or communication that it is to constitute a part of this Agreement and is to be attached as a right to this Agreement and is signed by both parties hereto.

16) SUCCESSORS AND ASSIGNS

Subject to the limitations on assignments as provided in Section 10, this Agreement shall be binding on the successors and assigns of the parties hereto.

17) PARAGRAPH HEADINGS

The paragraph headings contained in this Agreement are inserted only for convenience and they are not to be construed as part of this Agreement.

18) AUTHORIZATION AND REPRESENTATION

Each party represents to the others that it has been authorized to execute and deliver this Agreement through the persons signing on its behalf.

IN WITNESS WHEREOF, the parties have executed this Agreement in duplicate on the date first herein above written.

ACCEPTED AND AGREED:

University of East Anglia:

BY: _____

HANNAH C. WAKELY

Title: TRAINEE CLINICAL PSYCHOLOGIST

DATE: 07.06.2019

ACCEPTED AND AGREED:

PAR:

BY: _____

KRISTIN GRECO, MBA

Title: CEO

DATE: June 17, 2019

PAYMENT RECEIVED: VISA

PAR CUSTOMER No.: 260923

SIGNATURE OF PROFESSOR REQUIRED:

I hereby agree to supervise this student's use of these materials. I also certify that I am qualified to use and interpret the results of these tests as recommended in the *Standards for Educational and Psychological Testing*, and I assume full responsibility for the proper use of all materials used per this Agreement.

BY: _____

Printed Name: Dr Fergus Gracey, ClinPsyD

SCHEDULE A

The Test licensed to Licensee pursuant to the above license consist of PAR's items, scoring keys, scales, profiles, standard-score conversion tables, norms tables, and related materials created, prepared, devised, and combined by PAR for the administration, scoring, reporting, and analysis of the Test, and include the words, symbols, numbers, and letters used to represent the Test. However, PAR and Licensee acknowledge and agree that Licensee may use only the PAR items and scoring information for the Test as appropriate for the Limited Purpose. The Test referred to in the body of this Agreement is defined as follows:

1) Frontal Systems Behavior Scale (FrSBe)
   Self-Rating Test Booklet
   Self-Rating Profile Forms / Score Report

Appendix M: Responses to health questions in the empirical paper

Table 3.

Participant responses to health questions (n 140)

Health Condition		*n (%)*
Brain Injury		
	Stroke	1 (0.7)
	Traumatic Brain Injury	7 (5)
	Other	0 (0)
Neurodegenerative		
	Parkinsons disease	0 (0)
	Dementia	0 (0)
	Huntington disease	0 (0)
	Multiple Sclerosis	0 (0)
	Other	
	Epilepsy	1 (0.7)
Neurodevelopmental		
	Autism Spectrum Disorder	1 (0.7)
	Attention Deficit Disorder	2 (1.4)
	Learning disability	1 (0.7)
	Other	
	Dyslexia/ mild dyspraxia	1 (0.7)
	Dyspraxia	1 (0.7)
Mental Health		
	Bipolar Disorder	1 (0.7)
	Psychotic Disorder	0 (0)
	Schizophrenia	0 (0)
	Other	
	Depression	4 (2.9)
	Anxiety	2 (1.4)
	PTSD	1 (0.7)
	OCD	1 (0.7)

Note. Some participant reported multiple health conditions

Appendix N: Frequency of responses on DEX-R items

Table 4.

DEX-R item distribution for time 1.

Item Number and Description	Frequency (%) (n = 125)				
	Never	Occasionally	Sometimes	Fairly often	Very often
1. Impulsivity	32 (25.6)	59 (47.2)	29 (23.2)	5 (4.0)	0 (0)
2. Prospective memory	32 (25.6)	59 (47.2)	24 (19.2)	6 (4.8)	4 (3.2)
3. Apathy	32 (25.6)	57 (45.6)	24 (19.2)	11 (8.8)	1 (0.8)
4. Initiation	21 (16.8)	56 (44.8)	29 (23.2)	15 (12.0)	4 (3.2)
5. Planning	54 (43.2)	43 (34.4)	17 (13.6)	8 (6.4)	3 (2.4)
6. Social disinhibition	28 (22.4)	65 (52.0)	21 (16.8)	9 (7.2)	2 (1.6)
7. Intention	29 (23.2)	49 (39.2)	22 (16.6)	20 (16.0)	5 (4.0)
8. Verbal aggression	3 (2.4)	38 (30.4)	47 (37.6)	33 (26.4)	4 (3.2)
9. Verbal fluency	19 (15.2)	56 (44.8)	35 (28.0)	13 (10.4)	2 (1.6)
10. Anger	36 (28.8)	62 (49.6)	19 (15.2)	7 (5.6)	1 (0.8)
11. Perseveration	77 (61.6)	32 (25.6)	11 (8.8)	4 (3.2)	1 (0.8)
12. Performance monitoring	57 (45.6)	57 (45.6)	11 (8.8)	0 (0)	0 (0)
13. Abstract thinking	66 (52.8)	45 (36.0)	9 (7.2)	5 (4.0)	0 (0)
14. Metaworry	44 (35.2)	48 (38.4)	17 (13.6)	12 (9.6)	4 (3.2)
15. Lack of concern	57 (45.6)	34 (27.2)	20 (16.0)	13 (10.4)	1 (0.8)
16. Blunted affect 1	60 (48.0)	42 (33.6)	16 (12.8)	6 (4.8)	1 (0.8)
17. Working memory	40 (32.0)	53 (42.4)	23 (18.4)	6 (4.8)	3 (2.4)
18. Lack of social composure	46 (36.8)	43 (34.4)	25 (20.0)	8 (6.4)	3 (2.4)
19. Insight	86 (68.8)	27 (21.6)	7 (5.6)	4 (3.2)	1 (0.8)
20. Inertia	44 (35.2)	54 (43.2)	16 (12.8)	8 (6.4)	3 (2.4)
21. Temporal sequencing	80 (64.0)	32 (25.6)	8 (6.4)	5 (4)	0 (0)
22. Cognitive control	33 (26.4)	59 (47.2)	14 (11.2)	16 (12.8)	3 (2.4)
23. Variable motivation	72 (57.6)	36 (28.8)	12 (9.6)	4 (3.2)	1 (0.8)
24. Physical aggression	102 (81.6)	17 (13.6)	3 (2.4)	2 (1.6)	1 (0.8)
25. Organisational ability	45 (36.0)	42 (33.6)	21 (16.8)	14 (11.2)	3 (2.4)
26. Inability to inhibit responses	53 (42.4)	57 (45.6)	7 (5.6)	7 (5.6)	1 (0.8)
27. Confabulation	107 (85.6)	18 (14.4)	0 (0)	0 (0)	0 (0)
28. Emotional lability	91 (72.8)	23 (18.4)	10 (8.0)	1 (0.8)	0 (0)
29. Distraction	17 (13.6)	73 (58.4)	17 (13.6)	13 (10.4)	5 (4.0)
30. Restlessness	52 (41.6)	51 (40.8)	18 (14.4)	4 (3.2)	0 (0)
31. Cognitive confidence	53 (42.4)	56 (44.8)	14 (11.2)	1 (0.8)	1 (0.8)
32. Knowing doing dissociation	68 (54.4)	46 (36.8)	9 (7.2)	2 (1.6)	0 (0)
33. Blunted affect 2	57 (45.6)	49 (39.2)	12 (9.6)	7 (5.6)	0 (0)
34. Information processing	76 (60.8)	28 (22.4)	14 (11.2)	7 (5.6)	0 (0)

35. No concern for social rules	75 (60.0)	31 (24.8)	14 (11.2)	5 (4.0)	0 (0)
36. Complex attention	44 (35.2)	59 (47.2)	14 (11.2)	8 (6.4)	0 (0)
37. Decision making	31 (24.8)	54 (43.2)	20 (16.0)	15 (12.0)	5 (4.0)

Appendix O: Means, Standard Deviations and Interquartiles

Table 5.

Means, Standard Deviations and Interquartiles for all questionnaires

			Interquartile		
	Mean	S.D	25	50	75
DEX-R, time 1	33.06	17.80	19.50	30.00	42.00
DEX-R, time 2	35.95	20.05	21.00	28.50	47
FrSBe	82.77	19.41	69.75	77.50	93.50
GAD-2	1.55	1.53	0.00	1.00	2.00
PHQ-2	0.92	1.37	0	0	1.25