THE UNIVERSITY OF HULL

The Neuropsychological Assessment of Cognitive Decline Following Brain Injury and in a Cross-Cultural Sample

being a Thesis submitted in partial fulfilment
of the requirements for the degree of Doctor of Clinical Psychology
in the University of Hull

By

Hayley Gould,
Bsc. (Hons) Psychology, University of York

August 2022

Acknowledgements

I owe an abundance of thanks to a great many people who have helped me throughout the process of conducting and writing this research. To all the participants who gave up their time to take part, thank you for being so generous and willing. To all the people who assisted in spreading the word about my research and helped me to recruit participants, I could not have completed the thesis without you. To all of the course staff and my placement supervisors, thank you for your support and guidance throughout the three years on the course.

Thank you to all of my friends and fellow course mates who have been an incredible support when things have been tough and have kept me going towards the end goal. Thank you to my friend Thomas, who has always shown unwavering support throughout all of my endeavours and was an *incredible* help with recruitment.

Thank you to my supervisors, Dr Pete Fleming and Dr Stephen Evans, for their guidance throughout the process of the research and for helping to shape the thesis into what it is today.

To Fintan, thank you for all of your support throughout the last three years and for believing in me when I did not believe in myself.
Thank you for your understanding, your care and love.

To my family, Gill, Helen and Helena-Clare, I am so grateful for your unwavering confidence, unconditional love, compassion and support. Thank you for proof reading countless essays and course work over the years, and helping me to carry on when times were hard. I would not be where I am today without you.

Finally, I would like to thank my Grandfather, Roy, who always believed in me and encouraged my curiosity and passion for learning.

I hope that I have made him proud.

This thesis is dedicated to his memory.

"If you've a chance to do some good, don't put it off, just do it"

Roy Ponting

(1933-2020)

Overview

This portfolio thesis comprises three parts. Part one is a systematic literature review and part two is an empirical paper. The overall aims of these parts are to evaluate the literature and add to the evidence base regarding the prediction of premorbid functioning during neuropsychological assessment. Part three forms the associated appendices.

Part One: A systematic quantitative literature review looking at the use of 'hold' tests and demographic variables to predict premorbid functioning, cross-culturally, in non-English speaking populations. The review looked at regression-based methods and identified twenty articles. The review demonstrated that several cross-cultural 'hold' tests have been developed using various methods that are described. It notes several limitations to the current evidence base and discusses the limitations in methodologies used. Clinical implications and avenues for further research are discussed.

Part Two: An empirical study looking to investigate the predictability of the RBANS from demographic variables and $TOPF^{UK}$ score to assist in the assessment of cognitive decline in clinical services. Multiple linear regression was used to analyse data obtained from a sample without neurological conditions (n=56) to derive regression models. The predictive power of these models was then assessed using Leave-One-Out Cross Validation. The models were, also, applied to a clinical sample (n=10) to assess their sensitivity to cognitive decline. Implications are discussed for neuropsychological assessment and further research.

Part Three contains the accompanying appendices for the previous two sections

Total word count (including tables, figures, references and appendices): 35,616

Table of Contents

Acknowledgements 2
Overview 4
Table of Contents 5
List of Tables and Figures 6
Part One: Systematic Literature Review 8
Abstract 10
Introduction 11
Method 15
Results 20
Discussion 49
References 58
Part Two: Empirical Paper 65
Abstract 67
Introduction 68
Method 74
Results 79
Discussion 99
References 104
Part Three: Appendices 112
Appendix A. Reflective Statement 113
Appendix B. Epistemological Statement 119
Appendix C. Submission Guidelines for Neuropsychology Review 121
Appendix D. Adapted AXIS tool (Downes et al., 2016) 125
Appendix E. Quality Assessment 126
Appendix F: Formatting guidelines for The Clinical Neuropsychologist. 129
Appendix G: SOC 2020 Major Categories and Skill Levels 135
Appendix H: Letter detailing permission from Pearson for online testing 137
Appendix I. Histograms plotting frequency against subtest score 138
Appendix J. Study Information Sheet for Clinical Participants 144
Appendix K. Non-Clinical Participants Information Sheet 148
Appendix L. Consent Form for Clinical Participants 152
Appendix M. Consent Form for Non-Clinical Participants 153
Appendix N. Recruitment Process Flow Charts 154
Appendix O. University Ethics Approval 156
Appendix P. University Sponsorship 157
Appendix Q. IRAS Approval 158
Appendix R. Hospital Anxiety and Depression Scale 162

List of Tables and Figures

Part One: Systematic Literature Review

List of Tables

Table 1. Derivation sample characteristics **21**

Table 2. Results of regression analyses for measures of current functioning from 'Hold' tests and demographic variables across all studies **26**

Table 3. The inclusion of demographic variables in predictor models across the studies **38**

Table 4. Cross validation methods and key findings across the studies **45**

List of Figures

Figure 1. PRISMA flow diagram showing a summary of article selection..... **18**

Figure 2. The countries and languages represented in the literature..... **24**

Part Two: Empirical Paper

List of Tables

Table 1. Derivation Sample Characteristics (mean ± SD)..... **80**

Table 2. Comparison of means between data collected online and in-person **81**

Table 3. Hierarchical regression model summaries for each RBANS index..... **84**

Table 4. Predicted and observed scores for each index **87**

Table 5. Predictive accuracy of the index regression models..... **88**

Table 6. Hierarchical Regression Models for RBANS subtests **91**

Table 7. Observed and predicted scores, and standardised absolute residuals for subtest models..... **93**

Table 8. Results and predictive accuracy of the leave-one-out cross validation analyses **95**

Table 9. Range of diagnoses and number of participants included..... **97**

Table 10. Demographic variables and scores for normative and clinical cross validation samples ... **97**

Table 11. Paired-sample t-tests across RBANS indexes for normative and clinical samples..... **98**

List of Figures

Figure 1. Line graph of absolute residuals plotted against observed qualitative category..... **89**

Figure 2. Illustration of the leave-one-out cross validation methodology..... **94**

Figure 3. Illustration of the clinical cross validation methodology **96**

Part One: Systematic Literature Review

Regression-Based Approaches to Predicting Premorbid Functioning in non-English Speaking Populations Using 'Hold' Tests and Demographic Variables: A Systematic Literature Review

Hayley Gould*[a], Dr Pete Fleming[a] & Dr Stephen Evans[b]

[a]Faculty of Health Sciences, School of Health and Social Work,

Aire Building, University of Hull, Hull, United Kingdom, HU5 7RX

[b] Psychological Medicine, The Old Chapel, Bootham Park, York, YO31 8HE

*Corresponding Author Email: hjigould@outlook.com

This paper is written in the format ready for submission to the journal

Neuropsychology Review

Please see Appendix C for submission guidelines

Word Count (including tables, figures and references): 13,074

Abstract

Introduction: Estimation of premorbid cognitive functioning is essential when quantifying cognitive decline. Commonly, word-based tests are used to predict premorbid functioning that is thought to be relatively resistant to cognitive change. These are termed 'hold' tests. One such 'hold' test paradigm is oral word reading tests that consist of reading words that have an irregular pronunciation in the English language. The pronunciation irregularities within the English language are not always present in other languages and alternative methods have been developed. The aim of this paper is to provide an up-to-date systematic review of the state of the literature that looks at regression-based, cross-cultural methods of predicting premorbid functioning within non-English speaking populations.

Method: The literature was searched systematically in April 2022. A Systematic Quantitative Literature Review Methodology was adopted. Twenty studies were identified and included in the review. The results are presented using a narrative design.

Results: The review identified a broad range of methodologies to predict premorbid functioning cross-culturally, in non-English languages, such as lexical decision tests, irregular word reading tests and accentuation tasks. Regression models were developed to predict several cognitive domains- for instance, executive functioning, fluid intelligence and memory. Cross validation methods were varied between studies.

Conclusions: Several adaptions to English-based 'hold' tests have been created. However, the need for further research is discussed to move towards adequate and accessible neuropsychological provision for all populations and countries. Heterogeneity between the studies was discussed, particularly in relation to methodological approach and clinical utility considered.

Key Words: premorbid functioning, neuropsychological assessment, cross-cultural, irregular word reading, cognitive decline, regression model

Introduction

Neuropsychological psychometric assessment comprises a key role in the identification of cognitive deficits and the formulation of appropriate, tailored support (Franzen et al., 1997). To examine the extent of cognitive decline, it is important to have knowledge of a patient's functioning prior to Aquired Brain Injury (ABI), Traumatic Brain Injury (TBI), or a form of dementia. This measure of premorbid functioning acts as a point of comparison to which current performance can be compared (Crawford, 1989; Lezak et al., 2012). Without this baseline, individuals can be misdiagnosed or deficits can be overlooked (Crawford, 1989).

Pre-injury psychometric assessment, however, is seldom available in clinical practice and, thus, various methods have been developed to estimate prior intellect. These are, for instance, demographic based approaches (e.g. Barona et al., 1984; Wilson et al., 1978); so called 'hold' tests which are psychometric measures thought to be relatively resistant to cognitive decline (Franzen et al., 1997) such as lexical decision tasks (e.g. Baddeley et al., 1993), reading tests (e.g. Nelson & McKenna, 1975) and particular Wechsler's Adult Intelligence Scale (WAIS; Wechsler, 2008) subtests such as Information and Vocabulary (Vanderploeg & Schinka, 1995); combined approaches using demographic variables and 'hold' tests (e.g. Crawford et al., 1990; Krull et al., 1995).

The National Adult Reading Test (NART; Nelson et al., 1975), was one of the first reading tests which was developed to predict premorbid intellectual ability and co-normed with the WAIS-Revised (WAIS-R; Willshire et al., 1991). Over time this was superseded by the Wechsler Test of Adult Reading (WTAR; Wechsler, 2001) and most recently by the Test of Premorbid Functioning (TOPF; Wechsler 2011) which is the current measure widely used in clinical practice and co-normed with the WAIS-IV (Wechsler, 2008).

Reading tests are based on the principles that reading ability is correlated with intelligence level and is relatively resistant to cognitive decline (Willshire et al., 1991). The TOPF for instance, consists of 70

English words which have irregular grapheme to phoneme translation and participants are scored on their ability to read each word aloud with correct pronunciation. As these words are irregular, responses cannot be deduced or guessed and, thus, correct pronunciation is based upon prior knowledge and is impacted less by cognitive decline. Reading tests are well evidenced to be valid measures of premorbid functioning. The NART, the original predecessor for the TOPF, was shown to predict 66% of the variance of the Full Scale Intelligence Quotient (FSIQ) when applied to a neurologically healthy sample (Crawford, 1989). Subsequently, the TOPF accounted for 72 percent of the variance observed in the FSIQ (Wechsler, 2008). Additionally, there were no significant differences found between TOPF measurements in a sample of individuals with a TBI and those without (Pitman et al., 2015) suggesting that the cognitive mechanisms utilised were relatively well preserved.

Despite this, the generalisable applicability of reading tests is limited. The NART, for instance, was developed in the United Kingdom (UK) for an English-Speaking population and its applicability outside of the UK and to those who are not fluent in English, is minimal. Even in English-Speaking countries, such as the United States, adaptions were required to the scoring rules which were based on British pronunciations (Franzen et al., 1997). To this end, revisions of the NART have been developed including the North American Adult Reading Test (NAART; Blair and Spreen, 1989; Spreen & Strauss, 1991) and the American version of the NART (AMNART; Grober and Sliwinski, 1991).

Whilst relatively minor adaptions are required when adapting reading tests to an alternative English-speaking population, further difficulties are encountered when translating tests to alternative languages. The tests are built on the principle that within the English language, word pronunciation irregularities are common. Languages such as Turkish, Hungarian, Finnish and Spanish, however, have high grapheme and phoneme correspondence with very few exceptions to this (e.g. Liberman and Shankweiler, 1979; Cuetos & Suárez-Coalla, 2009). Additionally, languages outside of the Indo-European family have a diverse range of alphabetic, syllabic and logographic systems (Gelb, 1952). In these cases, the concept of irregular word

reading, that English-based reading tests are built upon, are more difficult to translate. For instance, Korean is based on the writing system of *Han'gul* which, whilst still based on an alphabetic system, is made up of orthographic blocks which correspond to phonetic syllables rather than utilising a linear string of letters, as in English (Yi et al., 2017).

Similar difficulties are encountered with other 'hold' test methods based on vocabulary. For example, Lexical decision tasks must be created and validated in other languages, as opposed to being translated from English versions. This is due to the need for sufficient variation and range in frequency of use for the included words in order to have an effective scoring system.

Alternative methods of premorbid estimation such as demographic-based approaches are more easily translated. Commonly, these methods are based on regression models that allow variables such as age, years of education, occupation, and geographic locality to be used to create algorithms to predict premorbid functioning (e.g. Barona et al., 1984; Wilson et al., 1978; Crawford and Allan, 1997). Demographic methods are bolstered by the independence of variables from cognitive decline. Whilst methodologically translatable to diverse populations, regression-based algorithms experience shrinkage when applied to new populations due to the difference in relationships between demographics and Intelligence Quotient (IQ) in different cultures and countries (Franzen et al., 1997). Thus, in order to be clinically utilised, regression equations must be validated to the population in question.

Demographic-based approaches are also limited by their reliance on general patterns in a population and neglect of individual differences. The models tend to overestimate the IQ in a normative sample and underestimate IQ in a sample with above average intelligence (e.g. Griffin et al., 2002; Eppinger et al., 1987; Ritchie et al., 1996). Thus, more recently, researchers have focused on combining demographics with a reading test or an alternative 'hold' test to predict premorbid functioning, often yielding a better estimate than from either variable alone. For instance, Crawford et al. (1990) identified that the inclusion of

demographic variables alongside the NART significantly increased the variance accounted for in FSIQ, Verbal IQ (VIQ), and Performance IQ (PIQ), than with NART alone. Additionally, demographic variables increased the variance accounted for on the FSIQ alongside the TOPF (Wechsler, 2011).

Regression equations are beneficial to this method as they allow for a higher amount of predictor variables in comparison to norm tables which are often corrected by only one variable, such as age. Additionally, regression equations allow alternative cognitive tests to be validated for use in conjunction with premorbid measures without the need for relatively large samples to co-norm measures -for example, Jenkinson et al., (2018) investigated an actuarial method to predict alternative cognitive measures of verbal fluency and naming ability using the TOPF. This benefits clinical practice by providing an estimated baseline for tests commonly used in clinical practice other than general IQ.

In the same way, regression algorithms can allow for tests to be investigated for validity and cross-cultural use. Watt et al. (2018), for example, developed a regression equation using the NART and demographics to predict WAIS-IV indexes for an Australian sample.

There is a substantial need for cross-culturally generalisable, robust and evidence-based neuropsychological tests that address biases and under-representation in normative samples, and support equal access to healthcare (Pedraza & Mungas, 2008). The translation of premorbid estimation methods to different languages and populations poses challenges and requires both investigation and validation prior to clinical use. Thus, this literature review aims to investigate the use and validation of 'hold' tests within regression-based methods of estimating premorbid functioning in non-English speaking populations and the use of demographic variables within these models.

Method

Information Sources

The search took place using four electronic databases which were accessed and searched using EBSCOhost on the 1st April 2022. The databases were the following: MEDLINE, Cumulative Index to Nursing and Allied Health Literature (CINAHL), PsychINFO and Academic Search Premier. These were chosen to include both psychological and broader health literature.

Search Strategy

The search strategy was determined using an initial scoping search. Words were included relating to "Translation" and "Adaption" to encapsulate studies that adapted tests for use with different populations. This part of the strategy was adapted from a review looking at the cross-cultural applicability of the Montreal Cognitive Assessment (MoCA; O'Driscoll & Madiha, 2017). The following strategy was used:

Premorbid*

N3

Function* OR intelligen* OR estimat* OR IQ OR Cognit* OR Abilit*

AND

Norm* OR Adapt* OR Regress* OR equation* OR algorithm* OR validat* OR translat* OR reliab*

Search Limiters

Due to the nature of the review, there were no limiters placed on the language of the published paper. The reviewer took reasonable steps to accurately translate papers such as sourcing translated copies and using 'Google Translate' where necessary. The only limiter used was "peer reviewed journals" to ensure the academic rigour of the literature review.

Study Selection and Eligibility criteria

The first author was the sole reviewer, conducting the search and assessing the search results to select eligible articles based on the inclusion and exclusion criteria.

Studies were included if they met the following criteria: (1) Published any time up to 1st April 2022; (2) Regression equations were derived from a normative sample; (3) Regression equations were derived using a non-English based 'hold' test only or, alternatively, a non-English based 'hold' test and demographic variables; (4) They were published in a peer review journal to ensure papers were of sound quality; (5) Appropriate statistics were reported.

Studies were excluded based on the following criteria: (1) The premorbid tests did not exclusively measure premorbid cognitive functioning and included measures of social functioning (e.g. the Premorbid Adjustment Scale; Cannon-Spoor et al., 1982); (2) Demographic variables were only included as predictors; (3) The study was a literature review; (4) The sample in question was looking exclusively at under 18 year olds; (5) The normative sample or method was not adequately described in order to protect academic rigour of the studies included.

Summary of Selection Process

The literature search was carried out on 1st April 2022 and the initial search identified 3,349 papers. When limiters and duplications were removed, the total amount of papers screened initially was 2,274. Using the titles, papers were assessed for relevance to the research question. This left 104 papers of which the abstracts were screened, for the inclusion and exclusion criteria. This resulted in 42 full text papers which were read and screened. Of these, 16 papers were identified and included in the review.

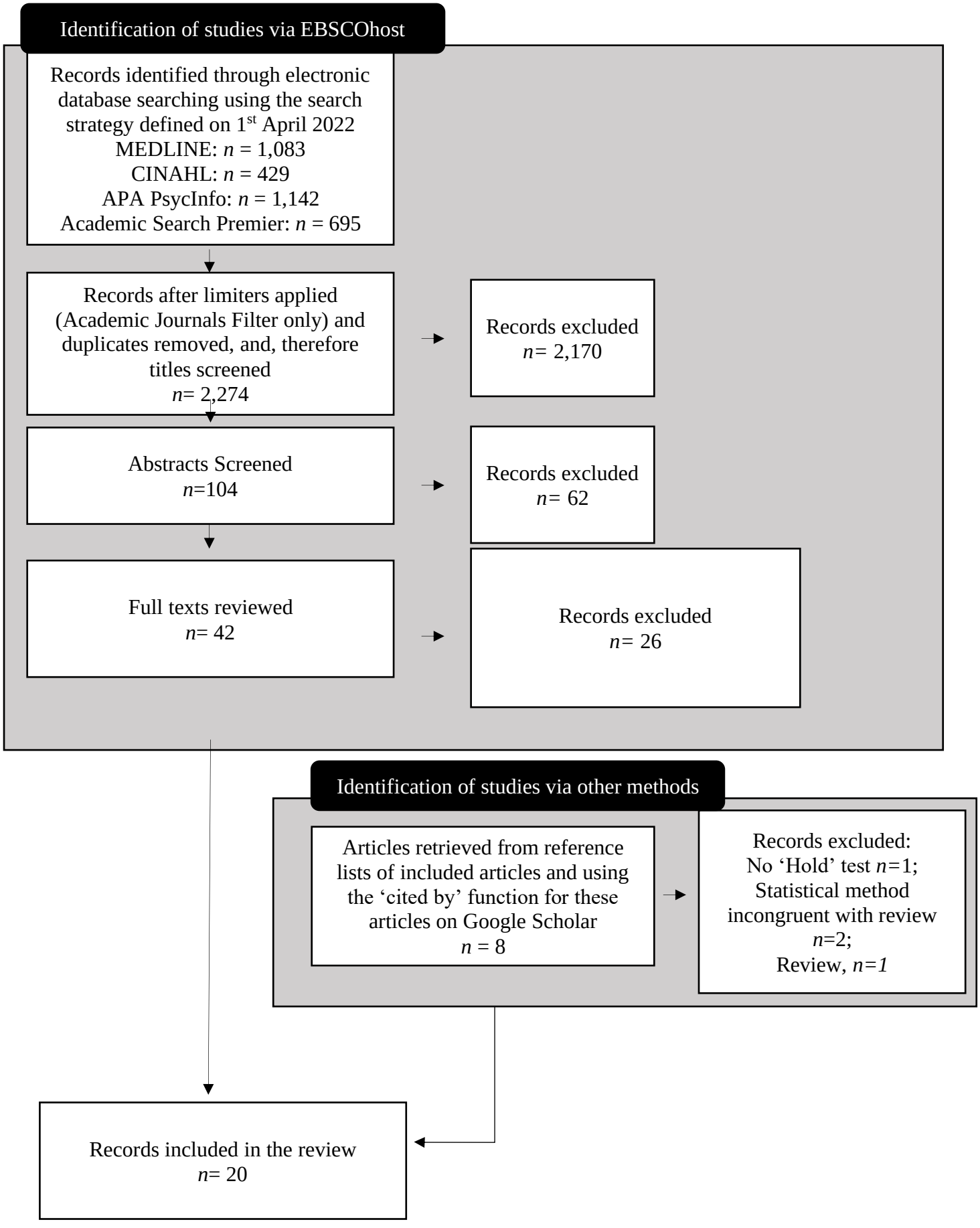
The references of the remaining articles were then reviewed to identify further relevant studies. Additionally, using the 'cited by' function on google scholar, articles that cited these papers were reviewed. Using this process, 8 further articles were identified and screened, and 4 were excluded.

In total, 20 papers were included in the review.

Figure 1 depicts this process using a Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Page et al., 2021) guidelines.

Figure 1

Flow diagram depicting the process of systematic article selection following the PRISMA reporting guidelines.



Quality Review

The Appraisal tool for Cross-Sectional Studies (AXIS; Downes et al., 2016; Appendix D) was adapted and used to assess methodological quality of each study. This tool was selected due to the studies being exclusively quantitative studies and of cross-sectional design. The tool covers common issues with cross-sectional studies across the Introduction, Methods, Results, Discussion and Ethical Concerns. The question "Were the outcome variables measured correctly using instruments/measurements that had been trialled, piloted or published previously?" was removed due to several studies presenting new measures. Thus, the highest achievable score was 19. 25% of the studies, selected randomly, were assessed by the researcher and a peer. Ratings were 86% in agreement. Any disagreements discussed and resolved, full agreed scoring for all studies is shown in Appendix E. Overall, all studies were of reasonable quality with 15 studies being rated as high quality and 5 as moderate. No studies were excluded during this process.

Data Analysis

A systemic review was carried out following the Systemic Quantitative Literature Review methodology described in Pickering et al. (2015) and took a narrative review design. Meta-analysis was not deemed to be appropriate due to the heterogeneity of the studies and the aims of the review.

Data was analysed using a process of abstraction and synthesis. The studies were read multiple times in order to obtain an overall understanding of the material. Data was then extracted into a database that detailed characteristics of the studies such as the geographic location, sample size, and measures used.

Results

Study Characteristics

Twenty studies were identified and included in the review. Studies were published between the years of 1997-2022. The studies and derivation sample demographics are shown in table 1.

Derivation Sample Characteristics

The derivation sample is defined as the participant sample on which the regression equations were modelled. Sample sizes for the derivation sample were heterogeneous across studies and ranged from 30 to 1021. The mean sample size was not calculated due to the impact of outliers on the mean. However, the median and inter-quartile range of the sample size was calculated as 105 and 87 respectively.

Five studies included individuals aged 16-18 within their control sample (Chen et al., 2009; Alves et al., 2012; Al-Ghantani et al., 2011; Tang & Yao, 2012; Karakula-Juchnowicz & Stecka, 2017). Only one study (Alves et al., 2012) included a younger sample (16-25) in an additional separate regression analysis. Six studies (Del Ser et al., 1997; Isella et al., 2005; Rolstad et al., 2009; Sarrao et al., 2015; Yi et al., 2017; Matsuoka et al., 2006) did not include individuals below 50 years in their control sample. Years of education were varied across the studies. Mean years of education ranged from 5.8 (Del Ser et al., 1997) to 14.26 (Pluck et al., 2017).

Table 1

Derivation Sample Characteristics

				Age (Years)			Years of Education		
Author (s)	Year	*n*	Sex (*n*) *(Male/Female)*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*
Al-Ghantani et al.	2011	198	99/99	NR	NR	16-65	NR	NR	Primary education- 7+ years at University
Almkvist et al.	2007	109	51/58	49.5	19.4	NR	12.8	3.3	NR
Alves et al.	2012	124	64/60	48.2	4.7	16-86	10.3	4.4	4-20
Chaurasiya,et al.	2022	207	NR	NR	NR	NR	NR	NR	NR
Chen et al.	2009	296	142/154	43.3	19.1	16-93	12.9	3.3	1-18
Colombo et al.	2002	127	56/71	58.2[a]	NR	30-80	10.0[a]	NR	3-degree level
Del Ser et al.	1997	81	39/42	72.2	5.1	‘elderly’	5.8	4.3	NR
Gomar et al.	2011	103	57/46	39.2	NR	18-65	NR	NR	NR
Isella et al.	2005	145	49/96	63.9	8.6	50-92	10.4	3.9	3-21

				Age (Years)			Years of Education		
Author (s)	Year	*n*	Sex (*n*) *(Male/Female)*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*
Karakula-Juchnowicz et al.	2017	28	14/15	Male: 37.1 [b] Female 37.5[b]	12.1 [b] 15.6[b]	16-60	NR	NR	NR
Kim et al.	2015	607	283/324	34.3	6.0	NR	NR	NR	NR
Krueger et al.	2006	45	22/23	45.2	11.9	29-73	11.4	4.1	0-18
Matsuoka et al.	2006	50	17/33	69.6	5.3	'elderly'	11.5	2.6	NR
Pluck & Ruales-Chieruzzi	2021	53	29/24	37.5	20.7	18-65	13.9	3.9	NR
Pluck et al.	2017	51	NR	38.9	18.0	18-82	14.3	3.7	6-26
Rolstad et al.	2008	53	18/35	66.1	7.7	50-78	12.3	3.1	NR
Sanjurjo et al.	2015	120	60/60	49.1	14.9	20-74	10.6	5.3	1-24
Schrauf et al.	2006	80	39/41	69.4	5.2	NR	8.5	4.2	NR

			Age (Years)				Years of Education		
Author	Year	*N*	Sex *(n)* *(Male/Female)*	*Mean*	*SD*	*Range*	*Mean*	*SD*	*Range*
Serrao et al.	2015	38	NR	67.4	5.9	60-88	11.9	5.1	4-24
Tang & Yao	2012	1021	510/511	41.1	20.7	16-92	8.9	4.0	NR
Yi et al.	2017	30	13/17	67.9	6.3	'elderly'	12.0	4.0	NR

Note: All values rounded to 1 decimal place due to heterogeneity in reporting between studies. **NR=** Not Reported. **SD=** Standard Deviation.

[a] Calculated from two reported values

[b] Value reported for complete control group which was later split into computational and validation groups

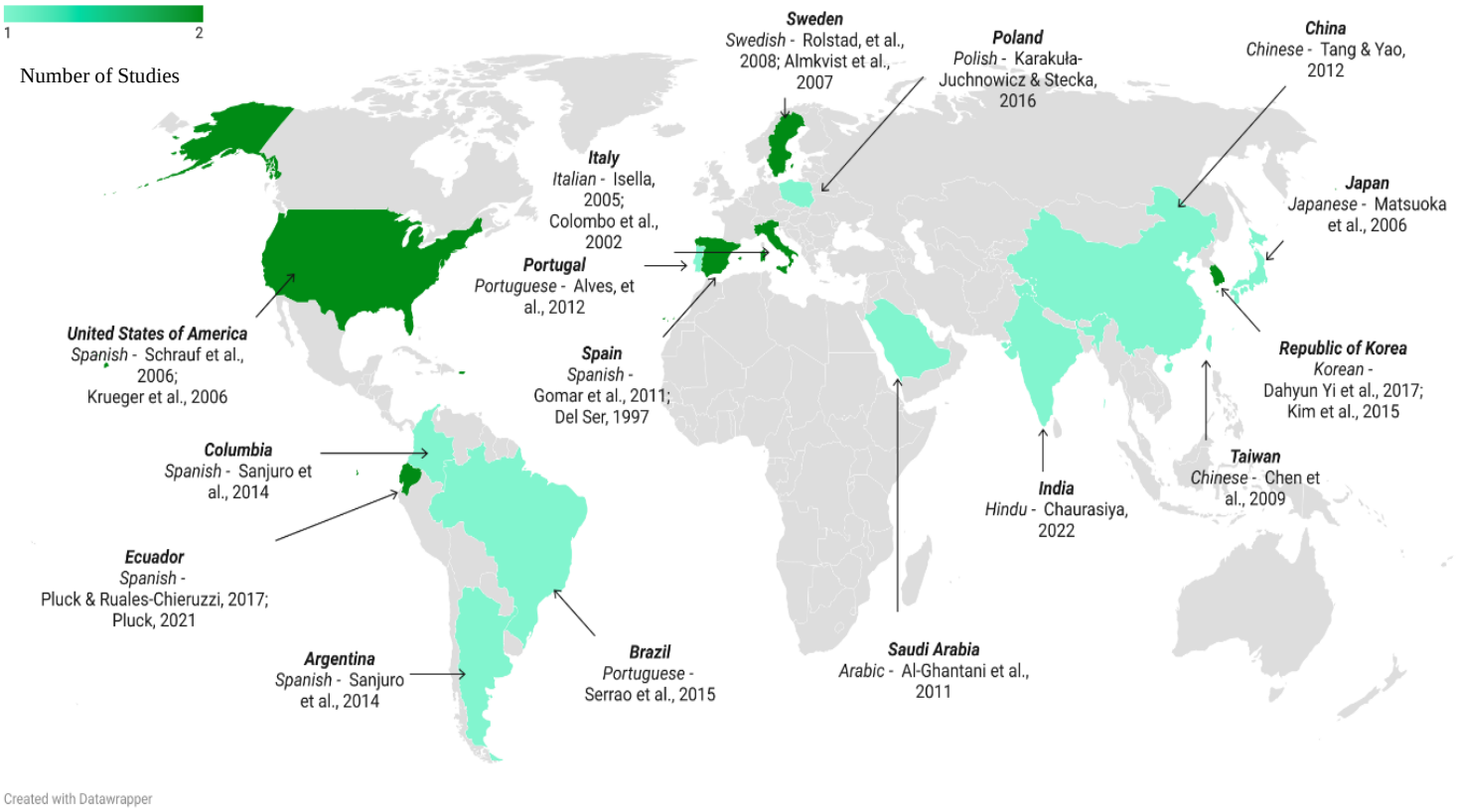
Representation of Languages and Populations

Figure 2 illustrates the representation of countries and languages included in the literature. Eight studies investigated Spanish-speaking populations and samples were included from Argentina, Columbia, Ecuador, Spain and the United States of America (USA). Two studies looked at Spanish-speaking immigrants residing in the USA (Schauf et al., 2006; Krueger et al., 2006) who originated from the following countries: Argentina, Colombia, Cuba, Dominican Republic, Ecuador, El-Salvador, Guatemala, Honduras, Mexico, Peru, Puerto Rico and Uruguay.

Two studies looked at Portuguese-speaking populations. Alves et al. (2012) considered individuals residing in Portugal and Serrao et al. (2015) investigated Brazilian-Portuguese speakers in Brazil.

Figure 2

The countries and languages represented in the literature



Countries in Asia were investigated in seven studies. Yi et al. (2017) and Kim et al. (2015) explored the prediction of premorbid functioning in the Republic of Korea in a Korean speaking population. Tang & Yao (2012) and Chen et al. (2009) researched a Chinese-speaking population in China and Taiwan respectively. India, Japan and Saudi Arabia were also represented in the literature.

In addition to studies completed in Spain and Portugal, five studies were completed in European countries. The countries were Italy, Poland and Sweden.

'Hold' Test Methods

Table 2 illustrates the created regression models, predictor variables and included neuropsychological measures across the studies for each cognitive domain.

Table 2

Results of regression analyses for measures of current functioning from 'Hold' tests and demographic variables across all studies

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Executive Functioning	Stroop-int	Stroop Test	Subtest scores	WAIS-V+PC	-	Al-Ghatani et al., 2011	Arabic	Saudi Arabia	0.45
	WCST-ppe	WCST	Subtest scores	WAIS-V+PC	-	Al-Ghatani et al., 2011	Arabic	Saudi Arabia	0.09
Memory	RAVLT-DR	RAVLT	Accentuation Test, Word Reading	TIB	Age	Isella et al., 2005	Italian	Italy	0.27
	RAVLT-IM	RAVLT	Accentuation Test, Word Reading	TIB	Age	Isella et al., 2005	Italian	Italy	0.25
	WMI	K-WAIS-IV	Irregular Word Reading	KART	Years of Education	Yi et al., 2017	Korean	South Korea	0.46
Overall Functioning	BWM-R - CK	BWM-R	Accentuation Test, Word Reading	WAT	-	Schrauf et al., 2005	Spanish	United States	0.77
	FSIQ	ISCA	-	-	Age, Occupation, Sampling Location, Sex, Years of Education	Tang & Yao, 2012	Chinese	China	0.38
	FSIQ	WAIS-III (Spanish)	-	-	Place of Birth, Years of Education	Sanjuro et al., 2014	Spanish	Columbia, Argentina	0.63
	FSIQ	WAIS-R (Swedish Version)	-	-	Sex, Years of Education	Almkvist et al., 2007	Swedish	Sweden	0.32

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Overall Functioning	FSIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT	-	Sanjuro et al., 2014	Spanish	Columbia, Argentina	0.68
	FSIQ	WAIS-IV (Spanish)	Accentuation Test, Word Reading	WAT	-	Pluck, 2021	Spanish	Ecuador	0.61
	FSIQ	WAIS-IV	Accentuation Test, Word Reading	WAT	-	Pluck et al., 2017	Spanish	Ecuador	0.68
	FSIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT	-	Gomar et al., 2011	Spanish	Spain	0.57
	FSIQ	WAIS (Italian)	Accentuation Test, Word Reading	TIB - errors	Age, Sex, Years of Education	Colombo et al., 2002 [c]	Italian	Italy	0.60
	FSIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT-Chicago	Age, Years of Education	Krueger et al. 2006	Spanish	United States-Chicago	0.59
	FSIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT	Place of Birth, Years of Education	Sanjuro et al., 2014	Spanish	Columbia, Argentina	0.76
	FSIQ	WAIS-R	Irregular Word Reading	JART-errors	-	Matsuoka et al., 2006	Japanese	Taiwan	0.78
	FSIQ	WAIS-R-PL	Irregular Word Reading	PART	-	Karakula-Jucknowicz & Stecka 2017	Polish	Poland	0.39
	FSIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI - errors	-	Alves et al., 2012	Portuguese	Portugal	0.53/0.54 [b]
	FSIQ	K-WAIS-IV	Irregular Word Reading	KART	Years of Education	Yi et al., 2017	Korean	South Korea	0.63
	FSIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI - errors	Years of Education	Alves et al., 2012	Portuguese	Portugal	0.6/063[b]

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Overall Functioning	FSIQ	WAIS-IV (Spanish)	Lexical Decision	SpanLex	-	Pluck, 2021	Spanish	Ecuador	0.40
	FSIQ	WAIS-R (Swedish Version)	Lexical Decision	SLDT – Correct responses real words, Incorrect responses pseudo words	-	Almkvist et al., 2007	Swedish	Sweden	0.49
	FSIQ	WAIS-R (Swedish Version)	Lexical Decision	SLDT – Correct responses real words, Incorrect responses pseudo words	Age, Years of Education	Almkvist et al., 2007	Swedish	Sweden	0.62
	FSIQ	WAIS-IV (Spanish)	Stem Completion Task	SCIRT	-	Pluck, 2021	Spanish	Ecuador	0.62
	FSIQ	K-WAIS-IV	Subtest Scores	WAIS- IN, MR	Age	Kim et al., 2015	Korean	South Korean	0.65
	FSIQ	K-WAIS-IV	Subtest Scores	WAIS-IN, V	Age	Kim et al., 2015	Korean	South Korean	0.63
	FSIQ	K-WAIS-IV	Subtest Scores	WAIS- MR, VP	Age, Sample Location Years of Education,	Kim et al., 2015	Korean	South Korean	0.57
	FSIQ	K-WAIS-IV	Subtest Scores	WAIS- IN, MR, VP, V	Age, Years of Education	Kim et al., 2015	Korean	South Korean	0.76
	FSIQ	ISCA	Subtest Scores	ISCA- IN	Age, Occupation	Tang & Yao, 2012	Chinese	China	0.67
	FSIQ	ISCA	Subtest Scores	ISCA-IN, FR	Age, Occupation, Sampling Location	Tang & Yao, 2012	Chinese	China	0.78

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Overall Functioning	FSIQ	ISCA	Subtest Scores	ISCA-IN, PC	Age, Occupation, Sampling Location	Tang & Yao, 2012	Chinese	China	0.77
	FSIQ	ISCA	Subtest Scores	ISCA-FR, VS	Age, Occupation, Sampling Location, Sex	Tang & Yao, 2012	Chinese	China	0.77
	FSIQ	ISCA	Subtest Scores	ISCA-FR	Age, Occupation, Sampling Location, Sex, Years of Education,	Tang & Yao, 2012	Chinese	China	0.68
	FSIQ	ISCA	Subtest Scores	ISCA-PC	Age, Occupation, Sampling Location, Sex, Years of Education,	Tang & Yao, 2012	Chinese	China	0.66
	FSIQ	ISCA	Subtest Scores	ISCA- IN, FR, PC, VC,	Age, Occupation, Sampling Location, Years of Education	Tang & Yao, 2012	Chinese	China	0.87
	FSIQ	ISCA	Subtest Scores	ISCA- IN, VC	Age, Occupation, Sex	Tang & Yao, 2012	Chinese	China	0.74
	FSIQ	ISCA	Subtest Scores	ISCA-VC	Age, Occupation, Sex, Years of education	Tang & Yao, 2012	Chinese	China	0.64
	FSIQ	K-WAIS-IV	Subtest Scores	WAIS-IN	Age, Years of Education	Kim et al., 2015	Korean	South Korean	0.53
	FSIQ	WAPIS (Indian)	Vocabulary test	Hindi Vocabulary Test	Age, Sampling Location, Sex, Years of Education	Chaurasiya et al., 2022	Hindi	India	0.49
	Subtests V + MR	WAIS-III (Portuguese)	Lexical Decision Test	LDT-Brazilian Portuguese	Years of Education	Sarrao et al., 2015	Brazilian	Brazil	0.66
	Subtests V + PC	WAIS	Accentuation Test, Word Reading	WAT	-	Del Ser et al., 1997	Spanish	Spain	0.70

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Verbal IQ and Subtests	Subtest VF	WAIS	Subtest scores	WAIS-VOC+PC	-	Al-Ghatani et al., 2011	Arabic	Saudi Arabia	0.58
	VIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT	-	Gomar et al., 2011	Spanish	Spain	0.60
	VIQ	WAIS (Italian)	Accentuation Test, Word Reading	TIB – errors	Sex, Years of Education	Colombo et al., 2002	Italian	Italy	0.43
	VIQ	WAIS-R	Irregular Word Reading	JART-errors	-	Matsuoka et al., 2006	Japanese	Taiwan	0.84
	VIQ	WAIS-R-PL	Irregular Word Reading	PART	-	Karakula-Jucknowicz & Stecka et al., 2017	Polish	Poland	0.42
	VIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI – errors	-	Alves et al., 2012	Portuguese	Portugal	0.48/0.51[b]
	VIQ	WAIS-III (Short form; Swedish)	Irregular Word Reading	NART-SWE	-	Rolstad et al., 2008	Swedish	Sweden	0.54
	VIQ	K-WAIS-IV	Irregular Word Reading	KART	Years of Education	Yi et al., 2017	Korean	South Korea	0.48
	VIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI – errors	Years of Education	Alves et al., 2012	Portuguese	Portugal	0.57/0.62 [b]

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Verbal IQ and Subtests	VIQ	WAIS-R (Swedish)	Lexical Decision	SLDT – Correct responses real words, Incorrect responses pseudo words	Age	Almkvist et al., 2007	Swedish	Sweden	0.84
	VIQ	VAIS (Indian)	Hindi Vocabulary Test	Hindi Vocabulary Test	Age, Sex, Locality, Years of Education	Chaurasiya et al., 2022	Hindi	India	0.49[c]
Non-Verbal Intelligence and Fluid Reasoning	PIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT	-	Gomar et al., 2011	Spanish	Spain	0.27
	PIQ	WAIS (Italian)	Accentuation Test, Word Reading	TIB – errors	Age	Colombo et al., 2002	Italian	Italy	0.37
	PIQ	WAIS-R	Irregular Word Reading	JART-errors	-	Matsuoka et al., 2006	Japanese	Taiwan	0.46
	PIQ	WAIS-R-PL	Irregular Word Reading	PART	-	Karakula-Jucknowicz & Stecka et al., 2017	Polish	Poland	0.24
	PIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI – errors	-	Alves et al., 2012	Portuguese	Portugal	0.42/0.43 [b]
	PIQ	WAIS-III (Short form; Swedish)	Irregular Word Reading	NART-SWE	-	Rolstad et al., 2008	Swedish	Sweden	0.21

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Non-Verbal Intelligence and Fluid Reasoning	PIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI – errors	Years of Education	Alves et al., 2012	Portuguese	Portugal	0.45/0.47 [b]
	PIQ	WAIS-R (Swedish)	Lexical Decision	SLDT – Correct responses real words, incorrect responses pseudo words	Age, Years of Education	Almkvist et al., 2007	Swedish	Sweden	0.54
	PIQ	WAIS-III (Spanish)	Accentuation Test, Word Reading	WAT	-	Gomar et al., 2011	Spanish	Spain	0.27
	PIQ [d]	WAIS (Italian)	Accentuation Test, Word Reading	TIB – errors	Age	Colombo et al., 2002	Italian	Italy	0.37
	PIQ	WAIS-R	Irregular Word Reading	JART-errors	-	Matsuoka et al., 2006	Japanese	Taiwan	0.46
	PIQ	WAIS-R-PL	Irregular Word Reading	PART	-	Karakula-Jucknowicz & Stecka et al., 2017	Polish	Poland	0.24
	PIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI – errors	-	Alves et al., 2012	Portuguese	Portugal	0.42/0.43 [b]

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Non-Verbal Intelligence and Fluid Reasoning	PIQ	WAIS-III (Short form; Swedish)	Irregular Word Reading	NART-SWE	-	Rolstad et al., 2008	Swedish	Sweden	0.21
	PIQ	WAIS-III (Portuguese)	Irregular Word Reading	TeLPI – errors	Years of Education	Alves et al., 2012	Portuguese	Portugal	0.45/0.47 [b]
	PIQ	WAIS-R (Swedish Version)	Lexical Decision	SLDT – Correct responses real words, incorrect responses pseudo words	Age, Years of Education	Almkvist et al., 2007	Swedish	Sweden	0.54
	PIQ	WAPSI (Indian adapted version)	RPSM	RPSM	Age, Sex, Locality, Years of education	Chaurasiya et al., 2022	Hindi	India	0.40[c]
	PRI	K-WAIS-IV	Irregular Word Reading	KART	Years of Education	Yi et al., 2017	Korean	South Korea	0.25
	PSI	K-WAIS-IV	Irregular Word Reading	KART	Years of Education	Yi et al., 2017	Korean	South Korea	0.33

	Measure of Current Functioning		Significant Predictor variables						
Cognitive domain	**Index Score**	**Psychometric Measure**	**'Hold' Test Type**	**'Hold' Test**	**Demographics**	**Study**	**Language**	**Country**	**R^2**
Non-Verbal Intelligence and Fluid Reasoning	RPSM	RPSM	-	-	Age, Years of Education	Chen et al., 2009	Chinese	Taiwan	0.62/0.55[a]
	RPSM	RPSM	Accentuation Test, Word Reading	WAT	-	Del Ser et al., 1997	Spanish	Spain	0.43
	RPSM	RPSM	Accentuation Test, Word Reading	WAT	-	Schrauf et al., 2005	Spanish	United States	0.58
	RPSM	RPSM	Irregular Word Reading	CGWRT	Age	Chen et al., 2009	Chinese	Taiwan	0.61/0.50[a]
	RPSM	RPSM	Irregular Word Reading	CGWRT	Age, Years of Education	Chen et al., 2009	Chinese	Taiwan	0.64/0.58 [a]
	RPSM	RPSM	Vocabulary test	Vocabulary Subtest	-	Del Ser et al., 1997	Spanish	Spain	0.44

[a]*Equations were calculated on two separate normative samples – both are reported respectively*
[b]*Regression analysis was completed both for the sample and in a sample excluding those aged 16-25, both are reported respectively.*
[c]*Adjusted R^2 Reported*
Note. PIQ refers to non-verbal subtests as measured by the WAIS (Wechsler, 1955); VIQ refers to verbal subtests as measured by the WAIS
BWM-R = Batería Woodcock-Muñoz (BWM-R. Woodcock & Muñoz-Sandoval= 1996). **BWM-R-CK =** Batería Woodcock-Muñoz- Comprehensive Knowledge. **CGWRT =**Chinese Graded Word Reading Test. **FR** =Figure Reasoning. **FSIQ=** Full Scale Intelligence Quotient. **IN** =Information. **VP** =Visual Puzzles. **ISCA =**Intelligence Scale for Chinese Adults. **JART** =Japanese Adult Reading Test. **KART** =Korean Adult Reading Test. **K=WAIS=IV** =Korean Wechsler Adult Intelligence Scale =fourth edition. **LDT** =Lexical Decision Test. **MR=** Matrix Reasoning. **NART=SWE** =National Adult Reading Test= Sweden. **PART** =Polish Adult Reading Test. **PC**= Picture Completion. **PIQ=** Performance Intelligence Quotient. **PRI=** Perceptual Reasoning Index. **PSI=** Processing Speed Index. **RAVLT-DR=** Rey Auditory Verbal Learning Task Delayed Recall. **RAVLT-IR=**Rey Auditory Verbal Learning Task Immediate Recall. **RPSM**= Raven's Progressive Matrices. **SCIRT =** Stem Completion Implicit Reading Test. **SLDT =**Swedish Lexical Decision Task. **TeLPI** =Portuguese Irregular Word Reading Test. **TIB** = Test d'Intelligenza Breve. **V=** Vocabulary. **VAIS =**Verbal Adult Intelligence Scale. **VF=** Verbal Fluency. **WAIS =** Wechsler Adult Intelligence Scale. **WAIS-III** = Wechsler Adult Intelligence Scale =third edition. **WAIS-R=** Wechsler Adult Intelligence Scale =Revised. **WAIS-R-PL =** Wechsler Adult Intelligence Scale-Revised- Polish edition. **WAPIS =**Wechsler Adult Performance Intelligence Scale. **WAT** =Word Accentuation Test. **WCST-ppe=** Word Wisconsin card sorting test percent perseverative errors. **WMI =**Working Memory Index

Irregular Word Reading

Six studies used irregular word reading as a 'hold' test to predict pre-morbid functioning. These included: Chinese Graded Word Reading Test (CGWT; Chen et al., 2009), The Irregular Word Reading Test (TeLPi; Alves et al., 2012), National Adult Reading Test Sweden (NART-SWE; Rolstad et al., 2008); Japanese Adult Reading Test (JART; Matsuoka et al., 2006), Polish Adult Reading Test (PART; Karakula-Juchnowicz & Stecka, 2017), Korean Adult Reading Test (KART; Yi et al., 2017).

Two studies utilised loan words (words adopted from foreign languages) from alternative languages within irregular word reading tests due to fixed pronunciation rules in the required language and scarcity of words with irregular pronunciation. The PART and NART-SWE used words that were loaned from countries such as English, French and Italian (Karakula-Juchnowicz & Stecka, 2017; Rolstad et al., 2008).

Three studies developed irregular reading tasks in languages that used alternative lexical scripts to Latin Script. Yi et al. (2017), for example, used the KART which was developed using existing irregular reading rules in the Korean alphabet *hangul* which is written in syllable blocks. The KART is based on the discrepancies which exist in phonological pronunciation of syllable blocks within compound words. Similarly, Matsuoka et al. (2006) utilised existing discrepancies between phonetic components and phonological pronunciation within Japanese *Kanji* script to develop the JART. Within *Kanji* script, a single character can have various pronunciations depending on the orthographic context. The JART is based on the principle that the correct pronunciation of the symbol, within the context of the compound word, uses the same lexical process as the reading of irregular English words. Additionally, Chen et al. (2009) utilised existing discrepancies in written Chinese to create the CGWT.

Accentuation Task

Accentuation tasks were used in eight studies. Similar to irregular word reading tasks, accentuation tasks are based on the pronunciation of words. However, scoring is based on appropriate accentuation and stress when pronouncing words.

Six studies used the Word Accentuation Test (WAT) which is developed for Spanish speakers. Adaptions to the WAT were included in two studies. Krueger et al. (2006) adapted the WAT for a population residing in Chicago by changing included words to adapt to the frequency of their use in Spanish-speaking population in Chicago. Similarly, Schrauf et al. (2006) investigated the need to revise the WAT for a further sample of Spanish-speaking American immigrants but did not find that adaptions were needed for this population.

Two studies used the Test Breve di Intelligenza (TIB; Colombo et al., 2002; Isella et al., 2005) to predict premorbid functioning. Italian does not have ambiguity in pronunciation of phonemes; however, it does contain irregularities in lexical stress. Dominant stress patterns are replaced in particular words with an alternative stress pattern. Thus, correct pronunciation can only be known through stored phonological knowledge.

Lexical Decision Task

Three studies investigated the use of Lexical Decision Tasks (LDT) to predict premorbid functioning which requires participants to identify which are real words and which are pseudo words. Serrao et al. (2015) studied the applicability of this test in the Brazilian Portuguese language. Pseudo words were created to begin and end with the same letter as their real-word counterparts and to match the number of syllables. Similarly, Almkvist et al. (2007) used the Swedish LDT (SLDT) in which words were randomly sampled from a dictionary and screened for frequency and length. Pseudo words in this study were created by substituting letters, adding an extra letter or by substituting syllables. Alternatively, Pluck (2021) adapted

the traditional format of a lexical decision task and investigated the use of The SpanLex. This is a Spanish lexical decision task requiring a participant to select a real word from a triplet of words.

Vocabulary Test and Intelligence Scale Subtests

Five studies used adaptions of intelligence scale subtests to predict premorbid functioning.

Three studies adapted combinations of WAIS subtests. Al-Ghantani et al. (2011) used a combination of the vocabulary and picture completion sub-test of the WAIS-R to investigate their use in predicting premorbid functioning in an Arabic speaking population in Saudi Arabia. Kim et al. (2015) looked at the use of WAIS-IV subtests (Vocabulary, Matrix Reasoning, Information and Visual Puzzles), thought to be resistant to cognitive decline, alongside demographic variables to predict premorbid functioning. Similarly, Tang & Yao (2012) used comparative subtests (Vocabulary test, Information, Picture completion and Figural reasoning) from the Intelligence Scale for Chinese Adult (ISCA).

Two studies used Vocabulary Test scores as measures of premorbid functioning, within Hindu-speaking individuals (Chaurasiya et al, 2022) in an Indian population, and within a Spanish-speaking population (Del Ser et al., 1997) respectively.

Other included tests

Two further tests were included within the studies. Pluck (2021) included a Stem Completion Implicit Reading Test (SCIRT) which required a participant to complete the word using the Lexical Stem in the context of a sentence and a picture. This is based on the principles of the irregular word reading tasks and requires participants to draw from lexical knowledge.

Chaurasiya et al. (2022) adopted an alternative approach and used Raven's Progressive Matrices (RSPM), a non-verbal measure of intelligence, to predict PIQ as measured by the WAIS.

Inclusion of Demographic Variables as predictor variables

Table 3

The inclusion of demographic variables within predictor models across the studies

Study	Age	Occupation	Place of birth	Sampling region/ Area of Country/ Locality	Sex	Years of Education
Almkvist et al., 2007	✓				✓	✓
Alves et al., 2012						✓
Chaurasiya et al., 2022	✓			✓	✓	✓
Chen et al., 2009	✓					✓
Colombo et al., 2002	✓				✓	✓
Kim et al., 2015	✓			✓	✓	✓
Krueger et al., 2006	✓	✓			✓	✓
Rolstad et al., 2008	✓				✓	✓
Sanjuro et al., 2014	✓	✓	✓		✓	✓
Serrao et al., 2015						✓
Tang & Yao, 2012	✓	✓		✓	✓	✓
Yi et al., 2017						✓

Studies that included demographics within the regression model are shown in table 3. Twelve studies included demographics within the analyses. Years of education was the most investigated as a predictor variable and was included in all ten studies.

Cognitive domains

Table 2 illustrates the models used to predict cognitive functioning that were created for each cognitive domain across the languages and populations within the literature.

Executive functioning

Only one study looked at predicting executive functioning measures. Al-Ghantani et al. (2011) predicted executive functioning using two WAIS subtests, VOC and PC, as measures of premorbid functioning (Al-Ghantani et al., 2011) in an Arabic-speaking sample. Two tests were investigated- the Wisconsin Card Sorting Test (WCST) and the Stroop test. The variance of the WCST explained by the model was just 9% and in the Stroop test, 45% of variance was explained. No demographics were included in the analysis.

Memory

Three models were created to predict performance on memory measures. Two models used the Rey Auditory Verbal Learning Test (RAVLT) indexes, delayed memory and immediate memory respectively, to measure current functioning and included an accentuation test (WAT) and age, as predictor variables in both cases (Isella et al., 2005).

One model used the working memory index, as measured by the Korean-WAIS-IV (K-WAIS-IV), to measure memory functioning including irregular word reading and years of education as predictor variables. The level of variance explained ranged from 25%-46%.

Overall Functioning

Overall cognitive functioning was defined in this literature review as an index/measure that included non-verbal and verbal abilities presented as a single score. Sixteen papers included a model predicting a measure of overall cognitive functioning. Of these, 91.67% of models presented included FSIQ as the dependent variable, as measured by adaptions of the WAIS and the ISCA. One model (Shrauf et al., 2005) used the Spanish revision of the Woodcok-Johnson Psycho-educational Battery-Revised: The Batería Woodcock-

muñoz Revised (BWM-R; Woodcok & Johnson, 1989) as a measure of current functioning. The BWM-R consists of batteries testing cognition (e.g. visual processing, auditory processing*)* and achievement (e.g. writing, reading, numeric and science skills). Finally, two models taken from two studies (Sarrao et al., 2015; Del Ser et al., 1997) used a combination of two subtests to assess current global cognition. Both studies used a combination that included the vocabulary subtest. Studies combined this score with the Matrix Reasoning and Picture completion respectively to estimate current functioning.

Two studies used subtest scores from the ISCA and K-WAIS-IV, combined with various demographic variables, to predict premorbid cognitive functioning (Kim et al., 2015; Tang & Yao, 2012). Fourteen models were created. In each study, the model that explained the highest variation in FSIQ included all four subtests. Years of education and age significantly contributed to the WAIS four subtest model to explain 76% of the variation whereas sex and sampling location did not contribute significantly to the model. On the other hand, age, occupation, sampling location and years of education significantly contributed to ISCA model to explain 87% of the variance. Only sex was found to not be a significant predictor.

LDTs were used in 10.8% of models predicting overall functioning, 13.5% of models utilised irregular word reading and 24.3% of models used accentuation tests including one that validated the WAT against BWM-R in United States. One model used the Stem completion task and explained 62% of the variance of FSIQ.

To predict the WAIS subtests measure of general functioning, Sarrao et al. (2015) and Del Ser et al. (1997) utilised a LDT, alongside years of education, and WAT, respectively, to predict WAIS verbal and nonverbal subtests together. R^2 values were 0.66 and 0.70 respectively.

Three models only included demographic variables and R^2 values ranged from 0.32-0.63. Overall, models including ‘hold’ tests and demographic variables explained a range of 59%-76% of the variance of FSIQ.

Verbal Ability

VIQ is a value that includes two indexes- Verbal Comprehension Index (VCI) and Working Memory Index (WMI). It is considered to measure abilities that are based on verbal functioning. Ten models investigated the prediction of VIQ across ten studies.

Irregular word reading was used in 60% of tests, including two models that also used years of education as a predictor variable. Two models used Accentuation tests from Italy and Spain. One model used LDT which, combined with age, explained 84% of the variance of the VIQ. Finally, Chaurasiya et al. (2022) used a Hindi vocabulary test, akin to the vocabulary subtest of the WAIS, combined with age, sex, locality, and years of education to predict VIQ which explained 49% of the variance.

'Hold' test only models' R^2 value ranged from 0.42 to 0.84 (non-adjusted). Dual models including demographic and 'hold' test R^2 ranged from 0.48 to 0.62.

One further model investigated the predictability of an Arabic verbal fluency test from the vocabulary subtest and picture completion subtest which explained 58% of the variance (Al-Ghatani et al., 2011).

Non-Verbal Intelligence and Fluid Reasoning

Non-verbal Intelligence and fluid reasoning were measured by two cognitive tests, WAIS and RSPM.

PIQ is a measure included in the WAIS up to and including the WAIS-III. It is generally considered as a measure of fluid intelligence and relies on non-verbal abilities. Nine models used the PIQ as a measure of current functioning; 55.6% of the models used irregular word reading as predictor variables; 22.2% used accentuation tasks; 11.1% of models utilised LDT to predict premorbid functioning; 11.1% of models used RPSM to predict premorbid functioning.

The models predicting PIQ generally explained a lower percentage of variation relative to other cognitive domains. R^2 values ranged from 0.21 to 0.46 in 'hold' test only models. Models that included demographic variables in addition to 'hold' tests, explained a range of 37-54% of the variance observed.

Only age and years of education were included in the models as significant predictors. Age was included in two models (Colombo et al., 2002; Almkvist et al., 2007). Years of education was the sole significant demographic variable in one model (Alves et al., 2012) and included alongside age in one model (Almkvist et al., 2007). Sex was found to be a non-significant predictor in two models (Almkvist et al., 2007; Colombo et al., 2002).

PIQ is a relatively broad value and was removed in lieu of using the indexes, Perceptual Reasoning Index (PRI) and Processing Speed Index (PSI). Yi et al. (2017), thus, included measures of PRI and PSI which were modelled with KART and years of education as predictive variables. The model explained 25% and 33% of the variance of PRI and PSI respectively.

Six models utilised RPSM across three studies (Chen et al., 2009; Del Ser et al., 1997; Schrauf et al., 2005). RPSM is a non-verbal test which specifically measures abstract reasoning and fluid intelligence. One demographic-only model was created using age and years of education as co-variates. This explained 55-62% of the variation observed (Chen et al., 2009). Two further models, taken from the same study, included the CGWRT into the model which increased the variation observed to 58-64%. WAIS vocabulary subtest and the WAT were also modelled without demographic variables as co-variates.

Cross validation methods

Eleven studies used a cross validation method for the developed equations. The methods are presented in table 4. Three studies (Chen et al., 2009; Kim et al., 2015; Tang et al., 2012) used only external, normative samples for the validation of the models.

Clinical Samples were included in eight studies and compared alongside a normative validation sample in seven of these.

In five studies, models were validated on clinical participants who had Alzheimer's disease (AD) and Dementia. Almkvist et al. (2007) used a case-study methodology on one patient to demonstrate the utility of the developed model in assessing premorbid functioning in above average intelligence.

Alternatively, Colombo et al. (2002), Matsuoka et al. (2006), Yi et al. (2017) and Del Ser et al. (1997) applied the regression equations to a sample of patients with AD and compared the means of the predicted and actual IQ. This was then compared to the results of the same methodology on a normative sample. Yi et al. (2017) completed a similar analysis in an extended sample including patients with AD and Mild Cognitive Impairment (MCI). They looked at the differences between predicted IQ and observed IQ. In this case, the observed IQ was measured by a measure of cognitive reserve. In all cases, significant differences between groups were only found within observed IQs and not predicted IQs. This suggests a degree of stability within the predicted value in AD and MCI.

Chaurisiya et al. (2022) applied the equations to a sample with brain injuries and investigated the correlations between patients' cognitive reserve score and estimated VIQ and PIQ respectively. The values were highly correlated indicating a strong relationship.

Only one study, Gomar et al. (2011) included participants with Schizophrenia in both hospital and community settings. Descriptive statistics were compared showing that mean observed FSIQ was lower than estimated FSIQ across both samples.

Finally, Pluck et al. (2021) opted to simulate clinical data by reducing the score on the premorbid 'hold' tests by various standard deviations to investigate whether the regression equation was robust to this change. They found that the lexical decision task was most resilient to fluctuations in score.

Table 4

Cross validation methods and key findings across the studies

Study	Clinical Population (*n*)	Cross Validation Method	Key Findings
Almkvist et al., 2007	AD (*n*=1)	Case study	SLDT helped to identify deficit in participant with above average intelligence.
Chaurasiya et al., 2022	Normative (*n=100)*	Equations applied to validation sample. Discrepancies between predicted and actual VIQ and PIQ were analysed through descriptive statistics for normative sample.	Discrepancies within 10 points in "high" number of individuals.
		Correlations were calculated between actual and predicted VIQ and PIQ for Normative sample.	Estimated PIQ and VIQ were significant correlated with Actual PIQ and VIQ.
	Brain Injury (*n=39)*	Correlations were calculated between the estimated VIQ and PIQ and Cognitive Reserve Index of patients.	Cognitive Reserve Index of Patients significantly correlated with estimated VIQ (r=0.87) and PIQ (r=0.91).
Chen et al., 2009	Normative (*n=130)*	Equations applied to validation samples Correlation between predicted and obtained RPSM score Discrepancies between predicted and obtained compared in each group through one-way ANOVA	Comparable correlation coefficients between groups. No significant differences in residuals between groups.

Study	Clinical Population (*n*)	Cross Validation Method	Key Findings
Colombo et al., 2002	Normative *(n=20)*	Equations applied to validation samples.	Significant difference between actual WAIS scores measured from each sample.
	AD (*n=20)*	T-test completed for to compare each sample estimated and actual scores.	No significant difference found between the sample's estimated values of verbal, performance and FSIQ.
	Normative (*n=104)*	Equations applied to second validation sample. Mean Squared Prediction Regression was calculated. Control data added to initial regression analysis	
Del Ser et al., 1997	Normative *(n=40)* Dementia *(n=20)*	Scores compared between two validation samples (t-test) Equations applied to validation samples. Discrepancies were examined using descriptive.	WAIS, RPSM and MMSE scores sig lower in Dementia group. WAT scores not statistically different between validation groups. Discrepancies best suited for clinical diagnosis deemed to be RAVEN actual and WAT-predicted.
Gomar et al., 2011	Patients with Schizophrenia: Chronic Hospitalised *(n=86)* Community Resident *(n=72)*	Equations applied to validation samples. Descriptive statistics compared.	Mean observed FSIQ was lower than estimated FSIQ across both samples.

Study	Clinical Population (*n)*	Cross Validation Method	Key Findings
Kim et al., 2015	Normative Sample (*n=609)*	Equations applied to validation sample. Equations were compared using analysis of the percentage of scores that (1) were ±5 actual FSIQ (2) ±10 of actual FSIQ (3) Same ability classification level (4) ability change in classification by one level.	KPIE-4(4ST) and KPIE-4 (2ST) were deemed most accurate compared to other equations based on analysis.
Matsuoka et al., 2006	Normative Sample (*n=50)* AD (*n=74)*	Equations applied to validation samples. Pearson's *r* correlations calculated for FSIQ, VIQ and PIQ between predicted and actual scores. *T* tests were used to analyse mean difference between the normative and AD group for both predicted and actual scores.	Correlations were all significant for normative group. Observed FSIQ, VIQ, PIQ were significantly different between the AD and control group. Predicted FSIQ, VIQ, PIQ were not significantly different.
Pluck & Ruales-Chieruzzi, 2021	Normative *(n=53)*	Equations applied to validation sample. Discrepancies examined.	Significant correlations were found between estimated and predicted values.
	Simulated Clinical Data	Simulated focal and global cognitive impairments by reducing WAT, SCIRT and SpanLex by various SD to investigate impact on premorbid estimation using the median estimated IQ as the cut-off for 'clinical impairment'.	Regression equations were found to be relatively stable to fluctuations in score Lexical decision task was most robust.

Study	Clinical Population (*n*)	Cross Validation Method	Key Findings
Tang & Yao 2012	Normative Sample (*n=1014)*	Equations applied to validation sample. ANOVA between actual and predicted mean values.	No significant difference between actual and predicted values.
Yi et al., 2017	Normative (*n=30)* AD (*n=31)*	Equations applied to validation samples Correlation between residual of regression and IQ across sample	No significant patterns between residual and Kart-errors. Significant correlation between KART-Predicted IQs and observed IQs in Normative sample
		Correlation between KART observed and estimated IQ	Observed IQs were significantly different between normative and AD group, KART-predicted IQs were non-significant between the two groups.
		Observed and Estimated IQ compared individually between AD and normative groups	Higher KART predicted FSIQ than observed IQ in AD group.
	Extended validation: Normative (*n*=80) AD (*n*=43) MCI (*n=56*)	Extended validation on further sample: KART-Predicted compared between groups Global Cognition (CERAD-K) compared between groups	KART-predicted IQs did not sig. differ between groups. Current CERAD-K scores sig. differed between groups.

Note. **AD=** Alzheimer's Disease. **ANOVA=** Analysis of variance. **CERAD-K=** Consortium to Establish a Registry for Alzheimer's Disease – Korean**. FSIQ=** Full Scale Intelligence Quotient. **KART=** Korean Adult Reading Test**. KPIE-4 (2ST)=**Korea Premorbid Intelligence Estimation two-subtest (Vocabulary and Information) formula. **KPIE-4 (4ST)=** Korea Premorbid Intelligence Estimation four-subtest (Vocabulary= Information= Matrix Reasoning= and Visual Puzzle) formula. **MCI=** Mild Cognitive Impairment. **PIQ=** Performance Intelligence Quotient. **RPSM=** Raven's Progressive Matrices**. VIQ=** Verbal Intelligence Quotient. **WAT=** Word Accentuation Task. **SCIRT=** Stem Completion Implicit Reading Test.

Discussion

Overview of Research Findings

This review aimed to investigate the current state of the literature looking at regression-based methods of estimating premorbid functioning using 'hold' tests and demographic variables, in non-English speaking populations. Studies were found to include regression-based models that predict premorbid functioning across several domains such as executive functioning, memory and fluid reasoning. Languages that did not have the same grapheme to phoneme irregularities, such as those utilised within English word reading tests, were able to use alternative methods that were thought to tap into the same mechanisms as stored lexical knowledge. Methods included: removing accentuation marks; using loan words; using alternative irregularities within a written script such as symbol pronunciation within compound words. Other methods included tests based on lexical decision and subtests thought to be resistant to cognitive decline within versions of the WAIS and ISCA.

Whilst the studies identified within this review included countries spanning Asia, Europe, North America and South America, absences were identified. Notably, no studies were completed in the UK to investigate or validate methods to predict premorbid functioning in the residing non-English speaking population. There is a clear need for studies to assess the validity of use within a different country to that where the original validation sample resides. Schrauf et al. (2006), for example, showed adaptions were required to the Spanish words included within the WAT for a Spanish-speaking immigrant population in the USA. This finding is mirrored in studies within Australia, New-Zealand, and Canada (Hennessy & Mackenzie, 1995; Starkey & Halliday, 2011; Blair & Spreen, 1989) that show the need for adaptions cross-culturally due to the discrepancies in dialect and word-use across populations. Due to acculturation, words which may be frequently used in one country may be less well widely used in another. This impacts the interpretation of the test score and the validity of the norm-based data.

There has been a growing consensus in the neuropsychology community for the need to look at the provision for accurate neuropsychological assessment for non-English speaking and ethnic diverse populations (Brickman et al., 2006). This review identifies the need for further norms and research for under-represented populations, particularly in those residing outside of their country of origin. In light of this, careful consideration of the limitations of current premorbid functioning measures is needed when being used within this population, particularly with those where English is their second language. This will help to improve accuracy of the assessment of need and the quality of care for under-represented samples in clinical settings.

In addition to the heterogeneity of languages and cultures investigated, other differences were identified between studies. One such variation was the large variety in sample sizes. Several studies used relatively small sample sizes to derive the equations, such as Karakula-Juchnowicz & Stecka (2017) and Yi et al. (2017). Due to this, caution is needed regarding the conclusions that are made and the generalisability of the findings. The applicability of regression models is limited by the degree of error observed when they are applied to external data sets to that from which the data was derived. This is termed 'Shrinkage' (Copas, 1997). Regression models derived on smaller sample sizes may experience higher levels of 'shrinkage' due to the fact that the general population will have a larger amount of variation (Copas, 1997). Further research may be beneficial in these cases, to bolster the evidence for the relationships found between variables.

A further heterogeneity between studies was the age range included. Three studies included only 'elderly' participants within their norm sample. Age matching is helpful when aiming to use the regression equations within a particular population. It is important, however, to assess the normative participants for cognitive decline which can occur with aging (Murman, 2015) so that bias is limited within the models. Six studies included a lower age limit of 16. Alves et al. (2012) noted that crystallised intelligence is still developing until the age of 25 years of age and is assumed more stable following this age. Alves et al. (2012) excluded participants aged 16-25 in a second regression model and noted an improvement in the variation explained.

Thus, the inclusion of younger participants across studies may have added 'noise' to the models. This should be considered for further research development.

As previously noted, there is a particular need for the sample to be representative of the general population when deriving a predictive model for clinical use. The representation of education levels included within the normative samples varied across the studies. As discussed by Alves et al (2012), the general level of education is low is some countries; thus, a representative sample should have similar mean years of education. The use of years of education as a measure can be limited due to the differences in education provision between countries and localities. Additionally, countries can have alternative education systems and differ in the value placed on traditional schooling. It is important, therefore, to consider the validity of years of education as a predictive measure in certain cases, particularly when attempting to represent those with heterogeneous levels of education. Tests must be sensitive to distinguish between cognitive deficits and lower exposure to formal education which can have an impact performance on neuropsychological testing (Oliveira et al., 2014).

'Hold' Test Methods

As previously discussed, 'Hold' Test methods were varied between studies and irregular word reading tests extended beyond the premise of the NART and TOPF to adapt to different language rules. Attempts were made to utilise the same lexical mechanism within diverse languages and scripts that do not have the same pronunciation irregularities as those within the English language – for instance, accentuation marks were removed in the WAT (Del Ser et al., 1997). Whilst the purpose of this review was not to compare methods, careful investigation is needed into the resistance of these adapted methods to clinical decline and to ensure that a similar mechanism was being measured to the NART. In order to do this, Yi et al. (2017) demonstrated during cross validation, that KART was adequate at identifying cognitive decline in a sample with AD and MCI. This suggests that the irregularities used within the *hangul* script tapped into the same lexical mechanisms and were resistant to cognitive decline. Whilst attempts were made to cross-validate on

clinical populations in some cases, several studies did not include a clinical sample within their analysis. The resistance to cognitive decline for further measures, and for a wider range of neurological disorders, should be considered prior to clinical use.

Demographic predictor variables

Age and years of education were most commonly retained in the models as significant contributors. Thus, these variables are important to be considered in future development of premorbid regression algorithms. The heterogeneity of significant demographic predictors across the models, highlights the need for individual validation of premorbid estimates for different populations and in those residing outside of their country of origin. Due to population differences, relationships between the demographic variables and cognitive performance can differ, as illustrated within this review. Inclusion of demographic variables within regression models should be carefully considered and chosen based on previous research and observed relationships within particular countries.

Cognitive Domains

Whilst FSIQ was used most frequently in the models, other cognitive domains were also included such as memory and executive function. It is important to note that the purpose of the review was not to compare models and their ability to predict cognitive functions. It would not be meaningful to compare the 'best' method of premorbid functioning for each cognitive domain due to the heterogeneity of the data in terms of covariates included, 'hold' test methods and samples analysed. Despite this, general trends can be identified.

To this end, models predicting PIQ tended to be lower comparative to VIQ and FSIQ. Generally, studies have shown that vocabulary tests correlate less well with fluid intelligence, of which PIQ could be thought to be a measure (Bright & van der Linde, 2017). This is important to consider when looking at cognitive decline in measures of fluid intelligence.

Several models used WAIS-III or earlier to measure current cognitive functioning and, thus, use the PIQ and VIQ values within their analysis. Later versions of the WAIS remove these values in lieu of index values that provide a more detailed profile of cognitive functioning. PIQ combines PSI and PRI that may have very different relationships to reading tests and premorbid estimates (Bright & van der Linde, 2017). Additionally, VIQ includes measures of WMI and VCI which can be considered to be two separate cognitive domains. This may have an impact on the predictability of VIQ. For example, Isella et al. (2005) argues that the TIB's ability to predict memory functioning is not comparable to other domains. Whilst this could be a comment on the test itself, this may also indicate that memory domains are less well predicted by 'hold' tests and demographic variables. This would, therefore, impact models predicting VIQ and it would, thus, be more meaningful to investigate the predictability of PSI and PRI individually.

Additionally, it is salient that recent research used earlier versions of the WAIS, such as the WAIS-R and WAIS-III, despite being completed following the release of the WAIS-IV in 2008. This further highlights inequalities to neuropsychological provision across the world.

Cross validation methods

Various cross validation methods were used within the studies. Three studies only used an independent normative sample to validate the equations. This is helpful to assess the level of shrinkage that is experienced by the model when applied to an independent data set. However, prior to clinical use, validation is also required with clinical participants. This will assess how robust the test is to cognitive decline which is essential when evaluating premorbid measures.

In studies that include a clinical sample, t-tests were the most common methodology used to identify whether there is a significant difference between the actual and observed IQ in clinical and normative samples. In all cases, the t-tests identified a significant difference between the two values in the clinical sample only. It should be noted, however, that a non-significant p-value does not provide evidence to

determine the null hypothesis to be true, despite common misconception in the literature (Greenland et al., 2016). Thus, using a p-value alone, it cannot be determined that there is not a difference found for the normative sample within the general population. Further analysis using Bayesian statistics could be meaningful which can provide evidence for the null hypothesis (Rouder et al., 2009). Alternatively, correlations were used within one study (Chaurisiya et al., 2022) to assess the relationship between patients' cognitive reserve score and estimated VIQ and PIQ respectively which were highly correlated and indicated a strong relationship. This is positive, but does not take into account residuals for each individually predicted score. Thus, the regression equation could consistently over predict the scores, but still yield a high correlation.

One study implemented a case-study methodology (Almkvist et al., 2007). Whilst the aim here was to demonstrate the utility of the regression model in a participant with above average intelligence, the method is limited in the degree of generalisability to the population as a whole. Crucially, it excludes participants performing at the more extreme ends of the scales such as those performing at the extremely high and low ends of the scales. This is important to assess due to regression methods being found to over and under predict IQ for these individuals (Veiel & Koopman, 2001).

Pluck et al. (2021) took a different approach and attempted to simulate clinical data by reducing the score on the 'hold' tests by a certain standard deviation. Whilst ultimately testing how robust the regression equations are to fluctuations in scores, 'hold' test scores should, by nature, be relatively stable despite cognitive decline. Thus, this method is limited and may not represent a clinical neuropsychology profile accurately. Further analyses would be required on a clinical sample prior to clinical use.

Overall, validation methods were limited in different ways across studies. It is important that further research considers the clinical utility of measures prior to clinical use.

Methodological Limitations

Due to the nature and aims of this review, it was important to include articles written and published in alternative languages. Thus, there were limitations in the author's ability to identify studies for inclusion through database searching. The search was based on English terms only, due to the variety of appropriate terms across language and to avoid bias. Thus, reference list searching and 'cited by' searching was deemed essential to identify missed papers. Additionally, the literature search was completed in the UK and thus, papers may have been missed that were not published or accessible in the UK. Similarly, some of the studies included in the literature review had to be translated by the first author, to the best of their ability. Whilst every care was taken to avoid misinterpretation through the translation process, this remains a possibility.

The literature review did not include alternative methods of predicting premorbid functioning alternative to regression-based approaches, such as norm tables. Regression-based approaches have been criticised for inaccurate estimation of current functioning in some populations (Bright & van der Linde, 2018) and for the degree of error introduced by shrinkage when the model is applied to alternative populations. Attempts at correcting for this bias have been developed (Veiel & Koopman, 2001) and thus, further research could look at the utility of these corrections in non-English speaking regression methods.

Due to the heterogeneity of studies, the literature review was limited in its ability to draw conclusions about the 'best' method of predicting premorbid functioning. Whilst outside the scope of this review, further research may benefit from comparing methods of predicting pre-morbid estimation methods within non-English speaking populations for different cognitive domains.

Clinical Implications

The issue of under representation in neuropsychological norms is gaining more attention in recent years (e.g. Brickman et al., 2006). The use of universal norms leads to error, misdiagnosis and can cause inaccurate profiles of an individual's cognitive ability (O'Driscoll & Shaikh et al., 2017). It is, therefore, important to be aware of the current available and validated methods for estimating premorbid functioning cross-

culturally and within linguistically-diverse populations. Additionally, it is important to identify the current gaps in the literature. Very few studies investigated the validity of using measures with a culturally diverse, non-native population. This is especially important to promote equal, non-discriminatory access to accurate neuropsychological assessment in a clinical environment. There is scope for further research to look at validating methods of predicting premorbid functioning in non-English speaking and non-native populations within the UK and other western countries.

Various ways of adapting English-based reading tests were identified. Despite differences in the frequency of irregular words within languages, alternative ways were found to measure stored lexical knowledge- the core mechanism of the NART and TOPF that is thought to be resistant to cognitive decline. With this in mind, research is needed in further languages to develop novel ways and tests to predict premorbid functioning and measure stored lexical knowledge where the existing methods may not be appropriate.

Finally, the main function of estimating premorbid functioning is for use with clinical populations. This review illustrates the current methodological limitations and gaps within the literature for assessing cognitive decline within non-English speaking populations. It also highlights the important considerations for clinicians utilising these methods.

Conclusion

This review aimed to assess the current state of the literature for the development of regression models to predict premorbid functioning in non-English speaking populations. Adaptions to English-based 'hold' tests have been created and reading tests have been adapted for languages that do not include the same lexical irregularities. Despite this, further validation is needed to assess the need to adapt these measures for use in further countries and non-native populations, for accurate and meaningful clinical use. Demographic variables may be more easily translated, but, validation of their relationship in different cultures is required. Clinicians should be mindful of the adaptions needed and error introduced when using regression-based

approaches in populations that are different to that in which the model was validated. The further need for research has been discussed in order to move towards adequate and accessible neuropsychological provision across all populations and countries.

References

Al-Ghatani, A. M., Obonsawin, M. C., Binshaig, B. A., & Al-Moutaery, K. R. (2011). Saudi normative data for the Wisconsin card sorting test, Stroop test, test of non-verbal intelligence-3, picture completion and vocabulary (subtest of the Wechsler adult intelligence scale-revised). *Neurosciences Journal*, *16*(1), 29-41.

Almkvist, O., Adveen, M., Henning, L., & Tallberg, I. M. (2007). Estimation of premorbid cognitive function based on word knowledge: The Swedish Lexical Decision Test (SLDT). *Scandinavian Journal of Psychology*, *48*(3), 271-279.

Alves, L., Simões, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: the Irregular Word Reading Test (TeLPI). *Archives of Clinical Neuropsychology*, *27*(1), 58-68.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*(1), 55-65.

Barona, A., Reynolds, C., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology*, 52(5), 885–887.

Blair, J. R., & Spreen, O. (1989). Predicting premorbid IQ: a revision of the National Adult Reading Test. *The Clinical Neuropsychologist*, *3*(2), 129-136.

Brickman, A. M., Cabo, R., & Manly, J. J. (2006). Ethical issues in cross-cultural neuropsychology. *Applied Neuropsychology*, *13*(2), 91-100.

Bright, P., & van der Linde, I. (2018). Comparison of methods for estimating premorbid intelligence. *Neuropsychological Rehabilitation*. *30(1),* 1-14.

Cannon-Spoor, H. E., Potkin, S. G., & Wyatt, R. J. (1982). Measurement of premorbid adjustment in chronic schizophrenia. *Schizophrenia Bulletin*, *8*(3), 470-484.

Chaurasiya, A., Ranjan, J. K., Pandey, N., & Asthana, H. S. (2022). Estimation of premorbid intelligence: Demographical and current neurocognitive functioning based algorithms. *Asian Journal of Psychiatry*, *72*, 103065.

Chen, Y. J., Ho, M. Y., Chen, K. J., Hsu, C. F., & Ryu, S. J. (2009). Estimation of premorbid general fluid intelligence using traditional Chinese reading performance in Taiwanese samples. *Psychiatry and Clinical Neurosciences*, *63*(4), 500-507.

Colombo, L., Sartori, G., & Brivio, C. (2002). Stima del quoziente intellettivo tramite l'applicazione del TIB (test breve di Intelligenza). *Giornale Italiano di Psicologia*, *29*(3), 613-638.

Copas, J. B. (1997). Using regression models for prediction: shrinkage and regression to the mean. *Statistical Methods in Medical Research*, *6*(2), 167-183.

Crawford, J. R. (1989). Estimation of premorbid intelligence: A review of recent developments. *Developments in Clinical and Experimental Neuropsychology*, 55-74.

Crawford, J. R., & Allan, K. M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, *11*(2), 192-197.

Crawford, J. R., Nelson, H. E., Blackmore, L., Cochrane, R. H. B., & Allan, K. M. (1990). Estimating premorbid intelligence by combining the NART and demographic variables: An examination of the NART standardisation sample and supplementary equations. *Personality and Individual Differences*, *11*(11), 1153-1157.

Cuetos, F., & Suárez-Coalla, P. (2009). From grapheme to word in reading acquisition in Spanish. *Applied Psycholinguistics*, *30*(4), 583-601.

Del Ser, T., González-Montalvo, J. I., Martínez-Espinosa, S., Delgado-Villapalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition*, *33*(3), 343-356.

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

Eppinger, M. G., Craig, P. L., Adams, R. L., & Parsons, O. A. (1987). The WAIS—R index for estimating premorbid intelligence: Cross validation and clinical utility. *Journal of Consulting and Clinical Psychology*, *55*(1), 86.

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology*, *12*(8), 711-738.

Gelb, I. J. 1. (1952). *A study of writing: The foundations of grammatology*. London: Routledge and Kegan Paul.

Gomar, J. J., Ortiz-Gil, J., McKenna, P. J., Salvador, R., Sans-Sansa, B., Sarró, S., ... & Pomarol-Clotet, E. (2011). Validation of the Word Accentuation Test (TAP) as a means of estimating premorbid IQ in Spanish speakers. *Schizophrenia Research*, *128*(1-3), 175-176.

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337-350

Griffin, S. L., Mindt, M. R., Rankin, E. J., Ritchie, A. J., & Scott, J. G. (2002). Estimating premorbid intelligence comparison of traditional and contemporary methods across the intelligence continuum. *Archives of Clinical Neuropsychology*, *17*(5), 497-507.

Grober, E., & Sliwinski, M. (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. *Journal of Clinical and Experimental Neuropsychology*, 13, 933-949

Hennessy, M., & Mackenzie, B. D. (1995). AUSNART: The development of an Australian version of the NART. In *The 18th Annual Brain Impairment Conference* (pp. 183-188).

Isella, V., Villa, M. L., Forapani, E., Piamarta, F., Russo, A., & Appollonio, I. M. (2005). Ineffectiveness of an Italian NART-equivalent for the estimation of verbal learning ability in normal elderly. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 618-623.

Jenkinson, T. M., Muncer, S., Wheeler, M., Brechin, D., & Evans, S. (2018). Estimating verbal fluency and naming ability from the test of premorbid functioning and demographic variables: Regression equations derived from a regional UK sample. *British Journal of Clinical Psychology*, *57*(2), 135-147.

Karakuła-Juchnowicz, H., & Stecka, M. (2017). Polish Adult Reading Test (PART)-construction of Polish test for estimating the level of premorbid intelligence in schizophrenia. *Psychiatr Pol*, *51*(4), 673-85.

Kim, S. G., Lee, E. H., Hwang, S. T., Park, K., Chey, J., Hong, S. H., & Kim, J. H. (2015). Estimation of K-WAIS-IV premorbid intelligence in South Korea: development of the KPIE-IV. *The Clinical Neuropsychologist*, *29*(sup1), 19-29.

Krueger, K. R., Lam, C. S., & Wilson, R. S. (2006). The Word Accentuation Test–Chicago. *Journal of Clinical and Experimental Neuropsychology*, *28*(7), 1201-1207.

Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist*, *9*(1), 83-88.

Lezak, M., Howieson, D., Bigler, E., & Tranel, D. (2012). *Neuropsychological assessment.* (5th ed.). New York: Oxford University Press

Liberman, L Y., & Shankweiler, D. (1979). Speech, the alphabet and teaching to read. In L. B. Resnick & P. A. Weaver (Eds.), *Theory and Practice of Early Reading* (Vol. 2, pp. 109-132). Hillsdale, NJ: Erlbaum

Matsuoka, K., Uno, M., Kasai, K., Koyama, K., & Kim, Y. (2006). Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test. *Psychiatry and Clinical Neurosciences*, *60*(3), 332-339.

Murman, D. L. (2015, August). The impact of age on cognition. In *Seminars in Hearing* (Vol. 36, No. 03, pp. 111-121). Thieme Medical Publishers.

Nelson, H. E., & McKenna, (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology, 14*(3), 259-267.

O'Driscoll, C., & Shaikh, M. (2017). Cross-cultural applicability of the Montreal Cognitive Assessment (MoCA): a systematic review. *Journal of Alzheimer's Disease*, *58*(3), 789-801.

Oliveira, M. O. D., Nitrini, R., Yassuda, M. S., & Brucki, S. M. D. (2014). Vocabulary is an appropriate measure of premorbid intelligence in a sample with heterogeneous educational level in Brazil. *Behavioural neurology*, 2014.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1), 1-11.

Pedraza, O., & Mungas, D. (2008). Measurement in cross-cultural neuropsychology. *Neuropsychology Review*, *18*(3), 184-193.

Pickering, C., Grignon, J., Steven, R., Guitart, D. and Byrne. J. (2015). Publishing not perishing: How research students transition from novice to knowledgeable using systematic quantitative literature reviews. *Studies in Higher Education*. 40: 1756-1769.

Pitman, I., Haddlesey, C., Ramos, S. D., Oddy, M., & Fortescue, D. (2015). The association between neuropsychological performance and self-reported traumatic brain injury in a sample of adult male prisoners in the UK. *Neuropsychological Rehabilitation*, 25(5), 763-779.

Pluck, G., & Ruales-Chieruzzi, C. B. (2021). Estimation of premorbid intelligence and executive cognitive functions with lexical reading tasks. *Psychology & Neuroscience*. *14*(3), 358-377.

Pluck, G., Almeida-Meza, P., Gonzalez-Lorza, A., Muñoz-Ycaza, R. A., & Trueba, A. F. (2017). Estimación de la función cognitiva premórbida con el Test de Acentuación de Palabras. *Revista Ecuatoriana de Neurología*, *26*(3), 226-234.

Ritchie, A. J., Lam, M., & Rankin, E. J. (1997). Estimating Premorbid Intelligence: Comparisons of the Barona Formula, North American Adult Reading Test (NAART), Wide Range Achievement Test-3 Reading Subtest (WRAT-3), and Oklahoma Premorbid Intelligence Estimate (OPIE). *Archives of Clinical Neuropsychology*, *4*(12), 395.

Rolstad, S., Nordlund, A., Gustavsson, M. H., Eckerström, C., Klang, O., Hansen, S., & Wallin, A. (2008). The Swedish National Adult Reading Test (NART-SWE): A test of premorbid IQ. *Scandinavian Journal of Psychology*, *49*(6), 577-582.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225-237.

Sanjurjo, N., Montañes, P., Matamoros, F. A., & Burin, D. (2015). Estimating intelligence in Spanish: regression equations with the word accentuation test and demographic variables in Latin America. *Applied Neuropsychology: Adult*, *22*(4), 252-261.

Schrauf, R. W., Weintraub, S., & Navarro, E. (2006). Is adaptation of the Word Accentuation Test of Premorbid Intelligence necessary for use among older, Spanish-speaking immigrants in the United States? *Journal of the International Neuropsychological Society*, *12*(3), 391-399.

Serrao, V. T., Brucki, S. M. D., Campanholo, K. R., Mansur, L. L., Nitrini, R., & Miotto, E. C. (2015). Performance of a sample of patients with Mild Cognitive Impairment (MCI), Alzheimer's Disease (AD) and healthy elderly on a lexical decision test (LDT) as a measure of pre-morbid intelligence. *Dementia & Neuropsychologia*, *9*, 265-269.

Spreen, O., & Strauss, E. (1998). *A Compendium of Neuropsychological Tests.* New York, NY: Oxford University press.

Starkey, N. J., & Halliday, T. (2011). Development of the New Zealand Adult Reading Test (NZART): Preliminary Findings. *New Zealand Journal of Psychology*, *40*(3).

Tang, X. R., & Yao, S. Q. (2012). Estimation of premorbid intelligence quotient of the Intelligence Scale for Chinese Adult. *Chinese Journal of Clinical Psychology*. *20*(6), 743–747.

Vanderploeg, R. D., & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid ability: Combining subtest performance and demographic variable predictors. *Archives of Clinical Neuropsychology*, *10*(3), 225-239.

Veiel, H. O., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment*, *13*(3), 356.

Watt, S., Ong, B., & Crowe, S. F. (2018). Developing a regression equation for predicting premorbid functioning in an Australian sample using the National Adult Reading Test. *Australian Journal of Psychology*, *70*(2), 186-195.

Wechsler, D. (2001). *Wechsler Test of Adult Reading: WTAR.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition Administration and Scoring Manual.* San Antonio, TX: Pearson

Wechsler, D. (2011). *The test of premorbid function (TOPF).* San Antonio, TX: The Psychological Corporation.

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: a cross validation. *Journal of Clinical and Experimental Neuropsychology*, *13*(2), 204-216.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978), An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology*, 46, 1554-1555

Woodcock, R. W., & Sandoval, A. F. M. (1996). *Batería Woodcock-muñoz: pruebas de aprovechamiento revisada*. Riverside Publishing Company.

Yi, D., Seo, E. H., Han, J. Y., Sohn, B. K., Byun, M. S., Lee, J. H., ... & Lee, D. Y. (2017). Development of the Korean Adult Reading Test (KART) to estimate premorbid intelligence in dementia patients. *Plos One*, 12(7), Article e0181523.

Part Two: Empirical Paper

An Investigation into the Predictability of the Repeatable Battery for the Assessment of Neuropsychological Status Using the Test of Premorbid Functioning and Demographic Variables: Regression Equations Derived from a UK sample

Hayley Gould*[a], Dr Pete Fleming[a] & Dr Stephen Evans[b]

[a]Faculty of Health Sciences, School of Health and Social Work,

Aire Building, University of Hull, Hull, United Kingdom, HU5 7RX

[b] Psychological Medicine, The Old Chapel, Bootham Park, York, YO31 8HE

*Corresponding Author Email: hjigould@outlook.com

This paper is written in the format ready for submission to the journal

The Clinical Neuropsychologist

Please see Appendix F for submission guidelines

Word Count (including tables, figures and references): 10,026

Abstract

Objective: The objective of this study is to investigate the predictability of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) from the Test of Premorbid Functioning (TOPF) and demographic variables. To do this, it will aim to derive regression models for each of the indexes and scores on the RBANS that may assist to inform clinicians when predicting premorbid performance on the RBANS.

Method: Fifty six community dwelling participants, who did not have a neurological disorder, made up the sample from which the regression models were derived. To create the models, multiple linear regression analysis was used. The models were then cross-validated using the Leave-One-Out Cross Validation method. Additionally, the models were applied to a clinical sample to assess how well they could identify cognitive decline.

Results: Significant models were found for all RBANS indexes apart from the Visuospatial index which is thought to be less well predicted by oral word reading tests. The TOPF was better at predicting verbal subtests comparative to non-verbal subtests.

Conclusions: Regression models are presented that assist in predicting premorbid functioning on the RBANS. The results show that caution is needed when estimating premorbid visuospatial functioning using the TOPF. The initial results seem promising and suggest that the RBANS premorbid scores are somewhat predictable using the TOPF and Demographic variables. However, further research is necessary to validate the models for clinical use.

Key Words: premorbid functioning, neuropsychological assessment, irregular word reading, cognitive decline, regression model, TOPF, RBANS

Introduction

The use of neuropsychological assessment to quantify cognitive decline is essential to evaluate the impact of traumatic brain injury, stroke, and other neuropsychological disorders. To do this meaningfully, it is important to have an idea of an individual's premorbid cognitive functioning (PCF). PCF refers to the level of functioning prior to cognitive difficulties emerging (Franzen et al., 1997).

PCF provides a baseline to which current performance can be compared. This baseline supports the assessment of cognitive change, the tracking of cognitive decline over a period of time, and the synthesis of a detailed neuropsychological profile to help to identify any support needed. Without an idea of a baseline, a single test score may represent substantial cognitive change for one person but may sit within a normal range for another (Crawford et al., 1998). This can lead to instances of misdiagnosis and missed-diagnoses.

In addition, PCF can assist in identifying and quantifying early changes to cognition. Early identification of cognitive decline can improve the success of treatments in cases of dementia (Tuokko et al., 1991). Moreover, the quantification of cognitive decline is imperative in litigation where damages are awarded upon the basis of an individual's loss of cognitive functioning (Reynolds, 1997). Knowledge that contributes to the accurate calculation of an individual's PCF is seldom available in clinical practice (Matsuoka et al., 2006). Thus, methods for clinicians to estimate PCF have been developed.

Qualitative approaches to predicting PCF in clinical practice are based upon the use of clinical judgement considering factors such as school attainment, occupational achievements, and family reporting (Crawford & Allan, 1997). However, the subjectivity of this method and susceptibility to bias has been criticised (Franzen et al., 1997). Thus, objective, actuarial methods are more commonly employed in research and in clinical use.

A regression-based approach to estimating PCF is one such method, and commands a compelling portion of the literature. Regression modelling is ameliorated by its ability to easily consider several predictive factors into a single equation, unlike traditional norm tables that often only incorporate a single variable such as age (Harnett et al., 2004). Regression models comprising of demographic characteristics such as age, sex, race, education, and occupation have been validated for use to predict PCF in countries including the United States (US; Wilson et al., 1978; Barona et al.,1984; Kirton et al., 2020) and the United Kingdom (UK; Crawford and Allan, 1997). These models were developed based on the well-established and accepted relationship between demographic variables and Intelligence Quotient (IQ; e.g. Wechsler, 2008). One such regression equation, the Barona equation, has been shown to account for 36% of the variance within the Full-Scale Intelligence Quotient (FSIQ) as measured by Wechsler's Adult Intelligence Scale (WAIS; Wechsler, 1955). More recently, this equation explained 32% of the variance observed in FSIQ when corrected for the Flynn effect (Kirton et al., 2020).

The benefit of using demographic characteristics within regression models is their independence from cognitive decline. Despite this, demographic models are confined to an oversimplification of the relationships between demographics and IQ (Franzen et al., 1997) which can overlook individual circumstance. For instance, traumatic experiences may affect learning and school engagement (e.g. Delaney-Black et al., 2002; van Os et al., 2017; Crouch et al., 2019) and have a lasting impact on cognition into adulthood (Hardcastle et al., 2018). These individual differences are particularly salient in neuropsychological assessment. To address this, 'hold' tests can be incorporated into regression models which allow for current testing of individual cognitive ability, to inform the estimation of PCF.

'Hold' tests are psychometric tests that are thought to be relatively resistant to cognitive decline (Franzen et al., 1997). WAIS subtests such as Vocabulary, Matrix Reasoning, Information and Picture Completion subtests have previously been used as 'hold' tests (e.g. Schoenberg et al., 2007) but have since been replaced

in favour of word reading tests that are thought to have a higher resistance to cognitive changes (Bright & van der Linde, 2018).

Word reading tests are based on the principle that reading is highly correlated with intelligence level (Willshire et al., 1991) and that certain mechanisms used in word reading are relatively resistant to cognitive decline (for review, see Franzen et al., 1997). This is particularly true for words that rely on stored lexical knowledge (Matsuoka et al., 2006) such as the reading of irregular words which are defined as words that have irregular grapheme to phoneme translation (Nelson & McKenna, 1975). Several reading tests have adopted this paradigm within the UK such as the National Adult Reading Test (NART; Nelson & McKenna, 1975), the Wechsler Test of Adult Reading (WTAR; Holdnack, 2001) and the more recently developed Test of Pre-morbid Functioning (TOPF^{UK}; Wechsler, 2011). Cross-cultural adaptions have also been developed based on similar principles in different languages, including, Chinese Graded Word Reading Test (CGWRT; Chen et al., 2009), National Adult Reading Test Sweden (NART-Swe; Rolstad et al., 2008), Japanese Adult Reading Test (JART; Matsuoka et al., 2006), Polish Adult Reading Test (PART; Karakula-Juchnowicz & Stecka, 2017), and Korean Adult Reading Test (KART; Yi et al., 2017).

Word reading tests compare well to demographic-only and WAIS subtest-based approaches (Bright & van der Linde, 2018). TOPF^{UK}, a relatively new measure, has been shown to account for 72 percent of the variance observed in the FSIQ (Wechsler, 2008) which is an improvement on that explained in demographic models. Although relatively resistant to cognitive decline, word reading and other cognitive 'hold' tests can still be impacted by changes in functioning, for instance, in cases of severe cognitive impairment (O'Carroll et al., 1995). Thus, combined regression algorithms including demographic variables and 'hold' tests are commonly adopted in clinical and research settings (e.g. Crawford et al., 1990; Krull et al., 1995).

Estimation of cognitive domains

Estimation methods of PCF have frequently focused on the prediction of FSIQ as measured by the WAIS. Whilst the WAIS is arguably the gold standard for IQ testing, it is not always appropriate for all patient groups or services, such as when briefer forms of assessment are necessary or when specific cognitive domains are of interest.

The predictability of premorbid performance on alternative psychometric tests and cognitive domains has received relatively less attention in the literature. Predictive equations have, however, been created for some alternative tests such as: verbal fluency test, and naming ability (Jenkinson et al., 2018); the trail making test, and mini-mental state exam (Knight et al., 2006); the Stroop test, and Wisconsin card sorting test (Al-Ghantani et al., 2011). Additionally, studies have looked at the predictability of alternative cognitive domains such as memory ability (e.g. Isella et al., 2005; Duff, 2010) and fluid intelligence (e.g. Chen et al., 2009; Shrauf et al., 2006) with varying success.

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998) is frequently used alongside PCF measures in neuropsychological assessment. The RBANS is a repeatable battery measuring several cognitive domains across five indexes: Immediate Memory, Visuospatial/Constructional, Language, Attention, and Delayed Memory. The RBANS is a popular test as it is relatively brief and is, thus, more tolerable than the WAIS for some individuals - for instance, with dementia patients (Randolph, 1998). It is also thought to be an efficient screening measure for cognitive decline across numerous neurological conditions, for instance, neuro-oncology and Parkinson ’s disease (Loughan et al., 2019; Carter et al., 2016). Despite being used in neuropsychology services alongside measures of PCF, the RBANS is not currently co-normed with a measure of PCF.

To address this, Duff & Ramezani (2015) looked at formulating regression-based normative formulae for the RBANS, using the demographic variables, age, sex, education and race. The variance accounted for by the

models ranged from 7-16% of the index scores and 8-28% of the subtest raw scores. Further to this, Duff et al. (2019) developed regression equations that included the WTAR and demographic variables. In these models, the variance in the RBANS indexes differed from 4% to 16%. In both cases the variation accounted for was relatively low. However, the normative sample was comprised of older adults (aged above 65 years). Due to changes in cognition across the lifespan, these older adults may have experienced cognitive change in varying degrees. For instance, normal aging has been found to be associated with a decline in processing speed, working memory, and executive function (Murman, 2015). In both cases, the studies did not take account of such impacts in the predictive assurance of their models. Additionally, the RBANS is validated for use with people between the ages of 12 and 89 (Randolph et al., 1998). Thus, an investigation that reflects the use of the RBANS in adult services by using an age representative normative sample would be beneficial as it will provide a more accurate clinical model.

Subsequent to previous studies, the TOPF^{UK} has been proposed as a replacement to the WTAR and is gaining popularity in clinical use and research (Wechsler, 2011). This is because the TOPF^{UK} addresses the shortcomings within the WTAR by widening prediction range, improving prediction accuracy, and reducing the effect of brain injury (Wechsler, 2011). Additionally, the TOPF^{UK} is co-normed with the most current version of the WAIS, the WAIS-IV. Thus, further investigation into the predictability of RBANS premorbid functioning from the TOPF^{UK} score is needed.

Despite the RBANS commonly being used alongside the TOPF^{UK} in services, there is currently no actuarial method of comparing the two scores. Clinical judgement is often used to determine cognitive decline from TOPF^{UK} predicted FSIQ. As discussed above, clinical judgement is unreliable due to the clinician's subjectivity and susceptibility to bias. This research, therefore, aims to quantitatively investigate the predictability of the RBANS scores from the TOPF^{UK} and demographic variables.

Aims

The study consists of two parts. The first, aims to investigate the predictability of the RBANS indexes and subtest scores from the TOPF^{UK} score and demographic variables. It will do this by creating regression models for each index and subtest respectively. The second part, aims to cross-validate the regression models using the Leave-One-Out Cross Validation (LOOCV) method on the original sample, and by applying the models to independent clinical data.

Method

Participants

Non-Clinical Sample

Sixty-seven community-dwelling adults were recruited to comprise the normative, non-clinical sample through opportunity sampling using online advertisement and word of mouth to advertise the research project across the United Kingdom. Participation was voluntary and no incentive was given to take part.

Exclusion criteria included: a previous diagnosis of a brain disorder, neurodevelopmental condition, neurological illness or head injury; historical or current severe mental health difficulties requiring inpatient admission or community mental health support; a score of 15 or above on the Hospital Anxiety and Depression Scale (HADS; considered severe; Zigmond & Snaith, 1994); a score lower than 70 on any RBANS index (rated as extremely low; Randolph, 1998). Participants were also required to be over 18 and proficient in English. Based on these criteria, eleven participants were excluded prior to analysis.

A power analysis determined that a sample size of fifty-five participants would be sufficient for multiple linear regression analysis, based on a desired power of 0.8, α of .05 and medium effect size (f^2=0.15). The non-clinical sample will henceforth be referred to as the derivation sample.

Clinical Sample

The second sample included ten community-dwelling participants who were experiencing cognitive difficulties and had sought neuropsychological assessment through National Health Service (NHS) services. Participants were included in the sample if they: had a diagnosed or suspected neurological disorder; had capacity to give consent and consented to taking part, as assessed by the clinician; were over 18 and proficient in English. The sample was not constrained to specific diagnosis so that it would reflect, as closely as possible, the organic use of the psychometric measures with patients by clinicians.

Procedure

Ethical approval for the study was granted by both the University of Hull and National Health Service research and ethics boards, in addition to the required local Research and development approvals for the NHS recruitment sites. All participants were given detailed information on the aims, rationale and procedure of the study prior to providing written, informed consent to participate.

Demographic information (age; years of education; biological sex; occupation/pre-retirement occupation) was collected from all participants. Diagnosis information was collected from the clinical sample. Age and years of education were presented in years. Dummy coding was used to code sex and occupation. Sex was defined as follows, 1=Male, 2=Female. Each participant's occupation was coded using the Standard Occupational Classification System (SOC; Office for National Statistics, 2020) into nine major categories (See Appendix G). Due to some of these categories not being represented in the data, this was further refined into the SOC 2020 four-tier skill level groups (See Appendix G).

Tests were administered either online or in-person for the derivation sample by one researcher. Online testing was administered with permission from Pearson (Appendix H) which was granted following the need for online provision due to the Covid-19 pandemic. For online testing, data collection was completed using video calling software and the coding subtest was sent in advance of testing in a sealed envelope which was confirmed as being unopened prior to the test being administered.

Clinical data was collected through routine assessment in NHS neuropsychology services. Consent was obtained, and tests were administered and scored by the clinicians working within the services.

All participants were administered the following instruments:

Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith, 1994)

The HADS is a 14-item measure which is a widely utilised measure in UK healthcare services and in research under a two-factor structure which includes an anxiety and depression subscale respectively (Stern, 2014). The HADS aims to provide the clinician with an understanding of the patient's experiences that relate to anxiety and depression over a limited time frame. It has been shown to have a sensitivity and specificity of approximately 0.8 for both scales (Bjelland et al., 2002).

The Test of Premorbid Functioning United Kingdom (TOPF^{UK}; Wechsler, 2011)

The TOPF^{UK} was administered according to the procedure outlined in the manual. The TOPF^{UK} is an oral word reading test consisting of 70 English words with irregular grapheme-phoneme translation, which has been anglicised. The test-retest stability has been shown to be good (r=.89-.95) and the internal consistency is considered excellent (r=.92-.99) (Holdnack & Drozdick, 2009). The TOPF^{UK} is scored out of 70 and the raw scores were used within the following analysis.

Repeatable Battery for the Assessment of Neuropsychological Status (RBANS; Randolph, 1998)

The RBANS is a widely used, brief, psychometric test which measures several cognitive domains, yielding five index scores (attention, language, visuospatial/constructional ability, immediate memory and delayed memory) and a total scaled score. Index scores are standardised values based on age (mean(M)=100, standard deviation (SD)=15). All twelve subtests were administered and scored as described in the manual.

Data Analysis

All data was analysed by using the statistical package for the social sciences version 27 (SPSS; IBM Corp, 2020) and Rstudio (RStudio Team, 2022) using the Caret package (Kuhn, 2008). Corrections for *p-values* to control against family-wise error were applied where appropriate, on a post-hoc basis, using Benjamini and

Hochberg's (1995) false discovery rate adjustment. This was chosen as a less conservative adjustment to manage the trade-off between type 1 and type 2 error. No corrections were completed on the regression models due to the risk of inflating type 2 error (Armstrong, 2014).

The derivation sample was first examined. Independent t-tests were calculated to identify if there was a significant difference between data collected online and in person. Pairwise correlations were performed between HADS scores and RBANS index scores to assess whether there was a significant relationship.

Hierarchical Linear Regression

To address the first primary aim of the paper, hierarchical linear regression was performed for each RBANS index score and subtest raw score using the derivation sample. TOPF^{UK} score and demographic variables were the predictor variables in each model. A two-step hierarchical regression analysis was used. TOPF^{UK} score was entered into the first block and demographic variables were entered together into the second block of the analyses. Pairwise correlation between predictor variables were calculated prior to analysis to assess for multicollinearity alongside the variance inflation factor and tolerance statistic. Cook's distance was used to assess the influence of individual cases in the model. The Durbin-Watson test was used to assess the independence of errors using a guideline value of 2. Residual plots were examined for heteroscedasticity and normality.

RBANS index predicted and observed scores for each model were analysed, described and residuals reported. For comparative purposes, TOPF^{UK} predicted FSIQ was calculated as per the regression equation within the TOPF^{UK} Manual (Wechsler, 2011), which includes demographic variables, for each participant in the derivation sample and residuals with each observed index were calculated. Predicted and observed subtest scores were presented alongside the residuals of the models.

Cross validation

Two methods of Cross validation were utilised to validate the RBANS index models.

LOOCV Cross validation, a form of k-fold Cross validation, was performed using the derivation data set only. LOOCV is a technique recommended for small sample sizes to assess the accuracy of a model applied to independent data (Wong, 2015).

A second Cross validation procedure was performed on a clinical sample (CLcv)- and compared with ten participants selected at random from the normative sample (Ncv). For the selection of the normative sample, random numbers were generated between 0 and 100 for each participant using SPSS RV.Uniform function to compute a variable. Ten participants assigned the smallest numbers were then selected to comprise the Ncv.

In order for both data sets to be 'external', further models were created using the derivation data set excluding the ten selected participants that made up Ncv. These models were then applied to the Ncv and CLcv samples. Paired-sample t-tests were then used to compare the obtained and estimated scores for each participant.

Results

Derivation sample

After the exclusion criterion was applied, the non-clinical sample comprised 56 participants- 30 females and 26 males. They had an average age of 47.07 (SD=16.87, Range=22-80). The average of years of education was 14.95 (SD=3.64, Range 8-23). The demographic characteristics of the final sample, divided into age bands, are shown in table 1. Participants aged 18-37 typically were educated longer than older participants. The average HADS depression score was 2.70 (SD=2.08, Range= 0-9) and the average HADS anxiety score was 6.05 (SD=2.84, Range=1-11). The mean $TOPF^{UK}$ score was 49.23 (SD=10.49, Range= 29-69). Distribution of participants across the occupation skill levels was as follows: 1 (*n*=1); 2 (*n*= 7); 3 (*n*= 16); 4 (*n*=32).

Pairwise correlations were not significant between the RBANS indexes and each HADS score respectively. The correlations coefficients showed small effect sizes according to Cohen (1988), ranging between r=.04 and r=.19.

Independent t-tests comparing the mean difference between data collected online and in person revealed no significant differences (see table 2).

Table 1

Derivation Sample Characteristics (mean ± SD)

	Age Band					
	18-27	28-37	38-47	48-57	58-67	68+
n	11	8	6	11	14	6
Female (*n)*	3	4	2	8	11	2
Years of Education	16.09 ± 2.77	16.69 ± 2.76	12.00 ± 3.85	15.86 ± 3.76	14.96 ± 3.40	11.83 ± 3.87
Age	23.81 ± 1.08	32.50 ± 2.45	39.67 ± 1.21	53.18 ± 3.66	60.64 ± 2.84	73.67 ± 5.24
HADS A	6.00 ± 2.76	3.88 ± 2.95	6.50 ± 3.73	6.82 ± 2.71	7.14 ± 2.18	4.66 ± 2.34
HADS D	2.91 ± 1.76	1.63 ± 1.41	3.17 ± 3.87	2.82 ± 2.23	2.79 ± 1.85	2.83 ± 1.60
TOPF UK	52.09 ± 7.12	46.38 ± 12.02	41.67 ± 11.31	49.31 ± 11.83	51.43 ± 8.32	50.00 ± 14.30

Note. **HADS A=**Hospital anxiety and depression scale anxiety score. **HADS D=**Hospital anxiety and depression scale depression score.

TOPF=Test of Premorbid Functioning.

Table 2

Comparison of means between data collected online and in-person for the derivation sample

RBANS indexes	Online		In-person		
	M	*SD*	*M*	*SD*	p^a
Total Score	109.08	16.78	106.98	11.89	.617
Immediate Memory	101.62	13.73	100.52	13.68	.803
Delayed Memory	104.69	12.80	103.21	9.81	.661
Attention	109.31	17.53	103.88	12.92	.231
Visuospatial/constructional	106.08	14.22	111.43	12.08	.212
Language	107.77	11.29	106.00	11.70	.633
$TOPF^{UK}$	54.23	8.64	48.17	10.33	.061

Note. **RBANS=**Repeatable battery for the assessment of neuropsychological status. **$TOPF^{UK}$=**Test of Premorbid Functioning. **M**=Mean. **SD=**Standard deviation.

[a]*p-value* was calculated using an independent t-test, equal variances assumed

Hierarchical Multiple Regression Analyses

Pairwise correlations, using Spearman's Rank Correlation Coefficient (r_s) to assess for multicollinearity, revealed a significant correlation between years of education and occupation skill level (r_s=.61, p<.001). It was considered that the two variables were closely related. Multicollinearity can have an impact on the stability of the coefficients (Daoud, 2017). Thus, occupation skill level was excluded from the models in favour of years of education to preserve richness of continuous data.

Distributions of each RBANS index and subtest score were examined. Skew and Kurtosis statistics were considered. Four subtests were identified as exhibiting negative skew (skew statistic>1; see Appendix I for histograms). These were List Recognition, Picture Naming, Figure Recall, and Figure Copy. Line Orientation was flagged for negative skew on visual inspection of the histogram. Additionally, Figure Copy, Picture Naming and List Recognition, were identified to be leptokurtic which suggests low variability in scores for these subtests that cluster around the mean. This may be indicative of ceiling effects.

Two regression models were created for each index score and subtest by entering the predictor variables in blocks. In all cases, the $TOPF^{UK}$ score was entered in the first block and the demographics- sex, age and years of education- were all entered in the second block. Non-significant predictors were left within the models as they can mediate relationships between other variables and provide important information (Rohlfs, 2018).

Models predicting RBANS indexes

RBANS indexes were first used as dependent variables respectively in each model. The results of the regression analysis and associated models are described in table 3. The regression models did not violate any assumptions. Only the Visuospatial index models were non-significant both using $TOPF^{UK}$, R^2 = .00, F(1,54)= 1.00, SEE=14.24, *p*=.421, and using $TOPF^{UK}$ and Demographics as predictors, R^2 = .00, F(4,51)=

0.58, SEE=14.46, p=.681. Significant models were found to explain between 7%-23% of the variance observed in the indexes and the standard errors of the estimate (SE_{est}) ranged between 9.69 to 13.51.

$TOPF^{UK}$ score significantly predicted the following indexes: Total Scale, F(1,54)= 16.15, SEE=11.74, p=<.001; Immediate Memory, F(1,54)= 17.68, SEE=11.96, p=<.001; Attention, F(1,54)= 5.33, SEE=13.51, p=<.025; Language, F(1,54)=, SEE=11.74, p=<.001 when individually entered into the model. Demographic variables only improved the variance explained in the model (adj R^2) for two indexes, Language (Adj. R^2 =.09 to Adj. R^2 =.17) and Attention (Adj. R^2 =.07 to Adj. R^2 =.11).

The predictive model using only $TOPF^{UK}$ as a predictive variable was used for subsequent analysis for the indexes Total Scale, Immediate Memory and Delayed Memory. This was due to demographics having no impact, or a negative impact, on the variance explained by the model for these indexes. Combined $TOPF^{UK}$ and demographic predictive models were used in subsequent analysis for the Language and Attention index. The Visuospatial model was not investigated further as $TOPF^{UK}$ and demographics did not have a significant impact on the model, and the value of R^2 was close to zero in both cases.

Table 3

Hierarchical regression model summaries for each RBANS index.

	Model Statistics							Coefficients			
RBANS Index	Model	R^2	Adj. R^2	SEE	F	df	*p*	Variable	B	SE	t	*p*
Total Scale	1	.23	.22	11.74	16.15	1,54	<.001	(Constant)	77.23	7.60	10.17	<.001
								$TOPF^{UK}$	0.61	0.15	4.02	<.001
	2	.28	.22	11.73	4.83	4,51	.002	(Constant)	65.29	10.38	6.29	<.001
								$TOPF^{UK}$	0.50	0.17	3.02	<.001
								Age	0.07	0.10	0.66	0.51
								Sex	2.29	3.32	0.69	0.49
								Years of Education	0.70	0.50	1.38	0.17
Immediate Memory	1	.25	.23	11.96	17.68	1,54	<.001	(Constant)	68.63	7.74	8.87	<.001
								$TOPF^{UK}$	0.65	0.15	4.21	<.001
	2	.25	.19	12.26	4.29	4,51	.005	(Constant)	64.32	10.86	5.92	<.001
								$TOPF^{UK}$	0.62	0.17	3.56	<.001
								Age	0.04	0.11	0.34	.734
								Sex	0.99	3.47	0.29	.777
								Years of Education	0.17	0.53	0.32	.752
Visuospatial/Constructional	1	.02	.00	14.24	1.00	1,54	.321	(Constant)	100.46	9.21	10.91	<.001
								$TOPF^{UK}$	0.18	0.18	1.00	.321
	2	.04	.00	14.46	0.58	4,51	.681	(Constant)	100.70	12.81	7.86	<.001
								$TOPF^{UK}$	0.14	0.21	0.70	.486
								Age	0.07	0.13	0.58	.568
								Sex	-4.48	4.09	-1.09	.279
								Years of Education	0.35	0.62	0.56	.582

	Model Statistics							Coefficients			
RBANS Index	Model	R^2	Adj. R^2	SEE	F	df	*p*	Variable	B	SE	t	*p*
Language	1	.10	.09	10.92	6.14	1,54	.016	(Constant)	89.25	7.07	12.63	<.001
								$TOPF^{UK}$	0.35	0.14	2.48	.016
	2	.23	.17	10.38	3.89	4,51	.008	(Constant)	81.06	9.19	8.82	<.001
								$TOPF^{UK}$	0.31	0.15	2.13	.038
								Age	-0.10	0.09	-1.14	.259
								Sex	7.86	2.94	2.68	.010
								Years of Education	0.19	0.45	0.42	.680
Attention	1	.09	.07	13.51	5.33	1, 54	.025	(Constant)	85.38	8.74	9.77	<.001
								$TOPF^{UK}$	0.40	0.17	2.31	.025
	2	.17	.11	13.25	2.69	4, 51	.041	(Constant)	69.84	11.73	5.95	<.001
								$TOPF^{UK}$	0.28	0.19	1.48	.144
								Age	0.03	0.12	0.26	.794
								Sex	5.45	3.75	1.45	.152
								Years of Education	0.79	0.57	1.39	.172
Delayed Memory	1	.18	.16	9.69	11.70	1,54	.001	(Constant)	82.30	6.27	13.12	<.001
								$TOPF^{UK}$	0.43	0.13	3.42	.001
	2	.22	.16	9.71	3.62	4,51	.011	(Constant)	73.89	8.60	8.59	.000
								$TOPF^{UK}$	0.37	0.14	2.68	.010
								Age	0.02	0.09	0.23	.819
								Sex	3.30	2.75	1.20	.235
								Years of Education	0.36	0.42	0.85	.398

Note. Adj. = Adjusted. SEE= Standard Error of the Estimate. df= Degrees of Freedom. SE= Standard error. $TOPF^{UK}$= Test of Premorbid Functioning.

To investigate the predictive accuracy, the observed and estimated scores for each participant were examined across the models (see table 4). The mean difference between observed RBANS index and predicted RBANS index was between 7.65 -9.94.

Residuals between $TOPF^{UK}$-predicted FSIQ and each observed index score were calculated and presented in table 4. For all models, the $TOPF^{UK}$-predicted RBANS score provided more accurate predictions and, overall, smaller residuals than the $TOPF^{UK}$-predicted FSIQ.

The cumulative percentage of cases in which the predicted score fell within +5, +10, +15, and +20 points of the observed score is described in table 5. 53.57% to 71.43% of cases were predicted within 10 points of the observed score. The model with the highest percentage of cases not predicted within 20 points of the observed score, was the Immediate Memory index. In this model, 10.71% of cases were not predicted within 20 points. Only 1.79% of cases were not predicted within 20 points of the observed score for the Delayed memory Index.

The predictive accuracy of the qualitative classification, as defined by the RBANS manual, (rated from extremely low to very superior) is, also, presented in table 5. Total Scale was correctly categorised in 50% of cases. 89.29% of cases were either correctly categorised or categorised within +/- one category.

Mean absolute residuals were then plotted against the qualitative classification (see figure 1). As per previous research (e.g. Alves, Simões & Martins, 2012) those scoring at the extreme ends of the classifications experienced greater residuals and were less well predicted.

Table 4

Predicted and observed scores for each index

	Observed Index Score			Predicted Index Score			Mean Absolute Error[a]			Mean absolute difference between TOPF^{UK} predicted WAIS score and RBANS actual index[b]		
RBANS indexes	*M*	*SD*	*Min-Max*	*M*	*SD*	*Min-Max*	*M*	*SD*	*Range*	*M*	*SD*	*Range*
Total Scale	107.09	13.14	82-142	107.09	6.30	94.82-119.08	9.35	6.72	0.72-27.37	9.36	6.72	0.70-27.31
Immediate Memory	100.46	13.53	73-136	100.46	6.72	87.38-113.25	9.29	7.18	0.51-32.98	13.80	8.85	0.86-33.45
Visuospatial	109.48	14.24	72-131	-	-	-	-	-	-	-	-	-
Language	106.38	11.32	79-127	106.37	5.47	93.97-117.29	7.99	5.86	0.02-27.07	8.88	7.32	0.52-30.84
Delayed Memory	103.29	10.50	78-127	103.29	4.43	94.66-111.71	7.65	5.67	0.22-22.21	8.63	6.97	0.16-31.03
Attention	105.13	13.91	75-132	105.13	5.80	92.30-116.22	9.94	7.81	0.01-29.95	10.10	8.76	0.25-34.31

Note. Visuospatial index excluded from further analyses due to a significant model not being found. **M**= Mean. **SD=** Standard Deviation. **RBANS=** Repeatable Battery for the Assessment of Cognitive Status. **TOPF^{UK}=** Test of Premorbid Functioning. **WAIS=** Wechsler's Adult Intelligence Scale.

[a] Residuals calculated by: Predicted Index Score – Actual Index Score

[b] TOPF^{UK} predicted WAIS calculated by the demographic regression equation presented in the manual (Wechsler, 2011)

Table 5

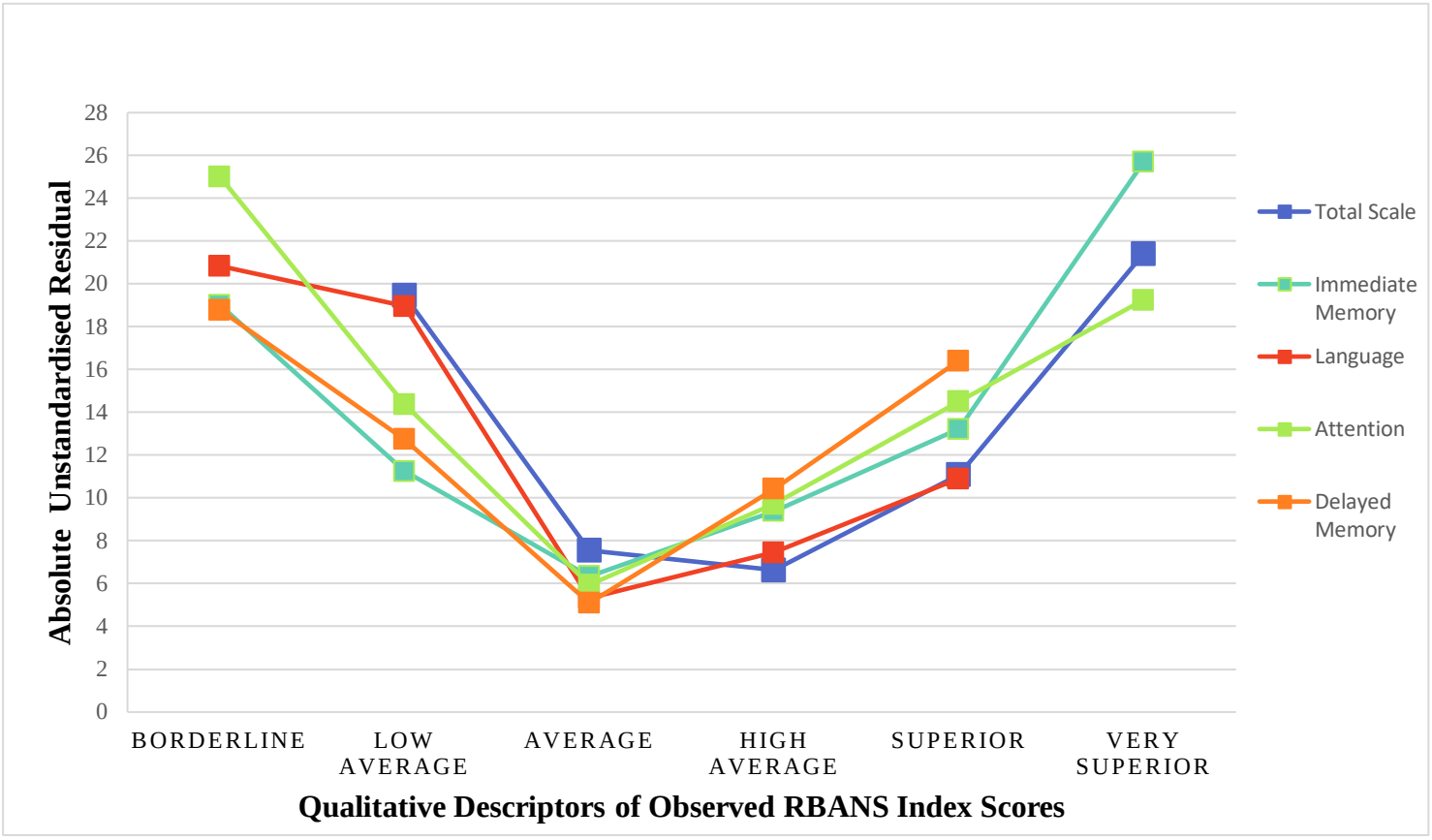
Predictive accuracy of the index regression models

	Percent of cases where TOPF^{UK}-predicted RBANS score is within 5, 10, 15 and 20 points of actual score				Percent within the same category (%)	Percent within previous/following category (%)
	± 5	±10	±15	±20		
Total Scale	35.71	58.93	78.57	91.07	50.00	39.29
Immediate Memory	35.71	62.50	82.14	89.29	60.71	32.14
Visuospatial	-	-	-	-	-	-
Language	37.50	71.43	87.50	94.64	50.00	39.29
Delayed Memory	35.71	69.64	85.71	98.21	66.07	26.32
Attention	33.93	53.57	71.43	91.07	46.43	37.50

Note. Visuospatial excluded due to a significant model not being found. Qualitative classifications were defined as follows: 130 and above=Very Superior. 120-129= Superior. 110-119=High Average. 90-109= Average. 80-89= Low Average. 70-79= Borderline. 69 and below= Extremely Low.

Figure 1

Line graph of the absolute unstandardized residuals plotted against RBANS observed qualitative category.



Subtest models

Subtest raw scores were also used as dependent variables in respective models. The results of these regression analyses and associated equations are shown in table 6. The assumption of residual normality was violated for three of the subtests, List Recognition, Picture Naming and Figure Copy. Guidance published by Knief & Forstmeier (2021) suggests that residual normality violations are relatively unproblematic within linear regression but that statistics should be interpreted with caution. Thus, linear regressions were tentatively performed for these subtests.

Adjusted variance explained by the significant models ranged from 8% to 42%. The TOPF^{UK} did not significantly predict scores on six subtests: Line Orientation, Coding, Figure Recall, List Recognition, Picture Naming and Figure Copy. Demographics added to the variance explained by the model, where

TOPF^{UK} was significantly contributed to the model, for the following subtests: List Learning (Adj. R^2 increase: .16); Story Memory (Adj. R^2 increase: .01); Semantic Fluency (Adj. R^2 increase: .12); Digit span (Adj. R^2 increase: .01); Coding (Adj. R^2 increase: .42); List Recall (Adj. R^2 increase: .27) and List Recognition (Adj. R^2 increase: .12).

Mean standardised residuals for each model were examined for significant models and ranged from 0.73-0.81 (see table 7). No outliers (Standard Residual>3) were reported. A correction for familywise error using the Benjamini-Hochberg procedure (Benjamini & Hochberg, 1995) showed no change in statistical significance.

Table 6

Hierarchical Regression Models for RBANS subtests

RBANS Subtests	Model	R^2	Adj R^2	SEE	F	df	*p*	Equation
List Learning	1	.13	.11	4.96	7.70	1,54	*.008*	=24.80+(0.19* $TOPF^{UK}$) + (-0.14*age) + (1.16*Sex)[a]
	2	.32	.27	4.91	6.07	4,51	*<.001*	
Story Memory	1	.24	.22	2.96	16.83	1,54	*<.001*	=11.06 + (0.15* $TOPF^{UK}$) +(-0.03*Age) + (-0.47*Sex) + (0.12*Years of Education)
	2	.29	.23	2.95	5.07	4,51	*.002*	
Figure Copy	1	.01	.00	1.63	0.75	1,54	.390	-
	2	.09	.02	1.61	1.30	4,51	.283	-
Line Orientation	1	.02	.00	2.12	1.06	1,54	*.309*	=19.13 + (0.03* $TOPF^{UK}$) + (-0.03*Age) + (-1.04*Sex) + (0.02*Years of Education)
	2	.16	.09	2.03	2.37	4,51	*.070*	
Picture Naming	1	.02	.01	0.39	1.31	1,54	*.257*	-
	2	.06	.00	0.39	0.80	4,51	*.534*	-
Semantic Fluency	1	.11	.09	5.08	6.60	1,54	*.013*	=13.11 + (0.15*$TOPF^{UK}$) + (-0.8*Age) + (3.47*Sex) + (0.14*Years of Education)
	2	.27	.21	4.74	4.67	4,51	*.003*	
Digit Span	1	.16	.14	2.45	10.22	1,54	*.002*	=5.94+(0.08* $TOPF^{UK}$) + (-0.11*Sex) + (0.17*Years of Education)[b]
	2	.21	.15	2.44	3.32	4,51	*.017*	
Coding	1	.00	.00	11.10	.00	1,54	*.954*	=59.79+(-0.42*Age) + (3.56*Sex) + (0.39*Years of Education)[c]
	2	.46	.42	8.39	10.85	4,51	*<.001*	
List Recall	1	.10	.08	2.09	5.74	1,54	*.020*	=4.98+(0.08* $TOPF^{UK}$) + (-0.07*Age) + (1.41*Sex) + (-0.05*Years of Education)
	2	.40	.35	1.75	8.42	4,51	*<.001*	

RBANS Subtests	Model	R^2	Adj R^2	SEE	F	df	*p*	Equation
List Recognition	1	.01	.00	0.78	0.60	1,54	.442	=19.27+(0.003* $TOPF^{UK}$) + (-0.01*Age) + (0.07*Sex) + (0.05*Years of Education)
	2	.18	.12	0.73	2.79	4,51	*.036*	
Story Recall	1	.11	.09	2.04	6.38	1,54	*.015*	=6.07 + (0.07* $TOPF^{UK}$)[d]
	2	.13	.06	2.06	1.95	4,51	*.117*	
Figure Recall	1	.01	.00	3.82	0.42	1,54	.518	=15.51 + (0.02* $TOPF^{UK}$) + (-0.07*Age) + (0.22*Sex) + (0.15*Years of Education)
	2	.15	.09	3.63	2.29	4,51	*.072*	

Note. Model 1= $TOPF^{UK}$ entered. Model 2= $TOPF^{UK}$ and Demographic variables entered.
[a]Years of Education negligible coefficient (.004)
[b] Age negligible coefficient (.004)
[C] $TOPF^{UK}$ negligible coefficient (.000)
[d] Significant model reported

Table 7

Observed scores, predicted scores and standardised absolute residuals for each subtest model

Subtest	Observed Scores			Predicted Scores			Standardised absolute Residual		
	M	*SD*	*Min-Max*	*M*	*SD*	*Min-Max*	*M*	*SD*	*Min-Max*
List Learning	29.23	5.26	17-37	29.23	2.99	21.96-35.71	0.74	0.60	0.01-2.85
Story Memory	18.00	3.36	9-24	18.00	1.79	13.75-20.81	0.73	0.62	0.04-2.34
Figure Copy	19.16	1.63	13-20	-	-	-	-	-	-
Line Orientation	18.09	2.13	12-20	18.09	0.84	16.03-19.71	0.80	0.53	0.03-2.60
Picture Naming	9.82	0.39	9-10	-	-	-	-	-	-
Semantic Fluency	24.18	0.54	11-35	24.18	2.76	17.56-30.07	0.78	0.55	0.09-3.00
Digit Span	12.00	2.64	7-16	12.00	1.20	9.38-13.78	0.81	0.51	0.12-2.02
Coding	51.23	11.00	24-76	51.23	7.45	36.30-64.21	0.78	0.56	0.01-2.12
List Recall	6.75	2.18	0-10	6.75	1.37	3.33-9.94	0.79	0.54	0.03-2.30
List Recognition	19.63	0.78	17-20	19.63	0.33	18.82-20.15	0.66	0.69	0.00-3.31
Story Recall	9.32	2.13	3-12	9.32	0.78	7.52-10.78	0.77	0.57	0.00-2.44
Figure Recall	15.68	3.80	5-20	15.68	1.48	12.00-18.20	0.76	0.78	0.01-2.60

Note. Due to no significant model being found, predicted scores and standardised absolute residuals are not presented for Figure Copy and Picture Naming. **M=**Mean. **SD**= Standard Deviation.

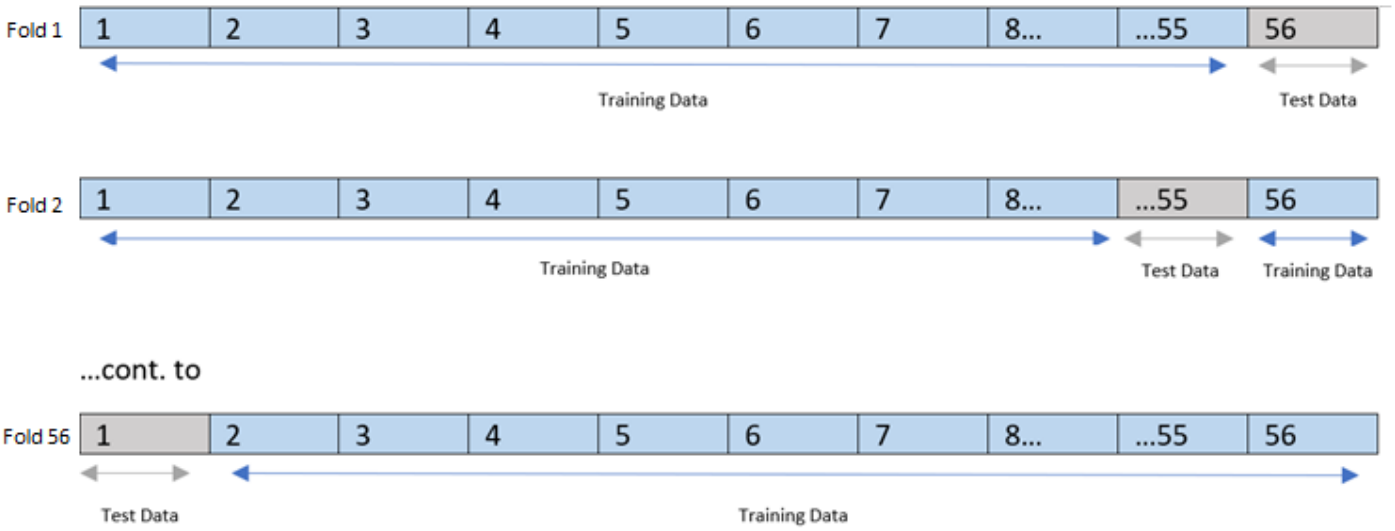
Cross validation

Leave-one-out Cross validation

The RBANS Index models were first cross-validated within the =derivation sample to perform LOOCV. To perform LOOCV, data was separated into a training and test set. This division is termed a 'fold'. In LOOCV, all but one of the data points (n=55) is extracted into the training set. The model is then trained with the training set and tested on the 'left out' data point (n=1) which simulates 'external data'. This process is then repeated so that k folds are created, where k=the total sample size (k=56). Thus, each data point becomes the test data set across the iterations. This process is illustrated in figure 2. The method benefits from not including any random selection due to all data points being tested systematically. This means that the validation error measures are stable.

Figure 2

Illustration of the Leave-One-Out Cross Validation methodology



The Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) averaged across the folds are reported in table 8 for each index. RMSE is more sensitive to large errors. Therefore, as RMSE > MAE there is a variation in error size across the cases in each model.

The percent predicted within the same qualitative category, and within one qualitative category, across all folds are also reported. Between 75% and 91% were predicted correctly or within one category of the observed score.

Table 8

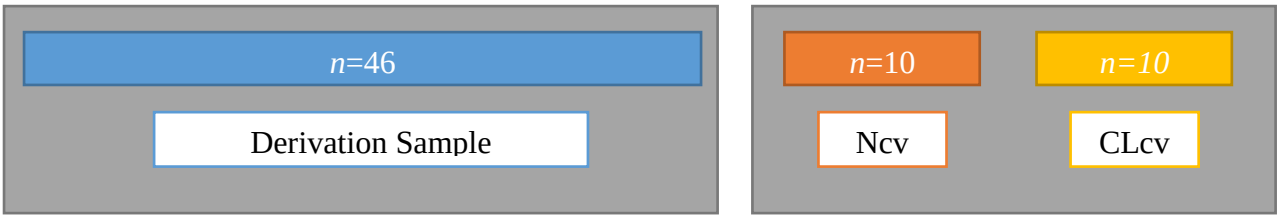
Results and predictive accuracy of the Leave-one-out Cross Validation analyses.

RBANS Indexes	RMSE	MAE	Percent within the same category (%)	Percent within previous/following category (%)
Total Scale	11.97	9.70	51.79	39.29
Immediate Memory	12.22	9.66	57.14	30.36
Visuospatial/Constructional	14.56	12.34	-	-
Language	10.87	8.78	51.79	33.93
Attention	13.99	10.98	42.86	32.14
Delayed Memory	9.90	7.94	67.86	23.21

Note. **RBANS** = Repeatable Battery for the Assessment of Neuropsychological Status. **RMSE=** Root Mean Squared Error. **MAE**= Mean Absolute Error.

Figure 3

Illustration of the clinical Cross validation methodology



Training Data

Test Data

Clinical Cross validation

Ten clinical participants were included in the CLcv sample. A range of diagnostic conditions were included in the sample and presented in table 9. There were three missing index scores within the data set due to significant impairment in one case meaning certain subtests could not be completed.

Ten participants were randomly selected from the derivation sample as a control. Demographic variables for the clinical sample and the derivation sub sample are described in table 10.

Models were derived using the remaining data from the derivation sample and applied to the CLcv and Ncv samples. Paired sample t-tests showed that there were no significant differences between predicted RBANS index scores and observed scores for the normative sample, but significant differences between the comparable values within the clinical sample (see table 11). When *p* values were adjusted for familywise error, delayed memory was no longer significantly different in the Clinical Sample.

Table 9

Range of diagnoses and number of participants included

Diagnosis	*n*
Stroke	*4*
Multiple Sclerosis	*3*
Dementia	*1*
Parkinson's disease	*1*
Hydrocephalus	*1*

Table 10

Demographic variables and scores across both normative and clinical cross validation samples

	N_{cv}	CL_{cv}
n	10	10
Female (%)	50.00	60.00
Years of Education *(Mean ± SD)*	14.90 ± 2.64	14.10 ± 2.88
Age *(Mean ± SD)*	40.80 ± 16.25	50.00 ± 10.38
HADS Depression	2.4 ± 1.43	9.10 ± 2.12
HADS Anxiety	6.1 ± 3.18	9.90 ± 5.51
$TOPF^{UK}$ Raw	46.00 ±8.46	50.70 ± 12.06

Table 11.

Paired-sample t-tests across RBANS indexes for the normative and clinical sample

		Predicted Index	Observed Index		
	RBANS indexes	*M* ± *SD*	*M* ± *SD*	*t*	*P*
Ncv	Total Scale	105.89 ± 5.53	100.70 ± 9.14	1.50	.168
	Immediate Memory	98.96 ± 5.43	95.80 ± 12.82	0.83	.429
	Attention	104.15 ± 7.33	100.70 ± 9.71	0.77	.462
	Language	106.62 ± 4.94	101.80 ± 9.37	1.73	.118
	Delayed Memory	102.18 ± 3.93	99.90 ± 4.95	1.13	.286
CLcv	Total Scale	108.21 ± 7.98	86.78 ± 12.85	5.06	<.001
	Immediate Memory	101.97 ± 7.74	86.20 ± 7.60	6.14	<.001
	Attention	106.01 ± 6.28	86.78 ± 11.18	4.17	.003
	Language	107.68 ± 5.41	90.70 ± 8.69	5.14	<.001
	Delayed Memory	103.83 ± 5.66	87.89 ± 19.46	2.58	.033

Note. **Ncv=** Normative cross validation sample. **CLcv=** Clinical cross validation sample. **M**= Mean. **SD=**Standard Deviation

Discussion

The aim of this study was to investigate whether the TOPF^{UK} and demographic variables could meaningfully predict performance on the RBANS indexes, in a neurologically healthy sample from the United Kingdom. To do this, multiple regression formulae were developed. Significant predictive models were found for the Total Scale index, Immediate Memory index, Language index, Attention index, and Delayed Memory index. The variance explained for these models was comparative and, in some cases, higher to that observed in previous research using the WTAR (Duff et al., 2019) which may reflect the use of a wider age range within the sample.

Despite the statistical significance of the models, there remains a large discrepancy between the degree of variance explained by the TOPF^{UK} on the RBANS total score performance relative to the WAIS FSIQ (e.g. Holdnack & Drozdick, 2009; Watt et al., 2018). One reason for this may be the inclusion of 'fluid' intelligence tests within the RBANS. The distinction between fluid and crystallised intelligence has commonly been discussed in the literature as separate and distinguishable cognitive skills (Blair, 2006). Fluid intelligence, commonly representative of analytic intelligence, has been found to be distinct from the language system (Woolgar et al., 2018) and, thus, when compared to crystallised intelligence, does not correlate as well with oral word reading tests (Bright & van der Linde, 2017). Takaiwa et al. (2018) considered all the RBANS indexes to be measures of fluid intelligence domains such as: current learning ability (Immediate and Delayed Memory indexes); visual attention (Visuospatial index); flexibility of thinking (Semantic Fluency subtest); executive functioning and processing speed (Attention index). In line with this inference, Bright & van der Linde (2018) found that correlations between the NART and WAIS 'fluid intelligence' indexes (Working Memory Index and Perceptual Reasoning Index) were modest compared to the more 'crystallised intelligence' index (Verbal Comprehension Index). This may suggest that there are alternative variables that could be considered in future research that better predict fluid premorbid functioning.

The Immediate and Delayed Memory indexes were significantly predicted by the TOPF^{UK} and demographic variables. The amount of variance accounted for is low (e.g. 16% to 23%) but comparable to previous research looking at the predictability of memory ability from oral word reading tests and demographic variables (e.g. Isella et al., 2005; Duff, 2010; Duff, 2015; Hilsabeck & Sutker, 2009). This result, in line with previous research, points to the importance of using caution when interpreting premorbid estimates, based on word reading, in relation to premorbid cognitive domains beyond IQ.

Similar caution is supported when predicting premorbid performance on the visuospatial index from word reading tests. Within this research, no significant models were found for this index. Along the same line, Duff et al. (2015) found that the visuospatial index was less well predicted by demographic variables and the WTAR comparative to the other indexes. As TOPF^{UK} is a verbal test, it reasons that its ability to predict non-verbal, visuospatial cognitive domains may be less robust. In some cases, alternative psychometric tests are suggested when predicting visuospatial or non-verbal domains. For instance, Chaurisiya et al. (2022) suggested that the use of a matrix reasoning test to predict premorbid functioning on non-verbal domains predicted 38% of the variance in Performance IQ, a non-verbal WAIS index. Patterns were found, consistent with this notion, at a subtest level when using a two-factor structure for the RBANS, as suggested by Duff et al. (2009). This comprised of a verbal index, made up of the subtests List learning, Story Memory, List Recall, List Recognition and Story Recall. It also included a revised visuospatial index which is made up of subtests Line orientation, Figure Copy, Coding and Figure Recall. Apart from List Recognition, the TOPF^{UK} significantly predicted all subtests contained within the verbal index and explained between 8-22% of the variance within these subtests. Alternatively, the TOPF^{UK} did not significantly predict scores within the revised visuospatial, non-verbal index. It may be beneficial for further research to assess whether an alternative predictor variable may improve these estimates.

An alternative explanation may be due to the sensitivity of the subtests within this index. The two subtests within the Visuospatial index, Figure copy and Line Orientation, were identified as exhibiting possible ceiling effects. Thus, this may impact the subtests', and therefore the indexes', sensitivity to performance variation within a healthy sample. This was not confined to the visuospatial index. Scores on the Picture Naming varied only between 9-10 and List Recognition scores ranged from 17-20. This observation is consistent with previous research (e.g. Bartels et al., 2010; Duff et al., 2008). These subtests are suggested to be 'deficit orientated' (Bartels et al., 2010). Therefore, premorbid functioning may be less important to consider in these cases due to the expectation of participant's scores to consistently cluster at full marks.

When evaluating the predictive accuracy of the model, the data illustrated that a high cumulative percentage was predicted within 10 points of the observed accuracy (53.57%-71.43%). This is consistent with previous research utilising alternative methods of estimating premorbid functioning such as Alves et al. (2012) and Schoenberg et al. (2002). Further to this, residuals were consistently smaller than those calculated between $TOPF^{UK}$-predicted FSIQ and observed indexes, emphasising the importance of creating test-specific normative data to predict premorbid functioning.

Two methods of cross validation were performed. In both cases this was to see how accurately models using $TOPF^{UK}$ and demographic variables would predict RBANS scores 'external' to the derivation sample. This is important due to the 'shrinkage' that can occur to the statistical fit of regression-based models to independent external data (Copas, 1997). LOOCV showed that the predictive accuracy when the models were applied to 'external' data was consistent with that observed in the original model. For instance, in the original model, 46.43-66.07% of cases were correctly qualitatively categorised, whereas the LOOCV correctly categorised between 42.86-67.86%. This supports a degree of generalisability for the models. However, the method is limited due to the methodology being confined to the original sample. Further

research is required to cross validate on larger, independent, samples to gain a more accurate sense of the generalisability of the findings.

When applied to a clinical sample, the regression models identified clinical decline. Significant differences were found between observed and estimated scores in only the clinical sample. When the t-tests were corrected for familywise error, the difference between observed and predicted delayed memory was no longer significant. This may be due to the heterogeneity of the diagnoses represented in the sample. A further cross validation is required to gain a clearer picture of the clinical utility of the equations.

Limitations

A number of limitations were carefully considered for this study. Firstly, the use of regression-based models for predicted PCF have been criticised for exhibiting a degree of bias in the produced estimates (Veiel & Koopman, 2001). Models tend to be susceptible to a large degree of error in the outer ranges of intellectual ability (e.g. Basso et al., 2000). This was evident in the data of the present study and encourages further consideration due to the potential impact on the accuracy of predictions, namely the over-estimation of PCF for those individuals performing in the below-average range and the under-estimation of PCF for those performing at an above-average level. Adjustments to reduce this bias have been proposed by Veiel & Koopman (2001), but criticised by Grove (2001) for inflating the error. This is a limitation of the use of regression equations as a whole and requires further research.

Additionally, it is important to note that this current study is limited to a relatively small sample size and, thus, the regression equations reported in this study were not created or validated for use in clinical settings. This was not within the scope of this paper. However, the results show promise in the use of $TOPF^{UK}$ and demographic variables to predict RBANS indexes and highlight important considerations when doing so.

Further research is needed with a larger normative sample to create and validate regression equations that can be used in clinical practice. Further clinical validation is also required. The clinical sample was limited by the heterogeneity of diagnoses and limited representation of neurological disorders. A larger sample is required to allow comparisons to be made across neurological conditions regarding the regression models' sensitivity to cognitive decline. This will guide their use in clinical settings.

Conclusions

This research aimed to investigate the predictability of RBANS performance from the $TOPF^{UK}$ and demographic variables. Consequently, it hoped to inform the use of the $TOPF^{UK}$ and demographic variables, alongside the RBANS, during neuropsychological assessment.

The results showed a high variability of predictive accuracy across models. In general, $TOPF^{UK}$ was better at predicting verbal subtests than non-verbal subtests. No significant model was found for the visuospatial index which suggests that caution should be utilised when predicting premorbid functioning from the $TOPF^{UK}$ using this index.

Across indexes, the models explained between 7-23% of the variance observed. This shows promise for using the $TOPF^{UK}$ and demographic variables alongside the RBANS in clinical services. Whilst equations are presented in this paper, further research using a larger sample size may help increase predictive accuracy of the models and assist in validating their use within clinical services. Additionally, further validation may be beneficial with a larger, more representative clinical sample to allow for the utility of the models to be investigated with different clinical populations.

References

Al-Ghatani, A. M., Obonsawin, M. C., Binshaig, B. A., & Al-Moutaery, K. R. (2011). Saudi normative data for the Wisconsin card sorting test, Stroop test, test of non-verbal intelligence-3, picture completion and vocabulary (subtest of the Wechsler adult intelligence scale-revised). *Neurosciences Journal*, *16*(1), 29-41.

Alves, L., Simões, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: the Irregular Word Reading Test (TeLPI). *Archives of Clinical Neuropsychology*, *27*(1), 58-68.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502-508.

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS—R. *Journal of Consulting and Clinical Psychology*, *52*(5), 885.

Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: a longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience*, *11*(1), 1-12.

Basso, M. R., Bornstein, R. A., Roper, B. L., & McCoy, V. L. (2000). Limited accuracy of premorbid intelligence estimators: A demonstration of regression to the mean. *The Clinical Neuropsychologist*, *14*(3), 325-340.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289-300.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: an updated literature review. *Journal of Psychosomatic Research*, *52*(2), 69-77.

Blair, C. (2006). How similar are fluid cognition and general intelligence? A developmental neuroscience perspective on fluid cognition as an aspect of human cognitive ability. *Behavioral and Brain Sciences*, *29*(2), 109-125.

Bright, P., & van der Linde, I. (2018). Comparison of methods for estimating premorbid intelligence. *Neuropsychological Rehabilitation*. *30(1),* 1-14.

Carter, K. R., Scott, J. G., Adams, R. L., & Linck, J. (2016). Base rate comparison of suboptimal scores on the RBANS effort scale and effort index in Parkinson's disease. *The Clinical Neuropsychologist*, *30*(7), 1118-1125.

Chaurasiya, A., Ranjan, J. K., Pandey, N., & Asthana, H. S. (2022). Estimation of premorbid intelligence: Demographical and current neurocognitive functioning based algorithms. *Asian Journal of Psychiatry, 72*, article 103065.

Chen, Y. J., Ho, M. Y., Chen, K. J., Hsu, C. F., & Ryu, S. J. (2009). Estimation of premorbid general fluid intelligence using traditional Chinese reading performance in Taiwanese samples. *Psychiatry and Clinical Neurosciences*, *63*(4), 500-507.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Copas, J. B. (1997). Using regression models for prediction: shrinkage and regression to the mean. *Statistical Methods in Medical Research*, *6*(2), 167-183.

Crawford, J. R., & Allan, K. M. (1997). Estimating premorbid WAIS-R IQ with demographic variables: Regression equations derived from a UK sample. *The Clinical Neuropsychologist*, *11*(2), 192-197

Crawford, J. R., Nelson, H. E., Blackmore, L., Cochrane, R. H. B., & Allan, K. M. (1990). Estimating premorbid intelligence by combining the NART and demographic variables: An examination of the NART standardisation sample and supplementary equations. *Personality and Individual Differences*, *11*(11), 1153-1157.

Crawford, J. R., Venneri, A., & O'Carroll, R. E. (1998). *Neuropsychological assessment of the elderly*. Elsevier.

Crouch, E., Probst, J. C., Radcliff, E., Bennett, K. J., & McKinney, S. H. (2019). Prevalence of adverse childhood experiences (ACEs) among US children. *Child Abuse & Neglect, 92*, 209-218.

Daoud, J. I. (2017, December). Multicollinearity and regression analysis. In *Journal of Physics: Conference Series* (Vol. 949, No. 1, p. 012009). IOP Publishing.

Delaney-Black, V., Covington, C., Ondersma, S. J., Nordstrom-Klee, B., Templin, T., Ager, J., & Janisse, J., & Sokol, R. J. (2002). Violence exposure, trauma, and IQ and/or reading deficits among urban children. *Archives of Pediatrics & Adolescent Medicine*, *156*(3), 280-285.

Duff, K. (2010). Predicting premorbid memory functioning in older adults. *Applied Neuropsychology*, *17*(4), 278-282.

Duff, K., & Ramezani, A. (2015). Regression-based normative formulae for the repeatable battery for the assessment of neuropsychological status for older adults. *Archives of Clinical Neuropsychology*, *30*(7), 600-604.

Duff, K., Clark, H. J. D., O'Bryant, S. E., Mold, J. W., Schiffer, R. B., & Sutker, P. B. (2008). Utility of the RBANS in detecting cognitive impairment associated with Alzheimer's disease: sensitivity, specificity, and positive and negative predictive powers. *Archives of Clinical Neuropsychology*, *23*(5), 603-612.

Duff, K., Dalley, B. C., Suhrie, K. R., & Hammers, D. B. (2019). Predicting premorbid scores on the repeatable battery for the assessment of neuropsychological status and their validation in an elderly sample. *Archives of Clinical Neuropsychology*, *34*(3), 395-402.

Duff, K., Langbehn, D. R., Schoenberg, M. R., Moser, D. J., Baade, L. E., Mold, J. W., ... & Adams, R. L. (2009). Normative data on and psychometric properties of verbal and visual indexes of the RBANS in older adults. *The Clinical Neuropsychologist*, *23*(1), 39-50.

Franzen, M. D., Burgess, E. J., & Smith-Seemiller, L. (1997). Methods of estimating premorbid functioning. *Archives of Clinical Neuropsychology*, *12*(8), 711-738.

Grove, W. M. (2001). Bias and error rates for premorbid IQ estimators: Comment on Veiel and Koopman (2001). *Psychological Assessment*, 13(3), 396–398.

Hardcastle, K., Bellis, M. A., Ford, K., Hughes, K., Garner, J., & Rodriguez, G. R. (2018). Measuring the relationships between adverse childhood experiences and educational and employment success in England and Wales: findings from a retrospective study. *Public Health, 165*, 106-116.

Harnett, M. A., Godfrey, H. P., & Knight, R. G. (2004). Regression equations for predicting premorbid performance on executive test measures by persons with traumatic brain injuries. *New Zealand Journal of Psychology*, *33*(2), 78.

Hilsabeck, R. C., & Sutker, P. B. (2009). Using an implicit memory task to estimate premorbid memory. *Archives of Clinical Neuropsychology*, *24*(2), 179-191.

Holdnack, H.A. (2001). *Wechsler Test of Adult Reading: WTAR.* San Antonio. The Psychological Corporation

Holdnack, J. A., & Drozdick, L. W. (2009). Advanced clinical solutions for WAIS-IV and WMS-IV: Clinical and interpretive manual. *Texas: Pearson*.

IBM Corp. (2021). IBM SPSS Statistics for Windows (28.0). Armonk, NY: IBM Corp

Isella, V., Villa, M. L., Forapani, E., Piamarta, F., Russo, A., & Appollonio, I. M. (2005). Ineffectiveness of an Italian NART-equivalent for the estimation of verbal learning ability in normal elderly. *Journal of Clinical and Experimental Neuropsychology*, *27*(5), 618-623.

Jenkinson, T. M., Muncer, S., Wheeler, M., Brechin, D., & Evans, S. (2018). Estimating verbal fluency and naming ability from the test of premorbid functioning and demographic variables: Regression equations derived from a regional UK sample. *British Journal of Clinical Psychology*, *57*(2), 135-147.

Karakuła-Juchnowicz, H., & Stecka, M. (2017). Polish Adult Reading Test (PART)-construction of Polish test for estimating the level of premorbid intelligence in schizophrenia. *Psychiatr Pol*, *51*(4), 673-85

Kirton, J. W., Soble, J. R., Marceaux, J. C., Messerly, J., Bain, K. M., Webber, T. A., ... & McCoy, K. J. (2020). Comparison of models of premorbid IQ estimation using the TOPF UK, OPIE-3, and Barona equation, with corrections for the Flynn effect. *Neuropsychology*, *34*(1), 43.

Knief, U., & Forstmeier, W. (2021). Violating the normality assumption may be the lesser of two evils. *Behavior Research Methods*, *53*(6), 2576-2590.

Knight, R. G., McMahon, J., Green, T. J., & Skeaff, C. M. (2006). Regression equations for predicting scores of persons over 65 on the Rey Auditory Verbal Learning Test, the mini-mental state examination, the trail making test and semantic fluency measures. *British Journal of Clinical Psychology*, *45*(3), 393-402.

Krull, K. R., Scott, J. G., & Sherer, M. (1995). Estimation of premorbid intelligence from combined performance and demographic variables. *The Clinical Neuropsychologist, 9,* 83–88

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software*, 28, 1-26.

Loughan, A. R., Braun, S. E., & Lanoye, A. (2019). Repeatable battery for the assessment of neuropsychological status (RBANS): preliminary utility in adult neuro-oncology. *Neuro-Oncology Practice*, *6*(4), 289-296.

Matsuoka, K., Uno, M., Kasai, K., Koyama, K., & Kim, Y. (2006). Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test. *Psychiatry and Clinical Neurosciences*, *60*(3), 332-339.

Murman, D. L. (2015). The impact of age on cognition. In *Seminars in hearing* (36(03), pp. 111-121). Thieme Medical Publishers.

Nelson, H. E., & McKenna, P. A. T. (1975). The use of current reading ability in the assessment of dementia. *British Journal of Social and Clinical Psychology*, *14*(3), 259-267.

O'Carroll, R., Prentice, N., Murray, C., Van Beck, M., Ebmeier, K., & Goodwin, G. (1995). Further Evidence that Reading Ability is not Preserved in Alzheimer's Disease. *British Journal of Psychiatry*, *167*(5), 659-662. doi:10.1192/bjp.167.5.659

Office for National Statistics. (2020). *Standard occupational classification.* Retrieved from https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassification soc/soc2020/soc2020volume1structureanddescriptionsofunitgroups

Randolph, C. (1998). *Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)*. San Antonio, TX: Psychological Corporation.

Reynolds, C. R. (1997). Postscripts on premorbid ability estimation: conceptual addenda and a few words on alternative and conditional approaches. *Archives of Clinical Neuropsychology***,** 12, 769–778.

Rolstad, S., Nordlund, A., Gustavsson, M. H., Eckerström, C., Klang, O., Hansen, S., & Wallin, A. (2008). The Swedish National Adult Reading Test (NART-SWE): A test of premorbid IQ. *Scandinavian Journal of Psychology*, *49*(6), 577-582

R Studio Team (2022). *RStudio: Integrated Development for R.* RStudio, PBC, Boston, MA

Rohlfs, C. (2018). Taking a Chance in the Classroom: Top 10 Reasons Not to Drop Insignificant Regressors: A Statistical Listicle. *CHANCE*, *31*(4), 51-56.

Schoenberg, M. R., Scott, J. G., Duff, K., & Adams, R. L. (2002). Estimation of WAIS-III intelligence from combined performance and demographic variables: Development of the OPIE-3. *The Clinical Neuropsychologist*, *16*(4), 426-438

Schrauf, R. W., Weintraub, S., & Navarro, E. (2006). Is adaptation of the Word Accentuation Test of Premorbid Intelligence necessary for use among older, Spanish-speaking immigrants in the United States? *Journal of the International Neuropsychological Society*, *12*(3), 391-399.

Stern, A. F. (2014). The hospital anxiety and depression scale. *Occupational Medicine*, *64*(5), 393-394.

Takaiwa, A., Kuwayama, N., Akioka, N., Kashiwazaki, D., & Kuroda, S. (2018). Discrepancy analysis between crystallized and fluid intelligence tests: a novel method to detect mild cognitive impairment in patients with asymptomatic carotid artery stenosis. *European Journal of Neurology*, *25*(2), 313-319.

Tuokko, H., Vernon-Wilkinson, R., Weir, J., & Beattie, B. L. (1991). Cued recall and early identification of dementia. *Journal of Clinical and Experimental Neuropsychology*, *13*(6), 871-879.

van Os, J., Marsman, A., van Dam, D., & Simons, C. J. (2017). Evidence that the impact of childhood trauma on IQ is substantial in controls, moderate in siblings, and absent in patients with psychotic disorder. *Schizophrenia Bulletin, 43*(2), 316-324.

Veiel, H. O., & Koopman, R. F. (2001). The bias in regression-based indices of premorbid IQ. *Psychological Assessment*, *13*(3), 356.

Watt, S., Ong, B., & Crowe, S. F. (2018). Developing a regression equation for predicting premorbid functioning in an Australian sample using the National Adult Reading Test. *Australian Journal of Psychology*, *70*(2), 186-195.

Wechsler, D. (1955). *Manual for the Wechsler adult intelligence scale.* Psychological Corp..

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale—Fourth Edition Administration and Scoring Manual.* San Antonio, TX: Pearson.

Wechsler, D. (2011). *The test of premorbid function (TOPF UK).* San Antonio, TX: The Psychological Corporation.

Willshire, D., Kinsella, G., & Prior, M. (1991). Estimating WAIS-R IQ from the National Adult Reading Test: a cross validation. *Journal of Clinical and Experimental Neuropsychology*, *13*(2), 204-216.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology*, 46, 1554-1555.

Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out Cross validation. *Pattern Recognition,* 48(9), 2839-2846.

Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nature Human Behaviour*, *2*(3), 200-204.

Yi, D., Seo, E. H., Han, J. Y., Sohn, B. K., Byun, M. S., Lee, J. H., ... & Lee, D. Y. (2017). Development of the Korean Adult Reading Test (KART) to estimate premorbid intelligence in dementia patients. *Plos One*, *12*(7), Article e0181523

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361-370.

Part Three: Appendices

Appendix A. Reflective Statement

Conducting this research and writing this thesis has been akin to reaching terminal velocity when sky diving. Throughout the last three years the pace, stress and work-load has slowly accelerated until reaching their maximum over the most recent few months. Despite moments where I have pushed to go faster and work harder, the need to prioritise my own wellbeing, which was supported by those around me, has been a welcome force that has pushed me to keep a manageable and sustainable pace. At times when I have been eager to set my feet on the ground, with the sky dive complete, I have been reminded to enjoy the views along the way. As I am writing this statement, I am where I was so eager to be, reflecting back on the process of the research with my feet firmly on the ground. However, it is with a heavy heart that I close the door on this chapter of my life and this project to which I have given so much of myself.

The Empirical Paper

The development of my research topic began with my undergraduate dissertation, which ignited my interest in neuropsychological testing. My project looked at developing a test of accelerated long term forgetting and I thoroughly enjoyed the topic and process of writing the report. Looking back, the development of my research skills, knowledge and approach since then, is significant. Despite enjoying my undergraduate project, it is clear to me that I have been able to immerse myself in the current research in a different way. Particularly gaining a larger sense of ownership over the project and, in turn, I have gained so much in terms of passion, knowledge and skills.

The beginning ideas behind the project were presented at the research fair and I immediately connected with the ideas, due to the similarities I drew with my undergraduate project. Since then, I have been able to make the project my own by developing the methodologies, procedure and analysis to best answer the question posited. I have always liked a conundrum, and I have felt the journey, since I undertook the research, has been about solving a puzzle to find the best possible version of the picture- that is, finding the best possible method and analysis to best address the question in hand. My family has often enjoyed puzzles themselves,

my Mum and Grandmother, Gill, often enjoying a jigsaw, my Grandad, Roy, commonly tackling Crosswords and Sudoku's and my sister enjoying reading and creating her own puzzles when writing stories. I often wondered about the draw to puzzles and why a quick 'google' of the answer did not give the same sense of achievement! However, on reflection of the process of this research, the parts in which I have learnt the most are the parts that caused the most frustration. These are the parts that took the most time, and that caused me to go to bed with my mind churning, trying to find the answers. An example of this is the data analysis section. Whilst there was an option to take the easy road, which was following word for word the process from previous studies without fully understanding the reasons why, this did not give me a sense of the puzzle being solved. Instead, I took the road that questioned each part of the analysis and required me to read around analysis techniques, the strengths and limitations, to decide on the best course of action and try a new approach.

One such decision was the use of 'stepwise' regression models, which have been commonly used in previous studies looking at developing premorbid functioning models. This is a method that tries each variable and excludes non-significant predictors in the final model. It is limited as the final model is not always the best possible model, and statisticians have broadly criticised the method. Additionally, variables can have a mediating effect on other variables within the model, thus, excluding them, can be limiting. Despite initially looking to follow this method, reading around the analysis suggested that it was not the most robust, or statistically sound, method to use. I then looked for an alternative method, and, after spending a time struggling to make sense of the complicated statistical alternatives offered by statisticians within several articles, I finally settled on inputting variables that were empirically supported to have a relationship, and leave non-significant predictors within the model. After struggling with the feeling of not-knowing, this method felt like it had a clear rationale that I understood and with which I agreed.

A further challenging decision was the use of Leave-One-Out Cross Validation. This was a method I came across after extensive research following a period of anxiety about the limitations of my conclusions; due to

not having a large enough sample size to test the models on independent data. This was a difficult decision to make, as I had to learn about the method and consider its applicability to my research question, as previous studies within the field had not used this approach. Additionally, it is completed on the program R which requires a basic level of coding which I had never done before, except an introductory module in my undergraduate degree. Whilst there were anxieties, once again I had a feeling that this was the right course of action and would be the best way to use and present my data. On reflection, the time taken to learn and complete this analysis was worth every second as it provided further evidence for my conclusions.

Ultimately, the statistics that were chosen were the product of careful consideration and time spent tolerating 'not knowing'. It was this resilience of trying to understand, staying curious and researching that taught me the most during this process. This is something that I also connected with, in terms of my clinical work and a something I will take forward into my post-qualified work. It is important to stay curious when things feel complicated in a therapy session and to work with the client to understand their difficulties together. It can sometimes be tempting to take a simpler approach- for instance, using a single term to define the struggles of an individual, as in diagnosis. However, the process of understanding the individual, whilst at times frustrating, can be the most rewarding for both the client and therapist, and teach us the most about ourselves.

A particular highlight during the research was completing the process of collecting data. I was overcome with the generosity of those around me, who, without recompense, reached out to help with my recruitment. When collecting data, I anticipated that, due to the nature of the assessments, people would begrudgingly take part. However, I was welcomed into peoples' homes, given home cooked brownies and was delighted with discussions about peoples' lives. Following testing, commonly, there was excited talk about the parts they found hard or easy. In most cases, I was left with the phrase "I know someone else who would like to do this study!". In turn, I was able to quickly achieve the desired sample size for my normative sample which is attributed to people taking time to assist me.

On reflection, I think that my cynicism going into participant recruitment was a product of my anxiety about the research. It may have, also, stemmed from the narratives in the news that inherently depict the negativity in the world and commonly overlook simple acts of kindness. Similarly, I think sometimes, as therapists, we often hear difficult stories, and are trained to look for the mental health 'difficulties' or the 'presenting problems'. In light of the growing movement of Positive Psychology, there is a need to bolster the true strengths that people show in response adverse situations and question the negative bias that exists within our society. This process has reminded me of the negative bias to which we can be susceptible, and that it is important not to underestimate the kindness and strength within humans.

In terms of my clinical sample, there were several challenges to recruitment. Firstly, the process of gaining ethical approval took a significant amount of time which was further delayed due to covid-19 related absences and reduced staffing. Additionally, in order to prioritise the wellbeing and safety of the clinical participants, and reducing the impact on services, I chose to extract data following the data collection which took place within services during routine neuropsychological assessment. The limitation to this, however, was the lack of control that I had about the rate of data collection which relied on patients coming to the services, who would be routinely assessed with the RBANS and TOPF^{UK}. In both cases, I had to manage with a reduced level of control over the research, and how quickly it progressed. This was hard to sit with, and caused anxiety and worry. However, it was also helpful to learn how to manage this, place trust in my colleagues, and use any time that I had waiting, productively. Despite having a smaller sample than I would have hoped, mainly due to time constraints, all of the sites helping me with recruitment, were extremely forthcoming and helpful, and I will proceed to collect more data in the future prior to attempting to publish.

Finally, the process of writing up the study caused a mixture of feelings. On the one hand, it was exciting to see the work coming together. On the other hand, particularly when writing my discussion, I had the anxiety

that I would be unable to do justice to the time and work I had put into the project. Despite this, once I started writing, I realised that 'good enough' was enough and that my work would show for itself.

Systematic Literature Review

The process of writing the literature review also had its challenges. A particular challenge was deciding on a question. I had settled on a topic quite early on, however, after getting quite far through the systematic search, I felt that it would be better to change topic to one more closely linked with my empirical paper. The question I chose felt quite different as it looked more at the *state* of the literature rather than addressing a psychological theory or question. It, also, consisted of only quantitative papers. This brought challenge, as the traditional methods such as thematic analysis, were not appropriate. Despite finding a methodology, I held a real concern all the way through the project process about whether I was doing it 'correctly'. This was only eased when I was writing my discussion, looking back at the work I had completed, and realising how the study can inform further research. I realised that there is no 'right' way to do research and that the important thing is to have a sound rationale behind all decisions. The heterogeneity of literature review styles that are published is broad, and thus, I hope that my literature review can contribute in its own way to the evidence base.

On writing up my literature review, I reflected on the hours that I had put into it. I initially had underestimated the time and thought needed, and the complexities of writing it. Despite this, I really enjoyed the process. Particularly, reading other people's research that directly linked to my own empirical study. The thoughts and considerations prompted by these papers were invaluable to informing my own empirical paper. This confirmed to me the true value of literature reviews.

Final Thoughts

Writing this thesis has been an experience that one would not forget easily, nor that I would wish to. It has been a unique experience that has taught me a huge amount. It has brought me a lot of stress, but also a lot of joy. As a final thought, my Grandad, Roy (to whom this thesis is dedicated) once said *"If you've a chance to do some good, don't put it off, just do it".* I have held this at my core during this research, particularly when I felt like giving up, with the hope that the result of this thesis will be to assist with positive change in some way.

Appendix B. Epistemological Statement

The purpose of this statement is to outline the ontological and epistemological positions that were taken during the conceptualisation of this research, and the process of conducting the research and analysing the data. The transparency of this is clear to allow for readers to understand how the data has been understood and how this may have been shaped.

Epistemology is concerned with what we define as knowledge and what the nature of this knowledge is considered to be (Cohen et al., 2007). At one pole of this continuum is the interpretivist orientation that posits that reality is a product of power imbalances within society and that knowledge is constructed by the interplay of these social dynamics. Saunders et al. (2007) described interpretivist as an "epistemology that it is necessary for the researcher to understand differences between humans in our role as social actors." (p106). At the other end is the positivist stance which describes an 'absolute truth' that can be discovered, measured and proven. It considers these social factors to be independent from the beliefs of individuals (Bahari et al., 2010).

Ontology, on the other hand, is defined as a theory of the nature of social entities (Bryman, 2004). Easterby-Smith et al. (2002) describe ontology to be assumptions that we make about the nature of reality. Subjectivism lends itself to qualitative research as it is focused on the idea that understanding is based upon social interactions and beliefs cultivated through these. On the other hand, objectivism discusses causation in social processes and objectivity within the social world. It posits that meaning has an element of independence from social interactions.

The goals of the presented research are guided by the epistemological and ontological stance of the researcher. The very nature of neuropsychological testing lends itself to the positivist epistemological stance as it aims to quantify and measure individual's IQ as a true value. This is also true of the idea of

predicting premorbid functioning which is based on a belief about causal relationships between constituent elements in the world. Similarly, the objectivity stance is in line with this idea due to accepting that IQ can be predicted using relationships that are independent from social understanding.

Whilst this is true, it is important to consider the limits to these ideas. Particularly that 'deficits' are relative to the societal norms with which they are compared. Deficits are only considered so, due to a shared understanding within a population. Additionally, heterogeneity, between people, calls into questioning the power of predictive models and their generalisability to all people. Thus, this researcher aligns more with a post-positivist stance. The post-positivist stance understands that whilst an absolute truth may be in existence, theory and research is limited in its ability to identify and prove this. In this way, possibility of a theory being disproven or improved is always kept in mind.

References

Bahari, S. F. (2010). Qualitative versus quantitative research strategies: contrasting epistemological and ontological assumptions. *Sains Humanika*, *52*(1).

Cohen, D. H. (2007). Virtue epistemology and argumentation theory. *Hansen HV (ed) Dissensus and the search for common ground.* OSSA, Windsor, 1–9

Saunders, M., P. Lewis, and A. Thornhill. (2007). *Research Methods for Business Students.* Prentice Hall: London.

Bryman, A. 2004. *Social Research Methods*. Second Edition. London: Oxford University Press.

Easterby-Smith, M., R. Thorpe, and A. Lowe. 2002. *Management Research an Introduction*. Thousand Oaks: Sage Publications

Appendix C. Submission Guidelines for Neuropsychology Review

Title Page

Title Page

Please make sure your title page contains the following information.

Title

The title should be concise and informative.

Author information

- The name(s) of the author(s)
- The affiliation(s) of the author(s), i.e. institution, (department), city, (state), country
- A clear indication and an active e-mail address of the corresponding author
- If available, the 16-digit ORCID of the author(s)

If address information is provided with the affiliation(s) it will also be published.

For authors that are (temporarily) unaffiliated we will only capture their city and country of residence, not their e-mail address unless specifically requested.

Abstract

Please provide an abstract of 150 to 250 words. The abstract should not contain any undefined abbreviations or unspecified references.

For life science journals only (when applicable)

- Trial registration number and date of registration for prospectively registered trials
- Trial registration number and date of registration, followed by “retrospectively registered”, for retrospectively registered trials

Keywords

Please provide 4 to 6 keywords which can be used for indexing purposes.

Statements and Declarations

The following statements should be included under the heading "Statements and Declarations" for inclusion in the published paper. Please note that submissions that do not include relevant declarations will be returned as incomplete.

- **Competing Interests:** Authors are required to disclose financial or non-financial interests that are directly or indirectly related to the work submitted for publication. Please refer to “Competing Interests and Funding” below for more information on how to complete this section.

Please see the relevant sections in the submission guidelines for further information as well as various examples of wording. Please revise/customize the sample statements according to your own needs.

Text

Text Formatting

Manuscripts should be submitted in Word.

- Use a normal, plain font (e.g., 10-point Times Roman) for text.
- Use italics for emphasis.
- Use the automatic page numbering function to number the pages.
- Do not use field functions.
- Use tab stops or other commands for indents, not the space bar.
- Use the table function, not spreadsheets, to make tables.
- Use the equation editor or MathType for equations.
- Save your file in docx format (Word 2007 or higher) or doc format (older Word versions).

Headings

Please use no more than three levels of displayed headings.

Abbreviations

Abbreviations should be defined at first mention and used consistently thereafter.

Footnotes

Footnotes can be used to give additional information, which may include the citation of a reference included in the reference list. They should not consist solely of a reference citation, and they should never include the bibliographic details of a reference. They should also not contain any figures or tables.

Footnotes to the text are numbered consecutively; those to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data). Footnotes to the title or the authors of the article are not given reference symbols.

Always use footnotes instead of endnotes.

Acknowledgments

Acknowledgments of people, grants, funds, etc. should be placed in a separate section on the title page. The names of funding organizations should be written in full.

References

Citation

Cite references in the text by name and year in parentheses. Some examples:

- Negotiation research spans many disciplines (Thompson, 1990).
- This result was later contradicted by Becker and Seligman (1996).
- This effect has been widely studied (Abbott, 1991; Barakat et al., 1995; Kelso & Smith, 1998; Medvec et al., 1999).

Authors are encouraged to follow official APA version 7 guidelines on the number of authors included in reference list entries (i.e., include all authors up to 20; for larger groups, give the first 19 names followed by an ellipsis and the final author's name). However, if authors shorten the author group by using et al., this will be retained.

Reference list

The list of references should only include works that are cited in the text and that have been published or accepted for publication. Personal communications and unpublished works should only be mentioned in the text.

Reference list entries should be alphabetized by the last names of the first author of each work.

Journal names and book titles should be ***italicized***.

If available, please always include DOIs as full DOI links in your reference list (e.g. “https://doi.org/abc”).

- Journal article Grady, J. S., Her, M., Moreno, G., Perez, C., & Yelinek, J. (2019). Emotions in storybooks: A comparison of storybooks that represent ethnic and racial groups in the United States. ***Psychology of Popular Media Culture, 8***(3), 207–217. https://doi.org/10.1037/ppm0000185
- Article by DOI Hong, I., Knox, S., Pryor, L., Mroz, T. M., Graham, J., Shields, M. F., & Reistetter, T. A. (2020). Is referral to home health rehabilitation following inpatient rehabilitation facility associated with 90-day hospital readmission for adult patients with stroke? ***American Journal of Physical Medicine & Rehabilitation***. Advance online publication. https://doi.org/10.1097/PHM.0000000000001435
- Book Sapolsky, R. M. (2017). ***Behave: The biology of humans at our best and worst***. Penguin Books.
- Book chapter Dillard, J. P. (2020). Currents in the study of persuasion. In M. B. Oliver, A. A. Raney, & J. Bryant (Eds.), ***Media effects: Advances in theory and research*** (4th ed., pp. 115–129). Routledge.
- Online document Fagan, J. (2019, March 25). ***Nursing clinical brain***. OER Commons. Retrieved January 7, 2020, from https://www.oercommons.org/authoring/53029-nursing-clinical-brain/view

Tables

- All tables are to be numbered using Arabic numerals.
- Tables should always be cited in text in consecutive numerical order.
- For each table, please supply a table caption (title) explaining the components of the table.
- Identify any previously published material by giving the original source in the form of a reference at the end of the table caption.
- Footnotes to tables should be indicated by superscript lower-case letters (or asterisks for significance values and other statistical data) and included beneath the table body.

Appendix D. Adapted AXIS tool (Downes et al., 2016)

1	Were the aims/objectives of the study clear?
2	Was the study design appropriate for the stated aim (s)?
3	Was the sample size justified?
4	Was the target/reference population clearly defined? (Is it clear who the research was about?)
5	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?
6	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?
7	Were measures undertaken to address and categorise non-responders?
8	Were the outcome variables measured appropriate to the aims of the study?
9	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. P-values, confidence intervals)
10	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?
11	Were the basic data adequately described?
12	Does the response rate raise concerns about non-response bias? (n=1, y=0)
13	If appropriate, was information about non-responders/exclusions described?
14	Were the results internally consistent?
15	Were the results presented for all the analyses described in the methods?
16	Were the authors' discussions and conclusions justified by the results?
17	Were the limitations of the study discussed?
18	Were there any funding sources of conflicts of interest that may affect the authors' interpretation of the results? (n=1, y=0)
19	Was ethical approval or consent of participants attained?

Apart from where indicated (Questions 12, 18) all responses were scored: Yes = 1; No= 0
For questions 12 and 18 responses were scored: No=1; Yes =0
Qualitative descriptors were defined as: 0-6 low, 7-13 moderate, 14-19 high

Appendix E. Quality Assessment

Study	Were the aims/objectives of the study clear?	Was the study design appropriate for the stated aim (s)?	Was the sample size justified?	Was the target/reference population clearly defined? (Is it clear who the research was about?)	Was the sample frame taken from an appropriate population base so that it closely represented the target/reference population under investigation?	Was the selection process likely to select subjects/participants that were representative of the target/reference population under investigation?	Were measures undertaken to address and categorise non-responders?	Were the outcome variables measured appropriate to the aims of the study?	Is it clear what was used to determined statistical significance and/or precision estimates? (e.g. P-values, confidence intervals	Were the methods (including statistical methods) sufficiently described to enable them to be repeated?	Were the basic data adequately described?	Does the response rate raise concerns about non-response bias? (n=1, y=0)	If appropriate, was information about non-responders/exclusions described?	Were the results internally consistent?	Were the results presented for all the analyses described in the methods?	Were the authors' discussions and conclusions justified by the results?	Were the limitations of the study discussed?	Were there any funding sources of conflicts of interest that may affect the authors' interpretation of the results? (N=1, Y=0)	Was ethical approval or consent of participants attained?	Score	Qualitative description
Al-Ghatani et al., 2011	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	1	15	High
Almkvist et al., 2007	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	ND	15	Moderate

Alves et al., 2012	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1	1	1	0	ND	13	Moderate
Chaurasiya et al., 2022	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	16	High
Chen et al., 2012	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	ND	14	Moderate
Colombo et al., 2002	1	1	0	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	ND	14	High
Del Ser. 1997	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1	ND	14	High
Gomar et al., 2011	1	1	0	1	1	1	0	1	1	0	0	1	0	1	1	0	0	1	ND	11	Moderate
Isella et al., 2005	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	ND	15	High
Karakula-Juchnowicz & Stecka, 2017	1	1	0	1	1	1	0	1	1	1	1	0	0	1	0	1	0	1	ND	12	Moderate
Kim et al., 2015	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	ND	15	High
Krueger et al., 2007	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND	18	High
Matsuoka et al., 2006	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	ND	1	17	High
Pluck, 2017	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	1	16	High
Pluck, 2021	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	0	ND	14	High
Rolstad et al., 2008	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	ND	ND	14	High

Sanjurjo et al., 2015	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	1	ND	16	High
Sarrao et al., 2015	1	1	0	1	1	1	0	1	1	0	1	1	0	1	0	1	1	1	1	14	High
Shrauf et al., 2005	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	18	High
Yi et al.,2017	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19	High

Note: Qualitative descriptors scored as 0-6 low, 7-13 moderate, 14-19 high
Abbreviations: ND, Not disclosed

Appendix F: Formatting guidelines for The Clinical Neuropsychologist.

About the Journal

The Clinical Neuropsychologist 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.

The Clinical Neuropsychologist accepts the following types of article: Original Articles, Review Articles, Grand Rounds Articles, Book Reviews..

Authors are strongly encouraged to consult the TCN reporting guidelines checklist when preparing or editing their manuscript. Gross disregard for the reporting guidelines could result in the manuscript being returned without a review.

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 95% more citations* and over 7 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 if you would like more information about our Open Select Program.

*Citations received up to 9th June 2021 for articles published in 2016-2020 in journals listed in Web of Science®. Data obtained on 9th June 2021, from Digital Science's Dimensions platform, available at https://app.dimensions.ai
**Usage in 2018-2020 for articles published in 2016-2020.

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. If you have shared an earlier version of your Author's Original Manuscript on a preprint server, please be aware that anonymity cannot be guaranteed. Further information on our preprints policy and citation requirements can be found on our Preprints Author Services page. Find out more about what to expect during peer review and read our guidance on publishing ethics.

Preparing Your Paper

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 all listed authors meet the Taylor & Francis authorship criteria. 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 a structured abstract of 250 words.

   A structured abstract should cover (in the following order):
   Objective: A brief statement of the purpose of the study.
   Method: A summary of the participants as well as descriptions of the study design, procedures, and specific key measures, to the extent that space allows.
   Results: A summary of the key findings.
   Conclusions: Clinical and theoretical implications of the findings.
   NOTE: If your manuscript is a critical review or a commentary, you can omit the Results portion of the abstract. However, retain that portion for systematic reviews and meta-analyses. Read tips on writing your abstract.

   Read tips on writing your abstract.

3. **Graphical abstract** (optional). This is an image to give readers a clear idea of the content of your article. For the optimal online display, your image should be supplied in landscape format with a 2:1 aspect ratio (2 length x 1 height). Graphical abstracts will often be displayed online at a width of 525px, therefore please ensure your image is legible at this size. Save the graphical abstract as a .jpg, .png, or .tiff. Please do not embed it in the manuscript file but save it as a separate file, labelled GraphicalAbstract1.
4. 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.
5. Between 5 and 10 **keywords**. Read making your article more discoverable, including information on choosing a title and search engine optimization.
6. **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].

7. **Disclosure statement.** This is to acknowledge any financial or non-financial interest that has arisen from the direct applications of your research. If there are no relevant competing interests to declare please state this within the article, for example: *The authors report there are no competing interests to declare*. Further guidance on what is a conflict of interest and how to disclose it.
8. **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.
9. **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.
10. **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.
11. **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.
12. **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.
13. **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.
14. **Equations.** If you are submitting your manuscript as a Word document, please ensure that equations are editable. More information about mathematical symbols and equations.
15. **Units.** Please use SI units (non-italicized).

Using Third-Party Material

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.

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 *The Clinical Neuropsychologist* uses Crossref™ to screen papers for unoriginal material. By submitting your paper to *The Clinical Neuropsychologist* 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.

Appendix G: SOC 2020 Major Categories and Skill Levels

Major Categories:

Major group	General nature of qualifications, training and experience for occupations in the major group
1 Managers, directors and senior officials	A significant amount of knowledge and experience of the production processes and service requirements associated with the efficient functioning of organisations and businesses.
2 Professional occupations	A degree or equivalent qualification, with some occupations requiring postgraduate qualifications and/or a formal period of experience-related training.
3 Associate professional occupations	An associated high-level vocational qualification, often involving a substantial period of full-time training or further study. Some additional task-related training is usually provided through a formal period of induction.
4 Administrative and secretarial occupations	A good standard of general education. Certain occupations will require further additional vocational training to a well-defined standard (e.g. office skills).
5 Skilled trades occupations	A substantial period of training, often provided by means of a work based training programme.
6 Caring, leisure and other service occupations	A good standard of general education. Certain occupations will require further additional vocational training, often provided by means of a work-based training programme.
7 Sales and customer service occupations	A general education and a programme of work-based training related to sales procedures. Some occupations require additional specific technical knowledge but are included in this major group because the primary task involves selling.
8 Process, plant and machine operatives	The knowledge and experience necessary to operate vehicles and other mobile and stationary machinery, to operate and monitor industrial plant and equipment, to assemble products from component parts according to strict rules and procedures and subject assembled parts to routine tests. Most occupations in this major group will specify a minimum standard of competence for associated tasks and will have a related period of formal training.
9 Elementary occupations	Occupations classified at this level will usually require a minimum general level of education (i.e. that which is acquired by the end of the period of compulsory education). Some occupations at this level will also have short periods of work-related training in areas such as health and safety, food hygiene, and customer service requirements.

Skill Level Definitions

The first skill level equates with the competence associated with a general education, usually acquired by the time a person completes his/her compulsory education and signalled via a satisfactory set of school-leaving examination grades. Competent performance of jobs classified at this level will also involve knowledge of appropriate health and safety regulations and may require short periods of work-related training. Examples of occupations defined at this skill level within the SOC 2020 include postal workers, hotel porters, cleaners and catering assistants.

The second skill level covers a large group of occupations, all of which require the knowledge provided via a good general education as for occupations at the first skill level, but which typically have a longer period of work-related training or work experience. Occupations classified at this level include machine operation, driving, caring occupations, retailing, and clerical and secretarial occupations.

The third skill level applies to occupations that normally require a body of knowledge associated with a period of post-compulsory education but not normally to degree level. Several technical occupations fall into this category, as do a variety of trades occupations and proprietors of small businesses. In the latter case, educational qualifications at sub-degree level or a lengthy period of vocational training may not be a prerequisite for competent performance of tasks, but a significant period of work experience is typical.

The fourth skill level relates to what are termed “professional” occupations and high-level managerial positions in corporate enterprises, or national or local government. Occupations at this level normally require a degree or equivalent period of relevant work experience.

Taken from SOC 2020 Volume 1: structure and descriptions of unit groups - Office for National Statistics

Appendix H: Letter detailing permission from Pearson for online testing

Pearson Education Limited
Clinical Assessments
80 Strand
London WC2R 0RL
PearsonClinical.co.uk

24 August 2020

To Whom It May Concern,

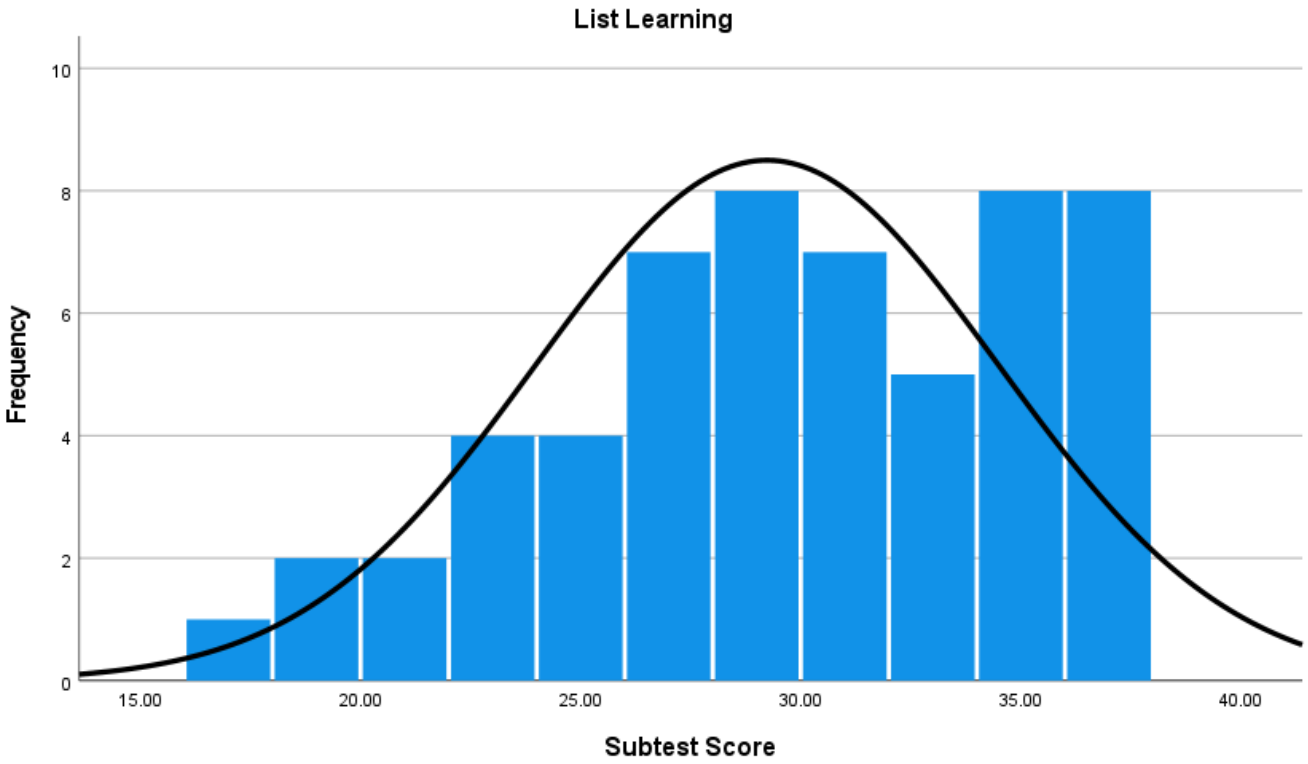
Due to restrictions in travel and personal interaction associated with the COVID-19 Pandemic and to support expanded capabilities for all service providers in their delivery of tele-health services, Pearson Clinical Assessment is issuing this letter of No Objection to permit the following use of our copyrighted materials:

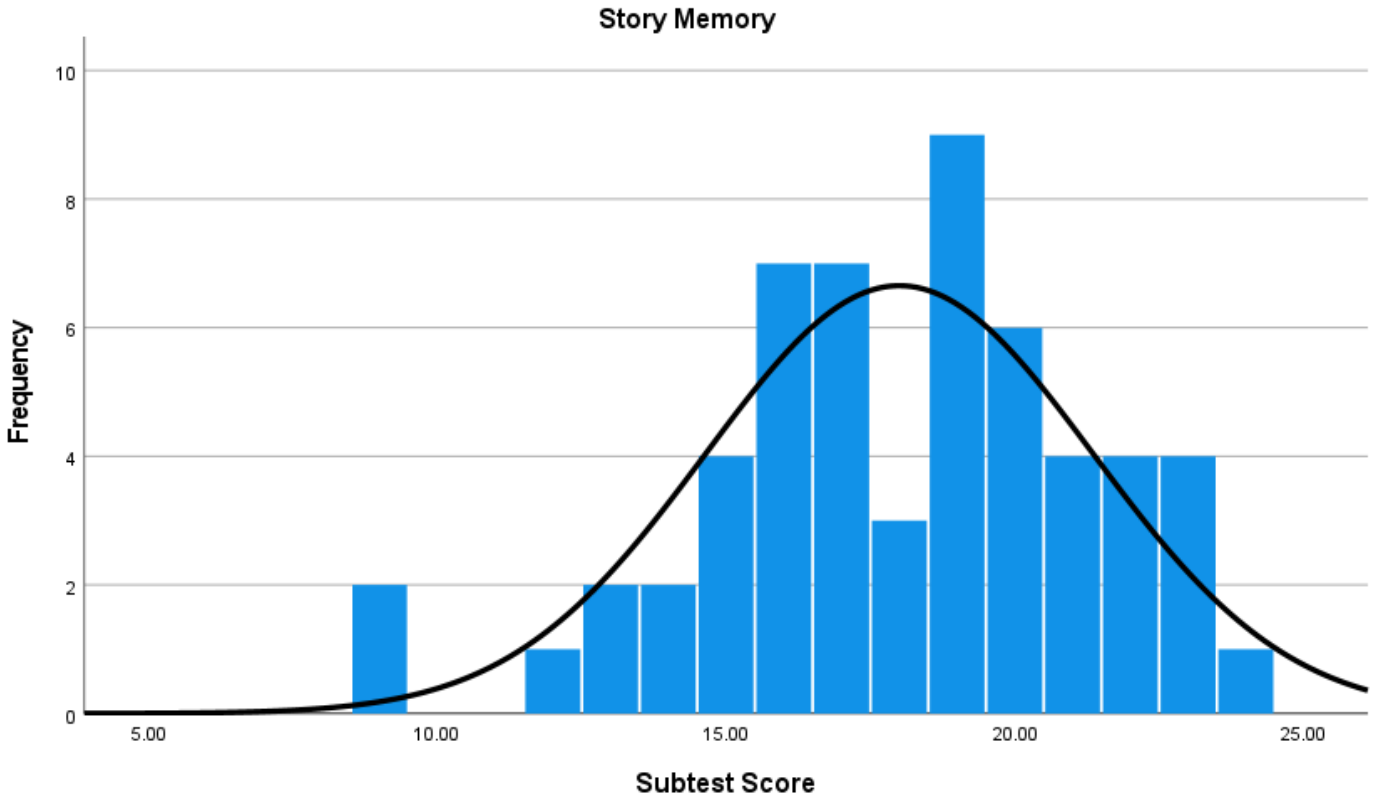
- Qualified professionals are hereby granted a limited and revocable permission to utilise appropriate non-public facing teleconferencing software and tools to assist in the remote administration of Pearson assessment content, consistent with this letter. Before test administration, the qualified professional must obtain documented agreement from the examinee that the session will not be recorded, reproduced or published, and that copies of the materials will not be made. Further, the qualified professional may not utilise recording capabilities to record live test administrations.
- This permission is intended to include the use of non-public facing screen-mirroring and screen share methods to remotely share test item content with examinees on a computer screen and capture responses either verbally or through other means.
- This permission is not intended to allow for use of photocopying, scanning, or duplication of test protocols, including any screen capture or session recording technology, but is merely intended to support practical live delivery of tele-health services.
- This permission is also not intended to allow for any modification to the original test content as it currently appears. Any use of original test content must include copyright notices as contained in the materials and show the copyright notice when displayed. Further requests to modify, translate, digitise, or alter test content should continue to follow Pearson's standard procedure for Licencing and Permissions requests located **here**.
- Pearson reserves the right to revoke this Limited Permission at any time in the interest of test content security and/or test validity and will review the ongoing status of this No Objection Letter as conditions require.

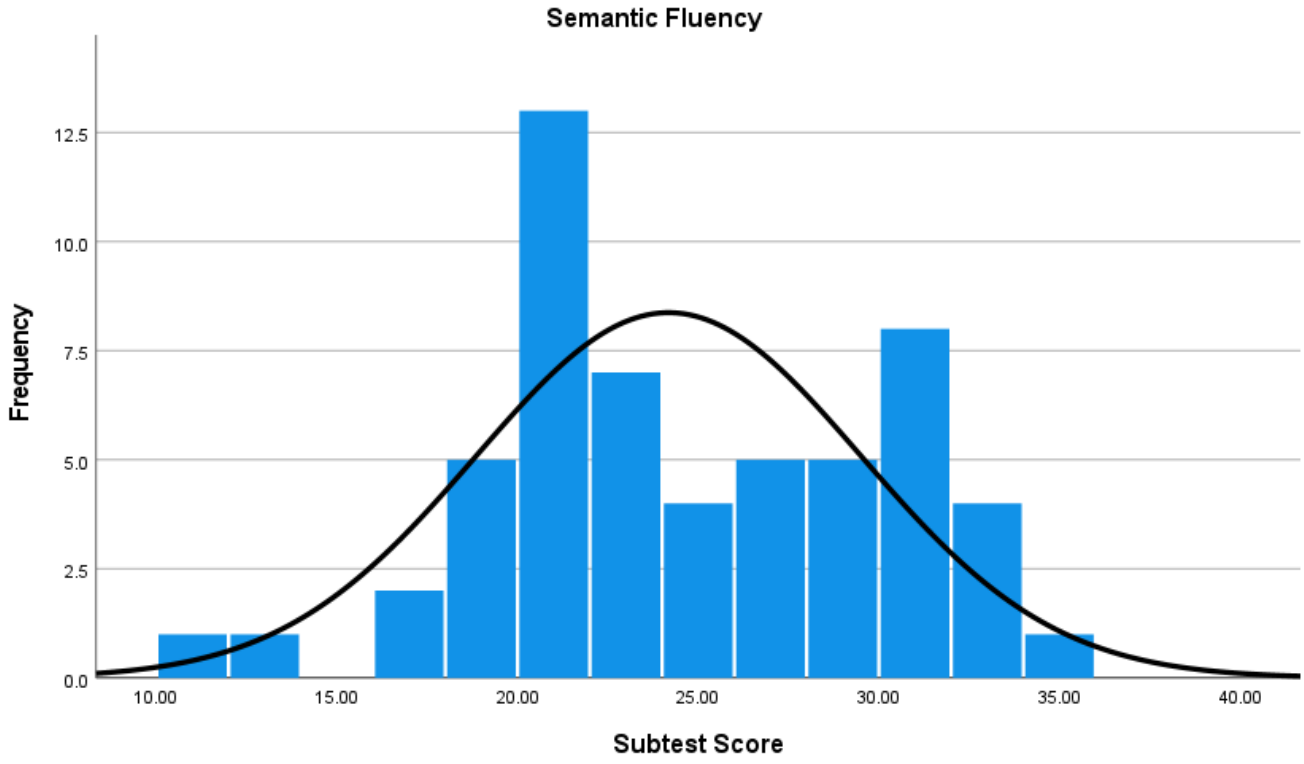
Thank you for your critical role in delivering essential services to your community. Pearson Clinical Assessments is equally committed to supporting your efforts to continue service however we can. Please let us know if you have additional questions on this topic or if we can help you in any other ways.

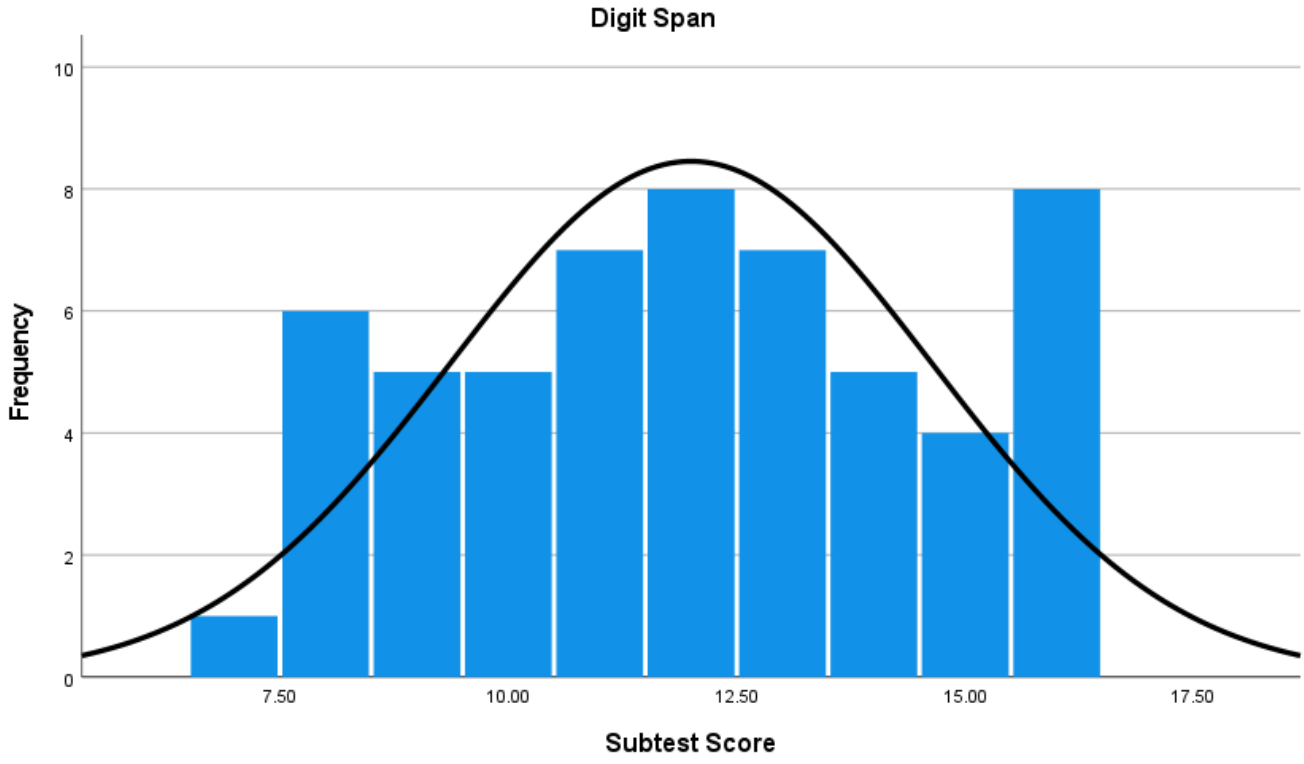
Registered office: 80 Strand, London WC2R 0RL
Pearson Education Limited is registered in England and Wales with company number 872828

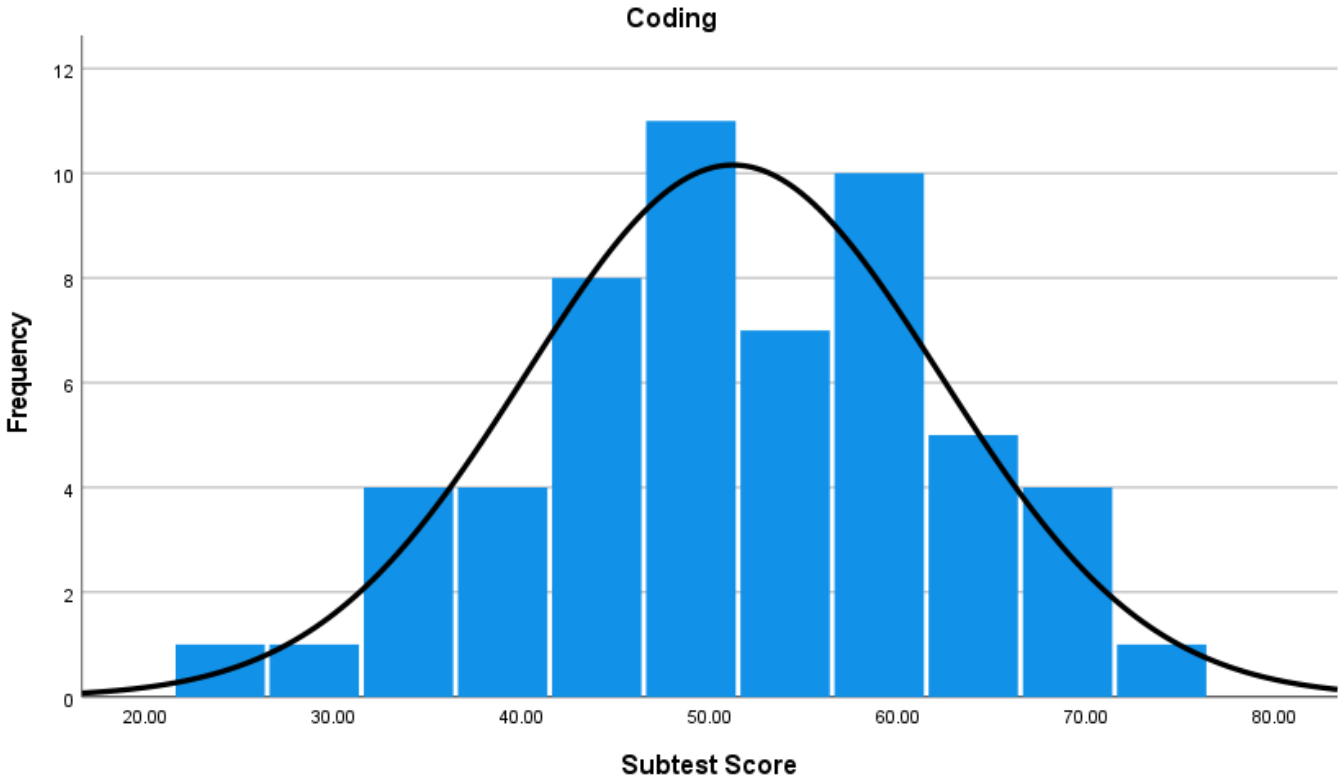
Appendix I. Histograms plotting frequency against subtest score

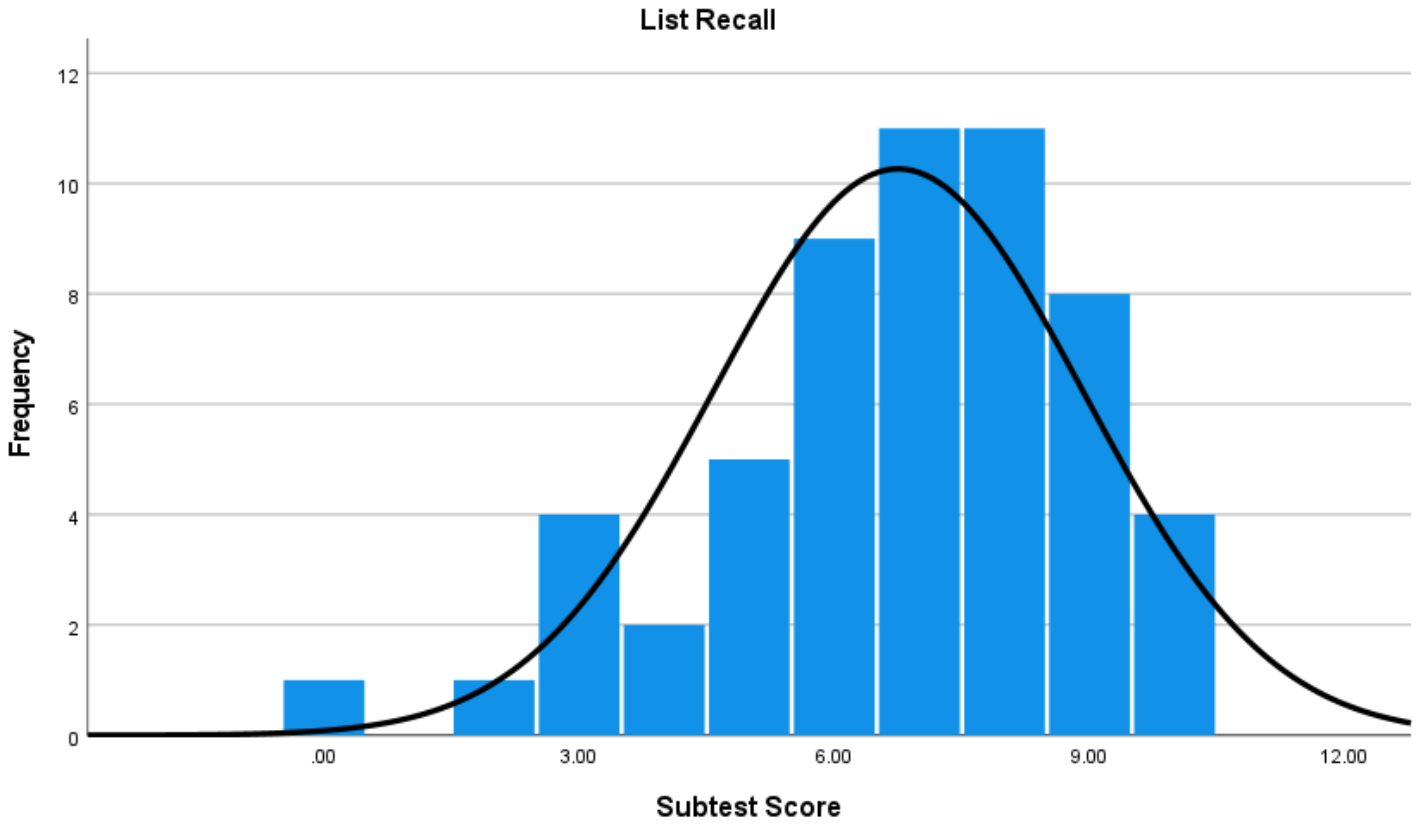




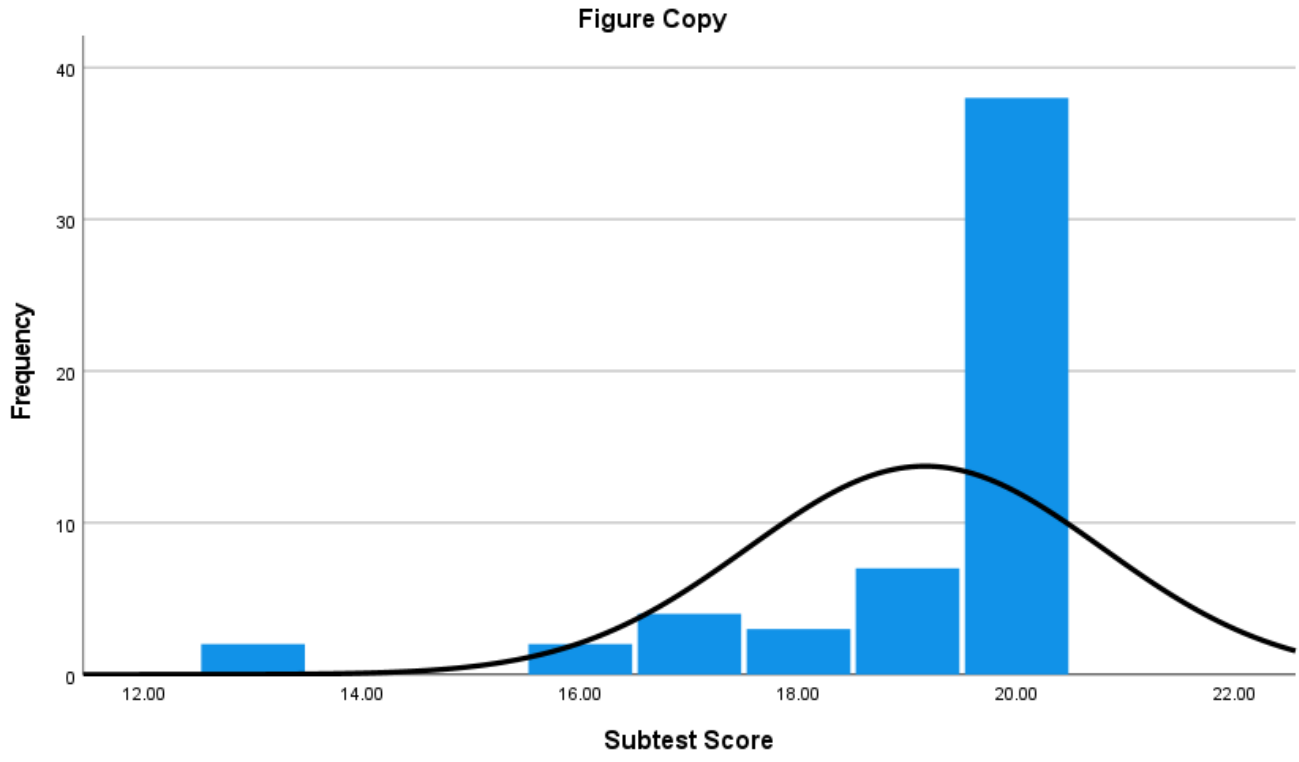

Semantic Fluency
Frequency
12.5
10.0
7.5
5.0
2.5
0.0
10.00
15.00
20.00
25.00
30.00
35.00
40.00
Subtest Score

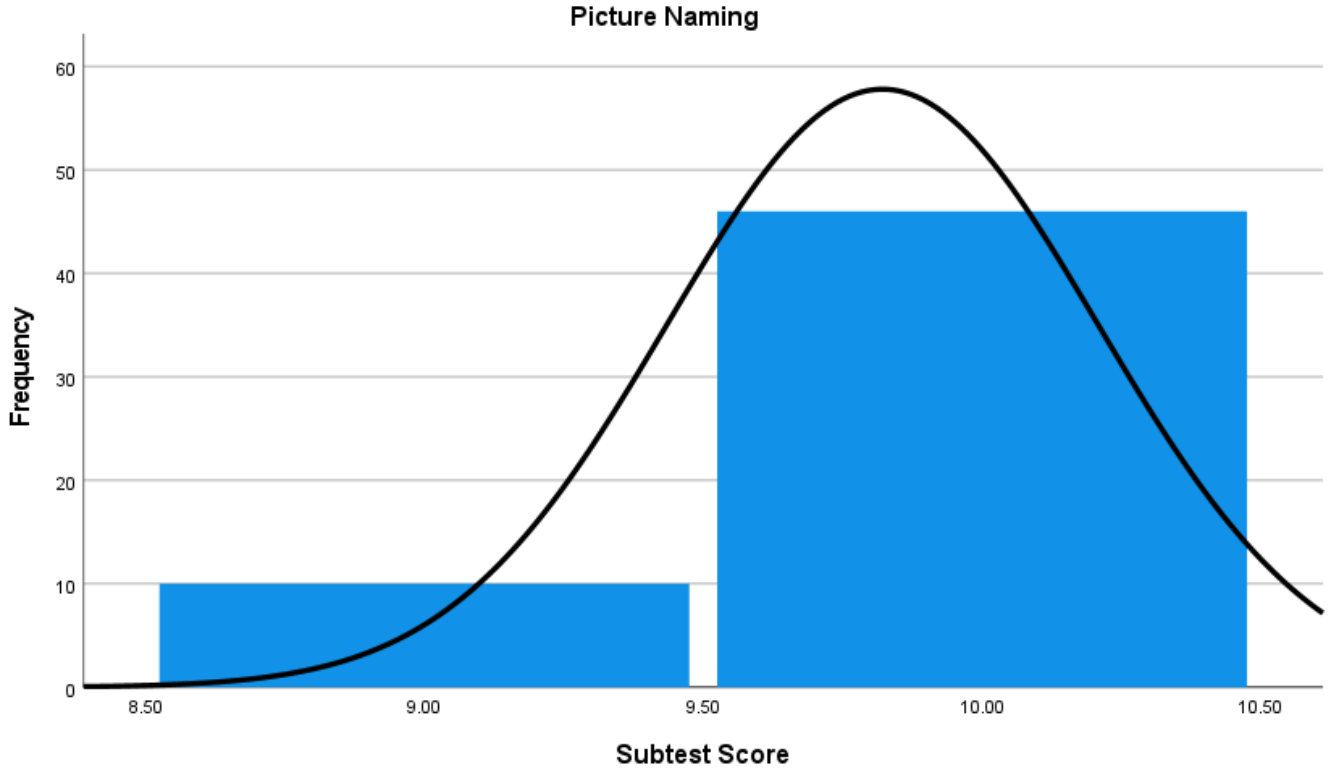

Digit Span
Frequency
10
8
6
4
2
0
7.50
10.00
12.50
15.00
17.50
Subtest Score

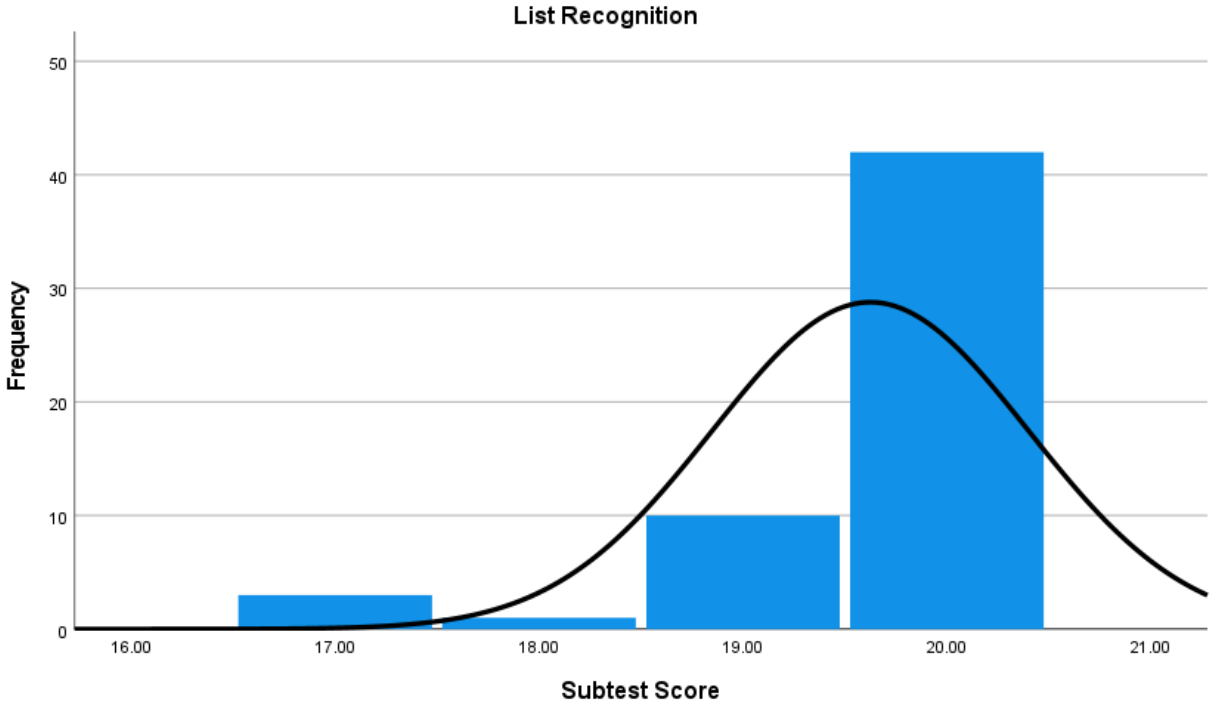

Coding
Frequency
12
10
8
6
4
2
0
20.00
30.00
40.00
50.00
60.00
70.00
80.00
Subtest Score


List Recall
Frequency
12
10
8
6
4
2
0
.00
3.00
6.00
9.00
12.00
Subtest Score

Flagged for Ceiling effects and Negative Skew




Picture Naming
Frequency
60
50
40
30
20
10
0
8.50
9.00
9.50
10.00
10.50
Subtest Score


List Recognition
Frequency
50
40
30
20
10
0
16.00
17.00
18.00
19.00
20.00
21.00
Subtest Score

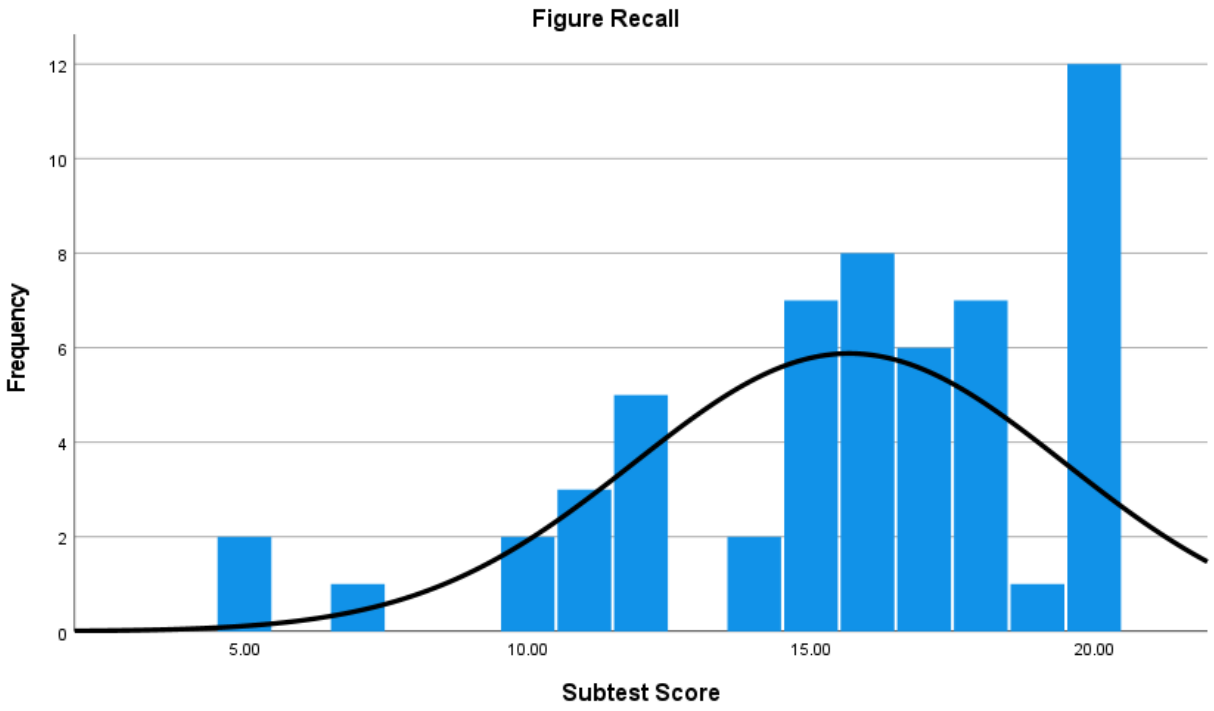
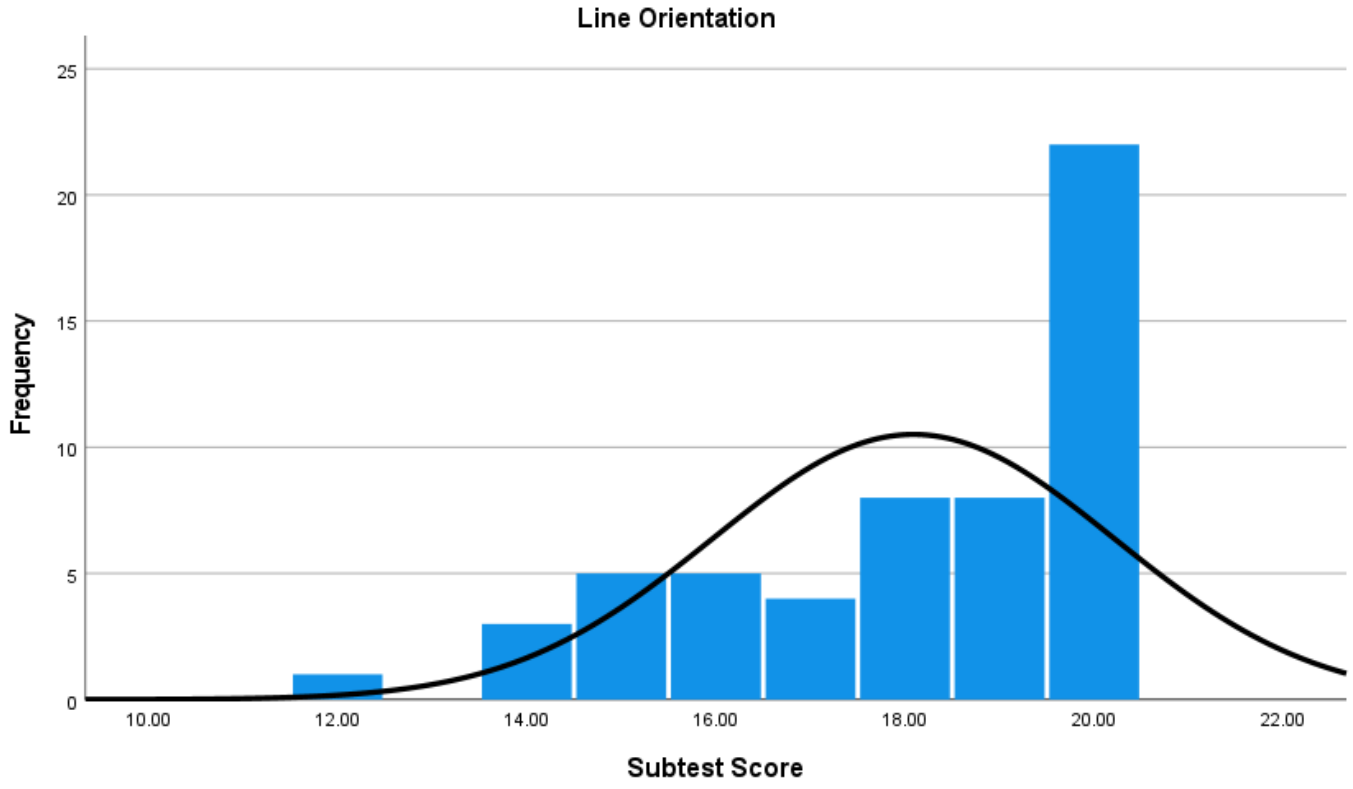


Figure Recall
Frequency
12
10
8
6
4
2
0
5.00
10.00
15.00
20.00
Subtest Score


Line Orientation
Frequency
25
20
15
10
5
0
10.00
12.00
14.00
16.00
18.00
20.00
22.00
Subtest Score

Appendix J. Study Information Sheet for Clinical Participants

Participant information sheet – Clinical Sample

This research is being completed as part of the requirements of the Doctorate in Clinical Psychology course at the University of Hull. The researcher, Hayley Gould, is a Trainee Clinical Psychologist and this study is part of her thesis project.

Title of study

An investigation into the predictability of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) using the Test of Premorbid Functioning (TOPF) and Demographic variables.

We would like to invite you to participate in this research which is investigating the ability of a neuropsychological test, which estimates pre-morbid functioning, and demographic variables to predict current cognitive functioning.

We are looking for two groups of people for this study:

1. People who have an acquired neurological condition or brain injury
2. People who have NOT had any neurological condition or brain injury

Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. If you have any questions, please use the contact details supplied at the end of this document to get in contact with us.

What is neuropsychological assessment?

If someone is noting that they are having difficulty with cognitive functions such as concentrating and making decisions, simple tests may be used to investigate whether there is anything wrong – these tests are called neuropsychological tests. These tests can look at a lot of different areas of cognitive functioning such as:

- Attention span
- Memory
- Motor function
- Problem-solving
- Verbal ability

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) is a brief neuropsychological test which looks at Memory, Language and Attention.

What is pre-morbid functioning and why is it important?

Pre-morbid functioning refers to the level of cognitive functioning prior to an event that may affect cognitive ability, such as a stroke or a head injury. It is important to have an idea of pre-morbid functioning to use as a baseline measurement in order to determine if there has been any decline from pre-morbid levels of functioning.

What is the purpose of the study?

Pre-morbid levels of cognitive functioning often have to be estimated. This is because cognitive ability is not often measured prior to a brain injury or cognitive impairment and therefore pre-morbid data is not available.

Reading tests such as the Test of Pre-morbid Functioning (TOPF) can be used to do predict pre-morbid functioning. Demographic variables, such as gender and age, can also be used.

To improve our ability to predict pre-morbid functioning, we can create equations that help clinicians to predict pre-morbid cognitive scores using reading tests such as the TOPF and demographic variables.

Currently there are no equations to predict pre-morbid RBANS scores using the TOPF and demographic variables. This research intends to develop an equation. This will help with neuropsychological assessment of cognitive decline when using the RBANS.

Why have I been asked to take part?

You have been identified by the service as needing a neuropsychological assessment. As part of this routine assessment, you will be tested using the RBANS and the TOPF.

What will happen if I agree to take part?

If you agree to take part in the research, you will be invited to take part in the routine assessment as usual. As part of this assessment, you will be asked to complete a short additional questionnaire. You will first be asked some short questions about yourself such as your gender, age and level of schooling. You will then be asked to complete a short questionnaire called the Hospital Anxiety and Depression Scale (HADS) this should take no longer than 5minutes.

Following this, the assessment will continue routinely and you will be tested using the RBANS and TOPF.

This data will then be anonymised and shared with the lead researcher to use in the research.

If you decide not to take part in the research, there will be no change to the care you receive. You will be invited to take part in the routine assessment as usual and no data will be shared with the researcher.

Your rights

- You do not have to take part
- You can withdraw from the study at any point without giving a reason

- You can contact the researcher via email and ask them to remove your data from the study within 72 hours of completing testing
- All your data will be kept safe and cannot be linked back to you
- You have a right to ask questions about the research before and after participating
- Participating or not participating will have no effect on your medical care

What are the possible risks of taking part?

Participating in the study will mean that your neuropsychological assessment may be up to 5minutes longer and you will be required to answer an additional questionnaire. Some people find cognitive assessments distressing if they struggle with the assessments. If you experience distress, you can withdraw from the research at any time. You can also contact the research supervisor or service lead with any concerns or worries.

What are the possible benefits of taking part?

We cannot promise that you will have any direct benefits from taking part in the study. However, it is hoped that the information you give us will mean that in the future, it will be easier for clinicians to assess cognitive decline.

What will happen to the results of the study?

The results of the study will be summarised in a written thesis as part of a Doctorate in Clinical Psychology. The thesis will be available on the University of Hull's on-line repository https://hydra.hull.ac.UK. The research may also be published in academic journals or presented at conferences. If you want to hear about the results of the study then do contact the researcher, Hayley Gould, who will be happy to provide you with a written summary of the research.

How will we use information about you?

We will need to use information from you for this research project. This information will include your:

- Name
- Contact details
- Date of birth
- Biological Sex
- Years of education and employment status

People will use this information to do the research or to check your records to make sure that the research is being done properly. People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead. We will keep all information about you safe and secure. Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study. The data controller for this project will be the University of Hull. The University will process your personal data for the purpose of the research outlined above. The legal basis for processing your personal data for research purposes under GDPR is a 'task in the public interest' You can provide your consent for the use of your personal data in this study by completing the consent form that has been provided to you**.** Information about how the University of Hull processes your data can be found at https://www.hull.ac.UK/choose-hull/university-and-region/key-documents/data-protection.aspx

You have the right to access information held about you. Your right of access can be exercised in accordance with the General Data Protection Regulation. You also have other rights including rights of correction, erasure, objection, and data portability. Questions, comments and requests about your personal data can also be sent to the University of Hull Information Compliance Manager (dataprotection@hull.ac.UK). If you wish to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.UK.

What are your choices about how your information is used?

You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have. We need to manage your records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about you.

Withdrawing from the study will not affect you in any way. Participant's data cannot be withdrawn from the study once the data has been anonymised and analysed. If you choose to withdraw from the study before this point the data collected will be destroyed. You have up to 72 hours after the completion of testing to withdraw your data from the research.

Where can you find out more about how your information is used?

You can find out more about how we use your information

- at www.hra.nhs.UK/information-about-patients/
- our leaflet available from www.hra.nhs.UK/patientdataandresearch
- by asking one of the research team
- by sending an email to

If you have any questions or require more information about this study, please contact me using the following contact details:

Personal details removed for publication

What if something goes wrong?

If you wish to make a complaint about the study, you can contact the University of Hull using the research supervisor's details below for further advice and information:

Personal details removed for publication

Thank you for reading this information sheet and for considering taking part in this research.

Appendix K. Non-Clinical Participants Information Sheet

Participant information sheet – Non-Clinical Group

This research is being completed as part of the requirements of the Doctorate in Clinical Psychology course at the University of Hull. The researcher, Hayley Gould, is a Trainee Clinical Psychologist and this study is part of her thesis project.

Title of study

An investigation into the predictability of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) using the Test of Premorbid Functioning (TOPF) and Demographic variables.

We would like to invite you to participate in this research which is investigating the ability of a neuropsychological test, which estimates pre-morbid functioning, and demographic variables to predict current cognitive functioning.

We are looking for two groups of people for this study:

1. People who have an acquired neurological condition or brain injury
2. People who have NOT had any neurological condition or brain injury

Before you decide whether you want to take part, it is important for you to understand why the research is being done and what your participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. If you have any questions, please use the contact details supplied at the end of this document to get in contact with us.

What is neuropsychological assessment?

If someone is noting that they are having difficulty with cognitive functions such as concentrating and making decisions, simple tests may be used to investigate whether there is anything wrong – these tests are called neuropsychological tests. These tests can look at a lot of different areas of cognitive functioning such as:

- Attention span
- Memory
- Motor function
- Problem-solving
- Verbal ability

The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) is a brief neuropsychological test which looks at Memory, Language and Attention.

What is pre-morbid functioning and why is it important?

Pre-morbid functioning refers to the level of cognitive functioning prior to an event that may affect cognitive ability, such as a stroke or a head injury. It is important to have an idea of pre-morbid functioning to use as a baseline measurement in order to determine if there has been any decline from pre-morbid levels of functioning.

What is the purpose of the study?

Pre-morbid levels of cognitive functioning often have to be estimated. This is because cognitive ability is not often measured prior to a brain injury or cognitive impairment and therefore pre-morbid data is not available.

Reading tests such as the Test of Pre-morbid Functioning (TOPF) can be used to do predict pre-morbid functioning. Demographic variables, such as gender and age, can also be used.

To improve our ability to predict pre-morbid functioning, we can create equations that help clinicians to predict pre-morbid cognitive scores using reading tests such as the TOPF and demographic variables.

Currently there are no equations to predict pre-morbid RBANS scores using the TOPF and demographic variables. This research intends to develop an equation. This will help with neuropsychological assessment of cognitive decline when using the RBANS.

What will happen if I agree to take part?

If you agree to take part, then I will contact you to arrange a convenient date and time for testing. Due to the current coronavirus restrictions, the research will be completed online through Microsoft Teams. I will send you an email explaining how to access the video call prior to the session.

At the session, you will be asked to complete a short additional questionnaire. You will first be asked some short questions about yourself such as your gender, age and level of schooling. You will then be asked to complete a short questionnaire called the Hospital Anxiety and Depression Scale (HADS) this should take no longer than 5minutes.

Following this, the TOPF, a reading test, will be administered. This will take no longer than 5minutes. The RBANS will then be administered, this will take no longer than 30minutes.

This data will then be anonymised and stored.

You are under no obligation to take part in the study.

Your rights

- You do not have to take part
- You can withdraw from the study at any point without giving a reason
- You can contact the researcher via email and ask them to remove your data from the study within 72 hours of completing testing
- All your data will be kept safe and cannot be linked back to you
- You have a right to ask questions about the research before and after participating
- Participating or not participating will have no effect on your medical care

What are the possible risks of taking part?

Participating in the study will take up to 45minutes. Some people find cognitive assessments distressing. If you experience distress, you can withdraw from the research at any time. You can also contact the research supervisor or service lead with any concerns or worries.

What are the possible benefits of taking part?

We cannot promise that you will have any direct benefits from taking part in the study. However, it is hoped that the information you give us will mean that in the future, it will be easier for clinicians to assess cognitive decline.

What will happen to the results of the study?

The results of the study will be summarised in a written thesis as part of a Doctorate in Clinical Psychology. The thesis will be available on the University of Hull's on-line repository https://hydra.hull.ac.UK. The research may also be published in academic journals or presented at conferences. If you want to hear about the results of the study then do contact the researcher, Hayley Gould, who will be happy to provide you with a written summary of the research.

How will we use information about you?

We will need to use information from you for this research project. This information will include your:

- Name
- Contact details
- Date of birth
- Biological Sex
- Years of education and employment status

People will use this information to do the research or to check your records to make sure that the research is being done properly. People who do not need to know who you are will not be able to see your name or contact details. Your data will have a code number instead. We will keep all information about you safe and secure. Once we have finished the study, we will keep some of the data so we can check the results. We will write our reports in a way that no-one can work out that you took part in the study. The data controller for this project will be the University of Hull. The University will process your personal data for the purpose of the research outlined above. The legal basis for processing your personal data for research purposes under GDPR is a 'task in the public interest' You can provide your consent for the use of your personal data in this study by completing the consent form that has been provided to you**.** Information about how the University of Hull processes your data can be found at https://www.hull.ac.UK/choose-hull/university-and-region/key-documents/data-protection.aspx

You have the right to access information held about you. Your right of access can be exercised in accordance with the General Data Protection Regulation. You also have other rights including rights of correction, erasure, objection, and data portability. Questions, comments and requests about your personal data can also be sent to the University of Hull Information Compliance Manager (dataprotection@hull.ac.UK). If you wish to lodge a complaint with the Information Commissioner's Office, please visit www.ico.org.UK.

What are your choices about how your information is used?

You can stop being part of the study at any time, without giving a reason, but we will keep information about you that we already have. We need to manage your records in specific ways for the research to be reliable. This means that we won't be able to let you see or change the data we hold about you.

Withdrawing from the study will not affect you in any way. Participant's data cannot be withdrawn from the study once the data has been anonymised and analysed. If you choose to withdraw from the study before this point the data collected will be destroyed. You have up to 72 hours after the completion of testing to withdraw your data from the research.

Where can you find out more about how your information is used?

You can find out more about how we use your information

- at www.hra.nhs.UK/information-about-patients/
- our leaflet available from www.hra.nhs.UK/patientdataandresearch
- by asking one of the research team
- by sending an email to

If you have any questions or require more information about this study, please contact me using the following contact details:

Personal details removed for publication

What if something goes wrong?

If you wish to make a complaint about the study, you can contact the University of Hull using the research supervisor's details below for further advice and information:

Personal Details removed for publication

Thank you for reading this information sheet and for considering taking part in this research.

Appendix L. Consent Form for Clinical Participants

CONSENT FORM

Title of study: **An investigation into the predictability of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) using the Test of Premorbid Functioning (TOPF) and Demographic variables.**

Name of Researcher: Hayley Gould

Please initial box

1. I confirm that I have read the information sheet dated 05/06/2021 (version 1.1) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

2. I understand that my participation is voluntary and that I am free to withdraw at any time during the study and up to 72hours following data collection without giving any reason, without my legal rights being affected. I understand that the data I have provided up to the point of withdrawal will be retained. ☐

3. I understand that after 72hours my data will be anonymised and it will no longer be possible to withdraw from the study. ☐

4. I give permission for the collection and use of my data to answer the research question in this study. ☐

5. I agree to take part in the above study. ☐

6. I confirm that the following exclusion criteria does not apply to me:. ☐

- I am over 18
- I have not previously completed the TOPF or RBANS
- I have not previously/ I am not currently being treated for a severe mental health problem (I have not been under the care of a CMHT or been an inpatient at a mental health facility).
- I am proficient in the English Language ☐

Name of Participant	Date	Signature
Name of Person taking consent	Date	Signature

Appendix M. Consent Form for Non-Clinical Participants

CONSENT FORM

Title of study: **An investigation into the predictability of the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) using the Test of Premorbid Functioning (TOPF) and Demographic variables.**

Name of Researcher: Hayley Gould

Please initial box

7. I confirm that I have read the information sheet dated 05/06/2021 (version 1.1) for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐

8. I understand that my participation is voluntary and that I am free to withdraw at any time during the study and up to 72hours following data collection without giving any reason, without my legal rights being affected. I understand that the data I have provided up to the point of withdrawal will be retained. ☐

9. I understand that after 72hours my data will be anonymised and it will no longer be possible to withdraw from the study. ☐

10. I give permission for the collection and use of my data to answer the research question in this study. ☐

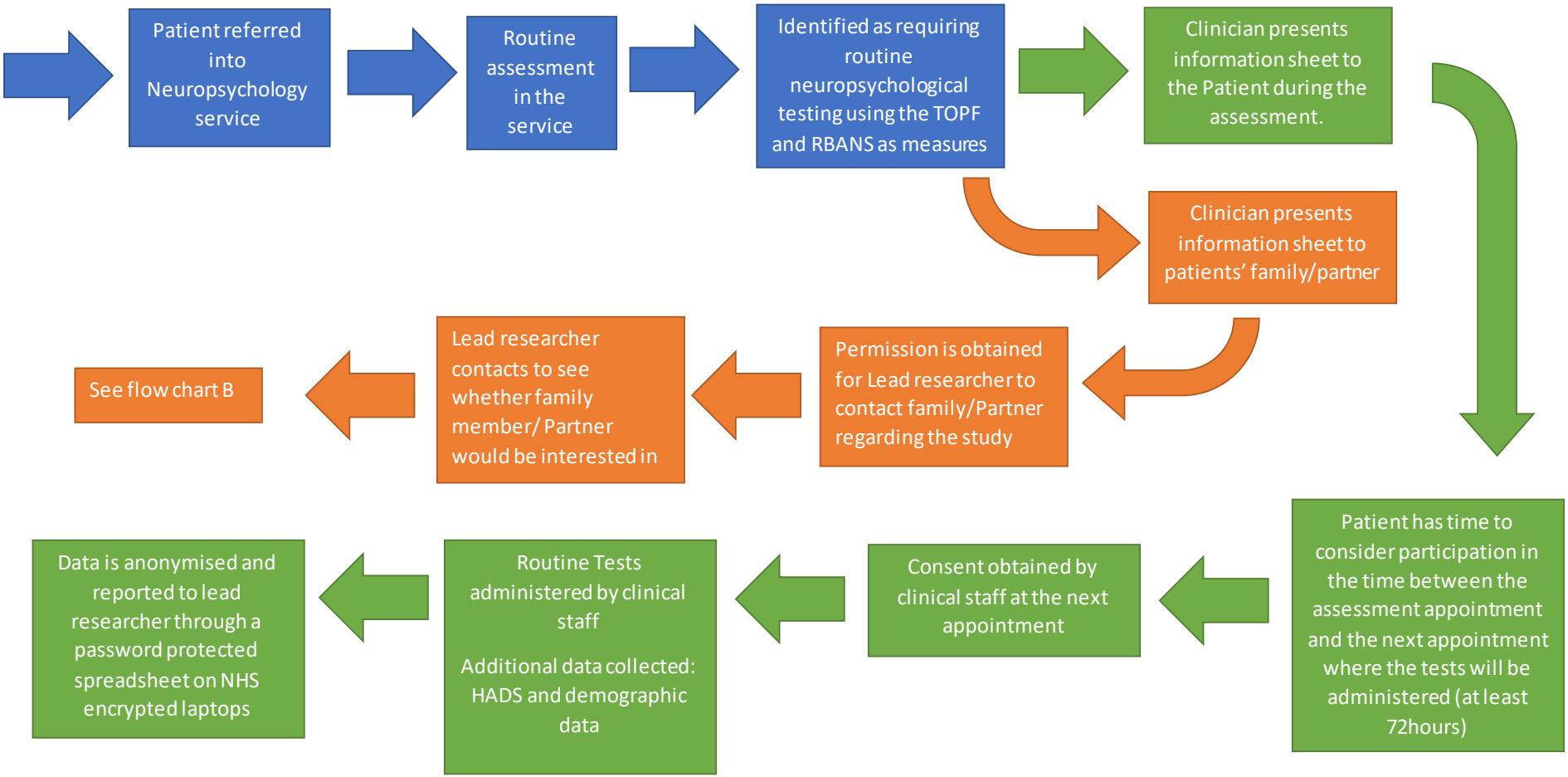
11. I agree to take part in the above study. ☐

12. I confirm that the following exclusion criteria does not apply to me:
    - I have not previously been diagnosed with a neurological condition or neurodevelopmental condition (e.g. ADHD, Autism, learning disability) nor am I undertaking assessment at the moment.
    - I am over 18
    - I have not previously completed the TOPF or RBANS
    - I have not previously/ I am not currently being treated for a severe mental health problem (I have not been under the care of a CMHT or been an inpatient at a mental health facility).
    - I am proficient in the English Language ☐

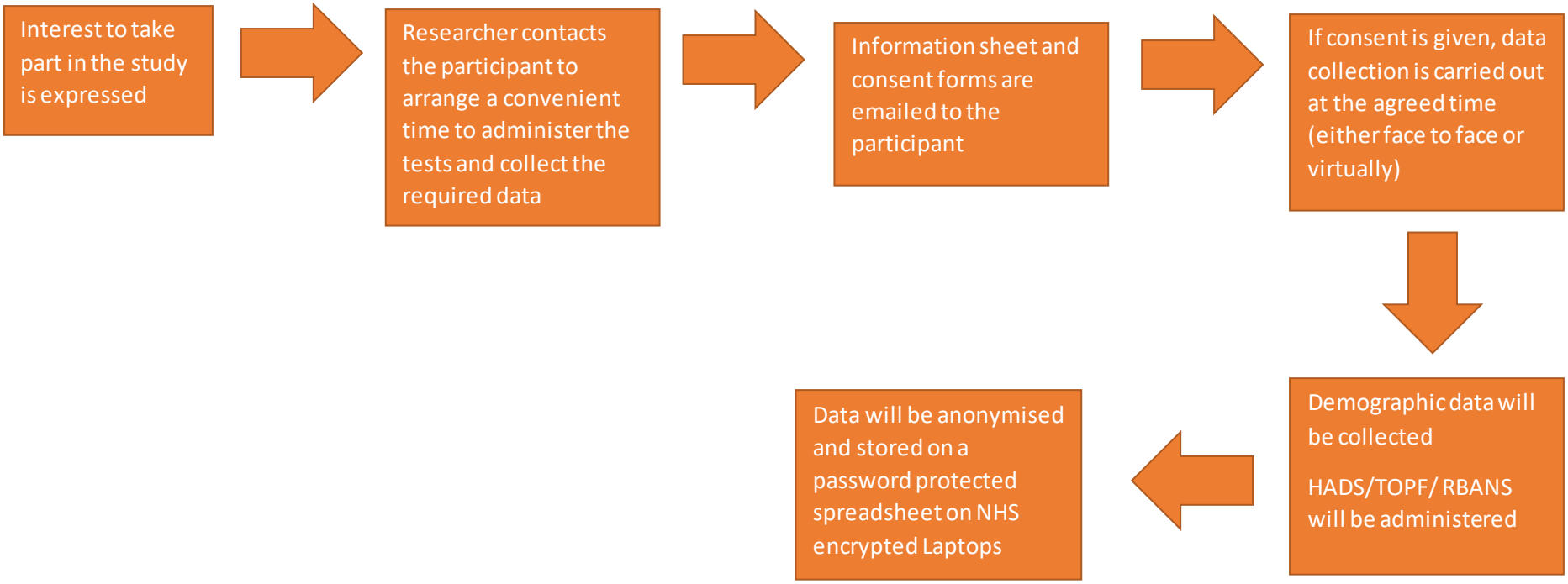
Name of Participant	Date	Signature
Name of Person taking consent	Date	Signature

Appendix N. Recruitment Process Flow Charts

Flow chart A: Recruitment and Data collection in services



Flow chart B : Recruitment and Data Collection of the Non-Clinical Sample



Appendix O. University Ethics Approval – Removed for publication

Appendix P. University Sponsorship – Removed for publication

Appendix Q. IRAS Approval

Removed

Removed

Removed

Removed

Appendix R. Hospital Anxiety and Depression Scale

Hospital Anxiety and Depression Scale (HADS)

Tick the box beside the reply that is closest to how you have been feeling in the past week.
Don't take too long over you replies: your immediate is best.

D	A		D	A	
		I feel tense or 'wound up':			**I feel as if I am slowed down:**
	3	Most of the time	3		Nearly all the time
	2	A lot of the time	2		Very often
	1	From time to time, occasionally	1		Sometimes
	0	Not at all	0		Not at all
		I still enjoy the things I used to enjoy:			**I get a sort of frightened feeling like 'butterflies' in the stomach:**
0		Definitely as much		0	Not at all
1		Not quite so much		1	Occasionally
2		Only a little		2	Quite Often
3		Hardly at all		3	Very Often
		I get a sort of frightened feeling as if something awful is about to happen:			**I have lost interest in my appearance:**
	3	Very definitely and quite badly	3		Definitely
	2	Yes, but not too badly	2		I don't take as much care as I should
	1	A little, but it doesn't worry me	1		I may not take quite as much care
	0	Not at all	0		I take just as much care as ever
		I can laugh and see the funny side of things:			**I feel restless as I have to be on the move:**
0		As much as I always could		3	Very much indeed
1		Not quite so much now		2	Quite a lot
2		Definitely not so much now		1	Not very much
3		Not at all		0	Not at all
		Worrying thoughts go through my mind:			**I look forward with enjoyment to things:**
	3	A great deal of the time	0		As much as I ever did
	2	A lot of the time	1		Rather less than I used to
	1	From time to time, but not too often	2		Definitely less than I used to
	0	Only occasionally	3		Hardly at all
		I feel cheerful:			**I get sudden feelings of panic:**
3		Not at all		3	Very often indeed
2		Not often		2	Quite often
1		Sometimes		1	Not very often
0		Most of the time		0	Not at all
		I can sit at ease and feel relaxed:			**I can enjoy a good book or radio or TV program:**
	0	Definitely	0		Often
	1	Usually	1		Sometimes
	2	Not Often	2		Not often
	3	Not at all	3		Very seldom

Please check you have answered all the questions

Scoring:
Total score: Depression (D) _____ Anxiety (A) _____
0-7 = Normal
8-10 = Borderline abnormal (borderline case)
11-21 = Abnormal (case)