

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Standardising and validating the Cleveland Verbal Working Memory with Interference (CVWMT-I)

Kapoor, Aparna

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

VOLUME I:
Main Research Project and Systematic
Literature Review

Aparna Kapoor

Institute of Psychiatry, Psychology and Neuroscience
King's College London

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

May 2019

Acknowledgments

I would like to express my sincere thanks to my supervisors, Dr Jessica Fish and Prof Robin Morris, for their continued support, encouragement and expertise, which has been invaluable in researching and writing this thesis.

A special thanks goes out to each of my participants who generously gave their time to take part in this study. I am grateful to Join Dementia Research who supported the recruitment of participants for this project. And also, to Mark Ballham who went out of his way to ensure the smooth running of my project, even dealing with my last-minute requests.

I would also like to thank Rukeya Khanam, who supervised my service-evaluation project, and the Peer Support volunteers who made this project possible.

A personal thank you to my fellow trainees, Mazda, Jheanell, Leo and Natasha, whose friendship and conversation helped contain my anxiety around completing this project.

To my mum, paps, D, Tasha and Rohit – for their unwavering faith through this long journey and for reminding me that this was possible and to keep my “chin up”.

To Sally, Smudge, Sal and Gina – for being there for me when my family couldn’t.

Finally, a massive thank you to Haig – for patiently listening to my rants, sorting out my technological difficulties, going above and beyond to help me recruit and keeping me fed and comforted, even from miles away – as well as baby Kapoor-Smith; we wrote this together!

I dedicate this to my grandfather – Dada, I wish you were here to see this.

Summary of Contents

Main Research Project

Standardising and Validating the Cleveland Verbal Working Memory Test with Interference (CVWMT-I)	4
---	---

Systematic Literature Review

Does the development of robotic and smart technology care interventions for people with dementia adhere to recommendations for good psychological care? A systematic review.....	102
--	-----

Main Research Project:

Standardising and validating the Cleveland Verbal Working Memory Test
with Interference (CVWMT-I)

Aparna Kapoor

Institute of Psychiatry, Psychology and Neuroscience
King's College London

Supervised by:

Dr Jessica Fish

Professor Robin Morris

Institute of Psychiatry, Psychology and Neuroscience
King's College London

Table of Contents

List of Tables	8
List of Figures	9
List of Abbreviations	10
1. Abstract.....	12
1.1 Background	12
1.2 Method	12
1.3 Results	13
1.4 Conclusions	13
2. Introduction.....	14
2.1 Definition of Executive Function.....	14
2.2 Models of EF and related concepts	15
2.3 EF in daily life.....	16
2.4 Deficits in EF and the impact of interference	17
2.5 Assessment of EF	20
2.6 Ecological validity of neuropsychological tests.....	21
2.7 The Brown Peterson Procedure (BPP).....	22
2.8 Demographic differences in performance on the BPP	24
2.9 Aims of the study	26
2.10 Research questions of the study	26
3. Method	27
3.1 Ethical Clearance	27

3.2	Design	27
3.3	Sample Size Calculations	27
3.4	Participants	30
3.4.1	Recruitment.....	30
3.4.2	Inclusion and exclusion criteria	31
3.4.3	Participant demographics.....	32
3.5	Measures	36
3.5.1	Cleveland Verbal Working Memory Test with Interference	36
3.5.2	Background measures of cognitive ability.....	39
3.5.3	Questionnaires	40
4.	Results.....	41
4.1	Scores on background neuropsychological measures and the CVWMT-I	45
4.2	Establishing test norms	46
4.2.1	Regression-based norms	46
4.2.2	Percentile ranks.....	49
4.3	Reliability	50
4.3.1	Parallel-form reliability.....	50
4.3.2	Test-retest reliability	51
4.3.3	Inter-rater reliability.....	51
4.4	Validity.....	52
4.4.1	Convergent validity	52

4.4.2	Construct Validity	53
5.	Discussion	55
6.	Strengths and Limitations of the study	61
7.	Conclusions and future recommendations	62
8.	References.....	64
9.	Appendices	77
	Appendix A: Ethical approval letter from PNM RESC at KCL.....	77
	Appendix B: Administration instructions for the CVWMT-I	78
	Appendix C: Scoring instructions for the CVWMT-I	82
	Appendix D: The Cleveland Verbal Working Memory Test with Interference (CVWMT-I; Mack, 2016) Versions A and B	85
	Appendix E: Raw data for Letter Recall, FSIQ-2 and age for the normative sample.	87
	Appendix F: Raw data for Reverse Counting, FSIQ-2 and age for the normative sample	89
	Appendix G: Instructions on using the RegBuild_MR_Raw.exe. programme	91
	Appendix H: Output for case example's performance on Letter Recall using RegBuild_MR_Raw.exe	92
	Appendix I: Output for case example's performance on Reverse Counting using RegBuild_MR_Raw.exe	96
	Appendix J: Percentiles of the distribution of scores on Letter Recall and Reverse Counting stratified by age group	100

List of Tables

Table 1.1	A-priori effect sizes and required sample sizes for each psychometric test.....	29
Table 1.2	Participant demographics (normative sample, $n = 96$).....	33
Table 1.3	Participant scores on GAD-7 and PHQ-8 ($n = 96$).....	35
Table 1.4	Participant demographics (clinical sample, $n = 55$).....	36
Table 1.5	List of trigrams included in the CVWMT-I (versions A and B).....	38
Table 1.6	Multiple regression analysis for effect of age and IQ on Letter Recall..	43
Table 1.7	Multiple regression analysis for effect of age and IQ on Reverse Counting.....	45
Table 1.8	Scores for participants on background questionnaires and neuropsychological measure ($n = 96$).....	46
Table 1.9	Percentiles of the distribution of scores on Letter Recall and Reverse Counting.....	49
Table 1.10	Performance on two parallel forms of the CVWMT-I.....	50
Table 1.11	CVWMT-I scores on first and second testing sessions.....	51
Table 1.12	Reliability coefficients for CVWMT-I measures ($n = 35$).....	52
Table 1.13	Pearson correlation matrix for performance on neuropsychological tests and CVWMT-I.....	53
Table 1.14	t -test results comparing demographics (age and education level) for the clinical and normative samples.....	54

List of Figures

Figure 1.1 Scatterplots displaying effects of IQ and age on Letter Recall.....42

Figure 1.2 Scatterplots displaying effects of age and IQ on Reverse Counting.....44

List of Abbreviations

ACT	Auditory Consonant Trigrams test
BADS	Behavioural Assessment of the Dysexecutive Syndrome
BPP	Brown-Peterson procedure
BRIEF	Behaviour Rating Inventory of Executive Function
CCC	Brown-Peterson Consonant Trigrams Auditory Memory Task
DEX-R	Dysexecutive Questionnaire Revised
CVWMT-I	Cleveland Verbal Working Memory Test with Interference
D-KEFS	Delis-Kaplan Executive Function System
EF	Executive function
FSIQ	Full Scale Intelligence Quotient
GAD-7	Generalized Anxiety Disorder – 7
ICC	Intra-class correlation
MCI	Mild cognitive impairment
ONS	Office for National Statistics
PASAT	Paced Auditory Serial-Addition Task
PHQ-8	Patient Health Questionnaire – 8
PI	Proactive interference
QUAND	Quantifying the Uncertainty Attached to Normative Data
RI	Retroactive interference
SAS	Supervisory attentional system
Stroop C-W	Stroop Colour-Word task
TBI	Traumatic brain injury
TEA	Test of Everyday Attention
TOPF	Test of Premorbid Functioning
TMT	Trail Making Test

WAIS-IV	Wechsler Adult Intelligence Scale
WASI-II	Wechsler Abbreviated Scale of Intelligence, Second Edition
WCST	Wisconsin Card Sorting Test
WM	Working memory

1. Abstract

1.1 Background

The current study addresses the need for improvement in the assessment of executive function (EF). An unpublished measure of working memory (WM) ability, the Cleveland Verbal Working Memory Test with Interference (CVWMT-I), incorporates an element of distraction, which enables more accurate prediction of how WM ability is used in daily life. This study presents psychometric properties and normative data for the CVWMT-I.

1.2 Method

Ninety-six volunteers from the general population, between the ages of 21-79 were recruited for participation. Participants were selected based on education level (degree, college or secondary school), age and biological sex, to ensure even representation across each category, within the sample. Exclusion criteria included current depression or anxiety, psychiatric or neurological diagnosis, or history of brain injury. Participants completed a neuropsychological test battery, which included the CVWMT-I. Forty-eight participants completed version A of the CVWMT-I, whilst 46 completed version B, in order to measure parallel-form reliability. Thirty-five participants were audio recorded with consent, in order to measure inter-rater reliability, and 35 were re-tested on alternate forms four weeks later, to measure test-retest reliability. Convergent validity was measured via comparison of participant performance on background neuropsychological measures and the CVWMT-I. Pre-existing data on the CVWMT-I from 55 participants with traumatic brain injury (TBI) were compared to a demographically matched sample of 40 participants from the normative sample, to measure construct validity.

1.3 Results

Performance on the CVWMT-I was found to be mediated by IQ and age. A regression-based norming method was used, stratifying for IQ and age, using the Reg_Build_MR_Raw.exe programme. Normative data is presented for use with this package, along with guidance on how to use it. Percentile rank tables are presented for information only. Two measures of the CVWMT-I, Letter Recall and Reverse Counting, were used for analyses. Letter Recall was found to be a sensitive measure of WM ability and correlated highly with other neuropsychological measures assessing similar constructs. Inter-rater and parallel-form reliability were high for both the Letter Recall and Reverse Counting measures, as was test-retest reliability for Reverse Counting. However, test-retest reliability was low for the Letter Recall measure. Normative participants performed better than those with TBI on the CVWMT-I, indicating good construct and ecological validity.

1.4 Conclusions

The psychometric properties of the CVWMT-I indicate that it is a good measure for use within clinical settings, with potential for wide accessibility, due to the employment of a regression-based norming method. Further research is required to assess clinical effectiveness of the CVWMT-I in identifying WM deficits within other clinical populations.

2. **Introduction**

This introduction explores our current knowledge of executive function (EF), with a focus on its overlap with working memory (WM), its assessment through neuropsychological testing and limitations of the same. First, a definition of EF is presented, followed by a summary of theoretical models, focusing primarily on the central executive. Related concepts of EF, including WM and attentional control, are defined and discussed. Following this, a summary of how we use EF skills in daily life is presented, alongside an exploration of factors that can hinder our ability to employ these skills effectively, particularly interference. This leads to a discussion about the clinical importance of robust assessment of EF deficits and a summary of current neuropsychological tests in clinical use, followed by a brief review of their ecological validity. Next, the Brown-Peterson procedure (BPP), a test of WM incorporating an interference component, is introduced as a test with high potential ecological validity. Adaptations of the BPP are presented and critiqued, and findings regarding the impact of demographic variables on BPP performance are considered. Finally, a new, unpublished adaptation, the Cleveland Verbal Working Memory Test with Interference (CVWMT-I) is described and the aims and research questions of the current study are presented.

2.1 **Definition of Executive Function**

Executive function (EF) is an umbrella term comprising a range of abilities and processes carried out by the frontal lobes and related networks, associated with planning, decision-making and goal-directed behaviour (Gioia, Isquith & Guy, 2001). The link between the frontal lobes and EF has been repeatedly demonstrated through the history of clinical neuropsychology (Stuss & Benson, 1984) and supported by findings from brain activation studies in participants completing EF tasks (Baker et al., 1996;

Crittenden, Mitchel & Duncan, 2016). EF encompasses the ability to tackle several competing tasks at once, often referred to as dual or multi-tasking, which is common in daily life (Shallice & Burgess, 1996). These processes develop throughout childhood, playing an important role in cognitive and social development (Anderson, 2002), and into adulthood, in the development of independent living skills (Lezak, 1995; Stuss & Levine, 2002) and work-related abilities (Bade, 2010).

2.2 Models of EF and related concepts

Numerous models of EF have been proposed over the years in an attempt to understand this complex area of function (Barkley, 1997; Lezak, 1995; Shallice & Burgess, 1996; Zelazo, Carter, Reznick & Frye, 1997). Shallice and Burgess (1996) proposed the “supervisory attentional system” (SAS), which is responsible for employing deliberate attention to novel situations which requires planning and organisation. These situations are different to those in which actions can be performed automatically, known as “contention scheduling”, in which deliberate attention is not required. They proposed that the SAS consists of three stages involving the construction of temporary schemas when faced with novel situations, the testing out of the schema to assess its effectiveness to cope with the new situation and finally, either rejecting or accepting the schema based on its success.

Another influential model proposed the presence of the “central executive” (Baddeley, 2002) which is responsible for functions such as selectively attending to important information whilst ignoring distracting stimuli, employing working memory (WM) ability to switch attention between tasks and carrying these out simultaneously, whilst selecting and retrieving information (or schemas) from long-term memory to assist with completion of tasks. Further, Barkley (1997) identified the role of WM in not only

planning and goal-setting, but also in ensuring that this process is not disrupted by distraction. Each of these models exist in an attempt to explain the processes behind this complex psychological construct, which is used routinely in everyday functioning.

Ongoing research continues in this area, to identify whether EF can be understood as a single or multicomponent construct, located in discrete regions versus broad networks (Rabinovici, Stephens, & Possin, 2015). Originally thought to be located solely in the frontal lobe, due to early studies looking at both adults and children with frontal lobe lesions (Finger & Almli, 1988; Damasio, Grabowski, Frank, Galaburda, & Damasio, 1994), later studies have demonstrated the role of several other regions of the brain. WM, for example, has been found to activate the ventrolateral prefrontal cortex and dorsolateral prefrontal cortex, at different stages of storage, retrieval and maintenance (Wager & Smith, 2003), but may also activate other regions located in the parietal cortex, thalamus, basal ganglia and cerebellum (Rabinovici et al., 2005). Further, in people with frontotemporal dementia (FTD), which is characterised by changes in decision making related to social situations and emotional management, the anterior temporal lobes may be affected, along with deficits in the frontal areas of the brain (Snowden, 2008). The complexity of the neurological processes involved in carrying out EF is of interest, particularly when assessing deficits in people's ability to carry these activities out in daily life.

2.3 EF in daily life

A variety of EFs are thought to be in constant use during our daily life across the lifespan, allowing us to live independently (Perna, Loughan & Talka, 2012). Likewise, a range of cognitive and behavioural EF skills are employed in order to complete activities of daily living independently (Lezak, 1995; Stuss & Levine, 2002). For

example, a task like grocery shopping involves a range of EF skills, from planning, targeted goal setting, recalling information, navigating and initiating behaviour to complete the task. EF skills like self-regulation, inhibition of responses and behaviour, and social skills are required to maintain healthy social relationships (McClelland, Cameron, Wanless & Murray, 2007), whilst cognitive EF skills like processing speed have been linked to academic achievement among school-age children (Gordon, Smith-Spark, Newton & Henry, 2018). A variety of cognitive and behavioural EF skills are used among adults to meet the requirements of employment (Bade, 2010), including time management skills, following directions, focussing attention and self-initiation. Across each of these areas of daily life, we often switch between numerous tasks, or carry out a number of tasks simultaneously (Pettigrew & Martin, 2016).

A key element in the process of task-switching and multi-tasking is WM ability; the capacity to temporarily store information in mind and manipulate that information (Baddeley & Logie, 1999). WM and deliberate attention are linked, with the need for information to be closely and adequately attended to, prior to it being processed successfully by our WM (Mack & Rock, 1998). Through selectively choosing and attending to important information presented via our environment, WM ability also plays an important role in the learning process (Fougnie, 2008). However, this learning process may be hindered by the presence of interference (Sakai, Rowe & Passingham, 2002), leading to the forgetting of important information.

2.4 Deficits in EF and the impact of interference

When learning new information, previously held information can hinder the learning process, known as “proactive interference” (PI), whilst newly learned information can hinder the recall of previously held information, known as “retroactive interference”

(RI; Pettigrew & Martin, 2016). PI has been shown to be affected by context effects, with greater interference being linked to memories learned in similar contexts (Edwards, 2010). Hence, altering the context within which new information is presented may help to remove PI and aid the encoding of information from WM into short-term memory. Further, RI, which is considered the more commonly observed of the two forms of interference, may occur as a result of distraction in the form of new information, which inhibits the ability to appropriately rehearse previously presented information, resulting in forgetting (Edwards, 2010). The extent of the impact of interference on learning and memory has been linked to mechanisms related to individual differences in executive control (Jonides & Nee, 2006) or deficits in the encoding and storage of information (McClelland, Naughton & O'Reilly, 1995). Thus, multi-tasking and task-switching can also affect performance on the primary task at hand, as a result of PI or RI (Pettigrew & Martin, 2016).

Although interference-control, or the ability to process information in the presence of competing stimuli, is a key function of WM (Kane & Engle, 2003), several factors can have an impact on a person's ability to carry this out successfully. These include low IQ (Dempster & Corkill, 1999), difficulties related to response inhibition (Friedman & Miyake, 2004) and age (De Luca & Leventer, 2008; Parasuraman & Greenwood, 1998), with young children and older adults performing poorly on such tasks. Further, WM ability and attentional capacity also differ across mental state including depression and anxiety (Dalglish & Watts, 1990) and in people with brain injury (Robertson & Rafal, 2000) or dementia (Morris & Baddeley, 1988). Neural networks may be diminished as a result of the brain injury or the process of aging, resulting in deficits in WM ability, due to its reliance on distributed neural networks (Robertson & Rafal, 2000). People with cognitive impairment related to dementia or brain injury may also be more negatively

affected by interference when carrying out EF tasks, as compared to healthy controls (Krawczyk et al., 2008). This is due to a decline in the ability to selectively attend to relevant information only, in the presence of competing stimuli (Aurtenetxe et al., 2016).

When Chao and Knight (1995) considered the impact of age-related decline of the pre-frontal cortex on performance on EF tasks, they found that performance was significantly more impaired in older participants in the presence of interference only, hypothesising that this may be attributed to an impairment in attentional control. A study by Aurtenetxe et al. (2016) found similar results. They considered the impact of two types of interference, distraction and interruption, on the ability to complete WM tasks, by healthy participants and those with mild-cognitive impairment (MCI). The ‘distraction’ task involved inclusion of additional information that participants were instructed to ignore, whilst the ‘interruption’ task involved inclusion of further information that participants were required to attend to and process. They found that distraction impacted on the performance of both groups; however, interruption had a significantly greater impact on the performance of participants with MCI. This study indicated that interference may impact negatively on people’s ability to carry out daily activities and decision-making tasks, particularly those with additional cognitive difficulties.

As WM is so functionally relevant, it is often clinically useful to obtain an estimate of a person’s WM capacity via neuropsychological testing, to contribute to a clinical formulation, make predictions about likely everyday functioning, and to direct support and plan interventions accordingly (Vakil, 2011).

2.5 Assessment of EF

There are a number of neuropsychological tests currently available for clinical use to assess EF performance, like the Stroop task (Trenerry, Crosson, DeBoe & Leber, 1989), the Wisconsin Card Sorting Test (WCST; Grant & Berg, 1948), trail making, and ‘tower’ tests, among several others. Although these measures provide useful information about specific functions within the context of a highly structured environment, they have typically lacked the presence of competing tasks which more accurately measure multi-tasking ability - a common feature of everyday life (Chaytor & Schmitter-Edgecombe, 2003). Furthermore, despite being used widely in clinical settings, these traditional neuropsychological tests do not provide adequate evidence to suggest that they predict functional ability in everyday life (i.e. have good predictive validity) and had only limited evidence of use with people with EF deficits, prior to being employed clinically (Burgess et al., 2006). Burgess et al. (2006) also noted that traditional neuropsychological assessments of WM, like the Digit Span subtest of the Wechsler Adult Intelligence Scale (WAIS-IV; Wechsler, 2008), attempt to assess everyday functioning via simplified versions of daily tasks. For example, the Digit Span subtest requires the participant to hold information presented verbally in mind for short periods of time and repeat them back in the correct order, as per guidelines presented at the beginning. Although a good measure of memory span, this test does not consider the use of interference control processes and selective attention, key skills employed in the daily use of WM in the presence of distracting or competing stimuli (Barkley, 1997). The use of these types of tests may lead to a type two error, with impairments in EF associated with daily functioning in the real world going undetected.¹

¹ Further, the standard assessment context often prevents people from utilising compensatory strategies they otherwise use in everyday life (Chaytor & Schmitter-Edgecombe, 2003), which may lead to a type one error, or an over-identification of EF deficits.

2.6 *Ecological validity of neuropsychological tests*

Burgess et al. (2006) proposed that to overcome the difficulties associated with the use of traditional neuropsychological assessments of EF, we should focus on the development of new tests which more realistically mimic everyday functioning. For example, the Multiple Errands Test proposed by Burgess et al. (2006), involves carrying out several tasks at once in the presence of distractions, switching between these tasks, prioritising tasks by urgency and effort, whilst assessing one's own performance regarding adequate task completion. Due to its similarities to how we carry out tasks in everyday life, Burgess et al. (2006) opined that this test may have better predictive and ecological validity.

Ecological validity refers to the “functional and predictive relationship between the patient's performance on a set of neuropsychological tests and the patient's behavior in a variety of real-world settings” (Sbordone, 1996, pp. 16). It consists of two concepts: ‘verisimilitude’, or the ability of a test to resemble a real-life task, and ‘veridicality’, or the the ability to predict impairment in everyday functioning (Chaytor & Schmitter-Edgecombe, 2003). These concepts are linked but may not overlap. Developments in neuropsychological testing appear to focus more on verisimilitude, rather than veridicality (Parsons & Courtney, 2011), such as tests incorporating virtual reality. However, this may be problematic as these tests may still lack the ability to accurately estimate a person's performance in real-life. Further, neuropsychological tests which may lack verisimilitude and may not have been constructed with the concept of ecological validity in mind, may still demonstrate veridicality (Jung, 2015). One potential such test of EF ability is the Brown Peterson Procedure (BPP; Brown, 1958; Peterson & Peterson, 1959).

2.7 *The Brown Peterson Procedure (BPP)*

The BPP (Brown, 1958; Peterson & Peterson, 1959) was developed to study the impact of interference on WM retrieval by incorporating a distraction task. The BPP tests the ability to hold information in mind and to maintain it, whilst attending to a competing task, presented during an interval period extending to a maximum of 20 seconds.

Although the BPP may not display good verisimilitude, owing to the unrealistic nature of the task, the inclusion of an interference task means that it may have good veridicality, by mimicking how we cope with the interruptions and distractions that characterise everyday life. Research has shown that the BPP is a good measure for identifying deficits in EF ability in people with dementia (Dannenbaum, Parkinson & Inman, 1988), brain injury (Stuss, Stethem, Hugenholtz, & Richard, 1989) and frontal lobe disorders (Winocur, Oxbury, Roberts, Agnetti & Davis, 1984).

One BPP measure currently in clinical use is the Auditory Consonant Trigrams test (ACT; Stuss et al., 1989). Subjects are aurally presented a list of three consonants or a trigram, followed by a competing task of counting backwards from a given number for a short period of time. Following this, they are required to recall the trigram presented initially. The ACT was found to be a good predictor of everyday performance, when measured against the Working Memory subscale of the Behaviour Rating Inventory of Executive Function (BRIEF; Gioia, Isquith, Guy, & Kenworthy, 2000), in an ASD population (Yerys, Wallace, Jankowski, Bollich, & Kenworthy, 2012). However, there is limited information regarding the psychometric properties of the ACT, as well as limited normative data due to large variation across studies in the administration of the task and the populations studied (Shura, Rowland & Miskey, 2016). Further, this test may be difficult to administer in a standardised manner across testing situations (Stuss et al., 1989), making results difficult to replicate and less reliable.

The current study presents another adaptation of the BPP, in the form of the Cleveland Verbal Working Memory Test with Interference (CVWMT-I; Mack, 2016). The CVWMT-I was developed separately from the ACT and incorporates the element of interference via distraction that is so common in everyday life into a simple WM task, to detect subtle impairments in EF. Unlike the ACT, the CVWMT-I incorporates interference trials of only one duration (15s) across 14 trials, whereas the ACT typically involves interference trials increasing in duration from zero, nine, 18 to 30 seconds, or a variation of the same (Stuss et al., 1989), making the CVWMT-I faster to administer and thus, a good candidate for wider clinical use. Further, trigrams in the CVWMT-I are presented visually, rather than aurally as is typical in the ACT. Grant and McCormack (1969) found a greater recency effect for information presented aurally, as compared to information presented visually, particularly when the interval between presentation of stimuli and recall was shorter. Further, interference of 15 to 30 seconds has been shown to reduce recency effects associated with free recall (Postman & Phillips, 1965). Additionally, Klingner, Tvrsky and Hanrahan (2011) found recall for visually presented stimuli to be better, as evaluated by accuracy of recall on a digit span task, than information presented aurally, possibly due to the greater cognitive load of aurally presented stimuli, or the dual-coding of visually presented verbal stimuli in WM (Baddeley, 2007). Similar findings were obtained by Bigelow and Poremba (2014), who noted better recall of visual, as opposed to auditory or tactile, information over longer intervals, suggesting that this may aid memory whilst reducing recency bias.

Hence, presenting CVWMT-I trigrams visually, along with a 15 second interference task, both of which are features that differentiate it from the ACT, may reduce any recency bias associated with recall. Further, due to the inclusion of an interference task which puts a similar strain on WM ability, as do real-life tasks employing WM ability,

the CVWMT-I was thought to demonstrate veridicality. However, at the time of writing, the CVWMT-I has only limited normative data based on an American population (Mack, 2016), which has not been published, limiting its use in clinical and research or experimental settings.

2.8 Individual differences in performance on the BPP

There is conflicting evidence regarding the effects of sex, age and education on BPP performance. A recent study considered sex difference in performance on high-demand verbal WM tasks (Reed, Gallagher, Sullivan, Callicott & Green, 2017). Findings from this study suggested better performance by males, as assessed by accuracy of recall, when compared to females. However, when sex differences in performance specifically on the BPP were tested, no interaction was found (Stuss, Stetham & Poirier, 1987). Bherer, Belleville and Peretz (2010), found education effects on performance on the BPP, with participants with lower educational attainment performing more poorly. However, they did not find any interaction for age-effects on performance. Shura et al. (2016) replicated these results, finding no age differences in performance on the ACT. However, neither study included participants older than 69 years. When considering an older population (18 – 86 years), Geurten, Vincent, Van der Linden, Coyette and Meulemans (2016) found age and education effects on BPP performance. With regards to IQ, Alloway and Alloway (2010) explored whether WM was simply a part of the concept of intelligence, or a separate cognitive function. Their findings suggested that WM was possibly a better predictor of academic achievement than IQ. Yet, contrary to these findings, Chooi and Thompson (2012) found no improvement in intellectual ability in a group of adults following WM training, suggesting a complex relationship between WM and IQ. Further, Melby-Lervåg and Hulme (2012), in their meta-analysis, found that WM training did not generalise to other areas of IQ (e.g. verbal or arithmetic

skills) and did not have lasting improvement on WM ability. Hence, it is of interest within the current study to consider the effects of each of these variables on performance on the CVWMT-I to contribute to this body of research.

Another variable of interest that may have an impact on WM performance is emotional arousal. As stated previously, anxiety and depression have been found to impair a person's WM ability. However, even where neither of these conditions are present, state anxiety caused by the testing situation itself has been found to impact on a person's performance (Retegui, 2006). More specifically, studies have found that negative emotional arousal reduces a person's WM capacity (Figueira et al, 2017). Hence, the current study monitored participants' mood prior to testing using mood measures to ensure this does not interfere with performance. Further, via observation, and feedback at the end of testing, state anxiety was monitored to identify if this played a role in participants' performance on the BPP.

From the literature reviewed here, it is clear that the BPP may possess good veridicality and that there is a need for UK normative data on a BPP adaptation which is reliable and valid, that can be readily applied to a clinical setting. The current study sought to standardise the CVWMT-I by developing normative data based on a larger, UK based population, and to examine its psychometric properties including reliability and validity, with the aim of improving future testing of EF to enable more accurate prediction and identification of everyday dysexecutive behaviour.

2.9 *Aims of the study*

The aims of the current study are as follows:

1. To develop normative data for the CVWMT-I, stratifying for age, biological sex and IQ, or as required
2. To assess the reliability of the CVWMT-I between testing occasions, between raters, and between two alternate forms
3. To assess the validity of the measure by:
 - (a) examining correlations between performance on the CVWMT-I and measures of related cognitive functions, and
 - (b) comparing CVWMT-I performance of a group of adults with TBI with that of the normative sample

2.10 *Research questions of the study*

The research questions addressed were:

1. Does the CVWMT-I possess adequate inter-rater, parallel form, and test-retest reliability?
2. Does the CVWMT-I possess convergent validity, as indicated by the presence of:
 - a. Significant relationships between scores obtained on the CVWMT-I and scores obtained on conceptually-related measures of cognitive functioning, within the normative sample
 - b. Weaker CVWMT-I performance by participants in the clinical group relative to performance of an age, sex and education level-matched normative group.
3. Is there an effect of age, IQ or biological sex on performance on the CVWMT-I?

3. Method

3.1 Ethical Clearance

The study was reviewed and approved by the Psychiatry, Nursing and Midwifery Ethics Subcommittee (PNM RESC) at King's College London (KCL). Approval for the study (reference number LRS-17/18-5157) was gained on 14th February 2018 (see Appendix A for ethical approval letter).

3.2 Design

The current project was a normative study, aiming to establish test norms from a group of UK residents, and incorporating additional smaller studies to investigate the essential psychometric properties of the CVWMT-I.

Ninety-six non-clinical participants from the general population provided the data for the main normative study, as well as parallel-form reliability. A sub-sample of 35 participants provided additional data for the studies examining inter-rater reliability (via audio recordings of testing sessions), and test-retest reliability (via re-testing after a four-week interval).

To assess the validity of the CVWMT-I, pre-existing data from 55 consenting participants with TBI attending the Oliver Zangwill Centre were compared to that of a demographically matched sub-sample of the normative sample.

3.3 Sample Size Calculations

The American Psychological Association (APA; 1999) recommends a minimum sample size of 100 for a norm-referenced test. Further, for the development of regression-based norms for a psychological test which is short in length, with a small number of item scores, a minimum sample size of 100 is recommended (Oosterhuis, van der Ark, &

Sijtsma, 2016). The previous unpublished norming study on the CVWMT-I (Mack, 2016) had a sample size of $N = 37$. Based on these recommendations, the previous study and time constraints on the current study, the statistical tool, Quantifying the Uncertainty Attached to Normative Data (QUAND; Crawford & Garthwaite, 2008) was used to compare the 95% confidence intervals for a sample where $N = 37$ and $N = 100$. At the 50th percentile, a substantially narrower confidence interval was obtained for $N = 100$ (CI: 15 points) compared with $N = 37$ (CI: 25 points). Hence, this larger normative reference group was thought to represent a substantial improvement relative to the previous study. Based on this, the current study aimed to collect data for a sample of 100 participants for the norming study.

The statistical package ICCSampleSize in R package version 1.0 (Rathbone, Shaw & Kumbhare, 2015) was used to identify the sample size needed for test-retest and inter-rater reliability using intra-class correlation (ICC). To detect an effect size of 0.8, a sample size of 8 was found to give 80% power to detect effects of $d = 0.8$ where $p = 0.05$.

A power analysis was conducted using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) to identify the sample size needed for parallel-form reliability. To detect a moderate effect size (0.6) a sample size of 45 for each group was found to give 80% power to detect effects of $d = 0.6$ or greater where $p = 0.05$.

A power analysis was conducted to identify the sample size required for a construct validity analysis (i.e. comparing test performance on the CVWMT-I between the clinical and non-clinical groups). To detect a moderate effect size (0.5), guided by a

preliminary study comparing scores on the ACT between controls and participants with severe TBI (Merkley, Larson, Bigler & Good, 2013), a sample size of 51 for each group was found to give 80% power to detect effects of $d = 0.5$ or greater, where $p = 0.05$.

To detect a correlation of 0.6 for convergent validity, a sample size of 47 was found to give 80% power to detect effects of $d = 0.6$ or greater where $p = 0.05$, using G*Power (Faul et al., 2007). Table 1.1 below outlines the statistical conventions for each of the tests of reliability and validity mentioned above.

Table 1.1. A-priori effect sizes and required sample sizes for each psychometric test:

Test	Analysis	Recommended effect size	<i>N</i>
Stability			
Parallel-form reliability	<i>t</i> -test (independent samples)	≥ 0.6	45x2
Test-retest reliability	ICC (mixed effects)	≥ 0.8	8
Inter-rater reliability	ICC (mixed effects)	≥ 0.8	8
Validity			
Convergent validity	Pearson's <i>r</i>	≥ 0.6	47
Construct validity	<i>t</i> -test (independent samples)	≥ 0.5	51x2

3.4 Participants

3.4.1 *Recruitment*

Participants were recruited on a voluntary basis from the local community via flyers advertising the project in local coffee shops and cafes, charity shops, GP surgeries, hospital waiting areas, supermarkets, online forums (i.e. Join Dementia Research, East Dulwich Forum, Gumtree and Craigslist) and through word of mouth. Participants were selected based on their age, gender and education level, to ensure that there was an equal spread of participants across each of these categories, in order to avoid over or under representation, which could bias the results.

Interested volunteers contacted the lead researcher via email or telephone and were then provided with an information sheet and given the opportunity for any queries to be addressed. Written informed consent was then gained from participants who wished to take part. Following this, an initial screening interview was completed via email or telephone, to ensure participants met the eligibility criteria. Four volunteers did not meet the inclusion criteria, due to a diagnosis of a major psychiatric disorder (i.e. bipolar affective disorder), history of TBI and post-concussion syndrome, history of epilepsy, and current low mood (assessed via questionnaires and brief interview; this volunteer was signposted to support via GP and IAPT services).

The lead researcher then met with participants for a 1.5-hour testing session. Written consent to audio record the administration of the CVWMT-I was gained on the day of testing. Furthermore, written consent to attend a second re-test session four weeks later was also gained on the day of testing. Participants received £10 for their participation for the first testing session. Thirty-five participants were selected for retest on the basis of convenience to the participants to attend a second testing session and were screened

to ensure variation across demographic groups. Of the 35 participants within the retest sample, half completed CVWMT-I form A at the first testing session, whilst half completed CVWMT-I form B. At retest after four weeks (Range = 3.4 – 4.4 weeks), the version not previously completed was administered. Seventeen participants completed form B first and A second, whilst 16 participants completed form A first and then form B during the retest.

3.4.2 Inclusion and exclusion criteria

Inclusion criteria for participants were as follows:

- Aged 18-80 years
- Living independently in the community or living at home for younger participants
- No history of head injury or other neurological or major psychiatric disorder
- Not currently depressed or anxious. Participants with a history of depression or anxiety or those currently on medication but not symptomatic were included.
- No current alcohol or substance use problems
- Adequate vision and hearing to complete the tests
- English spoken fluently

Exclusion criteria:

- Participants who had sustained a head injury resulting in loss of consciousness of longer than 5 minutes, requiring treatment, or with a diagnosis of post-concussion syndrome
- Participants whose daily functioning had been impaired because of a head injury

- Participants with a diagnosis of a major psychiatric disorder (e.g. bipolar affective disorder or schizophrenia)
- Participants with a diagnosis of a neurological condition (e.g. dementia, stroke, multiple sclerosis)

3.4.3 *Participant demographics*

Participants were 48 females, 47 males and one non-binary participant, ranging in age from 21 to 79 years ($M = 49.95$, $SD = 17.19$). The education level of participants ranged from secondary school, GCSE's, O-levels or below ($n = 29$) to college, A-levels or further education ($n = 33$) to degree or equivalent ($n = 34$). Six participants were currently engaged in ongoing education. Categories recommended by the Office for National Statistics (ONS; 2010) were utilised to classify participants' occupations. Participants fell within the following professional categories: higher managerial and/or professional roles ($n = 15$), lower professional or higher technical or self-employed roles ($n = 43$), intermediate roles like clerical, sales or administrative positions ($n = 23$), small industry employers like agriculturalists or craftsmen ($n = 3$), semi-routine workers like child care providers or support workers ($n = 10$) and full-time students ($n = 2$).

Current employment status of participants ranged from employed full-time ($n = 37$), employed part-time ($n = 20$), engaged in full-time or part-time education or training ($n = 6$), unemployed ($n = 5$), retired ($n = 20$) or self-employed ($n = 8$). As per the inclusion criteria, all participants were living independently in the community or at home with family, for younger participants. Table 1.2 presents detailed demographics of participants.

Table 1.2. Participant demographics (normative sample, $n = 96$)

	<i>n</i>	%
Age (in years)		
18 – 30	17	17.7%
31 – 40	17	17.7%
41 – 50	14	14.6%
51 – 60	17	17.7%
61 – 70	16	16.7%
71 – 80	15	15.6%
Gender		
Male	47	49%
Female	48	50%
Non-binary	1	1%
Biological sex		
Male	47	49%
Female	49	51%
Educational qualifications		
Degree or equivalent	34	35.4%
A-levels/ college/ further education	33	34.4%
GCSE/ O-levels/ secondary school	29	30.2%
Current employment status		
Employed full-time	37	38.5%
Employed part-time	20	20.8%
Education/ training	6	6.3%
Unemployed	5	5.2%
Retired	20	20.8%

Self-employed	8	8.3%
Professional categories (current or highest held if retired/ unemployed)		
Higher managerial/ professional	15	15.6%
Lower professional, higher technical, self-employed	43	40.6%
Intermediate (clerical, sales, administration)	23	28.1%
Small industry employers (agriculture, crafts)	3	3.1%
Semi-routine (child-care, support work)	10	10.4%
Full-time students	2	2.1%

Twenty-two participants reported having experienced depression, anxiety or another mental health condition during their lifetime. Of these participants, 11 had accessed services at some point. Of these, only two participants had received a formal diagnosis of depression ($n = 1$) and anxiety ($n = 1$). Three participants reported experiencing severe anxiety symptoms as indicated by scores above 15 on The Generalized Anxiety Disorder - 7 scale (GAD-7; Spitzer, Kroenke, Williams & Löwe, 2006) in the two weeks prior to testing, whilst three participants reported experiencing severe symptoms of depression as indicated by scores above 15 on the Patient Health Questionnaire – 8 (PHQ-8; Kroenke et al., 2009). Upon further exploration via a brief clinical interview, participants reported experiencing these symptoms in relation to family and/or work-related difficulties, which had resolved at the time of testing. None of the participants reported feeling anxious or depressed at the time of testing and did not report feeling at risk or requiring support. See Table 1.3 for further details.

Table 1.3. Participant scores on GAD-7 and PHQ-8 ($n = 96$)

	<i>M</i>	<i>SD</i>	Range	
			Minimum	Maximum
GAD-7	3.32	4.04	0	21
PHQ-8	3.40	4.19	0	23

Eight participants reported having sustained a head injury during their lifetime. None of these participants reported losing consciousness for longer than five minutes, and none had received a formal diagnosis (e.g. post-concussion syndrome, TBI). Of these, two participants required treatment following the head injury, including a visit to hospital for monitoring and dressing of superficial wounds. However, none of the participants reported any change in daily functioning following the injury.

Seventy participants required glasses and four required hearing aids. Each of these participants brought required equipment on the day of testing to ensure adequate vision and hearing to complete the test battery.

Following testing, two participants were found to have an FSIQ-2 below 70, and were hence excluded from subsequent analyses, leaving a sample of 94 participants.

The clinical sample consisted of 55 participants, who had attended the Oliver Zangwill Centre for an assessment following a TBI, resulting in impairment in everyday functioning to some extent. Each of these participants completed the CVWMT-I as part of their assessment and provided informed consent for their data to be used for the purpose of research. Demographic details for the clinical sample can be found in Table 1.4.

Table 1.4. Participant demographics (clinical sample, $n = 55$)

	<i>N</i>	%
Age (in years)		
18 – 30	27	49.1%
31 – 40	12	21.8%
41 – 50	10	18.2%
51 – 60	5	9.1%
61 – 70	1	1.8%
Biological Sex		
Male	47	85.5%
Female	8	14.5%
Educational qualifications		
Degree or equivalent	13	40%
A-levels/ college/ further education	20	36.4%
GCSE/ O-levels/ secondary school	22	23.6%

3.5 Measures

3.5.1 *Cleveland Verbal Working Memory Test with Interference* (CVWMT-I; Mack, 2016)

The CVWMT-I is a currently unpublished measure based on the BPP. There are two equivalent forms, A and B. Each form begins with a set of three practice trials in which participants count backwards from a specified number, as quickly as possible whilst avoiding errors, for a period of 15 seconds. This serves to acclimatise them to the procedure. Following completion of the practice trials, a set of 14 test trials begins, each with two components, trigram recall and reverse counting tasks together. On each test trial, participants are asked to read aloud a set of three consonants (i.e. a trigram)

presented in large bold font (Arial font size 100) on a stimulus card. The card is presented for no longer than three seconds. Following this, the stimulus card is taken away and the researcher gives the participant a number to commence backward counting verbally out loud for 15 seconds. While the participant counts, the researcher makes notes on any errors, omissions, repetitions, and the total number counted, and when 15 seconds have elapsed, asks the participants to recall the trigram (order of letters is not specified). No inter-trial intervals should be included. At the end of the test trials, participants are asked to recall as many three consonant combinations as they can. Verbatim responses are recorded on both practice and test trials (i.e. exact numbers spoken when counting backwards, exact trigrams of consonants recalled). Full instructions for administration of the CVWMT-I can be found in Appendix B.

Construction of the CVWMT-I involved the development of 28 different trigrams (14 for each of the two versions A and B), using only consonants. Each consonant appears only twice across the 28 trigrams and appears in the same position only once. In keeping with Witmer's recommendations (1935), none of the trigrams include the repetition of a letter, nor do they include letters that appear consecutively in the alphabet. Three independent judges selected trigrams from a pool which appeared to have the least potential for association with words and meaning. These were further adapted to decrease meaningfulness. Eighteen of the trigrams included in the CVWMT-I are those developed by Witmer (1935), whilst the rest are novel to this test. A full list of trigrams included are presented in Table 1.5.

Table 1.5. List of trigrams included in the CVWMT-I (versions A and B)

Trigrams	
Version A	Version B
ZBM	LVQ
FQV	BZK
HJN	YFP
YMR	RQM
SDG	NCX
KTP	HYV
WLX	SDB
CYQ	TGN
JRZ	PXH
XHL	WLR
PWF	GSJ
TKB	DKF
DVS	CTW
NGC	MJZ

With regards to the interference task, counting backwards by one was selected as appropriate for use with people with cognitive impairment, and a 15 second interference time was deemed long enough to establish interference effects, in the absence of further

guidance in this area. The length of the interference task is kept consistent throughout to allow for assessment of retentive ability in the presence of competing stimuli.

A number of interpretations can be made based on scores obtained on the CVWMT-I, including trigrams recalled in any order, trigrams recalled in correct order, total amount of numbers counted backwards, average amount of numbers counted backwards, number and type of counting errors made, number of trials with counting errors and total time taken to complete the task. Scoring instructions for the CVWMT-I can be found in Appendix C. For the purpose of this study, scores obtained on Letter Recall total (in any order) and Reverse Counting mean were selected as most relevant for analyses². See Appendix D for the full scale (versions A and B).

3.5.2 Background measures of cognitive ability

A range of measures assessing general intellectual ability and cognitive functions related to those required within the CVWMT-I were administered. These enabled the sample to be accurately characterised and for analyses of the test's validity to be undertaken. The battery comprised of:

- Test of Premorbid Functioning – UK Version (TOPF UK; Wechsler, 2011a):
This provides an estimate of optimal general intellectual ability.
- Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II; Wechsler, 2011b). These tests form the abbreviated version of this scale and are measures of verbal and perceptual

² Other measures on the CVWMT-I were considered for analyses (including Recall Order, Reverse Counting standard deviation, and Number of Counting Errors). However, results on these measures were similar to those obtained on the other measures. Hence, for ease of analyses and interpretation, two main measures, Letter Recall total and Reverse Counting mean were reported on.

reasoning, providing an overall estimate of current general intellectual ability in the form of the full-scale IQ-2 (FSIQ-2).

- Trail Making Test A and B (TMT; Reitan & Wolfson, 1985), measuring simple visuo-motor attention and attentional switching.
- Verbal Fluency - Phonemic and Semantic Fluency and switching trials of the D-KEFS (Delis-Kaplan Executive Function System; Delis, Kaplan & Kramer, 2001). This test provides a measure of verbal generativity and another index of switching ability.
- Stroop Task (Trenerry et al., 1989): This provides a measure of interference control.
- Digit Span subtest of the WAIS-IV (Wechsler, 2008): This provides a measure of working memory including storage (digits forwards) and manipulation (digits backwards and sequencing) components, along with an overall score.

3.5.3 Questionnaires

At the end of testing, participants completed the following questionnaires:

- The Dysexecutive Questionnaire Revised (DEX-R; Simblett, Ring, & Bateman, 2016), to obtain a self-reported estimate of executive abilities.
- The Patient Health Questionnaire-8 (PHQ-8; Kroenke et al., 2009) to estimate current symptoms of depression.
- The Generalised Anxiety Disorder-7 (GAD-7; Spitzer et al., 2006) to measure current symptoms of anxiety.

- A brief feedback form, consisting of 10 items, rated across a five-point Likert scale was administered at the end of the testing session, to gather participant's views of their experience of the testing session.³

4. **Results**

All data were collated and analysed using the statistical package for social sciences (SPSS, version 17.0). Further, four participants were colour-blind. Hence, these participants were excluded from any analysis involving performance on the Stroop task.

Preliminary analysis of the data via inspection of boxplots indicated that the CVWMT-I data were normally distributed. Data from the TMT A and B were negatively skewed, due to a ceiling effect. For TMT A, skewness of -3.7 ($SE = .25$) and kurtosis of -1.26 ($SE = .49$) was recorded, whilst for TMT B, skewness of -.71 ($SE = .25$) and kurtosis of -1.02 ($SE = .49$) was recorded. The TMT data were log transformed, which corrected this skew. Data from the Stroop Colour-Word (C-W) task were also negatively due to a ceiling effect, with skewness of -1.78 ($SE = .25$) and kurtosis of 2.85 ($SE = .50$).

However, no transformations applied to the Stroop task data improved the distribution of scores. Hence, non-parametric tests were used for all analyses that included this variable. Only one participant identified their gender as non-binary. Due to under-representation of this group, biological sex rather than gender was used for the purpose of analysis, in order to avoid any biases.

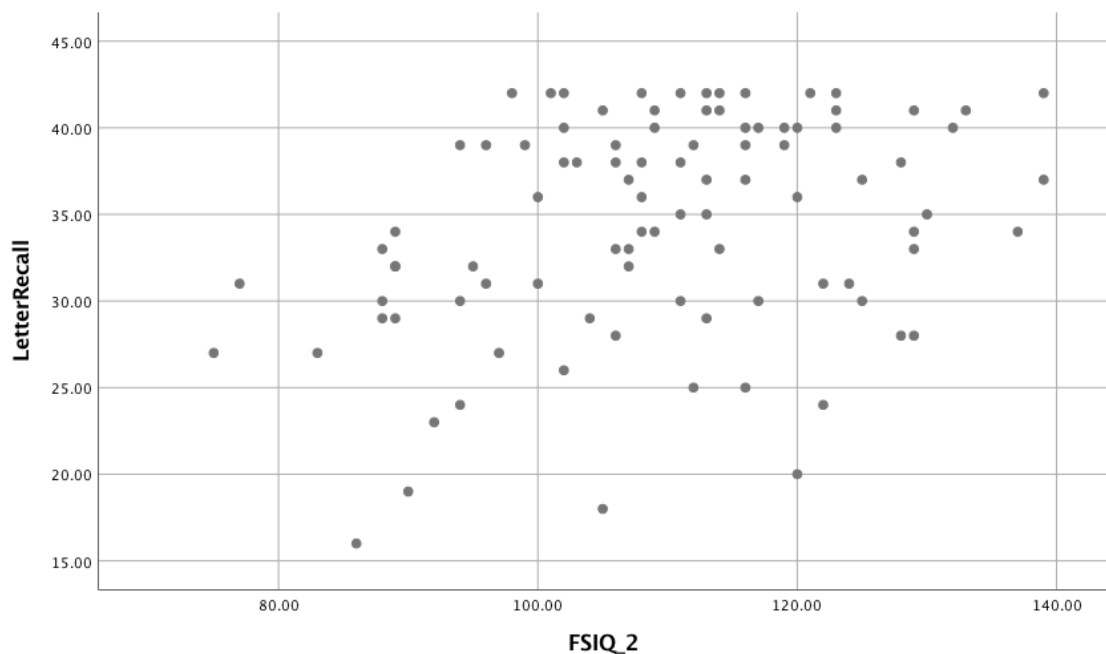
Multiple regression analysis was carried out to investigate the impact of age, biological sex and IQ (as estimated by the FSIQ-2 on the WASI-II) on performance on the

³ As all responses were close to the top of the scale, indicating a high degree of satisfaction with the process. Hence, these were not formally analysed.

CVWMT-I measures of Letter Recall and Reverse Counting. No significant effects of biological sex were identified; hence, biological sex was excluded as a variable from further analyses.

IQ and age were found to have a significant effect on performance on the CVWMT-I Letter Recall measure ($F(2,91) = 8.768, p < .001, R^2 = .162$). Together, IQ and age accounted for approximately 16% of the variability in performance across the sample, on Letter Recall. See Figure 1.1 for scatterplots and Table 1.6 for full results. Further, age and IQ were found to have a significant effect on performance on the Reverse Counting measure ($F(2,93) = 14.205, p < .001, R^2 = .238$), accounting for approximately 24% of variability in performance across the sample. See Figure 1.2 for scatterplot and Table 1.7 for full details.

Figure 1.1. Scatterplots displaying effects of IQ and age on Letter Recall



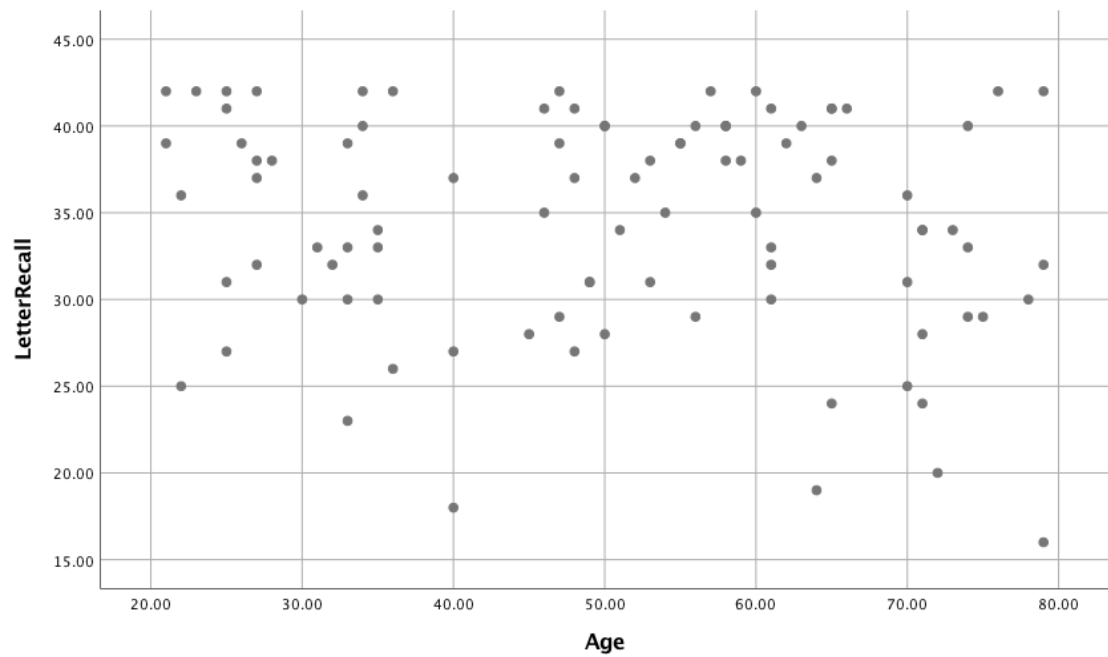


Table 1.6. Multiple regression analysis for effect of age and IQ on Letter Recall

	<i>B</i>	<i>SE B</i>	β
Step 1			
Constant	17.78	4.84	
IQ	0.15	0.04	0.34*
Step 2			
Constant	19.46	4.80	
IQ	0.17	0.04	.39*
Age	-0.08	0.04	-.22**

Note: $R^2 = .12$ for Step 1, $\Delta R^2 = .05$ for Step 2 ($p < .05$). * $p < .001$, ** $p < .05$

Figure 1.2. Scatterplots displaying effects of IQ and age on Reverse Counting

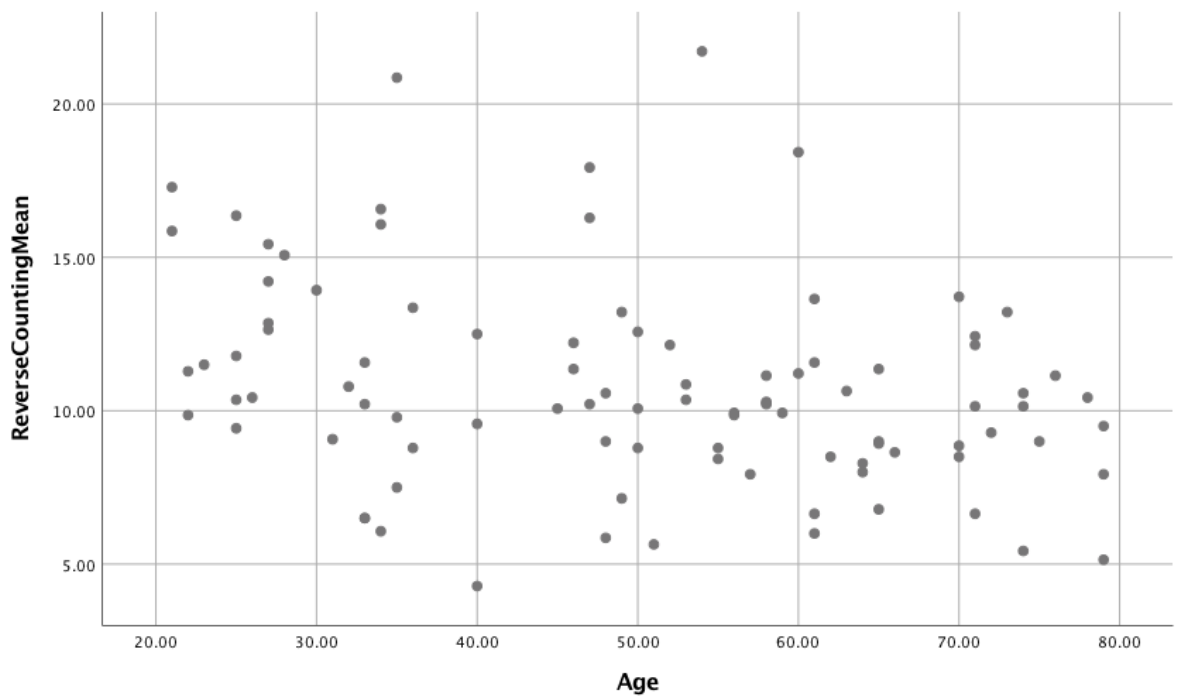
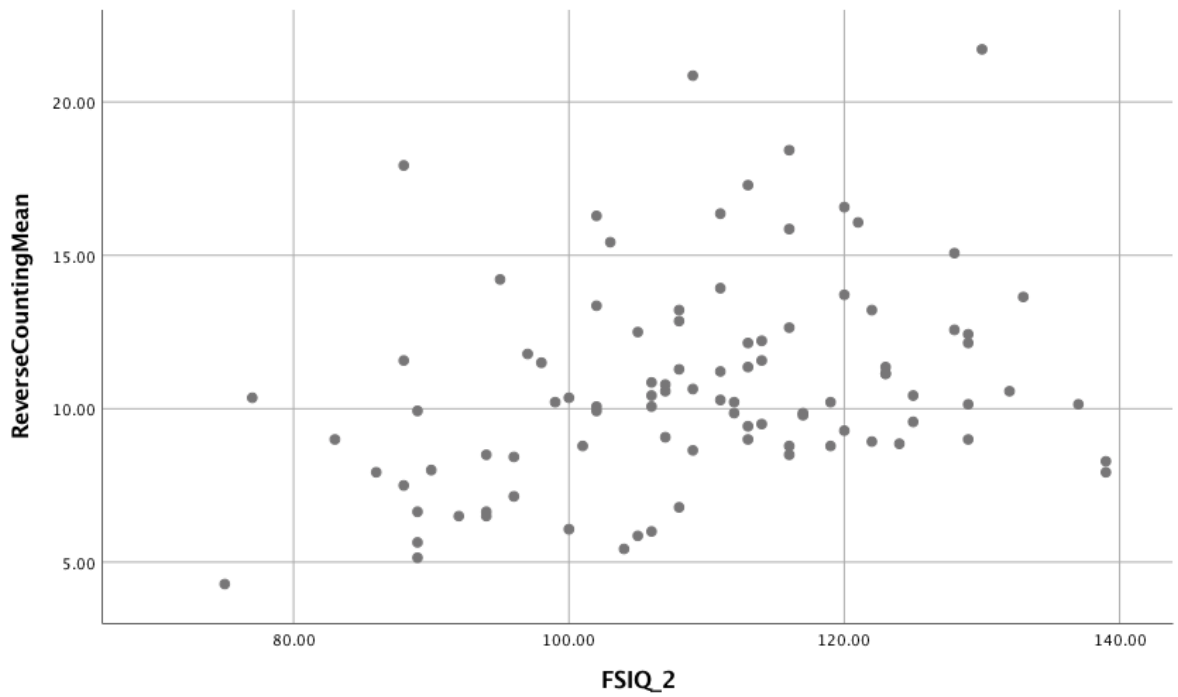


Table 1.7. Multiple regression analysis for effect of age and IQ on Reverse Counting

	<i>B</i>	<i>SE B</i>	β
Step 1			
Constant	13.76	1.03	
Age	-0.06	0.02	-.31*
Step 2			
Constant	4.41	2.44	
Age	-0.08	0.02	-.39*
IQ	0.09	0.02	.39*

Note: $R^2 = .09$ for Step 1, $\Delta R^2 = .15$ for Step 2 ($p < .05$). * $p < .001$

4.1 Scores on background neuropsychological measures and the CVWMT-I

Table 1.8 summarises scores of the full sample ($n = 96$) on each of the background questionnaires and neuropsychological measures and the CVWMT-I. Of note, the mean IQ based on reading scores was very close to the population mean (TOPF-estimated FSIQ = 103.61, SD = 8.71), indicating the intellectual ability of the sample reflected that of the broader population. However, scores for estimated current IQ using the FSIQ-2 and the other cognitive tests were slightly higher than the population mean.

Table 1.8. Scores for participants on background questionnaires and neuropsychological measures ($n = 96$)

Measure	<i>M</i>	<i>SD</i>	Range	
			Minimum	Maximum
DEX-R	30.28	17.75	1	85
GAD-7	3.42	4.09	0	21
PHQ-8	3.47	4.22	0	23
TOPF FSIQ	103.61	8.71	83	123
FSIQ-2	108.59	15.28	64	139
TMT A	54.27	29.01	10	90
TMT B	60.31	30.03	10	90
Letter Fluency	12.48	3.70	4	19
Category Fluency	12.65	3.88	3	19
Stroop C-W*	71.29	35.25	2	100
Digit Span	11.28	3.14	4	19

* $n = 92$ (four colour-blind participants excluded)

4.2 Establishing test norms

4.2.1 Regression-based norms

Significant effects of IQ and age were noted, as predictors of performance on the CVWMT-I. Hence, these factors needed to be considered when generating statistical norms. The computer programme RegBuild_MR_Raw.exe (Crawford, Garthwaite, Denham & Chelune, 2013) was used to generate regression equations using normative data from this study, which can then be applied to individual cases. The computer programme can be found at:

https://homepages.abdn.ac.uk/j.crawford/pages/dept/RegBuild_MR.htm. Appendix E

contains normative data for Letter Recall, IQ and age and Appendix F contains normative data for Reverse Counting, IQ and age, from this study, for use with the programme, within a .txt format. Only files within a .txt format will be accepted by the RegBuild_MR_Raw.exe programme. Once the normative data has been saved in this format, the data for an individual case can be added to the bottom⁴. See Appendix G for details on how to input data and use the RegBuild_MR_Raw.exe programme.

One case example is presented below, with hypothetical data for CVWMT-I performance, along with instructions in interpreting the output file. The case example is aged 22 years, with an FSIQ-2 of 100. They obtained a score of 25 on Letter Recall. This data was added to the bottom of the normative data file. See Appendix H for the full output file for this case.

The output file first provides a summary of the normative data, along with the descriptive statistics, followed by the regression equation, based on which the individual case was analysed.

The section titled 'Outputs: Result from analysis of the individual case' presents the main results of interest for the individual case. The 'case's predicted score from the regression equation' is 35.05, with a 'discrepancy (obtained minus predicted) between case's obtained and predicted scores' = -10.05, indicating that the case's score is 10 less than the estimate based on the regression equation.

⁴ Guidance on formatting of the raw data file is provided at the top of the RegBuild_MR_Raw.exe programme worksheet

The 'Effect size (Z-OP) for discrepancy between obtained and predicted scores (plus 95% CI)' is noted as $Z-OP = -1.741$ (95% CI = -2.186 to -1.252), indicating that the case example is performing close to 2 *SDs* below the normative prediction.

The next sub-section outlines the results of the significance test (i.e. *t*-test): $t(91) = -1.69$, $p = 0.047^5$, indicating that there is a significant difference between the case example's performance on Letter Recall, in comparison to the normative prediction.

The final sub-section provides an 'estimated percentage of population obtaining a discrepancy more extreme than the case' = 4.73% along with '95% confidence limits on the percentage' = 1.44% to 10.53%. This indicates that approximately 5% of the normative sample had observed-estimated discrepancies of this size.

The case example obtained a score of 18.4 on Reverse Counting. Their raw data is added to the bottom of the normative data file for Reverse Counting. Following the same steps outlined above, an output file is generated (see Appendix I), outlining the results in the same format as discussed above.

The 'case's obtained score on Task of Interest' (i.e. Reverse Counting) is 18.4, whilst 'case's predicted score from the regression equation' is 12.04, with a 'discrepancy (obtained minus predicted) between case's obtained and predicted scores' = 6.36. These scores indicate that the case example is performing better than expected, in comparison to the normative prediction.

⁵ A one-tailed test is recommended for use due to the ability to provide greater power to detect difficulties in task performance on the CVWMT-I.

The $Z\text{-OP} = 2.17$ (95% CI = 1.64 to 2.64), indicating that the case example is performing close to 2.5 *SDs* above the normative prediction.

There is a significant difference between the case example's performance on Reverse Counting, in comparison to the normative prediction, as identified by *t*-test results, $t(91) = 2.1$, $p = 0.02$ (one-tailed).

Finally, only approximately 1.9% of the population were found to have observed-predicted discrepancy scores as large as that found in the individual case, as indicated by an 'estimated percentage of population obtaining a discrepancy more extreme than the case' = 1.91 (95% CI = 0.41% - 5.07%).

4.2.2 Percentile ranks

Table 1.9 provides the percentiles of the distribution of scores on the Letter Recall measure and Reverse Counting measure of the CVWMT-I.

Table 1.9. Percentiles of the distribution of scores on Letter Recall and Reverse Counting

	Percentiles						
	5 th	10 th	25 th	50 th	75 th	90 th	95 th
Letter Recall	22.25	25.5	30	35.5	40	42	42
Reverse Counting	5.8	6.57	8.79	10.25	12.45	15.96	17.45

Percentiles of the distribution of scores on the Letter Recall measure and Reverse Counting measure of the CVWMT-I, stratified by age group, can be found in Appendix J.

These tables are included for information only. Given age and IQ were found to have significant impact on performance on both measures of the CVWMT-I, it is recommended that the RegBuild_MR_Raw.exe programme be referred to for clinical use.

4.3 Reliability

4.3.1 Parallel-form reliability

An independent samples *t*-test was carried out to establish any group difference in performance on the two versions of the CVWMT-I, versions A and B. Forty-eight participants completed version A at the initial testing session and 46 completed version B (two participants with an FSIQ-2 below 70 completed version B and were excluded from this analysis). There was no significant difference between performance on Letter Recall by participants who completed version A ($M = 34.25$, $SE = .85$) when compared to those who completed version B ($M = 34.63$, $SE = 1.00$), $t(92) = -.29$, $p = .77$. Further, there was no significant difference between performance on Reverse Counting by participants who completed version A ($M = 10.41$, $SE = .46$) as compared to those who completed version B ($M = 11.15$, $SD = .52$), $t(92) = -1.06$, $p = .29$. See Table 1.10 for details.

Table 1.10. Performance on two parallel forms of the CVWMT-I

Measure	<i>M</i> (<i>SD</i>)		<i>t</i> -test
	Version A	Version B	
Letter Recall	34.25 (5.88)	34.63 (6.78)	$t(92) = -.29$, $p = .77$
Reverse Counting	10.41 (3.16)	11.15 (3.56)	$t(92) = -1.06$, $p = .29$

4.3.2 Test-retest reliability

Two-way mixed effects ICC were calculated for both the CVWMT-I measures, to establish the consistency in participant performance on two separate testing occasions four weeks apart. Test-retest reliability for the Reverse Counting measure was high (.910, 95% CI = .822 – .954, $F(34,34) = 23.162, p < .001$). Test-retest reliability for the Letter Recall measure was low (.462, 95% CI from .157 – .687, $F(34,34) = 2.697, p = .002$).

A paired samples *t*-test identified a significant improvement in Reverse Counting scores between the first ($M = 11.71, SE = .62$) and second testing session ($M = 12.27, SE = .68, t(34) = -2.102, p = .043, r = 0.34$). An average improvement of 0.56 over time was noted, however, with a medium statistical effect size⁶ of 0.34. There was no significant improvement in Letter Recall scores between the first ($M = 35, SE = 1.09$) and second testing session ($M = 35.77, SE = .77, t(34) = -.79, p = .436, r = 0.13$). Table 1.11 presents CVWMT-I results across the first and second testing sessions.

Table 1.11. CVWMT-I scores on first and second testing sessions

Measure	<i>M (SD)</i>		<i>t</i> -test
	Time 1	Time 2	
Letter Recall	35.00 (6.42)	35.77 (4.54)	$t(34) = -.79, p = .44$
Reverse Counting	11.71 (3.67)	12.27 (4.03)	$t(34) = -2.102, p = .04$

4.3.3 Inter-rater reliability

Two-way mixed effects ICCs were calculated for each of the CVWMT-I measures, to establish the degree of agreement across raters, for data from 35 participants. Inter-rater

⁶ $r = .10$ (small effect size), $r \geq .30$ (medium effect size), $r \geq .50$ (large effect size; Cohen, 1992)

reliability was found to be high across each of the CVWMT-I measures, ranging from .999 for Reverse Counting (95% CI from .998 – 1.00, $F(34,34) = 2505.805$, $p < .001$), to .997 for Letter Recall (95% CI from .994 – .998, $F(34,34) = 621.828$, $p < .001$).

Results from each of the reliability analyses are presented in Table 1.12.

Table 1.12. Reliability coefficients for CVWMT-I measures ($n = 35$)

	Letter Recall	Reverse Counting
Test-retest reliability	.462	.910
Inter-rater reliability	.997	.999

4.4 Validity

4.4.1 Convergent validity

Pearson correlations (r) were calculated to explore the relationship between performance on neuropsychological tests in comparison to performance on the CVWMT-I.

A significant relationship was found between performance on the TMT A and Reverse Counting ($r = .386$, $p < .001$). Further, a significant relationship was found between performance on the TMT B and Reverse Counting ($r = .277$, $p < .007$).

There was a significant correlation between performance on the Letter Fluency subscale of the Verbal Fluency test and both measures on the CVWMT-I; Letter Recall ($r = .372$, $p < .001$) and Reverse Counting ($r = .441$, $p < .001$). Further, performance on the Category Fluency subtest of the Verbal Fluency test also correlated significantly with Letter Recall ($r = .349$, $p = .001$) and Reverse Counting ($r = .429$, $p < .001$) on the CVWMT-I.

Significant relationships were found between performance on Digit Span and Letter Recall ($r = .254, p = .013$), and Reverse Counting ($r = .501, p < .001$), on the CVWMT-I.

Spearman's rank-order correlations (r_s) were calculated to establish the nature of the relationship between performance on the Stroop C-W task and CVWMT-I subscales. A significant relationship was established between performance on the Stroop C-W task and the Reverse Counting measure of the CVWMT-I ($r_s = .418, p < .001$).

Full results can be found in Table 1.13.

Table 1.13. Pearson correlation matrix for performance on neuropsychological tests and CVWMT-I ($n = 94$)

Measure	Letter Recall	Reverse Counting
TOPF FSIQ	.089	.279**
TMT A	.130	.386**
TMT B	.182	.277**
Letter Fluency	.372**	.441**
Category Fluency	.349**	.429**
Digit Span	.254*	.501**
Stroop C-W***	.159	.418**

* $p < .05$, ** $p < .01$, *** Spearman's rank order correlation

4.4.2 Construct Validity

Forty participants from the normative sample were matched by age, gender and education level to 55 clinical participants from a pre-existing dataset to compare performance on the CVWMT-I between groups. An independent samples t -test was

carried out to identify any group differences between samples, for age and education level. No significant group differences were identified for age between the normative sample ($M = 38.75$, $SE = 12.49$) and clinical sample ($M = 34.04$, $SE = 1.61$), $t(93) = -1.87$, $p > .05$. Further, no significant differences were found for education level between the normative sample ($M = 2.03$, $SE = .13$) and clinical sample ($M = 1.84$, $SE = .11$), $t(93) = -1.15$, $p > .05$. Results can be found in Table 1.14.

Table 1.14. *t*-test results comparing demographics (age and education level) for the clinical and normative samples

	Clinical		Normative		<i>t</i> -test
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age	34.04	11.92	38.75	1.94	$t(93) = -1.87$, $p > .05$
	<i>n</i>	%	<i>n</i>	%	
Education					$t(93) = -1.15$, $p > .05$
Degree or equivalent	13	23.6%	13	32.5%	
A-level/ further education	20	36.4%	15	37.5%	
GCSE/ secondary school	22	40%	12	30%	

An independent samples *t*-test indicated that participants within the normative sample ($M = 34.65$, $SE = 0.97$) performed better than the clinical sample ($M = 30.53$, $SE = 1.02$) on the Letter Recall measure of the CVWMT-I, $t(93) = -2.84$, $p = .006$. However, clinical participants ($M = 11.25$, $SE = 0.48$) did not differ significantly from normative participants ($M = 11.68$, $SE = 0.63$) on the Reverse Counting measure of the CVWMT-I, $t(93) = -.559$, $p > .05$.

5. Discussion

This study has provided normative data from a UK population for the CVWMT-I measure of the impact of interference on WM (Mack, 2016). Our findings addressed first the impact of demographic factors on test performance; IQ and age were both found to predict performance on the CVWMT-I, as assessed by measures of Letter Recall and Reverse Counting, whilst biological sex was not found to be a predictor. These results are supported by previous research conducted in this area. For instance, Stuss et al. (1987) found no significant sex differences in performance on the Brown-Peterson Consonant Trigrams Auditory Memory Task (CCC; Brown, 1958; Peterson & Peterson, 1959). Moreover, in comparison to other tests of attention and WM, including the TMT (Army Individual Test Battery, 1947) and the Paced Auditory Serial-Addition Task (PASAT; Gronwall & Sampson, 1974), performance on the CCC was found to be the least affected by sex differences. Luders, Gaser, Narr & Toga (2009) corroborated these findings by studying sex differences in regional volumes of gray matter. Their findings suggested that higher volumes of gray matter predicted better performance on short-term memory tasks, however, the same was not true for performance on WM tasks, which was not mediated by sex. More recently, Solianik, Brazaitis and Skyrvydas (2016) tested this hypothesis, finding similar non-significant sex differences on WM tasks like the forced-choice recognition memory test and the free recall test (Roediger & Karpicke, 2005), among others.

With regards to age, the current study found better performance on the CVWMT-I among younger participants, as compared to older participants. Stuss et al. (1987) found contradictory results, indicating no differences across age groups on performance on the CCC. However, their oldest participants were aged only 69 years, within a small sample size of 60 participants. A more recent study looking at age effects on performance on

the BPP, across a wider age range (18 to 86 years) and a larger sample size ($n = 726$; Geurten et al., 2016) found similar findings to those of the current study, indicating significant decline in performance in older age participants. A systematic review of theoretical models of aging and memory presented overwhelming agreement across models in relation to age-related decline in WM ability (Park & Festini, 2017). Furthermore, Geurten et al. (2016) found greater age effects for performance with the inclusion of longer interference intervals (i.e. 10 and 20 seconds), as compared to no interval.

The current study also found effects of IQ on performance on the CVWMT-I. Bherer et al. (2010) noted effects of education level, as measured by years of education, on performance on the BPP, with performance improving with an increase in years of education. As a result, they advised on the importance of matching samples for future studies on education level, in order to avoid any biases. Taking this into account, the clinical analysis component of the current study matched participants on their highest qualification obtained, as a proxy for IQ (Tommasi et al., 2015). For the normative sample, FSIQ-2 scores were obtained during testing for each participant, based on the Vocabulary and Matrix Reasoning subtests of the WASI-II (Wechsler, 2011b). Participants with higher FSIQ-2 scores performed better on both measures of the CVWMT-I, and FSIQ-2 was also found to be the greater predictor of performance than age. These findings are also in conjunction with those obtained by Bherer et al. (2010), who proposed that some of the age effects on performance on the BPP were mediated by IQ. WM ability has also been found to be a strong predictor of academic achievement, even in comparison to IQ (Alloway & Alloway, 2010), suggesting that the two are strongly linked. The current study also found much poorer performance for participants with lower FSIQ-2 scores ($FSIQ-2 < 70$), as compared to those with higher

IQs, suggesting that use of this measure clinically with these groups must be done with caution, so as to avoid an overestimation of deficits. Hence, the normative sample did not include data for these participants.

Normative data has been developed for the CVWMT-I, stratifying for IQ and age. The current study used a regression-based norms calculator to allow for age to be classified as a continuous variable, due to a smaller sample size when classifying into age categories, along with the often-arbitrary nature of such categories. This was considered a more efficient method than traditional norming, as recommended by Oosterhuis et al. (2016). The ease and efficiency of using an easily accessible programme like Reg_Build_MR_Raw.exe (Crawford et al., 2013) to generate regression equations based on raw normative data provided by this study, was considered helpful for wider clinical use. A traditional norming method was also used, and normative tables offering percentile ranks for performance on the CVWMT-I, stratified by age, have also been provided, for information purposes only.

With regards to the second aim of this study relating to the assessment of reliability of the CVWMT-I, the psychometric properties of the CVWMT-I were evaluated, as this has not been done previously. The CVWMT-I displayed very high inter-rater reliability across both measures of the test (i.e. Letter Recall and Reverse Counting). Very high test-retest reliability was displayed for the Reverse Counting measure, however, participants performed significantly better on this measure at retest, with a medium effect size ($r = 0.34$). This may indicate presence of practice effects. Test-retest reliability was low for the Letter Recall measure, however, the practice effect was small ($r = 0.13$) and not statistically significant. This may indicate that although variation in performance on this measure was observed between testing occasions, it was not in a

consistent direction, with some participants performing better during the first testing session, and vice versa for other participants.

Participants were tested on alternate forms of the CVWMT-I during retest, which is one method to reduce practice effects associated with test-retest reliability (Lemay, Bédard, Rouleau & Tremblay, 2004). However, there may be several other reasons for these findings. Several tests of EF, including some sub-tests of the widely used Behavioural Assessment of the Dysexecutive Syndrome (BADS; Wilson, Emslie, Evans, Alderman & Burgess, 1996) have poor test-retest reliability (Basso, Bornstein, & Lang, 1999). In order to accurately gain an understanding of a person's EF ability, it is important to test their performance on novel tasks or within new contexts (Burgess, 1977). Hence, tests of EF are inherently not conducive to repeated administration, as the novelty of the task wears off, resulting in a reduction in their ability to assess performance accurately. Past research has suggested that, even where alternate forms are used across testing sessions, practice effects may impact on test performance, also known as the 'test sophistication effect' (Lemay et al., 2004), or the effect of participants learning new ways to cope with the novel task. Some participants in the current study provided feedback during the second testing session, indicating that they felt "more prepared" as compared to the first testing session. Others reported that they had "been practicing" and "thinking of new strategies to use" to aid their performance on the CVWMT-I, which may have resulted in some participants performing better during retest. These effects were still present despite the interval between the first and second testing sessions being approximately four weeks.

Youngjohn, Larrabee and Crook (1992) noted that test-retest reliability coefficients for divided attention tasks ranged from .49 to .72. Likewise, Hugdahl and Hammar (1997)

noted poorer test-retest reliability for performance on a divided attention condition of a dichotic listening test. Other commonly used tests of attention, like the Test of Everyday Attention (TEA; Robertson, Ward, Ridgeway, Nimmo-Smith & Anespie, 2001) has test-retest reliability coefficients ranging between 0.41 to 0.90, indicating that, for various reasons, some of which have been discussed above, measures of memory and attention may have varying reliability when administered repeatedly. Taking the above reasons into account, it is recommended that the CVWMT-I should not be administered to the same participants over short intervals of time, even if alternate forms are being administered. Hence, with clinical groups, a longer interval between testing sessions, greater than four weeks, may be recommended.

Other measures of internal consistency, including split-half reliability were not measured, due to a high degree of PI noted during administration of the CVWMT-I in prior clinical practice. Trigrams presented during early trials were noted to interfere during recall of later trigrams, as the test progressed, which may impact on the suitability of this metric in the context of this task. Furthermore, split-half reliability has been found to be problematic when measured for shorter scales, like the CVWMT-I, due to the decrease in reliability of the scale as a whole when split in half (Streiner, 2003). Hence, it was felt that this would not be a suitable measure for use with the CVWMT-I.

With regards to the third aim of this study considering the validity of the CVWMT-I, performance on the CVWMT-I measures correlated highly with performance on some of the additional neuropsychological tests administered to participants. Performance on the Letter Fluency and Category Fluency subtests of the Verbal Fluency tests correlated with both measures of the CVWMT-I. However, performance on both versions of the

TMT (A and B) and the Stroop task correlated only with the Reverse Counting measure of the CVWMT-I. This suggests that the Reverse Counting measure may be a better measure of attention, whilst the Letter Recall measure may be measuring WM, although this is difficult to say conclusively taking into account the limited neuropsychological battery used with the healthy sample within this study.

FSIQ-2 scores correlated with scores obtained on both the Letter Recall and Reverse Counting measures of the CVWMT-I. This finding is in conjunction with the previously discussed relationship between IQ and WM. However, the TOPF FSIQ was found to correlate only with the Reverse Counting measure. Further, construct validity, as measured by comparing performance on the CVWMT-I by clinical participants with a TBI and demographically-matched participants (on education level and age) from the normative sample, found significant results only on the Letter Recall measure.

Participants with TBI performed more poorly than non-TBI participants, even when IQ and age were accounted for, indicating that the Letter Recall measure of the CVWMT-I is a sensitive measure for identifying WM deficits in people with TBI, and one that displays verisimilitude. The FSIQ-2 was considered a better measure of IQ for use in the normative comparisons than the TOPF FSIQ, even though the latter may be an easier score to obtain clinically and one that estimates premorbid IQ, as opposed to IQ at the time of testing. However, this may not be so problematic, as the FSIQ-2 and TOPF FSIQ are highly correlated⁷ (i.e. $r = .86$; Wechsler, 2011b).

⁷ As tested on 182 participants.

6. Strengths and Limitations of the study

The current study had several strengths. Firstly, it gathered a larger sample to develop normative data than was previously done ($n = 37$; Mack 2016), making this a good preliminary study to explore the properties of the CVWMT-I and make recommendations for future exploration. Secondly, it included a clinical group of participants with TBI to compare performance of the normative sample, in order to identify the sensitivity of the tool to measure deficits, and hence, comment on its ecological validity. Thirdly, this study enables the use of regression-based norms, which have been found to be more efficient than traditional methods and utilises a freely accessible programme to generate regression equations, rather than doing this through a potentially much more error-prone manual process. Fourthly, this study presents normative data for a new test of WM with interference, which is easy to administer clinically and may be used as a quick estimate of functionally-relevant deficits. Finally, the selection of participants based on demographics (age, education level and sex) eliminated any possible biases related to the effects of these demographic variables on performance.

Some of the limitations of the current study include the still relatively small sample size, recruitment categories, a short retest period, and unknown effects of extraneous variables such as emotional arousal on performance. Firstly, although the normative sample gathered was 96, only 94 participants were included in the final study.

Moreover, when these were categorised by sex, age and education level, only approximately three participants represented each category. This placed limitations on the analyses that could be conducted, including the option of using the traditional norming method of presenting norms by subgroup, conclusively. Secondly, education level was used as an estimate of IQ at the recruitment stage, which may not have been

as accurate as using an IQ measure itself. However, due to time limitations, this was the only practical approach. Thirdly, although the four-week interval between first and second testing sessions served as a stringent test of reliability, this may not have been long enough to eliminate practice effects completely, as was reported by participants themselves. This may have contributed to poor test-retest reliability of the measure; however, this may be improved at longer intervals over four weeks. Finally, although mood states (anxiety and depression) were measured at the beginning of testing, state or performance anxiety during testing was not measured. This may have played a role in determining participant performance on the CVWMT-I. The sense from observations and feedback was that negative emotional reactions to testing were rare. Further, no relationship was found between scores on mood scales administered prior to testing and CVWMT-I performance. However, it must be noted that emotional arousal may have impacted performance and future studies should consider this.

7. Conclusions and future recommendations

The current study builds on previously unpublished work (Mack, 2016), providing evidence of the CVWMT-I as a robust, reliable and ecologically valid measure of WM, for use within clinical settings. It was found to be sensitive to WM deficits, as assessed by its use with clinical participants, and its scoring was found to be reliable when scored by different raters. Further, the use of regression-based norms allows for this test to be used widely, taking into account IQ and age as mediators of performance. Guidance on the process to generate regression equations and compare individual patient scores with those within the normative sample has also been provided. However, this study offers only preliminary data on the usefulness of the CVWMT-I within clinical settings. Further research on wider clinical and normative samples is called for in order to assess

clinical effectiveness of the CVWMT-I in identifying problems related to the impact of interference on WM.

8. References

- Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29.
- American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing. (1999). *Standards for educational and psychological testing*, USA.
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71-82.
- Aurtenetxe, S., Garcia-Pacios, J., del Rio, D., Lopez, M. E., Pineda-Pardo, J. A., Marcos, A., . . . Maestú, F. (2016). Interference impacts working memory in mild cognitive impairment. *Frontiers in Neuroscience, 10*, 1-9.
- Bade, S. (2010). Cognitive executive functions and work: Advancing from job jeopardy to success following a brain aneurysm. *Work, 36*(4), 389-398.
- Baddeley, A. D. (2002). The psychology of memory. In A. D. Baddeley, B. A. Wilson & M. Kopelman (Eds.), *Handbook of memory disorders, 2nd ed.* (pp. 3-15). Hove: Psychology Press.
- Baddeley, A. D. (2007). *Working memory, thought, and action*. Oxford University Press.
- Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake, & P. Shah (Eds.), *Models of working memory*. (pp. 28-61). New York: Cambridge University Press.

- Baker, S., Rogers, R., Owen, A., Frith, C., Dolan, R., Frackowiak, R., & Robbins, T. (1996). Neural systems engaged by planning: A PET study of the Tower of London Task. *Neuropsychologia*, 34, 515–526.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. *Psychological Bulletin*, 121(1), 65-94.
- Basso, M. R., Bornstein, R. A., & Lang, J. M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist*, 13, 283-292.
- Bherer, L., Belleville, S., & Peretz, I. (2001). Education, age, and the Brown-Peterson technique. *Developmental Neuropsychology*, 19(3), 237-251.
- Bigelow, J. & Poremba, A. (2014). Achilles' Ear? Inferior Human Short-Term and Recognition Memory in the Auditory Modality. *PloS ONE*, 9(2), 1-8.
- Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10, 12-21.
- Burgess, P. W., Alderman, N., Forbes, C., Costello, A., Coates, L. M-A., Dawson, D. R., ...Channon, S. (2006). The case for the development and use of “ecologically valid” measures of executive function in experimental and clinical neuropsychology. *Journal of the International Neuropsychological Society*, 12(2), 194-209.
- Burgess, P.W. (1997). *Theory and methodology in executive function research*. In P. Rabbitt (Ed.), *Methodology of frontal and executive functions* (pp. 81-115). Hove, UK: Psychology Press.

- Chao, L. L., & Knight, R. T. (1995). Human prefrontal lesions increase distractibility to irrelevant sensory inputs. *Neuroreport*, *6*(12), 1605-10.
- Chayter, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. *Neuropsychology Review*, *13*(4), 181-197.
- Chooi, W-T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531-542.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.
- Crawford, J. R., & Garthwaite, P. H. (2008). On the "optimal" size for normative samples in neuropsychology: Capturing the uncertainty associated with the use of normative data to quantify the standing of a neuropsychological test score. *Child Neuropsychology*, *14*, 99-117.
- Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2016). Task encoding across the multiple demand cortex is consistent with a frontoparietal and cingulo-opercular dual networks distinction. *The Journal of Neuroscience*, *36*(23), 6147-6155.
- Dalgleish, T., & Watts, F. N. (1990). Biases of attention and memory in disorders of anxiety and depression. *Clinical Psychology Review*, *10*, 589-604.
- Damasio, H., Grabowski, T., Frank, R., Galaburda, A. M., & Damasio, A. R. (1994). The return of Phineas Gage: Clues about the brain from the skull of a famous patient. *Science*, *264*(5162), 1102-1105.
- Dannenbaum, S., Parkinson, S., & Inman, V. (1988). Short-term forgetting: Comparisons between patients with dementia of the Alzheimer type, depressed, and normal elderly. *Cognitive Neuropsychology*, *5*(2), 213-233.

- Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System (D-KEFS)*. San Antonio, TX: Psychological Corporation.
- De Luca, C.R., Leventer, R.J. (2008). Developmental trajectories of executive functions across the lifespan. In P. Anderson, V. Anderson, & R. Jacobs (Eds.), *Executive functions and the frontal lobes: A lifespan perspective* (pp. 23-56). Washington, DC: Taylor & Francis.
- Dempster, F. N., & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educational Psychology Review, 11*, 1–88.
- Edwards, W. H. (2010). *Motor learning and control: From theory to practice*. Belmont, CA: Cengage Learning.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.
- Figueira, J. S. B., Oliveira, L., Pereira, M. G., Pacheco, L. B., Lobo, I., Motta-Ribeiro, G. C. ...& David, I. A. (2017). An unpleasant emotional state reduces working memory capacity: Electrophysiological evidence. *Social, Cognitive and Affective Neuroscience, 12*(6), 984-992.
- Finger, S., & Almli, C. R. (1988). Margaret Kennard and her "principle" in historical perspective. In S. Finger, T. E. Le Vere, C. R. Almli, & D. G. Stein (Eds.), *Brain injury and recovery: Theoretical and controversial issues* (pp. 117-132), New York: Plenum.
- Fougnie, D. Y. (2008). The relationship between attention and working memory. In Noah B. Johansen (Ed.), *New research on short-term memory* (pp. 1-45). Hauppauge: Nova Science Publishers.

- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, *133*; 101–135.
- Geurten, M., Vincent, E., Van der Linden, M., Coyette, F., & Meulemans, T. (2016). Working memory assessment: Construct validity of the Brown-Peterson Test. *Canadian Journal of Behavioural Sciences*, *48*(4), 328-336.
- Gioia, G. A., Isquith, P. K., & Guy, S. (2001). Assessment of executive functions in children with neurological impairment. In R. Simeonsson & S. Rosenthal (Eds.), *Psychological and developmental assessment: Children with disabilities and chronic conditions* (pp. 317–356). New York, NY: The Guildford Press.
- Gioia, G. A., Isquith, P.K., Guy, S., & Kenworthy, L. (2000). *BRIEF: Behavior Rating Inventory of Executive Function*. Odessa, FL: Psychological Assessment Resources.
- Gordon, R., Smith-Spark, J. H., Newton, E. J., & Henry, L. A. (2018). Executive function and academic achievement in primary school children: The use of task-related processing speed. *Frontiers in Psychology*, *9*, 582-585.
- Grant, D.A., & Berg, E.A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card sorting problem. *Journal of Experimental Psychology*, *38*, 404–411.
- Grant, K. W., & McCormack, P. D. (1969). Auditory and visual short-term memory with successive syllable presentation in both modalities. *Psychonomic Science*, *17*(6), 341-342.
- Gronwall, D. M. A., & Sampson, H. (1974). *The psychological effects of concussion*. Auckland: Oxford University Press.

- Hugdahl, K., & Hammar, A. (1997). Test-retest reliability for the consonant-vowel syllables dichotic listening paradigm. *Journal of Clinical and Experimental Neuropsychology*, *19*(5), 667-675.
- Jonides, J., & Nee, D. E. (2006). Brain mechanisms of proactive interference in working memory. *Neuroscience*, *139*(1), 181–193.
- Jung S. (2015). Ecological validity of traditional neuropsychological tests: role of memory, executive skills, and learning in predicting everyday functioning in a clinical population. *Theses and Dissertations (All)*, 908.
- Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to stroop interference. *Journal of Experimental Psychology: General*, *132*, 47-70.
- Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory and arithmetic tasks. *Psychophysiology*, *48*, 323-332.
- Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., ... Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, *46*(7), 2020-2032.
- Kroenke, K., Strine, T. W., Spitzer, R. L., Williams, J. B., Berry, J. T., & Makdad, A. H. (2009). The PHQ-8 as a measure of current depression in the general population. *Journal of Affective Disorders*, *114*(1-3), 163-73.
- Lemay, S., Bédard, M-A., Rouleau, I., & Tremblay, P-L. (2004). Practice effect and test-retest reliability of attentional and executive tests in middle-aged to elderly subjects. *The Clinical Neuropsychologist*, *18*(2), 284-302.

- Lezak, M. D. (1995). *Neuropsychological assessment*. New York, NY: Oxford University Press.
- Luders, E., Gaser, C., Narr, K. L., & Toga, A. W. (2009). Why sex matters: Brain size independent differences in gray matter distributions between men and women. *Journal of Neuroscience*, *29*(45), 14265-14270.
- Mack, A., & Rock, I. (1998). Inattention blindness: perception without attention. In Richard D. Wright (Ed.), *Visual perception* (pp. 55-76). New York, NY: Oxford University Press.
- Mack, J. L. (2016). *Verbal Working Memory with Interference*. Unpublished manuscript.
- McClelland, M. M., Cameron, C. E., Wanless, S. B., & Murray, A. (2007). Executive function, behavioural self-regulation, and social-emotional competence. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on social learning in early childhood education* (pp. 83-107). Charlotte, NC: Information Age Publishing Inc.
- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.
- Melby-Lervåg, M. & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270-291.
- Merkley, T. L., Larson, M. J., Bigler, E. D., & Good, D. A. (2013). Structural and functional changes of the cingulate gyrus following traumatic brain injury: relation to attention and executive skills. *Journal of the International Neuropsychological Society*, *19*(8), 899-910.

- Morris, R. G., & Baddeley, A. D. (1988). Primary and working memory functioning in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology*, 10(2), 279-96.
- Office for National Statistics (2010). *The National Statistics Socio-economic Classification (NS-SEC)*. London: UK.
- Oosterhuis, H. E. M., van der Ark, L. A., & Sijtsma, L. (2016). Sample size requirements for traditional and regression-based norms. *Assessment*, 23(2), 191-202.
- Parasuraman, R., & Greenwood, P. M. (1998). Selective attention in aging and dementia. In Raja Parasuraman (Ed.), *The attentive brain* (pp. 461-487). Cambridge, MA: MIT Press.
- Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology: Series B*, 72(1), 82-90.
- Perna, R., Loughan, A. R., & Talka, K. (2012). Executive functioning and adaptive living skills after acquired brain injury. *Applied Neuropsychology: Adult*, 19(4), 263-271.
- Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 58, 193-198.
- Pettigrew, C. & Martin, R. C. (2016). The role of working memory capacity and interference resolution mechanisms in task switching. *Quarterly Journal of Experimental Psychology*, 69(12), 2431-2451.
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 17(2), 132-138.

- Rabinovici, G. D., Stephens, M. L., & Possins, K. L. (2015). Executive dysfunction. *Continuum (Minneapolis, Minn.)*, 21(3 Behavioural Neurology and Neuropsychiatry), 646-659.
- Rathbone A., Shaw S., & Kumbhare, D. (2015). ICC. Sample.Size: calculation of sample size and power for ICC. R package version 1.0. <http://cran.r-project.org>Reitan.
- Reed, J. L., Gallagher, N. M., Sullivan, M., Callicott, J. H., & Green, A. E. (2017). Sex differenced in verbal working memory performance emerge at very high loads of common neuroimaging tasks. *Brain and Cognition*, 113, 56-64.
- Reitan, R. M., & Wolfson, D. (1985). *The Halstead–Reitan Neuropsychological Test Battery: therapy and clinical interpretation*. Tucson, AZ: Neuropsychological Press.
- Reteguz, J-A. (2006). Relationship between anxiety and standardized patient test performance in the medicine clerkship. *Journal of General Internal Medicine*, 21(5), 415-418.
- Robertson, I., Ward, T., Ridgeway, V., Nimmo-Smith, I., & Anespie, C. (2001). The Test of Everyday Attention. Manual. *Test Reviews*, 4(1), 51-55.
- Robertson, L. C., & Rafal, R. (2000). *Disorders of visual attention*. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences*, Second Edition (pp. 633-649). Cambridge, MA: MIT Press.
- Roediger, H. L., & Karpicke, J. D. (2005). *Learning and Memory*. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp.479-486). California, LA: Elsevier Science.

- Sakai, K., Rowe, J.B., & Passingham, R.E. (2002). Parahippocampal reactivation signal at retrieval after interruption of rehearsal. *Journal of Neuroscience*, 22(15), 6315–6320.
- Sbordone, R. J. (1996). *Ecological Validity: Some Critical Issues for the Neuropsychologist*. In R. J. Sbordone & C. J. Long (Eds.), *Ecological validity of neuropsychological testing* (pp. 15-42). Delray Beach, FL.: GR Press/St. Lucie Press.
- Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organisation of behaviour. *Philosophical Transactions of the Royal Society B. Biological Sciences*, 351, 1405-1412.
- Shura, R.D., Rowland, J.A., and Miskey, H.M. (2016). Auditory Consonant Trigrams: A Psychometric Update. *Archives of Clinical Neuropsychology*, 31, 47-57.
- Simblett, S. K., Ring, H., & Bateman, A. (2016). The Dysexecutive Questionnaire Revised (DEX-R): An extended measure of everyday dysexecutive problems after acquired brain injury. *Neuropsychological Rehabilitation*, 19, 1-18.
- Snowden, J. (2008). Frontotemporal dementia: Correlations between pathology and function. In P. Anderson, V. Anderson, & R. Jacobs (Eds.), *Executive functions and the frontal lobes: A lifespan perspective*. (pp.345-364). Washington, DC: Taylor & Francis.
- Solianik, R., Brazaitis, M., & Skurvydas, A. (2016). Sex-related differences in attention and memory. *Medicina*, 52(6), 372-377.
- Spitzer, R. L., Kroenke, K., Williams, J. B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166, 1092-97.

- Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, 80(1), 99-103.
- Stuss, D. T., & Benson, D. F. (1984). Neuropsychological studies of the frontal lobes. *Psychological Bulletin*, 95(1), 3-28.
- Stuss, D. T., & Levine, B. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, 53, 401-433.
- Stuss, D.T., Stethem, L.L., Hugenholtz, H., and Richard, M.T. (1989). Traumatic brain injury: A comparison of three clinical tests, and analysis of recovery. *Clinical Neuropsychologist*, 3(2), 145-156.
- Stuss, D. T., Stethem, L. L. & Poirier, C. A. (1987) Comparison of three tests of attention and rapid information processing across six age groups. *Clinical Neuropsychologist*, 1(2), 139-152.
- Tommasi, M., Pezzuti, L., Roberto, C., Abad, F. J., Saggino, A., & Orsini, A (2015). Increased educational level is related with higher IQ scores but lower g-variance: Evidence from the standardization of the WAIS-R for Italy. *Intelligence*, 50, 68-74.
- Trener, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test*. Odessa, FL: Psychological Assessment Resources.
- United States Army (1947). *Army individual test battery: Manual of directions and scoring*. Washington, D.C. Army Adjutant General's Office.
- Vakil, E. (2011). Neuropsychological assessment: Principles, rationale, and challenges. *Journal of Clinical and Experimental Neuropsychology*, 34(2), 135-150.

- Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory: A meta-analysis. *Cognitive, Affective and Behavioural Neuroscience*, 3(4), 255-274.
- Wechsler, D. (2008). *Wechsler Adults Intelligence Scale, Fourth Edition (WAIS-IV)*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2011a). *Test of Premorbid Functioning – UK Version (TOPF UK)*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (2011b). *Wechsler Abbreviated Intelligence Scale, Second Edition (WASI-II)*. San Antonio, TX: NCS Pearson.
- Wilson, B. A., Emslie, H., Evans, J. J., Alderman, N. & Burgess, P. W. (1996). *Behavioural Assessment of the Dysexecutive Syndrome (BADS)*. Bury St Edmunds, UK: Thames Valley Test Company.
- Winocur, G., Oxbury, S., Roberts, R., Agnetti, V, & Davis, C. (1984). Amnesia in a patient with bilateral lesions to the thalamus. *Neuropsychologia*, 22(2), 123-143.
- Witmer, L.R. (1935). The association value of three-place consonant syllables. *The Pedgogical Seminary and Journal of Genetic Psychology*, 47, 337-360.
- Yerys, B. E., Wallace, G. L., Jankowski, K. F., Bollich, A., & Kenworthy, L. (2011). Impaired consonant trigrams test (CTT) performance relates to everyday working memory difficulties in children with autism spectrum disorders. *Child Neuropsychology*, 17(4), 391-9.
- Youngjohn, J. R., Larrabee, G. J., & Crook, T. H. (1992). Test-retest reliability of computerized, everyday memory measures and traditional memory tests. *The Clinical Neuropsychologist*, 6(3), 276-286.

Zelazo, P. D., Carter, A., Reznick, J., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1*(2),198-226.

9. Appendices

Appendix A: Ethical approval letter from PNM RESC at KCL

Research Ethics
Office

Franklin Wilkins Building
5.9 Waterloo Bridge Wing
Waterloo Road
London SE1 9NH
Telephone 020 7848 4020/4070/4077
rec@kcl.ac.uk



Aparna Kapoor

14 February 2018

Dear Aparna,

LRS-17/18-5157

I am pleased to inform you that full approval for your project has been granted by the PNM Research Ethics Panel

- Ethical approval is granted for a period of **three years** from 14 February 2018
- You should report any untoward events or unforeseen ethical problems to the panel Chair, via the Research Ethics Office, within a week of occurrence. Information about the panel may be accessed at:

- <http://www.kcl.ac.uk/innovation/research/support/ethics/committees/ssh/reps/index.aspx>

If you wish to change your project or request an extension of approval, please complete and submit a Modification Request to crec-lowrisk@kcl.ac.uk.

Please quote your ethics reference number, found at the top of this letter, in all correspondence with the Research Ethics Office. Details of how to complete a modification request can be found at:

<http://www.kcl.ac.uk/innovation/research/support/ethics/applications/modifications.aspx>

All research should be conducted in accordance with the King's College London *Guidelines on Good Practice in Academic Research* available at:

<http://www.kcl.ac.uk/college/policyzone/assets/files/research/good%20practice%20Sept%2009%20FINAL.pdf>

Please note that we may, for auditing purposes, contact you to ascertain the status of your research.

We wish you every success with your research.

Best wishes,

Mr James Patterson

Senior Research Ethics Officer

For and on behalf of: PNM Research Ethics Panel

Appendix B: Administration instructions for the CVWMT-I

VERBAL WORKING MEMORY WITH INTERFERENCE

INSTRUCTIONS

Practice Trials.

Begin by saying:

I'M GOING TO SAY A NUMBER. AS SOON AS I SAY THE NUMBER, I WANT YOU TO BEGIN COUNTING BACKWARD FROM THAT NUMBER, LIKE THIS: IF I SAID, "28," YOU WOULD COUNT: 27, 26, 25, 24, 23, 22, 21, 20, 19, AND SO ON. BEGIN COUNTING THE MOMENT I SAY THE NUMBER, AND COUNT AS QUICKLY AS YOU CAN. DO YOU HAVE ANY QUESTIONS?

Give further explanation or examples as needed. When the subject appears to understand, begin practice trials. For each trial use the stimulus specified on the Response and Scoring Form. Each trial consists of 15 seconds of counting backward from the stimulus number.

Then say:

READY, 84 (or number for the current trial). At the end of the 15 second period say: **STOP.**

On both practice and test trials, record the subject's counting backward performance so the subject's exact performance can be reconstructed. If he or she counts backward without error, place after the number for the item the number the first and last counted numbers during the 15 second period separated by a hyphen (e.g., "84- 71"). If the subject makes an error, the error should be recorded. The most common errors are omissions, which can be recorded simply by placing a "/" in the place of the omitted number or numbers and then recording the next number the individual says. For example, if in counting from 84 to 71 the subject omitted the number 79, one should

record "84-80/78-71." When subjects repeat numbers, one should indicate the repeats as follows: "84-78, 78-71 (78 was repeated). Complex errors should be recorded verbatim.

Monitor the individual's performance carefully during practice trials to ensure that he or she understands the task and is counting as fast as possible. It is important to ascertain how quickly the individual can count backward so that on test trials one can determine whether he or she is slowing down while, presumably, silently repeating the letters. Most individuals quickly understand the task and complete the practice trials without great difficulty, though occasional counting errors are not unusual. If an individual is confused about what is required or seems unable to count backward, make up additional two-digit trials to ensure the individual is able to perform the task. In such a case additional practice trials may be carried out until the subject grasps the task. If the individual remains unable to count backward, modify the instructions so that he or she carries out the practice and test trials counting forward, although the trigram recall performance may be interpretable only if the accuracy score is low.

Test Trials.

When the practice trials have been completed, the examiner should say:

NOW I'M GOING TO ADD SOMETHING. I'M GOING TO SHOW YOU A CARD WITH THREE LETTERS ON IT... I WANT YOU TO QUICKLY READ THE LETTERS OUT LOUD... AS SOON AS YOU'VE READ THE LETTERS ALOUD, I'LL TAKE THE CARD AWAY AND GIVE YOU A NUMBER. YOU BEGIN COUNTING BACKWARD FROM THAT NUMBER AS QUICKLY AS YOU CAN, JUST AS YOU'VE BEEN DOING. BUT THIS TIME WHEN I SAY STOP, I WANT YOU TO TELL ME THE LETTERS YOU READ ALOUD BEFORE.

Pause to let the individual reflect. Then say:

LET ME EXPLAIN... I WANT TO SEE IF THE LETTERS WILL REMAIN IN YOUR MIND, EVEN THOUGH YOU'VE BEEN COUNTING BACKWARD. DON'T REPEAT THE LETTERS TO YOURSELF WHILE YOU'RE COUNTING. CONCENTRATE ON COUNTING BACKWARD AS QUICKLY AS YOU CAN UNTIL I TELL YOU TO STOP. THEN TELL ME THE LETTERS. IF YOU'RE NOT SURE, TAKE A GUESS. ALWAYS TELL ME THREE LETTERS EACH TIME I TELL YOU TO STOP.

Pause to give the individual time to grasp the task. Then say:

FIRST, I'LL SHOW YOU THE CARD WITH THREE LETTERS ON IT. YOU READ THE LETTERS QUICKLY OUT LOUD. THEN, I'LL TAKE THE CARD AWAY AND GIVE YOU A NUMBER... YOU BEGIN COUNTING BACKWARD FROM THAT NUMBER AS QUICKLY AS YOU CAN UNTIL I TELL YOU TO STOP. THEN, TELL ME THE LETTERS.... DO YOU UNDERSTAND?

Explain further if necessary. As soon as the individual seems to understand the task, begin by starting a stop watch to time to the total length of time on the task and then immediately present the first stimulus card. Use another stop watch to time the 15 second intervals of counting backward. Present each stimulus card (with the letters) for three seconds.

While administering the test, watch for attempts by the subject to delay the task in a manner that facilitates storing the letters in secondary memory, and take steps to minimize such tactics. For example, if an individual reads the letters slowly (presumably to rehearse the trigram or think of a mnemonic aid, the examiner should hurry the subject by removing the stimulus card (being sure he or she does say the letters aloud) and indicating the importance of saying the letters quickly. Delay when recalling the letters, of course, does not help and is not discouraged.

At the end of the last test trial record Total Time on Task, turn over Response Form, and say:

NOW TRY TO REMEMBER AS MANY THREE LETTER COMBINATIONS FROM THE TEST AS BEST YOU CAN. IT'S ALL RIGHT IF YOU CAN'T REMEMBER ANY, BUT I'M INTERESTED IN WHETHER ANY OF THEM ARE STILL IN YOUR MIND.

After subject is unable to recall further letter combinations, for each combination ask subject why those letters were recalled (i.e. determine the association). Record subject's explanation.

Appendix C: Scoring instructions for the CVWMT-I

VERBAL WORKING MEMORY WITH INTERFERENCE

SCORING

Accuracy. Accuracy in recalling the letters is the critical score for this test. If a letter is recalled, it is scored + regardless of whether it was recalled in the order in which it was presented. Each row of blanks in the Correct Response by Position columns corresponds to a letter recalled, i.e., the first blank is for the first letter recalled, and so on. That is, the scores for Correct Responses by Position refer to the position in which a letter was presented, not its position at recall. Thus, a "+" in the first blank space in the Correct Responses by Position columns indicates that the first stimulus letter presented was recalled correctly, even if it was not recalled as the first letter. Each of the three columns is summed in the cells for Total Correct Recall by Position, and the three totals in turn are summed in the cell Grand Total Accuracy, which has a maximum score of 42.

Number of Items Counted Backward per Trial. The purpose of this score is to indicate the extent to which the subject was successfully distracted. Any number spoken by the subject is, in effect, considered a distraction. Scoring the number of items counted is easy when a subject makes no errors in counting. When errors occur, however, one must decide whether a particular number is to be counted. In general, intrusions of incorrect numbers are counted as additional numbers, regardless of their erroneous nature. They are counted since even an error presumably reflects a subject's focus of attention and thus constitutes a distraction. Thus, if a subject counted, "8382-81/79-78/75/40/93-92-91/38," ("/" indicates a break in the sequence of correctly counted numbers) the total number of items counted would be 11; the fact that 81/79, 78/75, 75/40, 40/43, and 41/38 were errors would not affect the total items counted (although the trial would be counted as a counting trial with an error).

There is, however, an exception to the rule that erroneous numbers are counted as if they were correctly counted: If a subject immediately repeats a single number he or she has just said, no matter how many times, do not include this (these) repetition(s) in the total numbers counted. Thus, if a subject counts, "83-82/82-81-80/80/80-79-78-77-76-75-74-73-72-71," the total would be 13; the immediately repeated numbers, 82 and 80, are not counted. On the other hand, if the subject counted, "83-82-81-80/82-81-80-79-78-77-76-75-74-73-72-71," the total number of items counted would be 16, since the second "82" and "81" were not immediate repeats of a single number. The numbers of items counted backward for each of the fourteen trials are totaled and the mean and s (using $N-1$ as an error term) are calculated.

Number of Trials with Counting Errors. The number of trials on which one or more errors in counting backward was made is also recorded. Note that no matter how many errors are made on a single trial, it is counted as only one error. The maximum number of trials with an error is 14.

Time on Task. The total time on task extends from the moment the first test stimulus is presented until the response to the 14th test trial is produced. Time on task is a check on the extent to which the subject is attempting to use mnemonic strategies to store the stimuli in secondary memory and thus compensate for working memory deficits. It is important, however, to note the basis of unusually long times. If the time is great because the subject took a long time to retrieve the letters, the accuracy score is not called into question. Only when the elongated time was produced by the subject extending the time to read the letters aloud can one conclude that he or she was attempting to focus on the letter storage task while counting backward. Thus, the examiner should keep careful notes regarding the subject's behavior during testing.

Trigrams Recalled. The final information recorded is the number of three-letter or even correct two-letter combinations recalled after the last test trial. This information is

not formally scored but is recorded as a further check regarding an individual's tendency to use verbal mnemonics to facilitate storage of the letters in secondary memory. Without mnemonic storage strategies individuals are likely to recall only the last and, perhaps, the next to last trigram they produced. Most subjects usually produce one or two other associations to earlier trigrams. Occasionally, however, a subject is quite adept at coming up with mnemonic associations, so that the test may not be a sensitive measure of his or her ability to carry on two tasks in primary memory. The number of trigrams recalled is a check on this type of performance. Once an individual has produced all the three- or two-letter combinations he or she can, the examiner should inquire how each of the correct three- or two-letter combinations were remembered. Do not inquire if only one letter is recalled.

Appendix D: The Cleveland Verbal Working Memory Test with Interference

(CVWMT-I; Mack, 2016) Versions A and B

VERBAL PRIMARY MEMORY WITH INTERFERENCE

Name _____

Form A (v. 12-94)

Date _____ Exmnr _____

Practice Trials	Counting Backwards	Total Correctly Counted
a	84 -	
b	291 -	
c	816 -	

Start timer for total time on task when first test stimulus is presented.

Trial	Counting Backwards	Total Counted Correctly	Stim	Stimulus Recall	Correct Responses by Position		
					1	2	3
1	630 -		ZBM				
2	243 -		FQV				
3	804 -		HJN				
4	523 -		YMR				
5	389 -		SDG				
6	972 -		KTP				
7	565 -		WLX				
8	179 -		CYQ				
9	417 -		JRZ				
10	758 -		XHL				
11	186 -		PWF				
12	647 -		TKB				
13	962 -		DVS				
14	491 -		NGC				

Mean Number Counted per Trial

Correct Recall by Position

Stop timer for time on task immediately after subject's last response. Then ask subject to recall all three-letter combinations he or she can. When all recalls are recorded, then inquire the mnemonic for each.

s

Accuracy Grand Total

Trigrams Recalled Mnemonics

Total Time in Seconds

Number of Trials with Counting Errors

VERBAL PRIMARY MEMORY WITH INTERFERENCE

Name _____

Form B (v. 12-94)

Date _____ Exmnr _____

Practice Trials	Counting Backwards	Total Correctly Counted
a	84 -	
b	291 -	
c	816 -	

Start timer for total time on task when first test stimulus is presented.

Trial	Counting Backwards	Total Counted Correctly	Stim	Stimulus Recall	Correct Responses by Position		
					1	2	3
1	630 -		LVQ				
2	243 -		BZK				
3	804 -		YFP				
4	523 -		RQM				
5	389 -		NCX				
6	972 -		HYV				
7	565 -		SDB				
8	179 -		TGN				
9	417 -		PXH				
10	758 -		WLR				
11	186 -		GSJ				
12	647 -		DKF				
13	962 -		CTW				
14	491 -		MJZ				

Mean Number Counted per Trial

Correct Recall by Position

Total Correctly Recalled

Stop timer for time on task immediately after subject's last response. Then ask subject to recall all three-letter combinations he or she can. When all recalls are recorded, then inquire the mnemonic for each.

Trigrams Recalled Mnemonics

Total Time in Seconds

Number of Trials with Counting Errors

Appendix E: Raw data for Letter Recall, FSIQ-2 and age for the normative sample

(N.B. These data should be pasted into a document with .txt file extension, adding the comparison case's score, IQ estimate and age, in order for them to be used in the programme RegBuild_MR_Raw.exe)

39 112 47
36 120 70
40 123 58
37 116 27
41 113 25
42 111 25
40 102 50
27 97 25
31 77 53
34 108 73
38 128 28
40 119 58
42 108 27
26 102 36
38 106 53
34 109 35
30 94 33
40 109 63
33 88 35
29 113 75
16 86 79
32 95 27
39 94 62
38 111 58
42 113 21
34 137 71
29 89 56
27 83 48
40 120 34
42 121 34
28 129 71
35 130 54
32 107 32
42 102 47
42 98 23
39 116 21
39 99 33
36 108 22
41 123 65
38 102 59
34 129 71
41 133 61
41 129 65
33 107 31

24 122 65
35 111 60
29 104 74
40 132 74
42 116 60
25 116 70
41 114 46
42 123 76
37 125 40
42 114 79
28 106 45
41 109 66
28 128 50
20 120 72
37 139 64
33 129 74
34 89 51
42 139 57
19 90 64
30 125 78
39 119 55
32 89 79
24 94 71
31 124 70
38 108 65
37 107 48
25 112 22
36 100 34
37 113 52
41 105 48
31 96 49
35 113 46
30 117 35
27 75 40
42 101 36
23 92 33
31 122 49
38 103 27
39 106 26
33 114 33
18 105 40
40 116 50
30 88 61
40 117 56
30 111 30
31 100 25
29 88 47
39 96 55
32 89 61
33 106 61

Appendix F: Raw data for Reverse Counting, FSIQ-2 and age for the normative

sample (N.B. These data should be pasted into a document with .txt file extension,

adding the comparison case's score, IQ estimate and age, in order for them to be used in

the programme RegBuild_MR_Raw.exe)

10.21 112 47
13.71 120 70
11.14 123 58
12.64 116 27
9.43 113 25
16.36 111 25
10.07 102 50
11.79 97 25
10.36 77 53
13.21 108 73
15.07 128 28
10.21 119 58
12.86 108 27
13.36 102 36
10.86 106 53
20.86 109 35
6.5 94 33
10.64 109 63
7.5 88 35
9 113 75
7.93 86 79
14.21 95 27
8.5 94 62
10.29 111 58
17.29 113 21
10.14 137 71
9.93 89 56
9 83 48
16.57 120 34
16.07 121 34
12.14 129 71
21.71 130 54
10.79 107 32
16.29 102 47
11.5 98 23
15.86 116 21
10.21 99 33
11.29 108 22
11.36 123 65
9.93 102 59
12.43 129 71
13.64 133 61
9 129 65
9.07 107 31

8.93 122 65
11.21 111 60
5.43 104 74
10.57 132 74
18.43 116 60
8.5 116 70
12.21 114 46
11.14 123 76
9.57 125 40
9.5 114 79
10.07 106 45
8.64 109 66
12.57 128 50
9.29 120 72
8.29 139 64
10.14 129 74
5.64 89 51
7.93 139 57
8 90 64
10.43 125 78
8.79 119 55
5.14 89 79
6.64 94 71
8.86 124 70
6.79 108 65
10.57 107 48
9.86 112 22
6.07 100 34
12.14 113 52
5.86 105 48
7.14 96 49
11.36 113 46
9.79 117 35
4.29 75 40
8.79 101 36
6.5 92 33
13.21 122 49
15.43 103 27
10.43 106 26
11.57 114 33
12.5 105 40
8.79 116 50
11.57 88 61
9.86 117 56
13.93 111 30
10.36 100 25
17.93 88 47
8.43 96 55
6.64 89 61
6 106 61

Appendix G: Instructions on using the RegBuild_MR_Raw.exe. programme

On the worksheet of the RegBuild_MR_Raw.exe. programme, the ‘number of predictor (i.e. X) variables’ is noted as 2 (i.e. FSIQ-2 and age). ‘Sample size (N) of sample providing the summary data’ is 94 (i.e. the size of the normative sample in this study). Select the ‘95% credible limit’ required (two-sided is recommended). Then, the ‘continue’ tab is clicked, which opens a window to select the .txt file containing the normative data, as well as data for the individual participant at the bottom. The file is selected and opened, followed by a confirmation message from the programme stating the number of cases within the data file, as well as the number of variables. Clicking ‘OK’ gives us the output sheet for the analysis.

Appendix H: Output for case example's performance on Letter Recall using RegBuild_MR_Raw.exe

RegBuild_MR_Raw.exe: Builds a multiple regression equation and uses it to make inferences concerning a case

THIS PROGRAM IMPLEMENTS STATISTICAL METHODS DEVELOPED IN THE FOLLOWING PAPER:

Crawford, J.R., Garthwaite, P.H., Denham, A.K., & Chelune, G.J. (2012). Using regression equations built from summary data in the psychological assessment of the individual case: Extension to multiple regression. *Psychological Assessment*, 24, 801-814. (doi: 10.1037/a0027699).

INPUTS :

Number of predictor (i.e., X) variables = 2
 Sample size (n) for sample providing the summary data = 94
 Credible limit required: Two-sided

Raw data for controls:

[1]:	39.00000	112.00000	47.00000
[2]:	36.00000	120.00000	70.00000
[3]:	40.00000	123.00000	58.00000
[4]:	37.00000	116.00000	27.00000
[5]:	41.00000	113.00000	25.00000
[6]:	42.00000	111.00000	25.00000
[7]:	40.00000	102.00000	50.00000
[8]:	27.00000	97.00000	25.00000
[9]:	31.00000	77.00000	53.00000
[10]:	34.00000	108.00000	73.00000
[11]:	38.00000	128.00000	28.00000
[12]:	40.00000	119.00000	58.00000
[13]:	42.00000	108.00000	27.00000
[14]:	26.00000	102.00000	36.00000
[15]:	38.00000	106.00000	53.00000
[16]:	34.00000	109.00000	35.00000
[17]:	30.00000	94.00000	33.00000
[18]:	40.00000	109.00000	63.00000
[19]:	33.00000	88.00000	35.00000
[20]:	29.00000	113.00000	75.00000
[21]:	16.00000	86.00000	79.00000
[22]:	32.00000	95.00000	27.00000
[23]:	39.00000	94.00000	62.00000
[24]:	38.00000	111.00000	58.00000
[25]:	42.00000	113.00000	21.00000
[26]:	34.00000	137.00000	71.00000
[27]:	29.00000	89.00000	56.00000
[28]:	27.00000	83.00000	48.00000
[29]:	40.00000	120.00000	34.00000
[30]:	42.00000	121.00000	34.00000
[31]:	28.00000	129.00000	71.00000
[32]:	35.00000	130.00000	54.00000
[33]:	32.00000	107.00000	32.00000

[34]:	42.00000	102.00000	47.00000
[35]:	42.00000	98.00000	23.00000
[36]:	39.00000	116.00000	21.00000
[37]:	39.00000	99.00000	33.00000
[38]:	36.00000	108.00000	22.00000
[39]:	41.00000	123.00000	65.00000
[40]:	38.00000	102.00000	59.00000
[41]:	34.00000	129.00000	71.00000
[42]:	41.00000	133.00000	61.00000
[43]:	41.00000	129.00000	65.00000
[44]:	33.00000	107.00000	31.00000
[45]:	24.00000	122.00000	65.00000
[46]:	35.00000	111.00000	60.00000
[47]:	29.00000	104.00000	74.00000
[48]:	40.00000	132.00000	74.00000
[49]:	42.00000	116.00000	60.00000
[50]:	25.00000	116.00000	70.00000
[51]:	41.00000	114.00000	46.00000
[52]:	42.00000	123.00000	76.00000
[53]:	37.00000	125.00000	40.00000
[54]:	42.00000	114.00000	79.00000
[55]:	28.00000	106.00000	45.00000
[56]:	41.00000	109.00000	66.00000
[57]:	28.00000	128.00000	50.00000
[58]:	20.00000	120.00000	72.00000
[59]:	37.00000	139.00000	64.00000
[60]:	33.00000	129.00000	74.00000
[61]:	34.00000	89.00000	51.00000
[62]:	42.00000	139.00000	57.00000
[63]:	19.00000	90.00000	64.00000
[64]:	30.00000	125.00000	78.00000
[65]:	39.00000	119.00000	55.00000
[66]:	32.00000	89.00000	79.00000
[67]:	24.00000	94.00000	71.00000
[68]:	31.00000	124.00000	70.00000
[69]:	38.00000	108.00000	65.00000
[70]:	37.00000	107.00000	48.00000
[71]:	25.00000	112.00000	22.00000
[72]:	36.00000	100.00000	34.00000
[73]:	37.00000	113.00000	52.00000
[74]:	41.00000	105.00000	48.00000
[75]:	31.00000	96.00000	49.00000
[76]:	35.00000	113.00000	46.00000
[77]:	30.00000	117.00000	35.00000
[78]:	27.00000	75.00000	40.00000
[79]:	42.00000	101.00000	36.00000
[80]:	23.00000	92.00000	33.00000
[81]:	31.00000	122.00000	49.00000
[82]:	38.00000	103.00000	27.00000
[83]:	39.00000	106.00000	26.00000
[84]:	33.00000	114.00000	33.00000
[85]:	18.00000	105.00000	40.00000
[86]:	40.00000	116.00000	50.00000
[87]:	30.00000	88.00000	61.00000
[88]:	40.00000	117.00000	56.00000
[89]:	30.00000	111.00000	30.00000
[90]:	31.00000	100.00000	25.00000
[91]:	29.00000	88.00000	47.00000

```

[ 92]:      39.00000      96.00000      55.00000
[ 93]:      32.00000      89.00000      61.00000
[ 94]:      33.00000     106.00000      61.00000

```

INPUTS: Summary statistics for sample providing the data together with scores of the case:

Measure	Control	Control
Scores	Mean	SD
for Case		
Criterion (Y) : 25.000	34.436	6.305
Predictor (X) 1 : 100.000	109.500	14.089
Predictor (X) 2 : 22.000	50.106	17.186

INPUTS: Correlation(s) between criterion and predictor(s) in sample:

```

1.00000  0.33996  -0.13479
0.33996  1.00000  0.21907
-0.13479  0.21907  1.00000

```

OUTPUTS: Regression equation (alpha & betas) and squared semi-partial correlation for predictors:

Predictor	Beta	Squared
semi-partial		correlation
Intercept (alpha): -	19.457	
Predictor (X) 1 : 0.151	0.174	
Predictor (X) 2 : 0.048	-0.081	

OUTPUTS: FURTHER RESULTS FOR THE MULTIPLE REGRESSION MODEL:
 Standard error of estimate for regression equation =
 5.836
 Multiple R for regression equation =
 0.402
 R Squared for regression equation =
 0.162
 Adjusted (shrunken) R Squared for regression equation =
 0.143
 Significance test for overall regression: F [2, 91] =
 8.7685
 Significance test for overall regression: p value =
 0.0003

OUTPUTS: RESULTS FROM ANALYSIS OF THE INDIVIDUAL CASE:

Case's OBTAINED score on Task of Interest = 25.0000
 Case's PREDICTED score from regression equation = 35.0527
 Discrepancy (obtained minus predicted) between case's obtained and
 predicted scores = -10.0527

Effect size (Z-OP) for discrepancy between obtained and predicted scores (plus 95% CI):

Effect size (Z-OP) = -1.741 (95% CI = -2.186 to -1.252)

Standard error for an additional (i.e., N + 1th) case = 5.9525

Significance test (t) on the discrepancy between the case's obtained and predicted scores:

t value (on 91 df) = -1.6888

One-tailed probability = 0.0473

Two-tailed probability = 0.0947

Estimated percentage of population obtaining a discrepancy more extreme than the case = 4.734061%

95% confidence limits on the percentage = 1.4407% to 10.5293%

Appendix I: Output for case example's performance on Reverse Counting using**RegBuild_MR_Raw.exe**

RegBuild_MR_Raw.exe: Builds a multiple regression equation and uses it to make inferences concerning a case

THIS PROGRAM IMPLEMENTS STATISTICAL METHODS DEVELOPED IN THE FOLLOWING PAPER:

Crawford, J.R., Garthwaite, P.H., Denham, A.K., & Chelune, G.J. (2012). Using regression equations

built from summary data in the psychological assessment of the individual case: Extension to multiple

regression. *Psychological Assessment*, 24, 801-814. (doi: 10.1037/a0027699).

INPUTS :

Number of predictor (i.e., X) variables = 2
 Sample size (n) for sample providing the summary data = 94
 Credible limit required: Two-sided

Raw data for controls:

[1]:	10.21000	112.00000	47.00000
[2]:	13.71000	120.00000	70.00000
[3]:	11.14000	123.00000	58.00000
[4]:	12.64000	116.00000	27.00000
[5]:	9.43000	113.00000	25.00000
[6]:	16.36000	111.00000	25.00000
[7]:	10.07000	102.00000	50.00000
[8]:	11.79000	97.00000	25.00000
[9]:	10.36000	77.00000	53.00000
[10]:	13.21000	108.00000	73.00000
[11]:	15.07000	128.00000	28.00000
[12]:	10.21000	119.00000	58.00000
[13]:	12.86000	108.00000	27.00000
[14]:	13.36000	102.00000	36.00000
[15]:	10.86000	106.00000	53.00000
[16]:	20.86000	109.00000	35.00000
[17]:	6.50000	94.00000	33.00000
[18]:	10.64000	109.00000	63.00000
[19]:	7.50000	88.00000	35.00000
[20]:	9.00000	113.00000	75.00000
[21]:	7.93000	86.00000	79.00000
[22]:	14.21000	95.00000	27.00000
[23]:	8.50000	94.00000	62.00000
[24]:	10.29000	111.00000	58.00000
[25]:	17.29000	113.00000	21.00000
[26]:	10.14000	137.00000	71.00000
[27]:	9.93000	89.00000	56.00000
[28]:	9.00000	83.00000	48.00000
[29]:	16.57000	120.00000	34.00000
[30]:	16.07000	121.00000	34.00000
[31]:	12.14000	129.00000	71.00000
[32]:	21.71000	130.00000	54.00000
[33]:	10.79000	107.00000	32.00000

[34]:	16.29000	102.00000	47.00000
[35]:	11.50000	98.00000	23.00000
[36]:	15.86000	116.00000	21.00000
[37]:	10.21000	99.00000	33.00000
[38]:	11.29000	108.00000	22.00000
[39]:	11.36000	123.00000	65.00000
[40]:	9.93000	102.00000	59.00000
[41]:	12.43000	129.00000	71.00000
[42]:	13.64000	133.00000	61.00000
[43]:	9.00000	129.00000	65.00000
[44]:	9.07000	107.00000	31.00000
[45]:	8.93000	122.00000	65.00000
[46]:	11.21000	111.00000	60.00000
[47]:	5.43000	104.00000	74.00000
[48]:	10.57000	132.00000	74.00000
[49]:	18.43000	116.00000	60.00000
[50]:	8.50000	116.00000	70.00000
[51]:	12.21000	114.00000	46.00000
[52]:	11.14000	123.00000	76.00000
[53]:	9.57000	125.00000	40.00000
[54]:	9.50000	114.00000	79.00000
[55]:	10.07000	106.00000	45.00000
[56]:	8.64000	109.00000	66.00000
[57]:	12.57000	128.00000	50.00000
[58]:	9.29000	120.00000	72.00000
[59]:	8.29000	139.00000	64.00000
[60]:	10.14000	129.00000	74.00000
[61]:	5.64000	89.00000	51.00000
[62]:	7.93000	139.00000	57.00000
[63]:	8.00000	90.00000	64.00000
[64]:	10.43000	125.00000	78.00000
[65]:	8.79000	119.00000	55.00000
[66]:	5.14000	89.00000	79.00000
[67]:	6.64000	94.00000	71.00000
[68]:	8.86000	124.00000	70.00000
[69]:	6.79000	108.00000	65.00000
[70]:	10.57000	107.00000	48.00000
[71]:	9.86000	112.00000	22.00000
[72]:	6.07000	100.00000	34.00000
[73]:	12.14000	113.00000	52.00000
[74]:	5.86000	105.00000	48.00000
[75]:	7.14000	96.00000	49.00000
[76]:	11.36000	113.00000	46.00000
[77]:	9.79000	117.00000	35.00000
[78]:	4.29000	75.00000	40.00000
[79]:	8.79000	101.00000	36.00000
[80]:	6.50000	92.00000	33.00000
[81]:	13.21000	122.00000	49.00000
[82]:	15.43000	103.00000	27.00000
[83]:	10.43000	106.00000	26.00000
[84]:	11.57000	114.00000	33.00000
[85]:	12.50000	105.00000	40.00000
[86]:	8.79000	116.00000	50.00000
[87]:	11.57000	88.00000	61.00000
[88]:	9.86000	117.00000	56.00000
[89]:	13.93000	111.00000	30.00000
[90]:	10.36000	100.00000	25.00000
[91]:	17.93000	88.00000	47.00000

[92]:	8.43000	96.00000	55.00000
[93]:	6.64000	89.00000	61.00000
[94]:	6.00000	106.00000	61.00000

INPUTS: Summary statistics for sample providing the data together with scores of the case:

Measure Scores for Case	Control Mean	Control SD
Criterion (Y) : 18.400	10.773	3.363
Predictor (X) 1 : 100.000	109.500	14.089
Predictor (X) 2 : 22.000	50.106	17.186

INPUTS: Correlation(s) between criterion and predictor(s) in sample:

1.00000	0.30442	-0.30530
0.30442	1.00000	0.21907
-0.30530	0.21907	1.00000

OUTPUTS: Regression equation (alpha & betas) and squared semi-partial correlation for predictors:

Predictor semi-partial correlation	Beta	Squared
Intercept (alpha): -	4.411	
Predictor (X) 1 : 0.152	0.093	
Predictor (X) 2 : 0.153	-0.076	

OUTPUTS: FURTHER RESULTS FOR THE MULTIPLE REGRESSION MODEL:
 Standard error of estimate for regression equation =
 2.967
 Multiple R for regression equation =
 0.488
 R Squared for regression equation =
 0.238
 Adjusted (shrunk) R Squared for regression equation =
 0.221
 Significance test for overall regression: F [2, 91] =
 14.2129
 Significance test for overall regression: p value =
 0.0000

OUTPUTS: RESULTS FROM ANALYSIS OF THE INDIVIDUAL CASE:

Case's OBTAINED score on Task of Interest = 18.4000
 Case's PREDICTED score from regression equation = 12.0374
 Discrepancy (obtained minus predicted) between case's obtained and
 predicted scores = 6.3626

Effect size (Z-OP) for discrepancy between obtained and predicted scores (plus 95% CI):

Effect size (Z-OP) = 2.168 (95% CI = 1.638 to 2.643)

Standard error for an additional (i.e., N + 1th) case = 3.0265

Significance test (t) on the discrepancy between the case's obtained and predicted scores:

t value (on 91 df) = 2.1023

One-tailed probability = 0.0191

Two-tailed probability = 0.0383

Estimated percentage of population obtaining a discrepancy more extreme than the case = 1.914362%

95% confidence limits on the percentage = 0.4110% to 5.0710%

Appendix J: Percentiles of the distribution of scores on Letter Recall and Reverse**Counting stratified by age group**

		Age Groups		Percentiles				
				5 th	10 th	25 th	50 th	75 th
Letter Recall	21 – 40	21.5	25.4	30	36	39.5	42	42
	41 – 60	27.55	28.1	31	38	40	41.9	42
	61 – 70	19	22	30	37	41	41	-
	71 – 79	16	18.8	28.25	32.5	35.5	42	-
Reverse Counting	21 – 40	5.54	6.5	9.5	11.5	14.64	16.49	18.36
	41 – 60	5.76	7.22	8.95	10.25	12.16	17.76	19.91
	61 – 70	6	6.39	8	8.64	10.64	12.4	-
	71 – 79	5.14	5.34	8.2	10.14	11.89	13.36	-

Systematic Literature Review:

Does the development of robotic and smart technology care interventions for people with dementia adhere to recommendations for good psychological care?

A systematic review

Aparna Kapoor

Institute of Psychiatry, Psychology and Neuroscience

King's College London

Supervised by:

Dr Jessica Fish

Professor Robin Morris

Institute of Psychiatry, Psychology and Neuroscience

King's College London

Table of Contents

List of Tables	105
List of Figures.....	106
List of Abbreviations	107
1. Abstract.....	110
1.1 Background	110
1.2 Method	110
1.3 Results.....	110
1.4 Conclusions.....	110
2. Introduction.....	112
3. Aims of the study.....	118
4. Method.....	118
4.1 Literature search.....	119
4.2 Eligibility criteria for studies	119
4.3 Search criteria.....	120
4.4 Screening process.....	120
4.5 Data extraction process	120
4.6 Rating procedures and tools	121
4.7 Data analysis	123
4.8 Data presentation.....	123
5. Results.....	124
5.1 Search results	124

5.2	Data extraction and Inter-rater agreement.....	125
5.3	Participant characteristics.....	136
5.4	Sample Size.....	137
5.5	Study design.....	137
5.6	Type of RST studied	139
5.7	Main outcome measures.....	141
5.8	Reporting of outcomes	142
5.9	Nature of outcomes	143
5.10	Ratings on the QualSyst Measure	143
5.11	Ratings on the DeBs.....	144
6.	Discussion.....	146
7.	Limitations of the study.....	146
8.	Conclusions.....	153
9.	References.....	155
10.	Appendices	166
	Appendix A: PRISMA Checklist.....	166
	Appendix B: Data Extraction.....	169
	Appendix C: QualSyst Measure	170
	Quantitative studies	170
	Qualitative studies	171
	Appendix D: Distribution of Scores on the QualSyst Measure and DeBs	172
	Appendix E: Item-by-item ratings on QualSyst and DeBs.....	173

QualSyst (Quantitative Studies) 173

QualSyst (Qualitative Studies) 177

DeBs 178

List of Tables

Table 2.1	Data extraction summary and quality assessment ratings.....	126
Table 2.2	Mean, standard deviations and ranges for individual DeBs items.....	146

List of Figures

Figure 2.1	The Dementia Psychological Care Benchmarking Scale (DeBs).....	122
Figure 2.2	PRISMA flow diagram.....	124
Figure 2.3	PARO the seal.....	140
Figure 2.4	CuDDler.....	140
Figure 2.5	Robot Ed.....	140
Figure 2.6	SenseCam.....	140
Figure 2.7	Graph representing means and standard deviations for DeBs items....	146

List of Abbreviations

AAT	Animal-assisted therapy
ACE	Addenbrooke's Cognitive Examination
AD	Alzheimer's disease
AER	Apparent Emotion Rating Instrument
AES	Apathy Evaluation Scale
AI	Apathy Inventory
APADEM-NH	Apathy Scale for Institutionalized Patients with Dementia Nursing Home version
ARD	Alcohol-related dementia
AT	Assistive technology
AWS	Revised Algate Wandering Scale-Nursing Home version
BARS	Brief Agitation Rating Scale
BI	Barthel Index
BPSD	Behavioural and psychological symptoms of dementia
CMAI	Cohen-Mansfield Agitation Inventory-Long Form
CMAI-SF	Cohen-Mansfield Agitation Inventory-Short Form
CSDD	Cornell Scale for Symptoms of Depression in Dementia
DBD	Dementia Behaviour Disturbance Scale
DeBs	Dementia Psychological Care Benchmarking Scale
EVIBE	Instant Assessment of Wellbeing Tool
FTD	Fronto-temporal dementia
GDS	Geriatric Depression Scale
GDS-30	Geriatric Depression Scale-30
GIP-28	Behaviour Rating Scale for Psychogeriatric Inpatients
GSR	galvanic skin response

ICC	Intra-class correlation
IAFAI	Adults and Older Adults Functional Assessment Inventory
I-PANAS-SF	International Positive and Negative Affect Schedule Short-Form
IPPA	Individually Prioritized Problems Assessment
K-CMAI	Korean Cohen-Mansfield Agitation Inventory
LBD	Lewy-body dementia
MMSE	Mini-Mental State Examination
MMSE-K	Korean Mini-Mental State Examination
MPES	Menorah Park Engagement Scale
NDD	Neurodegenerative disease
NPI-ES	Neuropsychiatric Inventory for Health Staff
NPI-NH	Neuropsychiatric Inventory Nursing Home version
NPI-Q	Neuropsychiatric Inventory Questionnaire
OERS	Observed Emotion Rating Scale
OSBC	Oshkosh Social Behaviour Coding Scale
PD	Parkinson's disease
PwD	People with dementia
QOL-AD	Quality of Life in Alzheimer's Disease Scale
QOL-D	Quality of Life Questionnaire for Dementia
QUALID	Quality of Life in Late-Stage Dementia
RAID	Rating Anxiety in Dementia Scale
RSCSE	Revised Scale for Caregiving Self-Efficacy
RST	Robotic or smart technology
sMMSE	Severe Mini-Mental State Examination
SQI	Self-Identity Questionnaire
SUS	System Usability Scale

VAS	Visual Analogic Scale
VD	Vascular dementia
VTI	Vitality Index
WHOQOL-OLD	World Health Organization Quality of Life-OLD
ZBI	Zarit Burden Interview

1. **Abstract**

1.1 **Background**

This study reviews recent research exploring the use of robotic or smart technology (RST) for people with dementia (PwD). The purpose of the review was to assess the extent to which use of RST for PwD in recent publications adhered to psychological principles of person-centredness in relation to dementia care, as outlined by the British Psychological Society (BPS, 2016) guidelines.

1.2 **Method**

Two databases (Ovid and PubMed) and one database for ongoing clinical trials were searched for studies published between 2013 and 2018, which examined the effectiveness of RST for PwD. Only publications in English were included, and any studies focussing only on usability of RST or caregiver outcomes such as burden and stress without measuring outcomes for PwD were excluded. Each paper was evaluated using two rating scales; the QualSyst tool to assess the methodological quality of the studies, and the Dementia Psychological Care Benchmarking Scale (DeBs) to assess adherence to psychological care principles.

1.3 **Results**

Searches resulted in 1,029 publications, of which 37 met the inclusion criteria. Most studies considered the impact of the use of companion robots on mood, levels of engagement and quality of life of PwD. The majority were observational studies, in which researchers provided outcomes directly. Additionally, the studies mostly considered caregiver and/or staff views on RST for PwD. Only three studies tailored the intervention to each participant. Quality ratings on the QualSyst ranged from 93% (strong) to 23% (limited), with over half the studies ($n = 17$) being categorised as adequate (50–70%). Ratings on the DeBs ranged from 74% (good) to 20% (limited), with most of the studies being categorised as limited ($n = 33$).

1.4 Conclusions

The results of this review indicate a clear need for future studies on implementing RST in a person-centred, individualised manner, encouraging PwD to exercise choice and control over these interventions. The results also serve as a baseline from which progress in the field might be measured in the future.

2. **Introduction**

The steady increase in the world's aging population in recent times has brought dementia to the forefront of public health issues (World Health Organisation [WHO], 2017). Dementia is a broad term referring to a set of conditions characterised by a gradual decline in brain functioning, with symptoms that are mainly related to cognitive functions, including memory, attention and language (British Psychological Society [BPS], 2016). There are several causes of dementia, the most common being Alzheimer's disease (diagnosed in approximately 60–80% of cases), followed by vascular dementia (diagnosed in approximately 40% of cases). Other causes include dementia with Lewy bodies, fronto-temporal dementia and Parkinson's disease (Alzheimer's Association, 2018). In 2017, it was estimated that approximately 50 million people worldwide were living with a diagnosis of dementia (WHO, 2017), with a majority of those affected aged above 65 years (Alzheimer's Association, 2018). Dementia is an age-related diagnosis; however, in recent years, there has been an increase in the number of adults below the age of 65 being diagnosed (Alzheimer's Association, 2018), and this is expected to increase in the future. The high prevalence of dementia leads to higher societal costs pertaining to care. A 2015 estimate put the global cost of dementia at approximately US \$818 billion, estimated to rise to about US \$1 trillion by 2018 (Alzheimer's Disease International, 2015), making it important for research into understanding, treating and supporting people with this condition to progress.

Characterised by a decline in cognitive abilities alongside physical decline associated with the aging process, dementia can be a highly debilitating condition, resulting in deficits in memory, communication and the ability to perform everyday activities (Donaldson, Tarrier, & Burns, 1997; Moyle, Fetherstonagh, Greben, Beattie, &

AusQol Group, 2015; National Institute for Health and Care Excellence [NICE], 2013). As a result, PwD can gradually lose their independence, requiring more support with performing daily activities, becoming forgetful, disoriented to time and place, and at later stages even becoming bedridden (Alzheimer's Association, 2018). This loss in physical and cognitive abilities can result in people becoming increasingly reliant on family members or requiring admission to care facilities (Brodaty & Donkin, 2009), where they may become socially isolated (Brown, 2018). In 90% of cases, people also experience behavioural and psychological symptoms of dementia (BPSD), including agitation, wandering or aggressive behaviour (Chiberska, 2018; Feast et al., 2016; Stewart et al., 2014), which can be difficult for carers and staff to manage (Feast et al., 2016).

Pharmacological interventions have most commonly been used to treat BPSD (Kirkham et al., 2017; Maher et al., 2011). However, the sedative effects of antipsychotic medication can have serious repercussions on the quality of life of PwD, including increased risk of falls and further cognitive decline (Kirkham et al., 2017). A growing body of research now recommends non-pharmacological interventions including cognitive stimulation therapy, animal-assisted therapy (AAT), physical activity and behavioural therapy, amongst others, as first-line treatment for cognitive impairment and BPSD (NICE, 2017), due to their being more person-centred and focussed on improvement of quality of life, as compared to pharmacotherapy (Chiberska, 2018).

More recently, non-pharmacological interventions in the form of assistive technologies (AT) have been trialled for use with PwD (Bharucha et al., 2009). AT are any technological devices which support people with dementia to compensate for deficits in cognitive and/or physical abilities (Bharucha et al., 2009). These can include simple

memory aids like calendars, or physical aids like wheelchairs, or more advanced devices like robotic and smart technologies (RST). RST can take numerous forms (Bogue, 2013), ranging from hand-held devices and adapted computer software to robots in humanoid or animal form. Some popular assistive technologies include cognitive and memory aids, like SenseCam (Hodges et al., 2006), navigational tools like the HTC Touch smartphone (Lanza, Knörzer, Weber, & Riepe, 2014), activity monitoring and prompting technologies like Robot Ed (Begum, Wang, Huq, & Mihailidis, 2013; Begum, Huq, Wang, & Mihailidis, 2015; Wang, Sudhama, Begum, Huq, & Mihailidis, 2017) or companion or social robots like PARO the seal (Shibata et al., 2001) or the CuDDler (Limbu et al., 2013). RST were first introduced in dementia care to manage cognitive difficulties and BPSD (Moyle et al., 2011), while addressing the limitations of other traditional interventions. For example AATs, although effective in reducing agitation, carried the risk of infection or injury, and increased carer or staff burden (Cowling & McFadden, 2010), a limitation addressed by the introduction of robotic animals, like PARO the seal. More recently, psychosocial interventions utilising RST have been trialled in home and care settings, to support PwD experiencing both cognitive difficulties like memory impairment and BPSD symptoms like agitation.

RST have been found to have a positive impact on wellbeing and quality of life of PwD, by fostering independence (Lanza et al., 2014), increasing social interaction (Lancioni et al., 2016), supporting memory impairment (Woodberry et al., 2015) and encouraging physical activity (Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018). Further, they reduce the risk of injury to clients (Tamura et al., 2004) and support PwD to live independently in the community (Cahill, Begley, Faulkner, & Hagen, 2007). For the purpose of this review, only those technologies that fall under the umbrella term RST, as described above, have been included.

Currently, as there is no cure for dementia, enhancing the quality of life for people living with dementia becomes imperative to their care and wellbeing. RST has immense potential to improve the way we care for people with varying disabilities (Agree & Freedman, 2000), including dementia (Lilja, Mansson, Jahlenius, & Sacco-Petersen, 2003). However, only a few robust studies have considered their use with this specific population (Limbu et al., 2013), and until relatively recently, these studies have limited the application of RST to functional tasks (Sixsmith, Orpwood, & Torrington, 2007), like supporting completion of daily living activities. As Moyle et al. (2015) noted, quality of life for PwD is influenced not only by maintenance of independence, but also by engaging in meaningful activities and social interaction. Hence, for RST to be effective for PwD, it is important that it is used to address each person's individual needs and is implemented in a person-centred manner (Cahill et al., 2007; Chiberska, 2018).

In 2016, the BPS Dementia Advisory Group published a report entitled 'Psychological dimensions of dementia: Putting the person at the centre of care', outlining the psychological principles upon which care for PwD should be based. This report presented guidelines for carrying out effective assessment and treatment planning for PwD, considering it from a psychological perspective, and moving away from the medical model. These guidelines placed the individual at the centre of their care, emphasising that, without a current cure for dementia, the focus of all clinical care should be on ensuring the person lives a fulfilled and meaningful life. Drawing upon the four principles of person-centred care, that is, personalised, enabling, coordinated care, where the person is treated with dignity and respect (Collins, 2014), greater emphasis was placed on involvement of PwD in meaningful activities and dementia-friendly communities. The BPS recognised a gap in current service provision, and made specific

recommendations stressing the importance of tailoring care plans to individual needs, whilst involving PwD in decisions about their care. Biopsychosocial treatment plans were recommended, as opposed to the use of pharmacological therapies. These included life story work and positive behaviour support. Interestingly, the use of RST as a potential new and efficient support option for PwD was not mentioned in this report. There may be several reasons for this, as discussed below.

The intention with which researchers apply RST in dementia care remains unclear, given the need for RST to be directly applied in a way that improves people's quality of life (Orpwood et al., 2007). In their review of literature looking at the impact of robotic pets on the well-being of older adults with dementia, Leng et al. (2019) noted that most randomised controlled trials (RCTs) found significant reduction in BPSD, but no effect on cognitive function or quality of life as measured by standardised assessment. They made several recommendations based on these findings, including considering participants' cultural backgrounds, preferences and life histories when applying such technologies, and supplementing outcome measures with self-report measures to gain a better understanding of participants' own experiences. Another review of literature by Lynn et al. (2017) identified issues around consent when implementing electronic RST in dementia care. They noted that studies included within their review mentioned gathering informed consent from participants, but addressed this in a cursory fashion, with little information on how and from whom such consent was gained. These findings are problematic, as they indicate that the application of RST may not be sufficiently person-centred.

More recently, a literature review published by Whelan et al. (2018) considered acceptability and usability of social robots for older adults, and the factors that

contributed to this. They noted that individual characteristics and preferences of users were the greatest predictors of acceptability of social robots, suggesting that the use of such technology should be personalised.

There are, accordingly, indications that person-centred applications of RST have a range of benefits as compared to those that are less person-centred. These include adhering to ethical considerations like informed consent, being more consistent with BPS policy (2016) and being more acceptable to users. This in turn indicates that person-centred applications may be more effective than their alternatives in targeting cognitive and behavioural symptoms associated with dementia, as well as for improving the well-being and quality of life of PwD.

The current review explored attitudes towards RST in dementia care and how researchers in recent years have applied such technology in this population, and whether these applications have been consistent with the psychological principles outlined by the BPS (2016). The scale used focused on attitudes as indicated by exploring published research, thus providing a guide to practice-based approaches adopted in this field. An alternative approach that measures research attitudes directly would provide the basis for an alternative study. Ideas around the utility of AT for PwD are evolving rapidly alongside developments in technology (Bharucha et al., 2009). Hence, although earlier studies can be important in gaining a historical or comparative perspective, in order to accurately sample relatively 'current' attitudes towards RST, the review was restricted to papers published within the past five years, so reflecting current attitudes.

3. Aims of the study

The aims of this study were to review the current body of literature published during 2013 to 2018 on RST for PwD, with two primary objectives:

1. To assess the methodological quality of all included studies and identify any variation across the studies.
2. To identify current attitudes towards application of RST in dementia care and whether these are consistent with the recommendations made by the BPS (2016).

In relation to the second objective, the following questions are to be addressed by this systematic review:

1. Are robotic interventions for dementia person-centred (i.e., do they consider the individual's needs, choice and identity)?
2. Do robotic interventions foster independence in people with dementia by offering them choice and control?
3. Do robotic interventions support people with dementia to live more meaningful and fulfilling lives?
4. Do robotic interventions promote social inclusion in people with dementia, tackling loneliness and isolation?

In addition, a supplementary aim was to characterise the studies in terms of the study design, sample, technology used, analyses and outcomes.

4. Method

The body of literature was reviewed in line with PRISMA guidelines (Moher, Stewart, & Shekelle, 2016) and was registered on PROSPERO in February 2018 (registration

number: CRD42018087917). This report was also written in line with the PRISMA guidelines (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009). Please see Appendix A for the completed checklist.

4.1 Literature search

A literature search was conducted using two databases: PubMed and Ovid (PsycINFO, Global Health, MEDLINE, Embase), covering the past five years (2013 to 7th October 2018). Further, a database for ongoing clinical trials was searched and hand searches were completed to identify additional relevant studies and those within reference lists of included publications.

4.2 Eligibility criteria for studies

Studies were included within this review if they met the following criteria:

- English language publications
- Studies focusing specifically on RST and not on other forms of AT (i.e., wheelchairs, remote controls)
- Publications considering the use of RST for people with a diagnosis of dementia and/or Alzheimer's disease, irrespective of the age of participants (including social, assistive, cognitive or telepresence robots)
- Studies published within the last five years (i.e., 2013 to 2018)
- Single case studies and case series employing a systematic approach rather than a purely exploratory design

Exclusion criteria for studies included:

- Systematic reviews, meta-analyses, books, book chapters and conference abstracts
- Studies including participants without a diagnosis of dementia

- Studies whose only outcome measure is the impact of RST on staff and/or carer experiences (i.e., burden, stress, workload, etc.), as the current review was interested in the experience of the PwD, and outcomes specific to this population
- Studies focusing on design or usability of RST only
- Papers reporting trial protocols or studies in the design or planning stage, prior to implementation

4.3 Search criteria

A combination of the following search terms was used across databases: (Alzheimer's OR dementia OR vascular dementia OR Lewy body dementia OR Alzheimer* OR dement*) AND (assistive technology OR self-help device OR robotics OR gerontechnology OR health technology OR care technology OR service robot OR assistive robot OR robotic support OR social robot OR rehabilitation robot OR cognitive robot OR cognitive orthotics OR cognitive prosthetics OR cognitive prostheses OR telepresence OR robotic device OR robot*).

4.4 Screening process

The literature search identified 1,029 studies. These included 1,023 studies retrieved via the two databases and six from a database for ongoing clinical trials. As a first step in the screening process, any duplicate articles or those that clearly fell within the exclusion criteria (i.e., book chapters or conference abstracts) were removed. Following this, all remaining titles and abstracts were reviewed to exclude studies that were not relevant; for example, those that did not focus on PwD or the use of RST with this population. This resulted in 63 shortlisted publications. At the next stage of screening, full text articles were reviewed for eligibility, resulting in 37 articles being included in this review (see Figure 1.2 for details).

4.5 Data extraction process

Each of the 37 full text articles was read thoroughly and data on study design, methodology and outcome were extracted. Full details of the information extracted from each study can be found in Appendix C.

4.6 Rating procedures and tools

To address the two main objectives of this review, two assessment tools were used to rate each included study:

- The QualSyst tool (Kmet, Lee, & Cook, 2004): This tool was used to assess the quality of the methodology of each publication included in this review. It has separate scales for quantitative and qualitative studies. Questions relate to the suitability and relevance of the methodology, design and analyses employed. Each question is rated on a scale of 0 (“no”) to 2 (“yes”), with some including a 1 (“partial”) or N/A option. This is a widely used measure with strong reliability statistics (inter-rater agreement on qualitative studies ranged from 60–100%). See Appendix D for the full scale.
- The Dementia Psychological Care Benchmarking Scale (DeBs; Morris, 2018): This scale was used to assess researcher attitudes towards use of RST in dementia care and the extent to which each study adhered to psychological principles of dementia care. The scale includes 10 questions, each addressing a different area of psychological care for PwD, based on the recommendations made by the BPS (2016); for example, whether the approach is individualised and person-centred. Each question is rated on a 5-point scale, ranging from 1 (“not at all mentioned”) to 5 (“considered and strongly advocated”). See Figure 2.1 below for the full scale.

The Dementia Psychological Care Benchmarking Scale (DeBs)	
Taking into account the overall framework of care advocated in the paper or study, rate the following ten questions on the 1-5 scale provided below.	
Scale	
Not at all mentioned: No mention of issue in paper.	1
Minimal consideration: Slight mention in passing with no elaboration or clear opinion.	2
Partial consideration: Topic introduced and discussed but incomplete consideration or issue dealt with in a cursory fashion.	3
Considered: Main issues taken into account and discussed.	4
Considered and strongly advocated: Issues discussed and an advocacy position clearly taken.	5
Questions	
1	Is the approach advocating individualised care, taking into account the person's individual needs?
2	Is the approach person centred taking account the dementia journey and person identity?
3	Does the approach involving listening to people with dementia and ensuring their voice is heard?
4	Does the approach advocate involving the person with dementia in decision making regarding their care?
5	Does the approach advocate support for the person with dementia to exercise choices and control?
6	Does the approach advocate and facilitate access to meaningful activities?
7	Does the approach promote participation in local communities that are dementia friendly?
8	Is the approach focused on enabling the person to live lives that are meaningful and satisfying?
9	Does the approach consider the impact on the stigma of dementia?
10	Does the approach promote social inclusion and involvement in valued roles and activities?

Figure 2.1. The Dementia Psychological Care Benchmarking Scale (DeBs)

Each of the 37 included studies was assessed using both tools by the lead researcher. Fourteen publications included within this review were quality assessed by a second rater (the primary supervisor), to assess reliability of ratings.

4.7 Data analysis

A two-way mixed effects intra-class correlation (ICC) was calculated to establish the degree of inter-rater agreement for the 14 studies co-rated by a second assessor.

As suggested by Lee, Packer, Tang, and Girdler (2008), scores obtained on the QualSyst measure for each study were converted to percentages (excluding any items rated ‘not applicable’) and quality of methodology was interpreted as strong (scores >80%), good (scores ranging from 70–80%), adequate (scores ranging from 50–70%) or limited (scores <50%). The same categories for interpretation were used for the DeBs to assess the extent to which studies adhere to psychological principles, as outlined by the BPS (2016). These are presented in Table 2.1.

Additionally, the distribution of scores for each study, represented by mean (*M*), median (*Mdn*), standard deviation (*SD*) and range, on the QualSyst measure and DeBs have been presented in Appendix E.

4.8 Data presentation

The characteristics of each study are first summarised in terms of their design, methodology and outcomes, followed by the quality assessment ratings and finally the rating on the assessment of adherence to psychological care principles. These can be found in Table 1.

Additional information on individual ratings on the QualSyst measure and DeBs can be found in Appendices D, E and F.

5. Results

5.1 Search results

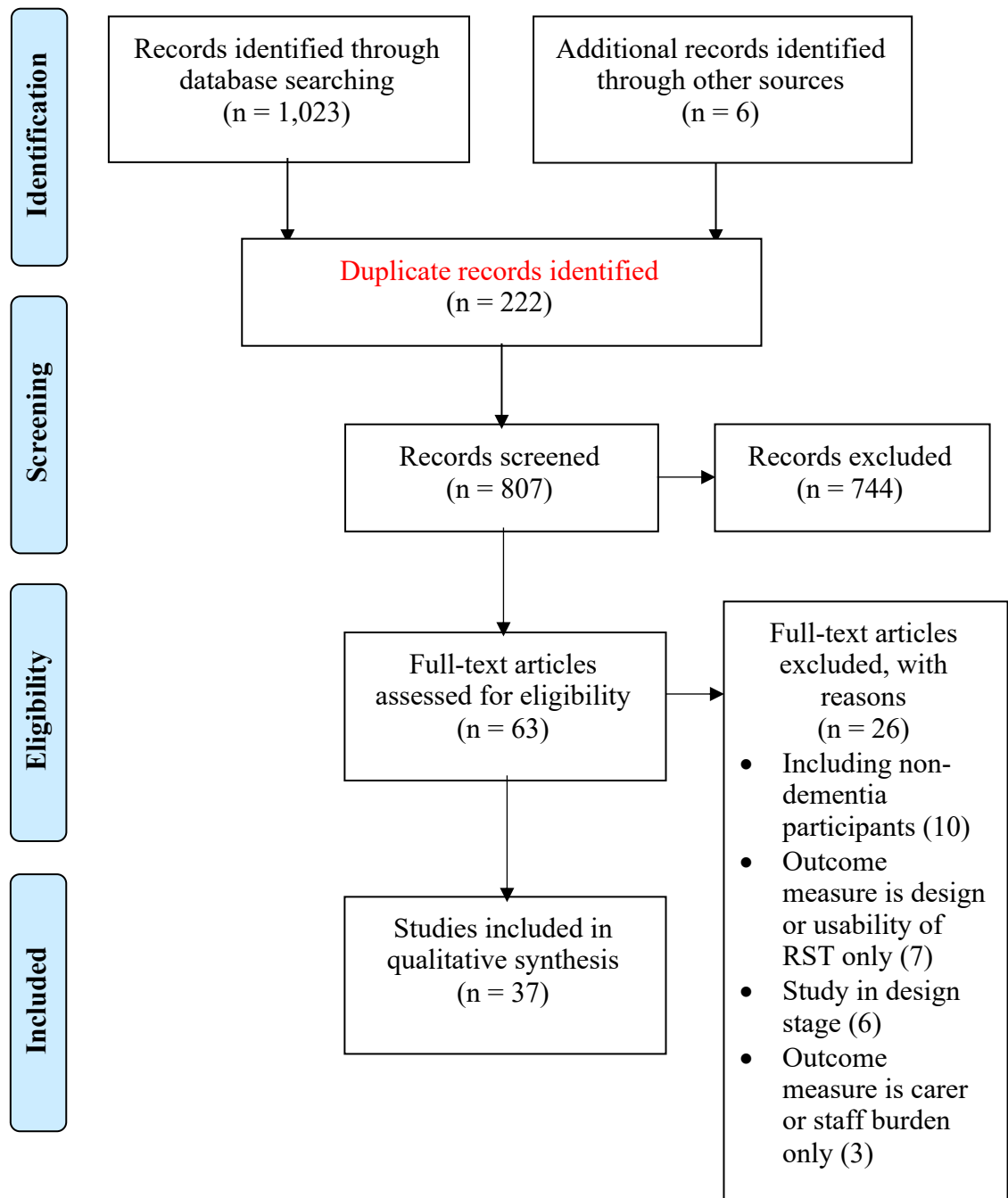


Figure 2.2. PRISMA flow diagram (Moher et al., 2009)

As seen in Figure 2 above, the literature search identified $n = 1,029$ papers, including six identified through a database for ongoing clinical trials. Following removal of duplicates ($n = 222$), 807 papers were screened through titles and abstracts. Of these, $n = 63$ papers were identified for review via full-text, resulting in exclusion of $n = 26$ papers. Reasons for exclusion were as follows: papers that included non-dementia participants ($n = 10$), papers that only considered usability or design of RST ($n = 7$), publications that were still in the study design or data collection stage ($n = 6$), and studies only considering the impact of RST on staff or carer burden ($n = 3$). Following this, 37 studies met the criteria for inclusion in this review. Table 1 provides a summary and quality assessment rating totals for each study.

5.2 Data extraction and Inter-rater agreement

Key information was extracted from each of the 37 studies included in this review. Details of this can be found in Table 1 and are discussed below.

For the 14 studies that were quality assessed by a second rater, ICC was calculated for both measures. Inter-rater reliability was good for both measures, being 0.82 for the QualSyst tool and 0.88 for the DeBs. Where further discrepancies were noted, these were discussed among the raters and a rating agreed upon. Detailed ratings on the QualSyst measure and DeBs for each of the 37 included papers can be found in Appendix F.

Table 2.1. Data extraction summary and quality assessment ratings

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis* *	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Arntzen et al. (2016)	Longitudinal Qualitative	39 devices for daily activities	Observation of use of RST to complete daily activities	Interview, observation Participant Carer Researcher	Nvivo 9, hermeneutic phenomenology	12 (10=AD, 1= not specified, 1=FTD) 14 Carers	Positive effect of RST	15/20 (75%) Good	29/50 (58%) Adequate
Begum et al. (2013)	Mixed methods observational	Robot Ed for daily activities	Completion of tea-making task with RST prompts	Interview, observation, survey Participant, Carer Researcher	Totals and percentages of adherence	5 AD 5 Carers	Inconclusive results due to small sample size	8/16 (50%) Adequate	14/50 (28%) Limited
Begum et al. (2015)	Exploratory observational usability study	Robot Ed for daily activities	Completion of tea-making task with RST prompts	Interview, observation, survey Researcher	Behavioural analysis	10 AD 10 Carers	No effect of RST	12/20 (60%) Adequate	21/50 (42%) Limited
Bemelmans, Gelderblom, Jonker, & de Witte (2015)	Quasi-experimental time series ABAB study	PARO seal robot	Therapeutic or care activity intervention with RST	IPPA, Mood scale, GIP 28 Carer	Within subject - Wilcoxon signed-rank test, Thematic Analysis	91 Dementia (not specified)	Positive effect of RST	12/22 (55%) Adequate	16/50 (32%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis* *	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Bemelmans, Gelderblom, Jonker, & de Witte (2016)	Feasibility study Qualitative	PARO seal robot	Therapeutic, care activity or social intervention with RST	Interview Participant Carer Staff	Added value ratings	23 Dementia (not specified)	Positive effect of RST	5/20 (25%) Limited	16/50 (32%) Limited
Chu et al. (2016)	Cross-sectional design, using observational method	Jack and Sophie social robots	Reactions to interaction with RST	Observation Researcher	KALPHA, two-way ANOVA	139 Dementia (not specified) 28 Staff	Positive effect of RST	20/22 (91%) Strong	12/50 (24%) Limited
Demange et al. (2018)	Quasi-experimental	PARO seal robot	Reactions to interaction with RST	I-PANAS-SF, MMSE, NPI-ES, SUS, Participant	ANOVA, Kruskal-Wallis, Pearson correlation, Thematic Analysis	17 Dementia (not specified)	Positive effect of RST	19/26 (73%) Good	17/50 (34%) Limited
Gustafsson, Svanberg, & Müllersdorf (2015)	Intervention mixed methods (single case ABA)	JustoCat social robot	Carer-led interaction with RST	QUALID, CMAI Researcher Carer	Visual analysis, qualitative description	4 Dementia (not specified) 11 Staff 3 Carers	Positive effect of RST	9/18 (50%) Adequate	10/50 (20%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis* *	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Jøranson, Pedersen, Rokstad, & Ihlebæk (2015)	Cluster RCT	PARO seal robot	Group interaction with RST v/s control	MMSE, medication use, BARS, CSDD Researcher	ANOVA, chi-square	52 Dementia (not specified) 1 Non-dementia control	Positive effect of RST	23/28 (82%) Strong	11/50 (22%) Limited
Jøranson, Pedersen, Rokstad, & Ihlebæk (2016a)	Cluster RCT	PARO seal robot	Group interaction with RST v/s control	BARS, QUALID, medication use Researcher	ANOVA, multiple linear regression	52 Dementia (not specified) 1 Non-dementia (control)	Positive effect of RST	21/28 (75%) Good	13/50 (26%) Limited
Jøranson et al. (2016b)	Systematic observational study	PARO seal robot	Group interaction with RST v/s control	Observation Researcher	ANOVA	30 Dementia (not specified)	Positive effect of RST	13/22 (59%) Adequate	12/50 (24%) Limited
Khosla, Nguyen, & Chu (2017)	Survey and observational study	Matilda social robot	Group activity with RST	Survey, observation Researcher Participant	MANOVA, ANOVA, post-hoc tests, reliability analysis	115 Dementia (not specified)	Positive effect of RST	16/22 (73%) Good	19/50 (38%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis*	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Koh & Kang (2018)	Non-equivalent control group pretest-posttest design	PARO seal robot	Group interaction with RST v/s control	MMSE-K, AER, K-CMAI, observation Researcher	Chi-square, Fisher's exact and independent <i>t</i> -tests, paired <i>t</i> -test	33 (20=AD, 6=VD, 2=LBD, 5=unknown)	Positive effect of RST	21/28 (75%) Good	11/50 (22%) Limited
Kuwamura et al. (2016)	Case series	Telenoid tele-operated social robot	Reaction to face-to-face v/s RST-facilitated interaction with speaker Researcher	MMSE, QOL-D, DBD, NPI-NH, BI, VTI Researcher	Within subjects paired <i>t</i> -test, Student's <i>t</i> -test, Welch's <i>t</i> -test	3 AD	Positive effect of RST	8/23 (35%) Limited	11/50 (22%) Limited
Lancioni et al. (2015)	Non-concurrent multiple baseline design	Computer with a micro-switch and sound amplifier	Activity completion with RST-prompts v/s control (no prompts) Researcher	Data recording, observation Researcher	Percentages of target responses	3 AD	Positive effect of RST	5/22 (23%) Limited	14/50 (28%) Limited
Lancioni et al. (2016)	Non-concurrent multiple baseline design	Computer with a micro-switch and sound amplifier	Activity completion with RST-prompts v/s control (no prompts) Researcher	Data recording, observation Researcher	Percentages, Kolmogorov-Smirnov test. Study 2 – Kolmogorov-Smirnov test	Study 1 – 8 AD Study 2 – 10 AD	Positive effect of RST	13/28 (46%) Limited	12/50 (24%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis*	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Lancioni et al. (2017)	Non-concurrent multiple baseline design	Computer with a micro-switch and sound amplifier	Activity completion with RST-prompts v/s control (no prompts)	Data recording, observation Researcher	Mean frequency, Nonoverlap of All Pairs, Tau _{novlap} indices	9 AD	Positive effect of RST	14/28 (50%) Adequate	13/50 (26%) Limited
Lancioni et al. (2018)	Non-concurrent multiple baseline design	Computer with a micro-switch and software	Exercise completion with RST v/s walker v/s chair	Data recording, observation Researcher	Frequency, range, percentages	10 AD 22 Staff (non-dementia)	Positive effect of RST	14/28 (50%) Adequate	11/50 (22%) Limited
Lanza et al. (2014)	Non-randomised, cross-sectional, mono-centric pilot study	HTC Touch Smart-phone navigator	Navigation using RST v/s aerial map	Cognitive tests, observation Researcher	Mean, SD, Pearson correlation	14 AD	Positive effect of RST	14/22 (64%) Adequate	13/50 (26%) Limited
Liang et al. (2017)	Block RCT	PARO seal robot	Interaction with RST (at day centre & home) v/s control	ACE, observation, cortisol, b/p, CMAI-SF, NPI-Q, CSDD Researcher Carer	ANOVA, pairwise comparisons, <i>t</i> -test, Mann-Whitney <i>U</i> test, ANCOVA, Thematic Analysis	30 Dementia (not specified) 30 Carers	Positive effect of RST	16/28 (57%) Adequate	14/50 (28%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis* *	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Lindqvist, Nygård, & Borell (2013)	Longitudinal study	Numerous	Observation of use of RST to complete daily activities	Interview, observation Participant	Constant comparative approach	10 AD	Positive effect of RST	14/20 (70%) Good	31/50 (62%) Adequate
Moyle et al. (2013)	Randomised crossover design (within subjects) - RCT	PARO seal robot	Group activity with RST v/s control	QOL-AD, RAID, AES, GDS, AWS, OERS Researcher	Standardised difference in means, Reliability Change Index	18 Dementia (not specified)	Positive effect of RST	21/28 (75%) Good	11/50 (22%) Limited
Moyle et al. (2014)	Mixed methods approach observational	Giraff tele-presence robot	RST-facilitated calls between participant and family member	Interview, observation Researcher Carer Staff	Thematic Analysis, video coding	5 Dementia (not specified) 6 Carers 7 Staff	Positive effect of RST	17/20 (85%) Strong	11/50 (22%) Limited
Moyle et al. (2016)	Case study design	CuDDler social robot	Research-facilitated individual interaction with RST	MMSE, CMAI, interview, OERS Researcher Participant	Thematic Analysis	5 Dementia (not specified)	Negative effect of RST	12/20 (60%) Adequate	16/50 (32%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis* *	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Moyle, Bramble, Jones, & Murfield (2017a)	Qualitative semi-structured individual interview	PARO seal robot	Interaction with RST v/s plush toy v/s control for BPSD	Interview Carer	Thematic Analysis	20 Dementia (not specified)	Positive effect of RST	12/20 (60%) Adequate	14/50 (28%) Limited
Moyle et al (2017b)	Parallel, 3-group, single-blind cluster RCT	PARO seal robot	Interaction with RST v/s plush toy v/s control for BPSD	CMAI-SF, observation Researcher Staff	Repeated measures mixed effects model	415 Dementia (not specified)	Positive effect of RST	26/28 (93%) Strong	15/50 (30%) Limited
Moyle et al. (2018)	Parallel 3 group single-blind Cluster RCT	PARO seal robot, SenseCam wearable camera	Individual interaction with RST v/s plush toy v/s control for motor activity and sleep	Data coding, observation Researcher	ICC, Cohen's d; chi-square, Fisher's Exact Tests, ANOVA	455 (149=AD, 73=VD, 7=LBD, 4=FTD, 4=ARD, 218=not specified)	Positive effect of RST	26/28 (93%) Strong	11/50 (22%) Limited
Navarro et al. (2016)	Within subject's design	Answer Board reminders, Answer Pad app	Effect of RST v/s control on memory	NPI-Q, AES, ZBI, RSCSE, diary Carer	Pearson correlation, Thematic Analysis	2 AD 2 Carer	Positive effect of RST	13/28 (46%) Limited	24/50 (48%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis*	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Nijhof, van Hoof, van Rijn, & van Gemert-Pijnen (2013)	Mixed methods (within subjects)	Chit-chatters (CC) leisure activity package	Engagement in RST-led leisure activity v/s control	OSBC, interviews Researcher	<i>t</i> -test, bootstrapping, coding for interviews	10 Dementia (not specified) 5 Facilitators	Positive effect of RST	17/28 (61%) Adequate	12/50 (24%) Limited
Petersen, Houston, Qin, Tague, & Studley (2017)	Randomized block design (RCT)	PARO seal robot	RST v/s control as biofeedback device	RAID, CSDD, GDS, pulse oximetry/rate, GSR Staff	<i>t</i> -test	61 Dementia (not specified)	Positive effect of RST	18/28 (64%) Adequate	10/50 (20%) Limited
Rouaix et al. (2017)	Exploratory study - uncontrolled experiment	NAO humanoid robot	RST-facilitated psychomotor therapy v/s control (therapist-led)	MMSE, NPI, SQL, I-PANAS-SF, EVIBE, MPES, VAS Participant Researcher	Time percentage, Wilcoxon test	9 (3=Mixed, 2=AD, 2=NDD, 1=PD, 1=LBD)	Positive effect of RST	12/20 (60%) Adequate	28/50 (56%) Adequate

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis* *	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Šabanović, Bennett, Chang, & Huber (2013)	Observation study	PARO seal robot	RST and therapist-led group multi-sensory behavioural therapy (MSBT)	Observation Researcher	<i>t</i> -test	7 Dementia (not specified)	Positive effect of RST	11/22 (50%) Adequate	10/50 (20%) Limited
Silva et al. (2017)	Single-blind RCT	SenseCam wearable camera	RST-assisted cognitive training v/s Memo+ (pen/paper) training	GDS-30, IAFAI, WHOQOL-OLD Participant	ANOVA, pairwise comparisons, <i>t</i> -test	51 AD	Positive effect of RST	23/28 (82%) Strong	13/50 (26%) Limited
Soler et al. (2015)	Block RCT	NAO humanoid robot, PARO seal robot	Phase 1 – Therapeutic intervention with NAO v/s PARO v/s control Phase 2 – Therapeutic intervention with PARO v/s dog v/s control	GDS, sMMSE, MMSE, NPI, APADEM-NH, AI, QUALID Blinded raters	Wilcoxon, Mann-Whitney, Kruskal-Wallis tests	Nursing home – 101 (Phase 1), 110 (Phase 2) Day Centre – 20 (Phase 1), 17 (Phase 2) Dementia (not specified)	Inconclusive results due to large number of comparisons	21/28 (75%) Good	11/50 (22%) Limited

Authors and year of publication	Study design	RST type	Intervention	Outcome measures* and reporter	Analysis	N and diagnosis*	Main results	Total QualSyst rating (%)	Total DeBS rating (%)
Takayanagi, Kirita, & Shibata (2014)	Within subject's design	PARO robot seal	Observation of interaction with RST v/s plush toy	Behavioural analysis Researcher	Time sampling method - Wilcoxon signed-rank test	19 Mild to moderate 11 Severe Dementia (not specified)	Positive effect of RST	13/28 (46%) Limited	11/50 (22%) Limited
Wang et al. (2017)	Descriptive mixed methods	Robot Ed for daily activities	RST-prompted completion of daily activity	Interview Participant Carer	Thematic Analysis	10 AD 10 Carer	RST not accepted by PwD	12/20 (60%) Adequate	37/50 (74%) Good
Woodberry et al. (2015)	Within subject's longitudinal design	SenseCam wearable camera	RST assisted v/s written diary assisted recall	Recall of events Participants	Chi-square	6 AD	Positive effect of RST	14/28 (50%) Adequate	17/50 (34%) Limited

* Measures: IPPA (Individually Prioritized Problems Assessment), GIP-28 (Behaviour Rating Scale for Psychogeriatric Inpatients), I-PANAS-SF (International Positive and Negative Affect Schedule Short-Form), MMSE (Mini-Mental State Examination), NPI-ES (Neuropsychiatric Inventory for Health Staff), SUS (System Usability Scale), QUALID (Quality of Life in Late-Stage Dementia), CMAI (Cohen-Mansfield Agitation Inventory-Long Form), BARS (Brief Agitation Rating Scale), CSDD (Cornell Scale for Symptoms of Depression in Dementia), MMSE-K (Korean Mini-Mental State Examination), AER (Apparent Emotion Rating Instrument), K-CMAI (Korean Cohen-Mansfield Agitation Inventory), QOL-D (Quality of Life Questionnaire for Dementia), DBD (Dementia Behaviour Disturbance Scale), NPI-NH (Neuropsychiatric Inventory Nursing Home version), BI (Barthel Index), VTI (Vitality Index), ACE (Addenbrooke's Cognitive Examination), CMAI-SF (Cohen-Mansfield Agitation Inventory-Short Form), NPI-Q (Neuropsychiatric Inventory Questionnaire), QOL-AD (Quality of Life in Alzheimer's Disease Scale), RAID (Rating Anxiety in Dementia Scale), AES (Apathy Evaluation Scale), GDS (Geriatric Depression Scale), AWS (Revised Algate Wandering Scale-Nursing Home version), OERS (Observed Emotion Rating Scale), ZBI (Zarit Burden Interview), RSCSE (Revised Scale for Caregiving Self-Efficacy), OSBC (Oshkosh Social Behaviour Coding Scale), GSR (galvanic skin response), SQI Self-Identity Questionnaire, EVIBE (Instant Assessment of Wellbeing Tool), MPES (Menorah Park Engagement Scale), VAS (Visual Analogic Scale), GDS-30 (Geriatric Depression Scale-30), IAFAI (Adults and Older Adults Functional Assessment Inventory), WHOQOL-OL, D (World Health Organization Quality of Life-OLD), sMMSE (Severe Mini-Mental State Examination), APADEM-NH (Apathy Scale for Institutionalized Patients with Dementia Nursing Home version), AI (Apathy Inventory).

** Diagnosis: AD (Alzheimer's disease), FTD (fronto-temporal dementia), VD (vascular dementia), LBD (Lewy-body dementia), ARD (alcohol-related dementia), NDD (neurodegenerative disease), PD (Parkinson's disease)

5.3 Participant characteristics

All studies within this review included participants with a diagnosis of dementia. Two studies included one participant with no diagnosis of dementia (Jøranson et al., 2015; Jøranson et al., 2016a). However, as these participants were within the control group, the two studies were included within this review. Most studies included participants who had a diagnosis of dementia, but did not specify the type ($n = 20$; Bemelmans et al., 2015; Bemelmans et al., 2016; Chu, Khosla, Khaksar, & Nguyen, 2016; Demange et al., 2018; Gustafsson et al., 2015; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Khosla et al., 2017; Liang et al., 2017; Moyle et al., 2013; Moyle et al., 2014; Moyle et al., 2016; Moyle et al., 2017a; Moyle et al., 2017b; Nijhof et al., 2013; Petersen et al., 2017; Šabanović et al., 2013; Soler et al., 2015; Takayanagi et al., 2014). Following this, Alzheimer's Disease was the most common diagnosis, with 13 studies including such participants only (Begum et al., 2013; Begum et al., 2015; Kuwamura et al., 2016; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Lindqvist et al., 2013; Navarro, Rodriguez, & Favela, 2016; Silva et al., 2015; Wang et al., 2017; Woodberry et al., 2015). Four studies included participants with various dementia diagnoses, including Lewy body dementia, vascular dementia, fronto-temporal lobe dementia, alcohol-related dementia, Parkinson's-related dementia, mixed dementia, and unspecified dementia (Arntzen, Holthe, & Jentoft, 2016; Koh & Kang, 2018; Moyle et al., 2018; Rouaix et al., 2017).

A number of studies also included carers or family members ($n = 8$; Arntzen et al., 2016; Begum et al., 2013; Begum et al., 2015; Gustafsson et al., 2015; Liang et al., 2017; Moyle et al., 2014; Navarro et al., 2016; Wang et al., 2017) and staff or therapy facilitators ($n = 5$; Chu et al., 2016; Gustafsson et al., 2015; Lancioni et al.,

2018; Moyle et al., 2014; Nijhof et al., 2013) as participants. These studies were included within this review as they did not only consider the impact of RST on carers and staff but studied these effects as supplementary outcomes.

5.4 Sample Size

The sample sizes of the included studies varied greatly. The largest sample included 455 participants with dementia (Moyle et al., 2018) and the smallest included two participants with dementia and two carers (Navarro et al., 2016). Two studies included only three participants with dementia, one of which was a case series (Kuwamura et al., 2016), and the other a preliminary study (Lancioni et al., 2015). Other studies included a maximum of 30 carers or family members (Liang et al., 2017) and 28 members of staff or therapy facilitators (Chu et al., 2016).

5.5 Study design

A majority of studies included in this review were quantitative studies ($n = 26$; Begum et al., 2013; Begum et al., 2015; Chu et al., 2016; Demange et al., 2018; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Khosla et al., 2017; Koh & Kang, 2018; Kuwamura et al., 2016; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Moyle et al., 2013; Moyle et al., 2017b; Moyle et al., 2018; Navarro et al., 2016; Nijhof et al., 2013; Petersen et al., 2017; Rouaix et al., 2017; Šabanović et al., 2013; Silva et al., 2017; Soler et al., 2015; Woodberry et al., 2015), whilst a small number were qualitative ($n = 7$; Arntzen et al., 2016; Bemelmans et al., 2016; Lindqvist et al., 2013; Moyle et al., 2014; Moyle et al., 2016; Moyle et al., 2017a; Wang et al., 2017) and mixed-methods ($n = 4$; Bemelmans et al., 2015; Gustafsson et al., 2017; Liang et al., 2017; Takayanagi et al., 2015) studies.

Eleven of the studies were observational (Begum et al., 2013; Begum et al., 2015, Chu et al., 2016; Jøranson et al., 2016b, Khosla et al., 2017; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Moyle et al., 2014; Šabanović et al., 2013). Nine studies were RCTs (Jøranson et al., 2015; Jøranson et al., 2016a; Liang et al., 2017; Moyle et al., 2013; Moyle et al., 2017b; Moyle et al., 2018; Petersen et al., 2017; Silva et al., 2017; Soler et al., 2015;), whilst five studies were quasi-experimental (Bemelmans et al., 2015; Demange et al., 2018; Koh & Kang, 2018) or uncontrolled experimental studies (Gustafsson et al., 2015; Rouaix et al., 2017). Three studies employed a within-subjects design (Navarro et al., 2016; Nijhof et al., 2013; Takayanagi et al., 2014). This review also included two case studies (Moyle et al., 2016) and a case series (Kuwamura et al., 2016), one which used a cross-sectional design (Lanza et al., 2014), one which was an interview study (Moyle et al., 2017a) and one which was a descriptive study (Wang et al., 2017). Each of these was included due to their use of systematic procedures to study the effects of RST, rather than being purely observational or exploratory. Three studies were longitudinal, with follow up ranging from one month (Woodberry et al., 2015) to six months (Lindqvist et al., 2013) or 12 months (Arntzen et al., 2016). One study was mainly a feasibility study (Bemelmans et al., 2016); however, as it also considered efficacy of RST within an individualised intervention, it was included within this review.

Most of the studies included in this review employed RST with participants as an individual intervention ($n = 23$; Arntzen et al., 2016; Begum et al., 2013; Begum et al., 2015; Bemelmans et al., 2015; Bemelmans et al., 2016; Demange et al., 2018; Gustafsson et al., 2017; Kuwamura et al., 2016; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Lindqvist et al.,

2013; Moyle et al., 2014; Moyle et al., 2016; Moyle et al., 2018; Navarro et al., 2016; Rouaix et al., 2017; Silva et al., 2017; Takayanagi et al., 2014; Wang et al., 2017; Woodberry et al., 2015), whilst the remaining 14 studies considered the use of RST for PwD within group-based interventions (Chu et al., 2016; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Khosla et al., 2017; Koh & Kang, 2018; Liang et al., 2017; Moyle et al., 2013; Moyle et al., 2017a, Moyle et al., 2017b; Nijhof et al., 2013; Petersen et al., 2017; Šabanović et al., 2013; Soler et al., 2015).

5.6 ***Type of RST studied***

A large number of studies included in this review considered the use of companion robots in animal form, like PARO the seal (Figure 2.3; $n = 14$; Bemelmans et al., 2015; Bemelmans et al., 2016; Demange et al., 2018; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Koh & Kang, 2018; Liang et al., 2017; Moyle et al., 2013; Moyle et al., 2017a; Moyle et al., 2017b; Petersen et al., 2017; Šabanović et al., 2013; Takayanagi et al., 2014) and the CuDDler, a stuffed bear (Figure 2.4; $n = 1$; Moyle et al., 2016) or JustoCat ($n = 1$; Gustafsson et al., 2015). Others considered the use of humanoid companion robots like NAO ($n = 1$; Rouaix et al., 2017), Jack and Sophie ($n = 1$; Chu et al., 2016), Matilda ($n = 1$; Khosla et al., 2017) or Telenoid ($n = 1$; Kuwamura et al., 2016). One study looked at the use of Giraff (Moyle et al., 2014), an RST that allows participants to remotely interact with family members through the use of a video conferencing unit attached to a mobile frame. The remaining studies considered the use of RST in the form of assistive devices like Robot Ed, an upright smart computer with a display screen at the top, top and bottom cameras, a microphone and speakers used to assist with and prompt completion of daily activities (Figure 2.5; $n = 3$; Begum et al., 2013; Begum et al., 2015; Wang et

al., 2017), computers with sound amplifiers and/or a pressure microswitch or button that can be pressed, used to promote exercise ($n = 4$; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018), SenseCam devices like a camera to enhance participants' memory of events (Figure 2.6; $n = 2$; Silva et al., 2017; Woodberry et al., 2015), HTC Touch Smartphone navigator ($n = 1$; Lanza et al., 2014), Chitchatters, which includes interactive objects like a telephone, a box, a television and a radio to support participants to engage in leisure activities through the triggering of memories ($n = 1$; Nijhof et al., 2013), or AnswerPad (a touchscreen computer) and AnswerBoard (an automated reminder board) to support independent completion of daily activities ($n = 1$; Navarro et al., 2016). Two studies considered the use of various types of RST in the form of assistive devices individualised to each participant's needs (Arntzen et al., 2016; Lindqvist et al., 2013).



Figure 2.3. PARO the seal (Shibata et al., 2012)



Figure 2.4. CuDDler (developed by Dr Tan Yeow Kee, Infocomm, Singapore)



Figure 2.5. Robot Ed (Begum et al., 2013)



Figure 2.6. SenseCam (developed by Microsoft Research Cambridge)

One study considered the use of PARO in comparison to NAO, to identify whether there was a difference in effects or preference across animal or humanoid RSTs (Soler et al., 2015), whilst another combined the use of PARO and SenseCam to study effects on day and night time behaviours (Moyle et al. 2018).

5.7 Main outcome measures

A range of outcome measures were used to study the effects of RST on cognitive and behavioural symptoms of participants with dementia. A large number of studies measured outcomes using observation of behaviour only ($n = 11$; Chu et al., 2016; Jøranson et al., 2016b; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Moyle et al., 2018; Nijhof et al., 2013; Šabanović et al., 2013; Takayanagi et al., 2014), via video recordings or direct observation of interactions with, or utilisation of, RST. Eleven of the 37 studies in this review also measured outcomes using standardised questionnaires only (Bemelmans et al., 2015; Demange et al., 2018; Gustafsson et al., 2015; Jøranson et al., 2015; Jøranson et al., 2016a; Moyle et al., 2013; Navarro et al., 2016; Petersen et al., 2017; Rouaix et al., 2017; Silva et al., 2017; Soler et al., 2015), measuring mood, cognitive state, quality of life and ability to perform daily activities. Semi-structured interviews were used in three studies as the main outcome measure only (Bemelmans et al., 2016; Moyle et al., 2017a; Wang et al., 2017).

Nine of the 37 studies included in this review evaluated outcomes using observation supplemented with one other outcome measure, including semi-structured interviews ($n = 4$; Arntzen et al., 2016; Lindqvist et al., 2013; Moyle et al. 2014; Woodberry et al., 2015), standardised questionnaires ($n = 4$; Koh & Kang, 2018; Kuwamura et al., 2016; Liang et al., 2017; Moyle et al., 2017b) and surveys ($n = 1$; Khosla et al.,

2017). Three studies included more than two main outcome measures; two studies included observations, interviews and surveys (Begum et al., 2013; Begum et al. 2015) and one included observations, interviews and standardised questionnaires (Moyle et al., 2016).

5.8 Reporting of outcomes

All studies in this review included participants with dementia. However, only four studies gathered outcome measures from the participants only (Demange et al., 2018; Lindqvist et al., 2013; Silva et al., 2017; Woodberry et al., 2015). In a large number of studies, reporting of outcomes was done only by the researchers themselves ($n = 18$; Begum et al., 2015; Chu et al., 2016; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Koh & Kang, 2018; Kuwamura et al., 2016, Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Moyle et al. 2013; Moyle et al., 2018; Nijhof et al., 2013; Šabanović et al., 2013; Soler et al., 2015; Takayanagi et al., 2014), followed by three studies where only carers or family members reported on outcomes (Bemelmans et al., 2015; Moyle et al., 2017a; Navarro et al., 2016). In one study, staff members alone completed outcome measures (Petersen et al., 2017).

Eleven of the 37 included studies gathered outcome measures from more than one reporter (Arntzen et al., 2016; Begum et al., 2013; Bemelmans et al., 2016; Gustafsson et al., 2015; Khosla et al., 2017; Liang et al., 2017; Moyle et al., 2014; Moyle et al., 2016; Moyle et al., 2017b; Rouaix et al., 2017; Wang et al., 2017). In two studies, participants and researchers reported on different outcome measures (Khosla et al., 2017; Moyle et al., 2016; Rouaix et al., 2017); for example, surveys and observations, respectively. In two studies, researchers and carers or family

members reported on outcomes (Gustafsson et al., 2015; Liang et al., 2017). In one study, researchers and staff members reported on outcomes (Moyle et al., 2017b), whereas in another, participants and carers or family members reported on outcomes (Wang et al., 2017). Outcomes were gathered from participants, carers or family members and staff in two of these studies (Bemelmans et al., 2016), and from researchers, carers or family members and staff in one study (Moyle et al., 2014).

5.9 Nature of outcomes

Most of the studies included in this review found positive outcomes for the impact of the studied RST on the cognitive or behavioural symptoms of participants with dementia ($n = 32$; Arntzen et al., 2016; Bemelmans et al., 2015; Bemelmans et al., 2016; Chu et al., 2016; Demange et al. 2018; Gustafsson et al., 2015; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Khosla et al., 2017; Koh & Kang, 2018; Kuwamura et al., 2016; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Liang et al. 2017; Lindqvist et al., 2013; Moyle et al., 2013; Moyle et al., 2014; Moyle et al., 2017a; Moyle et al., 2017b; Moyle et al., 2018; Navarro et al., 2016; Nijhof et al., 2013; Petersen et al., 2017; Rouaix et al., 2017; Šabanović et al., 2013; Silva et al. 2017; Takayanagi et al., 2014; Woodberry et al., 2015). Three studies found negative outcomes (Begum et al., 2015, Moyle et al., 2016; Wang et al., 2017). In two studies this related to the features of the RST themselves affecting usability (Begum et al., 2015; Moyle et al., 2016), whereas in one study this was related to participants' decision not to use the RST (Wang et al., 2017). Two studies had inconclusive results (Begum et al. 2013; Soler et al. 2015). For one study, this was due to the small sample size (Begum et al., 2013), whilst for the other this was due to a possible

Type 1 error associated with the use of multiple uncorrected statistical comparisons (Soler et al., 2015).

5.10 Ratings on the QualSyst Measure

An assessment of the methodological quality of each of the papers reviewed indicated that they ranged from strong (>80% compliance) to limited (<50% compliance), on the QualSyst measure. A large number of the studies ($n = 17$) were rated as adequate (50-70% compliance; Begum et al., 2013; Begum et al., 2015; Bemelmans et al., 2015; Gustafsson et al., 2015; Jøranson et al., 2016b; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Liang et al., 2017; Moyle et al., 2016; Moyle et al., 2017a; Nijhof et al., 2013; Petersen et al., 2017; Rouaix et al., 2017; Šabanović et al., 2013; Wang et al., 2017; Woodberry et al., 2015). Eight studies were rated as good (70–80% compliance; Arntzen et al., 2016; Demange et al., 2018; Jøranson et al., 2016a; Khosla et al., 2017; Koh & Kang, 2018; Lindqvist et al., 2013; Moyle et al., 2013; Soler et al., 2015), whereas six studies each were rated as strong (Chu et al., 2016; Jøranson et al., 2015; Moyle et al., 2014; Moyle et al., 2017b; Moyle et al., 2018; Silva et al., 2017) or limited (Bemelmans et al., 2016; Kuwamura et al., 2016; Lancioni et al., 2015; Lancioni et al., 2016; Navarro et al., 2016; Takayanagi et al., 2014).

5.11 Ratings on the DeBs

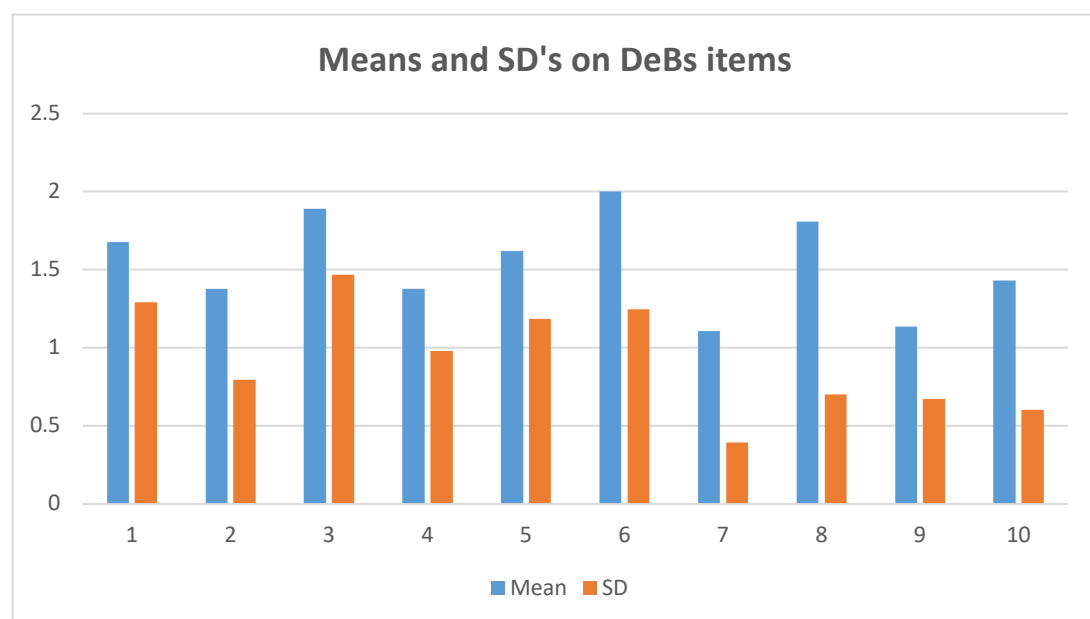
Assessment of the extent to which studies adhered to person-centred principles of psychological care, as measured by the DeBs, ranged from good (70–80% compliance) to limited (<50% compliance). Only one study received a good rating (Wang et al., 2017), whilst three studies were rated as adequate (50-70% compliance; Arntzen et al., 2016; Lindqvist et al., 2013; Rouaix et al., 2017). The

remaining 33 papers included in this review received a rating of limited on the DeBs (Begum et al., 2013; Begum et al., 2015; Bemelmans et al., 2015; Bemelmans et al., 2016; Chu et al., 2016; Demange et al., 2018; Gustafsson et al., 2015; Jøranson et al., 2015; Jøranson et al., 2016a; Jøranson et al., 2016b; Khosla et al., 2017; Koh & Kang, 2018; Kuwamura et al., 2016; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Liang et al., 2017; Moyle et al., 2013; Moyle et al., 2014; Moyle et al., 2016; Moyle et al., 2017a; Moyle et al., 2017b; Moyle et al., 2018; Navarro et al., 2016; Nijhof et al., 2013; Petersen et al., 2017; Šabanović et al., 2013; Silva et al., 2013; Soler et al., 2015; Takayanagi et al., 2014; Woodberry et al., 2015).

Overall, average scores on each of the items on the DeBs were low (ranging from 1.1 – 2). Scores on items such as ‘does the approach advocate and facilitate access to meaningful activities?’ (score of 2) and ‘does the approach involve listening to people with dementia and ensuring their voice is heard?’ (score of 1.9) were higher than those on items such as ‘does the approach promote participation in local communities that are dementia friendly’ (score of 1.1) and ‘does the approach consider the impact of stigma on dementia?’ (score of 1.1). Table 2.2 summarises mean (*M*), standard deviation (*SD*) and range across all 37 studies on each of the DeBs items. Scores are visually represented via a graph in Figure 2.7.

Table 2.2. Mean, standard deviations and ranges for individual DeBs items

	Individualised	Identity	Voice	Decision-making	Choice	Activities	Community	Meaningful	Stigma	Roles
<i>Mean</i>	1.7	1.4	1.9	1.4	1.6	2	1.1	1.8	1.1	1.4
<i>SD</i>	1.3	0.8	1.5	0.98	1.2	1.2	0.4	0.7	0.7	0.6
Range (min-max)	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	1 – 5	1 – 3	1 – 4	1 – 5	1 – 3

Figure 2.7. Graph representing means and standard deviations for DeBs items

6. Discussion

This review aimed to assess the current attitudes towards use of RST for dementia care and whether these are consistent with person-centred recommendations made by the BPS (2016). The use of the DeBs (Morris, 2018) as the primary measure of

person-centred attitudes makes this review the first of its kind. The inclusion of the QualSyst measure provides additional information on the methodological quality of the 37 included studies. Key information on study design, outcome measures, reporting of outcomes and findings was extracted. Together, these three sources of information were used to gain an understanding of the nature of the studies included in the review, and about author attitudes towards the use of RST in dementia care.

In order to provide a current perspective of attitudes among researchers and authors, rather than a summary of the existing body of research, this review focused on recently published literature, between 2013 and 2018. It was felt that this better met the aims of the review, considering the fast-evolving technology and ideas relating to RST. Further, BPS guidelines on person-centred dementia care were only published in 2016, suggesting that earlier studies may not have evolved in line with this thinking. Hence, most of these studies were felt to be outside the scope of the current review. Looking at the current body of research also serves as a baseline for future reviews, to consider a comparison of how attitudes have evolved over time.

Of particular interest is the type of RST implemented in the studies included in this review. Almost all the studies considered the use of companion or social robots like PARO or NAO on the social engagement, mood and quality of life of PwD. This appears to be a shift from previous research, which tended to focus more on the use of RST to carry out simple activities of daily living, rather than on the improvement of quality of life for PwD (Evans, Brown, Coughlan, Lawson, & Craven, 2015; Sixsmith et al., 2007). The current body of research, published between 2013 and 2018, appears to address the need for more studies considering the use of RST on well-being and quality of life of PwD.

Most of the studies included in this review found positive effects of the use of RST on a variety of outcomes for PwD, including mood, level of social engagement, fostering independence during completion of daily activities, memory and agitation. However, these results must be considered within the context of the methodological limitations outlined by the quality assessment, as most of the studies were rated either adequate or limited on the quality assessment. For example, for most studies where blinding of investigators and/or participants was possible, this was either not done or not reported (Demange et al., 2018; Gustafsson et al., 2017; Jøranson et al., 2015; Jøranson et al., 2016a; Koh & Kang, 2018; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Liang et al., 2017; Moyle et al., 2013; Navarro et al., 2016; Nijhof et al., 2013; Petersen et al., 2017; Silva et al., 2017; Takayanagi et al., 2014; Woodberry et al., 2015). This could indicate a possible bias in results, particularly for observational studies, where the risk of bias may be greater (Hammer, du Prel, & Blettner, 2009). Another methodological limitation in a large number of studies was the small sample size, making it difficult to generalise these results to the wider population. (Begum et al., 2013; Begum et al., 2015; Bemelmans et al., 2015; Chu et al., 2015; Demange et al., 2018; Jøranson et al., 2016a; Jøranson et al., 2016b; Khosla et al., 2017; Lancioni et al., 2015; Lancioni et al., 2016; Lancioni et al., 2017; Lancioni et al., 2018; Lanza et al., 2014; Liang et al., 2017; Moyle et al., 2013; Navarro et al., 2016; Nijhof et al., 2013; Petersen et al., 2017; Rouaix et al., 2017; Šabanović et al., 2013; Silva et al., 2017; Soler et al., 2015; Woodberry et al., 2015).

Overall, study ratings on the DeBs indicated low compliance with person-centred principles, with only one study receiving a good rating (Wang et al., 2017). Unlike other studies included in this review, this study focused on interviewing the study

participants (PwD) themselves to gain their perspective on the use of RST for carrying out daily activities. This study gained points on the DeBs for listening to PwD and giving them the opportunity to make choices, while also enabling them to engage in meaningful activities. Interestingly, this study noted differences between the opinions of family members or carers and the PwD, when it came to adopting the RST in daily life. Where both parties recognised the ease of use of the RST, PwD did not feel ready for RST and preferred the support of an “efficient human being” (p. 72). This contrasted with the readiness of carers or family members to implement RST immediately due to its perceived benefits, which they felt was an opinion shared by the PwD as well. Lilja, Bergh, Johansson, and Nygård, in their 2003 study, noted a similar dissonance between the perspectives of qualified practitioners and service users, with regard to the use of AT to meet their rehabilitation needs and levels of disability. Further, another study found that carers or family members were more likely to promote the use of RST, even compromising the independence and autonomy of the PwD, when they felt responsible for their safety (Landau, Ausländer, Werner, Shoval, & Heinik, 2010).

Other studies included in this review also considered the views of PwD, alongside those of carers or family members and staff. However, ratings on the DeBs suggest that this was often done in a cursory manner. A recent study conducted by Demange et al. (2018) focused mainly on self-report questionnaires as outcome measures provided by PwD. However, the intervention was not tailored as per preferences of participants. Due to this, the authors noted a high number ($n = 6$) of refusals to participate in the study. Reasons for attrition included a lack of interest in robots and finding the RST too “childlike” (p. 1307). Further, one participant experienced nightmares related to the RST, following exposure to the robot. This study also

identified greater positive effects of RST on negative affect for participants who were agitated, as compared to those who were depressed or apathetic, emphasising the importance of taking the PwD's choices and personal history into account when tailoring an intervention.

Wang et al. (2017) also pointed out the importance of timing when implementing RST for PwD, identifying that they may be more suitable for those with advanced dementia, or complex needs. Participants in this study with mild-moderate dementia reported feeling stigmatised by being offered an intervention they considered as being for those with greater needs than themselves. As per BPS guidelines (2016), it is important to adapt interventions to the context of where the person is within their dementia journey. A study conducted by Takayanagi et al. (2014) compared the effects of RST on people with mild-moderate dementia and severe dementia. They noted similar positive effects of RST on engagement and emotional expression across both groups. This finding appears to be significant, in light of the difficulties experienced by the participants with severe dementia in communicating their needs to staff members. These participants were observed to talk to the RST more spontaneously than to staff members, whilst those in the mild-moderate dementia group were able to interact with both equally.

This brings us to the long-standing question of whether the use of RST exacerbates social isolation and loneliness (Sharkey & Sharkey, 2010) or in fact decreases these. Takayanagi et al. (2014) opine that an observed increase in interaction with the RST leading to a reduction in the need and demand to interact with staff members was indicative of a decreased sense of loneliness. However, this study did not gain feedback from participants, so it is unclear whether this is the case, and there may be

a risk of RST interaction replacing interaction with staff members. Moyle et al. (2016) noted that PwD participating in their study viewed the RST as “a toy” (p. 151), and felt it would be more suitable for a child. Participants were observed to lose interest in the RST over time, preferring to engage with the facilitator instead. This raises the question of whether RST can ever replace human interaction, and whether this outcome is desirable. Wang et al. (2017) offer a word of caution relating to this. They identified the importance of finding a balance in encouraging engagement with an RST, ensuring that this does not replace regular human interaction with staff and carers or family members. PwD participating in this study noted that an RST could be useful; however, “you can never replace people” (p. 74). Another participant noted that developing an attachment to the RST would be difficult “because it’s a robot and you’re not” (p. 74). In this respect, carers or family members held similar views to their relative with dementia. In contrast, some participants with dementia were able to identify with the RST as a companion with whom they could form an attachment. This again emphasises the importance of tailoring interventions to the person’s needs and preferences, as clearly one size does not fit all. This is one of the main premises of the BPS guidelines (2016).

Three studies included within this review (Arntzen et al., 2016; Lindqvist et al., 2013; Rouaix et al., 2017) individualised the RST intervention to the PwD’s needs and preferences. Arntzen et al. (2016) and Lindqvist et al. (2013) carried out interviews and observations with participants and their carers or family members to identify difficulties they were experiencing at home, and to support them to choose a suitable RST to overcome these challenges. As a result, numerous forms of RST were implemented, specifically tailored to the problem at hand. Participants chose RST support mainly to carry out daily living activities, like remembering to take

their medication or orienting to time and place. A few participants chose support for keeping in touch with family members. These appear to be in line with previous findings that RSTs are not currently viewed as meeting social or companionship needs. Following this, participants' views on RST were gathered to identify factors contributing to successful adoption of RSTs in daily life. Rouaix et al. (2017) utilised a single RST, NAO, to carry out a psychomotor therapy session, but tailored the content of the intervention so it was more personalised to individual participants. Via pre-intervention interviews with participants, NAO was programmed to respond according to their preferences. For example, voice pitch and volume were adjusted as per participants' preferences and physical needs, NAO was programmed to address each participant by name and therapy content was adjusted to reflect participants' preferences of activities and music.

All three of these studies stated that these individualised outcomes led to better acceptance and implementation of the RST by participants. This is in conjunction with the BPS guidelines (2016) around ensuring PwD are given the opportunity to make decisions about their care. However, one drawback of these studies was that they did not include comparison groups, to identify if there was a difference in adoption where interventions were not personalised. The Rouaix et al. (2017) study compared an RST-led personalised intervention to a therapist-led intervention, resulting in no significant difference in engagement across the two conditions.

Arntzen et al. (2016) and Lindqvist et al. (2013) also found that participants in their studies would more readily accept those RST that allowed them to maintain their autonomy and independence. For example, the young PwD who were participants in the Arntzen et al. (2016) study preferred RST that they were able to operate

themselves, with minimal input from their caregivers. Similarly, Lindqvist et al. (2013) found that participants lost confidence in their skills and abilities when the RST did not function as it was supposed to, which led them to reject the RST. As identified by the Institute of Medicine (IOM, 1997) model of disability, the interaction of the person with their environment is a key factor in determining how disabled a person becomes. Drawing upon this, the BPS (2016) emphasised the importance of promoting dementia-friendly communities and addressing the stigma associated with the diagnosis, to ensure people can live fulfilled lives. Moyle et al. (2016) noted the importance of ensuring that up-to-date and fully functioning technology is used when implementing RST, as failure of the technology in their study led to poorer outcomes for participants and complete rejection of the RST.

7. Limitations of the study

As this was the first time the DeBs was used since its publication in 2018, some difficulties were experienced during its use. There were greater initial discrepancies among the researchers rating papers on this scale, as opposed to the QualSyst measure, particularly on those studies which met the requirements of the DeBs, at least to some extent. To address this, ratings were discussed, and papers rated again, with reliability then being good. Detailed guidance on using the DeBs, as is the case with the QualSyst manual, may be helpful in addressing such discrepancies in the future. Further, as this is the first review of its kind, an important target for future research may be to examine changes in attitudes towards implementation of RST over time, with this review providing a baseline.

8. Conclusions

Overall, there appears to have been a shift in how RST is being implemented currently, in comparison to findings from previous reviews. A larger number of studies are considering the use of RST for social engagement and companionship, addressing wider issues around well-being and quality of life of PwD, rather than focusing only on ease of completing daily life tasks.

However, only a few robust studies have been conducted recently, with most being observational studies, which may have a higher risk of researcher bias, or studies lacking comparison groups. Some RCTs have been conducted, and although these are methodologically stronger, they lack the individualisation of interventions to each participant's needs, and hence do not link research to clinical practice. The focus of the research community must shift from a 'one size fits all' approach to the use of RST for PwD.

Further, only a few studies consider the views and opinions of the PwD themselves, depending largely on caregiver or staff views. The difficulty is that caregiver and staff views may reflect their own needs more than those of the PwD themselves.

Hence, involving PwD in future research will be key to establishing a strong evidence base for use of RST in the future.

The need for more methodologically robust studies, which consider the preferences, needs and views of PwD is clear. Recommendations for future studies also include those that consider i) people at different stages of their dementia journey, ii) the stigma associated with using RST in dementia, and iii) usable technology to foster independence and autonomy.

9. References

- Agree, E., & Freedman, V. (2000). Incorporating assistive devices into community-based long term care. *Journal of Aging and Health, 12*(3), 426-450.
- Alzheimer's Association (2018). 2018 Alzheimer's disease facts and figures. *Alzheimer's Dementia, 14*(3), 367-429.
- Alzheimer's Disease International (2015). *The global impact of dementia: an analysis of prevalence, incidence, cost and trends*. Alzheimer's Disease International (ADI), London.
- Arntzen, C., Holthe, T., & Jentoft, R. (2016). Tracing the successful incorporation of assistive technology into everyday life for younger PwD and family carers. *Dementia, 15*(4), 646-662.
- Begum, M., Huq, R., Wang, R., & Mihailidis, A. (2015). Collaboration of an assistive robot and older adults with dementia. *Gerontechnology, 13*(4), 405-419.
- Begum, M., Wang, R., Huq, R., & Mihailidis, A. (2013). Performance of daily activities by older adults with dementia: the role of an assistive robot. *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICOR)*, Seattle, WA, pp. 1-8.
- Bemelmans, R., Gelderblom, G. J., Jonker, P. & de Witte, L. (2015). Effectiveness of robot PARO in intramural psychogeriatric care: a multicentre quasi-experimental study. *Journal of the American Medical Directors Association, 16*(11), 946-950.
- Bemelmans, R., Gelderblom, G. J., Jonker, P. & de Witte, L. (2016). How to use robotic interventions in intramural psychogeriatric care; a feasibility study. *Applied Nursing Research, 30*, 154-157.

- Berg-Weger, M., & Stewart, D. B. (2017). Non-pharmacological interventions for PwD. *Missouri Medicine*, *114*(2), 116-119.
- Bharucha, A. J., Anand, V., Forlizzi, J., Dew, M. A., Reynolds 3rd, C. F., Stevens, S., & Wactlar, H. (2009). Intelligent assistive technology applications to dementia care: Current capabilities, limitations, and future challenges. *American Association for Geriatric Psychiatry*, *17*(2), 88-104.
- Bogue, R. (2013). Robots to aid disabled and the elderly. *The Industrial Robot*, *40*(6), 519-524.
- British Psychological Society (2016). *Psychological dimensions of dementia: Putting the person at the centre of care*. Leicester: British Psychological Society.
- Brodsky, H., & Donkin, M. (2009). Family caregivers of PwD. *Dialogues in Clinical Neuroscience*, *11*(2), 217-228.
- Brown, E. (2018). How can care homes overcome social isolation? *Nursing and Residential Care*, *20*(3), 136-139.
- Cahill, S., Begley, E., Faulkner, J. P., & Hagen, I. (2007). "It gives me a sense of independence": Findings from Ireland on the use and usefulness of assistive technology for PwD. *Technology and Disability*, *19*, 133-142.
- Chiberska, D. (2018). The use of robotic animals in dementia care: Challenges and ethical dilemmas. *Mental Health Practice*, *21*(10), 23-28.
- Chu, M-T., Khosla, R., Khaksar, S. M. S., & Nguyen, K. (2016). Service innovation through social robot engagement to improve dementia care quality. *Assistive Technology*, *29*(1), 8-18.
- Collins, A. (2014). *Measuring what really matters: towards a coherent measurement system to support person-centred care*. The Health Foundation, London.

- Cowling, A., & McFadden, S. (2010). The benefits and limitations of pet therapy for PwD. *Oshkosh Scholar*, 5, 82-92.
- Demange, M., Lenoir, H., Pino, M., Cantegreil-Kallen, I., Rigaud, A. S., & Cristancho-Lacroix, V. (2018). Improving well-being in patients with major neurodegenerative disorders: Differential efficacy of brief social robot-based intervention for 3 neuropsychiatric profiles. *Clinical Interventions in Aging*, 13, 1303-1311.
- Donaldson, C., Tarrrier, N., & Burns, A. (1997). The impact of the symptoms of dementia on caregivers. *The British Journal of Psychiatry*, 170(1), 62-68.
- Evans, J., Brown, M., Coughlan, T., Lawson, G., & Craven, M. P. (2015). A systematic review of dementia focused assistive technology. In M. Kurosu (Ed.), *Human-computer interaction: Interaction technologies* (pp.406-417). Los Angeles, CA: Springer.
- Feast, A., Orrell, M., Charlesworth, G., Melunsky, N., Poland, F. & Moniz-Cook, E. (2016). Behavioural and psychological symptoms in dementia and the challenges for family carers: systematic review. *The British Journal of Psychiatry*, 208(5), 429-34.
- Gustafsson, C., Svanberg, C., & Müllersdorf, M. (2015). Using a robotic cat in dementia care: A pilot study. *Journal of Gerontological Nursing*, 41(10), 46-56.
- Hammer, G. P., du Prel, J. B., & Blettner, M. (2009). Avoiding bias in observational studies: Part 8 in a series of articles on evaluation of scientific publications. *Deutsches Arzteblatt International*, 106(41), 664-668.
- Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A.,... Wood, K. (2006). *Ubicomp 2006*, Orange County, CA, pp. 177-193.
- Institute of Medicine (1997). Models of disability and rehabilitation. In E. N. Brandt & A. M. Pope (Eds.), *Enabling America: assessing the role of*

rehabilitation science and engineering. (pp. 62-80). Washington (DC): Institute of Medicine National Academy Press.

- Jøranson, N., Pedersen, I., Rokstad, A. M. M., & Ihlebæk, C. (2015). Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. *Journal of the American Medical Directors Association, 16*(10), 867-873.
- Jøranson, N., Pedersen, I., Rokstad, A. M. M., & Ihlebæk, C. (2016a). Change in quality of life in older PwD participating in Paro-activity: A cluster-randomized controlled trial. *Journal of Advanced Nursing, 72*(12), 3020-3033.
- Jøranson, N., Pedersen, I., Rokstad, A. M. M., Aamodt, G., Olsen, C., & Ihlebæk, C. (2016b). Group activity with Paro in nursing homes: A systematic investigation of behaviors in participants. *International Psychogeriatrics, 28*(8), 1345-1354.
- Khosla, R., Nguyen, K., & Chu, M-T. (2017). Human robot engagement and acceptability in residential aged care. *International Journal of Human-Computer Interaction, 33*(6), 510-522.
- Kirkham, J., Sherman, C., Velkers, C., Maxwell, C., Gill, S., Rochon, P., & Seitz, D. (2017). Antipsychotic use in dementia: Is there a problem and are there solutions? *The Canadian Journal of Psychiatry, 62*(3), 170-181.
- Kmet, L. M., Lee, R. C., & Cook, L. S. (2004). Standard quality assessment criteria for evaluating primary research papers from a variety of fields. Edmonton: Alberta Heritage Foundation for Medical Research (AHFMR). AHFMR – HTA Initiative #13.
- Koh, I. S., & Kang, H. S. (2018). Effects of intervention using Paro on the cognition, emotion, problem behavior, and social interaction of elderly PwD. *Journal of Korean Academy of Community Health Nursing, 29*(3), 300-309.

- Kuwamura, K., Nishio, S., & Sato, S. (2016). Can we talk through a robot as if face-to-face? Long-term fieldwork using teleoperated robot for seniors with Alzheimer's disease. *Frontiers in Psychology, 7*, 1066-1078.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafos, J., D'Amico, F., Sasanelli, G.,...Signorino, M. (2015). Persons with Alzheimer's disease engage in leisure and mild physical activity with the support of technology-aided programs. *Research in Developmental Disabilities, 37*, 55-63.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafos, J., D'Amico, F., Renna, C., & Pinto, K. (2016). Technology-aided programs to support positive verbal and physical engagement in persons with moderate or severe Alzheimer's disease. *Frontiers in Aging Neuroscience, 8*(87), 1-8.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafos, J., D'Amico, F., Addante, L. M., & Pinto, K. (2017). Persons with advanced Alzheimer's disease engage in mild leg exercise supported by technology-aided stimulation and prompts. *Behaviour Modification, 41*(1), 3-20.
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., Sigafos, J., D'Amico, F., Pinto, K.,...Caffó, A. O. (2018). Promoting supported ambulation in persons with advanced Alzheimer's disease: a pilot study. *Disability and Rehabilitation: Assistive Technology, 13*(1), 101-106.
- Landau, R., Ausländer, G. K., Werner, S., Shoval, N., & Heinik, J. (2010). Families' and professional caregivers' views of using advanced technology to track PwD. *Qualitative Health Research, 20*(3), 409-419.
- Lanza, C., Knörzer, O., Weber, M., & Riepe, M. W. (2014). Autonomous spatial orientation in patients with mild to moderate Alzheimer's disease by using mobile assistive devices: A pilot study. *Journal of Alzheimer's Disease, 42*(3), 879-884.

- Lee, L., Packer, T. L., Tang, S. H., & Girdler, S. (2008) Self-management education programs for age-related macular degeneration: A systematic review. *Australasian Journal on Ageing*, 27(4), 170–176.
- Leng, M., Liu, P., Zhang, P., Hu, M., Zhou, H., Li, G.,... Chen, L. (2019). Pet robot intervention for PwD: A systematic review and meta-analysis of randomized controlled trials. *Psychiatry Research*, 271, 516-525.
- Liang, A., Piroth, I., Robinson, H., MacDonald, B., Fisher, M., Nater, U. M.,... Broadbent, E. (2017). A pilot randomized trial of a companion robot for PwD living in the community. *Journal of the Medical Directors Association*, 18(10), 871-878.
- Lilja, M., Bergh, A., Johansson, L., & Nygård, L. (2003). Attitudes towards rehabilitation needs and support from assistive technology and the social environment among elderly people with disability. *Occupational Therapy International*, 10(1), 75-93.
- Lilja, M., Mansson, I., Jahlenius, L. & Sacco-Petersen, M. (2003). Disability policy in Sweden: Policies concerning assistive technology and home modification services. *Journal of Disability Policy Studies*, 14(3), 130-135.
- Limbu, D. K., Anthony, W. C. Y., Adrian, T. H. J., Dung, T. A., Kee, T. Y., Dat, T. H.,...Jun, L. (2013), Affective social interaction with CuDDler robot. In: 6th *IEEE Conference on Robotics, Automation and Mechatronics (RAM)*, pp. 179-184.
- Lindqvist, E., Nygård, L., & Borell, L. (2013). Significant junctures on the way towards becoming a user of assistive technology in Alzheimer's disease. *Scandinavian Journal of Occupational Therapy*, 20(5), 386-396.

- Lynn, J. D., Rondón-Sulbarán, J., Quinn, E., Ryan, A., McCormack, B., & Martin, S. (2017). A systematic review of electronic assistive technology within supportive living environments for PwD. *Dementia (0)*0, 1-65.
- Maher, A. R., Maglione, M., Bagley, S., Suttorp, M., Hu, J. H., Ewing, B.,...Shekelle, P. G., (2011). Efficacy and comparative effectiveness of atypical antipsychotic medications for off-label uses in adults: A systematic review and meta-analysis. *Journal of American Medical Association, 306*, 1359-1369.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009), Preferred reporting of items for systematic reviews and meta-analyses: The PRISMA statement. *PloS Medicine, 6*(7), 1-6.
- Moher D., Stewart L., & Shekelle, P. (2016). Implementing PRISMA-P: Recommendations for prospective authors. *Systematic Reviews, 5*(15), 1-2.
- Morris, R. (2018). Ethical issues and robotic or artificial intelligence care for people with neuropsychological conditions. *The Neuropsychologist, 5*, 40-42
- Moyle, W., Bramble, M., Jones, C., & Murfield, J. E. (2017a). “She had a smile on her face as wide as the great Australian bite”: A qualitative examination of family perceptions of a therapeutic robot and plush toy. *The Gerontologist, 00*(00), 1-9.
- Moyle, W., Cooke, M., Beattie, E., Jones, C., Klein, B., Cook, G., & Gray, C. (2013). Exploring the effects of companion robots on emotional expression in older adults with dementia: A pilot randomized controlled trial. *Journal of Gerontological Nursing, 39*(5), 46-53.
- Moyle, W., Fetherstonhaugh, D., Greben, M., Beattie, E., & AusQoL Group (2015). Influencers on quality of life as reported by people living with dementia in long term care: A descriptive exploratory approach. *BMC Geriatrics, 15*(50), 1-10.

- Moyle, W., Jones, C., Cooke, M., O'Dwyer, S., Sung, B. & Drummond, S. (2014). Connecting the PwD and family: a feasibility study of a telepresence robot. *BMC Geriatrics*, *14*(7), 1-11.
- Moyle, W., Jones, C., Murfield, J. E., Thalib, L., Beattie, E. R. A., Shum, D. K. H.,...Draper, B. M. (2017b). Use of a robotic seal as a therapeutic tool to improve dementia symptoms: A cluster-randomized controlled trial. *Journal of the Medical Directors Association*, *18*, 766-773.
- Moyle, W., Jones, C., Murfield, J. E., Thalib, L., Beattie, E. R. A., Shum, D. K. H.,...Draper, B. M. (2018). Effect of a robotic seal on the motor activity and sleep patterns of older PwD, as measured by wearable technology: A cluster-randomized controlled trial. *Maturitas*, *110*, 10-17.
- Moyle, W., Jones, C., Sung, B., Bramble, M., O'Dwyer, S., Blumenstein, M., & Estivill-Castro, V (2016). What effect does an animal robot called CuDDler have on the engagement and emotional response of older PwD? A pilot feasibility study. *International Journal of Social Robotics*, *8*, 145-156.
- Moyle, W., Venturato, L., Griffiths, S., Grimbeek, P., McAllister, M., Oxlade, D., & Murfield, J. (2011). Factors influencing quality of life for PwD: a qualitative perspective. *Aging and Mental Health*, *15*, 970-77.
- National Institute for Health and Care Excellence (2013). *Dementia: Independence and wellbeing*. London: NICE.
- Navarro, R. F., Rodriguez, M. D., & Favela, J. (2016). Use and adoption of an assisted cognition system to support therapies for PwD. *Computational and Mathematical Methods in Medicine*, *2016*, 1-10.
- Nijhof, N., van Hoof, J., van Rijn, H., & van Gemert-Pijnen, J. E. W. C. (2013). The behavioural outcomes of technology-supported leisure activity in PwD. *Technology and Disability*, *25*, 263-273.

- Orpwood, R., Sixsmith, A., Torrington, J., Chadd, J., Gibson, G., & Chalfont, G. (2007). Designing technology to support quality of life of PwD. *Technology and Disability, 19*, 103-112.
- Petersen, S., Houston, S., Qin, H., Tague, C., & Studley, J. (2017). The utilization of robotic pets in dementia care. *Journal of Alzheimer's Disease, 55*, 569-574.
- Rouaix, N., Retru-Chavastel, L., Rigaud, A-S., Monnet, C., Lenoir, H., & Pino, M. (2017). Affective and engagement issues in the conception and assessment of a robot-assisted psychomotor therapy for persons with dementia. *Frontiers in Psychology, 8*, 960-974.
- Šabanović, S., Bennett, C. C., Chang, W-L., & Huber, L. (2013). PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICOR)*, Seattle, WA, pp. 1-6.
- Sharkey, A., & Sharkey, N. (2010). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology, 14*(1), 27-40.
- Shibata, T. (2012). Therapeutic seal robot as biofeedback medical device: Qualitative and quantitative evaluations of robot therapy in dementia care. *Proceedings of IEEE*, Orlando, United States, pp. 2527-2538.
- Shibata, T., Mitsui, T., Wada, K., Touda, A., Kumasaka, T., Tagami, K., & Tanie, K. (2001). Mental commit robot and its application to therapy of children. *Proceedings of IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, Como, Italy, pp. 1053-1058.
- Silva, A. R., Pinho, M. S., Macedo, L., Moulin, C., Caldiera, S., & Firmino, H. (2017). It is not only memory: Effects of SenseCam on improving well-being in patients with mild alzheimer disease. *International Psychogeriatrics, 29*(5), 741-754.

- Sixsmith, A., Orpwood, R., & Torrington, J. (2007). Quality of life technologies for PwD. *Topics in Geriatric Rehabilitation, 23*(1), 85-93.
- Soler, M. V., Agüera-Ortiz, L., Rodriguez, J. O., Rebolledo, C. M., Muñoz, A. P., Pérez, I. R.,... Martin, P. M. (2015). Social robots in advanced dementia. *Frontiers in Aging Neuroscience, 7*(133), 1-12.
- Stewart, R., Hotopf, M., Dewey, M., Ballard, C., Bisla, J., Calem, M., Fahmy, V., Hockley, J., Kinley, J., Pearce, H., Saraf, A., & Begum, A. (2014). Current prevalence of dementia, depression, and behavioural problems in older adult care home sector: The South East London Care Home Survey. *Age and Ageing, 43*, 562-567.
- Takayanagi, K., Kirita, T., & Shibata, T. (2014). Comparison of verbal and emotional responses of elderly people with mild/moderate dementia and those with severe dementia in responses to seal robot, PARO. *Frontiers in Aging Neuroscience, 6*(257), 1-5.
- Tamura, T., Yonemitsu, S., Itoh, A., Oikawa, D., Kawakami, K., Higashi, Y.,... Nakajima, K. (2004). Is an entertainment robot useful in the care of elderly people with severe dementia? *The Journals of Gerontology: Series A, 59*(1), M83-M85.
- Wang, R., Sudhama, A., Begum, M., Huq, R., & Mihailidis, A. (2017). Robots to assist daily activities: views of older adults with Alzheimer's disease and their carers. *International Psychogeriatrics, 29*(1), 67-79.
- Whelan, S., Murphy, K., Barrett, E., Krusche, C., Santorelli, A., & Casey, D. (2018). Factors affecting the acceptability of social robots by older adults including PwD or cognitive impairment: a literature review. *International Journal of Social Robotics, 10*, 643-668.

Woodberry, E., Browne, G., Hodges, S., Watson, P., Kapur, N., & Woodberry, K.

(2015). The use of a wearable camera improves autobiographical memory in patients with Alzheimer's disease. *Memory*, 23(3), 340-349.

World Health Organisation (2017). *Dementia Fact Sheet*. World Health Organisation, Geneva.

10. Appendices

Appendix A: PRISMA Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	101
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	110
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	112
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	118
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	118
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	119
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	119
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	120

Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	120
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	121
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	169
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	122
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	121
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	N/A

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	148
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	123
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	124
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	126
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	136

Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	172
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	173
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	146
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	153
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	154
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. *Note: This review received no external funding. Funding was only via the salaries of the NHS staff involved. However, the NHS had no role in this review.	N/A

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

For more information, visit: www.prisma-statement.org.

Appendix B: Data Extraction

- Study title
- Author names
- Publication date
- Study design
- Type of robotic intervention used
- Type of intervention / experiment
- Outcome measures used for evaluation
- Who is reporting on outcome measures (i.e. PwD, carer / family member, staff member or researcher)
- Quantitative or qualitative analysis
- Type of analysis employed
- Sample size
- Sample characteristics (i.e. did participants have a diagnosis of dementia)
- Main results of the study

Appendix C: QualSyst Measure***Quantitative studies***

Criteria		YES (2)	PARTIAL (1)	NO (0)	N/A
1	Question / objective sufficiently described?				
2	Study design evident and appropriate?				
3	Method of subject/comparison group selection or source of information/input variables described and appropriate?				
4	Subject (and comparison group, if applicable) characteristics sufficiently described?				
5	If interventional and random allocation was possible, was it described?				
6	If interventional and blinding of investigators was possible, was it reported?				
7	If interventional and blinding of subjects was possible, was it reported?				
8	Outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias? Means of assessment reported?				
9	Sample size appropriate?				
10	Analytic methods described/justified and appropriate?				
11	Some estimate of variance is reported for the main results?				
12	Controlled for confounding?				
13	Results reported in sufficient detail?				
14	Conclusions supported by the results?				

Qualitative studies

Criteria		YES (2)	PARTIAL (1)	NO (0)
1	Question / objective sufficiently described?			
2	Study design evident and appropriate?			
3	Context for the study clear?			
4	Connection to a theoretical framework / wider body of knowledge?			
5	Sampling strategy described, relevant and justified?			
6	Data collection methods clearly described and systematic?			
7	Data analysis clearly described and systematic?			
8	Use of verification procedure(s) to establish credibility?			
9	Conclusions supported by the results?			
10	Reflexivity of the account?			

Appendix D: Distribution of Scores on the QualSyst Measure and DeBs

<i>M</i>	<i>SD</i>	Range	<i>N</i> Strong (>80%)	<i>N</i> Good (70- 80%)	<i>N</i> Adequate (50-70%)	<i>N</i> Limited (<50%)
QualSyst Measure						
15.1	5.3	0 – 2	6	8	17	6
DeBs						
15.4	6.5	1 – 5	0	1	3	33

Appendix E: Item-by-item ratings on QualSyst and DeBs

QualSyst (Quantitative Studies)

Paper	Aims	Design	Selection	Subjects	Random allocation	Investigator blinding	Subject blinding	Outcomes	Sample size	Analysis	Variance	Confounding	Results	Conclusions	Total (%)
Begum et al. (2013)	2	1	1	1	N/A	N/A	N/A	1	1	N/A	N/A	N/A	0	1	50%
Begum et al. (2015)	2	2	1	2	N/A	N/A	N/A	2	0	0	0	N/A	2	1	60%
Bemelmans et al. (2015)	2	2	1	1	N/A	N/A	N/A	1	0	2	0	0	1	2	55%
Chu et al. (2016)	2	2	2	1	N/A	N/A	N/A	2	1	2	2	2	2	2	91%
Demange et al. (2018)	2	2	1	2	N/A	0	0	2	1	2	2	1	2	2	73%
Gustafsson et al. (2015)	1	1	N/A	1	N/A	0	N/A	2	N/A	0	2	N/A	1	1	50%
Jøranson et al. (2015)	2	2	2	2	1	0	0	2	2	2	2	2	2	2	82%

Jøranson et al. (2016a)	2	2	1	2	1	0	0	2	1	2	2	2	2	2	75%
Jøranson et al. (2016b)	2	1	1	1	N/A	N/A	N/A	1	1	2	1	0	1	2	59%
Khosla et al. (2017)	2	2	0	1	N/A	N/A	N/A	2	1	2	2	0	2	2	73%
Koh & Kang (2018)	2	1	2	1	2	0	0	2	2	2	2	2	2	1	75%
Kuwamura et al. (2016)	1	1	0	1	0	N/A	N/A	2	N/A	2	0	0	1	0	35%
Lancioni et al. (2015)	1	0	1	1	N/A	N/A	N/A	1	0	0	0	0	0	1	23%
Lancioni et al. (2016)	2	1	1	1	0	0	0	1	0	1	2	0	2	2	46%
Lancioni et al. (2017)	2	1	2	1	0	0	0	1	1	2	0	0	2	2	50%
Lancioni et al. (2018)	2	1	2	1	0	0	0	1	1	1	1	0	2	2	50%
Lanza et al. (2014)	2	2	1	1	N/A	N/A	N/A	2	1	2	0	0	2	1	64%

Liang et al. (2017)	2	2	1	1	2	0	0	2	0	1	1	1	2	1	57%
Moyle et al. (2013)	2	2	1	1	2	0	0	2	1	2	2	2	2	2	75%
Moyle et al (2017b)	2	2	1	2	2	1	2	2	2	2	2	2	2	2	93%
Moyle et al. (2018)	2	2	2	2	2	2	0	2	2	2	2	2	2	2	93%
Navarro et al. (2016)	2	2	1	1	0	0	0	1	0	1	2	0	1	2	46%
Nijhof et al. (2013)	2	2	2	1	0	0	0	2	0	2	2	0	2	2	61%
Petersen et al. (2017)	2	2	1	2	2	0	0	2	1	1	1	2	0	2	64%
Rouaix et al. (2017)	2	1	0	2	N/A	N/A	N/A	1	1	1	0	N/A	2	2	60%
Šabanović et al. (2013)	2	0	1	0	N/A	N/A	N/A	1	1	2	0	0	2	2	50%
Silva et al. (2017)	2	2	2	2	2	0	0	2	1	2	2	2	2	2	82%

Soler et al. (2015)	2	2	2	2	2	2	0	2	0	2	0	1	2	2	75%
Takayanagi et al. (2014)	1	1	1	1	0	0	0	1	2	1	2	0	1	2	46%
Woodberry et al. (2015)	2	2	1	1	0	0	0	2	1	1	0	0	2	2	50%

QualSyst (Qualitative Studies)

Paper	Aims	Design	Context	Theory link	Sampling	Data collection	Analysis	Verification	Conclusions	Reflexivity	Total (%)
Arntzen et al. (2016)	2	2	2	1	1	1	2	2	1	1	75%
Bemelmans et al. (2016)	1	0	1	1	1	1	0	0	0	0	25%
Lindqvist et al. (2013)	2	1	2	1	1	2	2	0	2	1	70%
Moyle et al. (2014)	2	2	2	2	1	2	2	2	2	0	85%
Moyle et al. (2016)	2	1	1	2	1	2	1	0	2	0	60%
Moyle et al. (2017a)	2	2	1	2	1	2	1	0	1	0	60%
Wang et al. (2017)	2	2	2	2	1	1	1	0	1	0	60%

DeBs

Paper	Individualised	Identity	Voice	Decision-making	Choice	Activities	Community	Meaningful	Stigma	Roles	Total (%)
Arntzen et al. (2016)	5	2	5	4	5	2	1	2	1	2	58%
Begum et al. (2013)	1	1	5	1	1	1	1	1	1	1	28%
Begum et al. (2015)	2	1	3	3	4	3	1	2	1	1	42%
Bemelmans et al. (2015)	3	1	1	1	1	3	1	3	1	1	32%
Bemelmans et al. (2016)	3	1	1	1	2	3	1	1	1	2	32%
Chu et al. (2016)	1	1	2	1	1	1	2	1	1	1	24%
Demange et al. (2018)	2	1	3	2	2	1	1	2	2	1	34%
Gustafsson et al. (2015)	1	1	1	1	1	1	1	1	1	1	20%

Jøranson et al. (2015)	1	1	1	1	1	1	1	2	1	1	22%
Jøranson et al. (2016a)	1	1	1	1	1	2	3	1	1	1	26%
Jøranson et al. (2016b)	1	1	1	1	1	1	1	2	1	2	24%
Khosla et al. (2017)	1	1	5	1	3	3	1	2	1	1	38%
Koh & Kang (2018)	2	1	1	1	1	1	1	1	1	1	22%
Kuwamura et al. (2016)	1	1	1	1	2	1	1	1	1	1	22%
Lancioni et al. (2015)	1	1	1	1	3	2	1	2	1	1	28%
Lancioni et al. (2016)	1	1	1	1	1	2	1	2	1	1	24%
Lancioni et al. (2017)	1	1	1	1	1	2	1	2	1	2	26%
Lancioni et al. (2018)	1	1	1	1	1	2	1	1	1	1	22%

Lanza et al. (2014)	1	1	1	1	1	2	1	2	1	2	26%
Liang et al. (2017)	1	2	1	1	1	2	1	2	1	2	28%
Lindqvist et al. (2013)	5	2	4	4	4	5	1	3	1	2	62%
Moyle et al. (2013)	1	1	2	1	1	1	1	1	1	1	22%
Moyle et al. (2014)	1	1	1	1	1	1	1	2	1	1	22%
Moyle et al. (2016)	1	1	5	1	1	1	1	2	1	2	32%
Moyle et al. (2017a)	1	2	1	1	1	2	1	2	1	2	28%
Moyle et al. (2017b)	2	2	1	1	1	2	1	2	1	2	30%
Moyle et al. (2018)	1	1	1	1	1	1	1	2	1	1	22%
Navarro et al. (2016)	5	3	1	1	3	5	1	2	1	2	48%

Nijhof et al. (2013)	1	1	1	1	1	2	1	2	1	1	24%
Petersen et al. (2017)	1	1	1	1	1	1	1	1	1	1	20%
Rouaix et al. (2017)	5	5	3	1	1	5	1	3	1	3	56%
Šabanović et al. (2013)	1	1	1	1	1	1	1	1	1	1	20%
Silva et al. (2017)	1	2	1	1	1	2	1	2	1	1	26%
Soler et al. (2015)	1	1	1	1	1	1	1	2	1	1	22%
Takayanagi et al. (2014)	1	1	1	1	1	1	2	1	1	1	22%
Wang et al. (2017)	2	2	5	5	5	5	1	4	5	3	74%
Woodberry et al. (2015)	1	2	3	2	1	2	1	2	1	2	34%

VOLUME II:
Service Evaluation Project

Aparna Kapoor

Institute of Psychiatry, Psychology and Neuroscience

King's College London

Thesis submitted in partial fulfilment of the degree of

Doctorate in Clinical Psychology

May 2019

Service Evaluation Project:

A peer-led evaluation of Lewisham IPTT

peer support groups

Aparna Kapoor

Institute of Psychiatry, Psychology and Neuroscience

King's College London

Supervised by:

Mrs Rukeya Khanam

Senior Counselling Psychologist

Lewisham IPTT

South London and Maudsley NHS Foundation Trust

Table of Contents

List of Figures	5
1. Abstract	7
1.1 Background	7
1.2 Method.....	7
1.3 Results	7
1.4 Conclusions	7
2. Introduction	8
2.1 Structure of the service	9
2.2 Leadership	11
2.3 Aims of the study	12
3. Method	12
3.1 Ethical approval.....	12
3.2 Design and procedure	12
3.3 Participants	14
3.4 Thematic Analysis	14
3.5 Demographic data.....	15
4. Results	15
4.1 Focus group	15
4.2 Evaluation questionnaire (by questions)	20
4.2.1 My mental health and well-being.....	21
4.2.2 My daily skills.....	28

4.3	Themes from qualitative data (by questions)	35
4.4	Demographic data.....	37
4.4.1	Ethnicity data	37
4.4.2	Gender data	38
4.4.3	Age data	39
5.	Discussion	39
5.1	Summary of results.....	39
5.2	Possible areas for improvement of Peer Support groups.....	42
6.	Limitations of the evaluation and future recommendations.....	43
7.	Conclusions	45
8.	References	46
9.	Appendices	49
	Appendix A: Peer Support group evaluation questionnaire.....	49
	Appendix B: Focus group questions	53

List of Figures

Figure 3.1	Structure of the Lewisham Peer Support Service.....	10
Figure 3.2	Service-user responses to improved self-esteem.....	21
Figure 3.3	Service-user responses to improved mood.....	21
Figure 3.4	Service-user responses to forming relationships.....	22
Figure 3.5	Service-user responses to sharing feelings.....	23
Figure 3.6	Service-user responses to feeling accepted.....	23
Figure 3.7	Service-user responses to feeling understood.....	24
Figure 3.8	Service-user responses to feeling isolated.....	25
Figure 3.9	Service-user responses to feeling comfortable.....	25
Figure 3.10	Service-user responses to developing ways of coping.....	26
Figure 3.11	Service-user responses to developing self-confidence.....	27
Figure 3.12	Overall service-user responses to questions on “mental health and well-being”.....	27
Figure 3.13	Service-user responses to developing work skills.....	28
Figure 3.14	Service-user responses to developing social skills.....	29
Figure 3.15	Service-user responses to carrying out activities.....	29
Figure 3.16	Service-user responses to developing hobbies.....	30
Figure 3.17	Service-user responses to getting out more.....	31
Figure 3.18	Service-user responses to trying new things.....	31
Figure 3.19	Service-user responses to socialising more.....	32
Figure 3.20	Service-user responses to accessing support.....	33
Figure 3.21	Service-user responses to taking care of needs.....	34
Figure 3.22	Overall service-user responses to questions on “daily skills”.....	34

Figure 3.23	Comparison of ethnicity data across Lewisham IPTT and Peer Support groups.....	37
Figure 3.24	Comparison of gender data across Lewisham IPTT and Peer Support groups.....	38
Figure 3.25	Comparison of age groups across Lewisham IPTT and Peer Support groups.....	39

1. Abstract

1.1 Background

Peer support services have been found to be beneficial, both to the service-users receiving support and to the peers offering it. This study aimed to evaluate the peer support groups offered within the Lewisham Integrated Psychological Therapies Team (IPTT). Accessibility to the groups was also explored.

1.2 Method

A focus group was conducted with eight service-users and peer support volunteers to pilot the evaluation survey. Following this, 41 evaluation surveys were completed between March to June 2-17, by service-users and peer support volunteers attending and/or facilitating the peer support groups.

1.3 Results

A high level of satisfaction was noted, particularly in relation to service-users finding the groups beneficial in relation to their personal well-being and mental health. Some suggestions for improvement were made, particularly around issues relating to practicalities of the group (i.e. timings and venues).

1.4 Conclusions

Benefits of the peer support groups, as reported by service-users, have been discussed. Suggestions for improvement of the groups will be considered in making recommendations for future development of the Lewisham Peer Support Service. Limitations of the current study have also been discussed, with recommendations made for future research in this area.

2. **Introduction**

The Together Charity and the National Survivor User Network (NSUN; 2014) define service- user involvement as “service-users involving themselves, as individuals or groups, to develop, deliver and evaluate mental health policy and practice”. Service-users may be involved in designing and developing new services, recruitment and selection of staff, delivering training, service delivery, or evaluating services (NHS England, 2015). Recently, the importance of service-user involvement in mental health services has become widely recognised. Service-user’s personal experiences of mental health services are profoundly valuable in supporting services to improve practice (NHS England, 2015). Further, service- users report indirect benefits of getting involved in service development including improvement in self-esteem and communication skills, and a reduction in social isolation (Omeni, Barnes, MacDonald, Crawford & Rose, 2014; Davidson, Bellamy, Guy & Miller, 2012). It is important to note, however, that for service-user involvement to be useful, it must be an active, on-going process, forming an integral part of every mental health provider (Together & NSUN, 2014).

Peer support is one of many ways in which service-users can be involved in mental health services. Peer support may be defined as "offering and receiving help, based on shared understanding, respect and mutual empowerment between people in similar situations" (Mead, Hilton & Curtis, 2001), and is often done through the delivery of various services. Peer support services have been found to be beneficial to service-users receiving support, as well to those offering support (Solomon, 2004), across several psychosocial areas, including improvement in mental health symptoms and quality of life (Davidson, Chinman, Kloos, Weingarten, Stayner & Tebes, 1999), increase in

empowerment and ability to cope with distress (Solomon, 2004), and reduced hospitalisation (Sherman & Porter, 1991).

The British Psychological Society (BPS; 2010) good practice guidelines around service-user and carer involvement outlines the main principles of carrying this out, including the importance of diversity (involving people from minority ethnic backgrounds, sexual minority groups and people with learning disabilities) and evaluation of services facilitating involvement. The current study draws upon these guidelines to identify a gap in the delivery of the Lewisham Integrated Psychological Therapies Team (IPTT) Peer Support Service – namely ‘evaluation’. BPS guidelines (2010) identify evaluating services as an essential part of ensuring the service is achieving its aims, fostering inclusivity and is constantly improving, through dissemination and sharing of results. Further, involvement of psychologists and trainee clinical psychologists in this process has been clearly outlined.

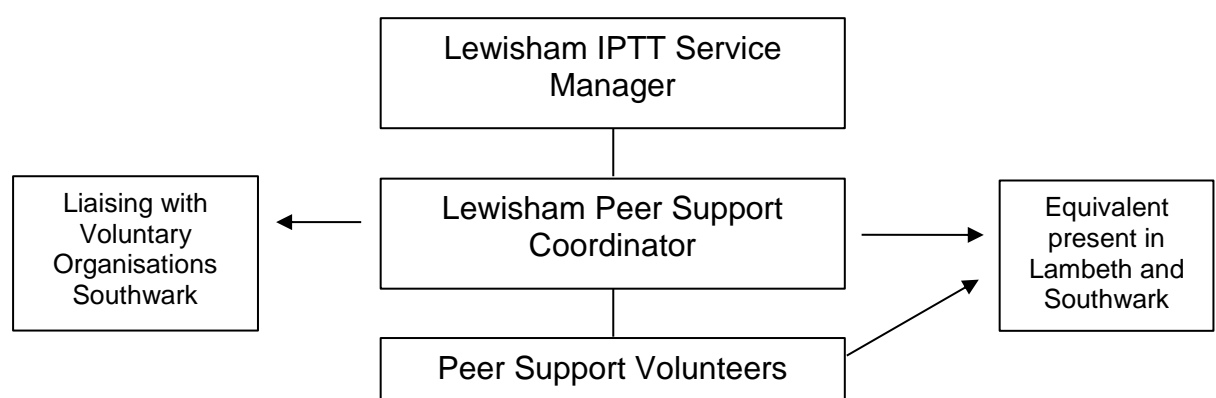
2.1 Structure of the service

The Integrated Psychological Therapies Team (IPTT) offers specialist psychological services across the boroughs of Lambeth, Southwark and Lewisham, for adults between the ages of 18 to 65, with a severe mental illness. A range of psychological therapies are offered, including Cognitive Behavioural Therapy (CBT), Family and couple therapy, perinatal psychotherapy, psychodynamic psychotherapy, Cognitive Analytic Therapy (CAT), Trauma-focused therapy and Mentalisation-Based Therapy (MBT), either individually, or within groups. Service-users often have complex needs, which can lead to them being out of work and socially isolated. To provide a holistic intervention, the Peer Support Service was set up in 2014, based on recommendations made by the BPS

(2010). This service operates across the three boroughs, offering service-users the opportunity to engage in group activities, facilitated by trained ex- service-users.

The Peer Support Service was set up through consultation with service-users and staff groups, keeping in mind the core principles, as identified by Repper, Aldridge, Gilfoyle, Gillard, Perkins and Rennison (2013). These include ensuring that the service aimed at being mutual, reciprocal, non-directive, strengths-based, inclusive, progressive, safe, and promoting recovery. Lewisham IPTT's Peer Support Service is led by a Peer Support Development Coordinator, who is an ex-service-user, supervised by a Senior Lewisham IPTT Therapist. Several current or ex-service-users, at different stages of their recovery, volunteer with the Peer Support Service, co-facilitating all groups with the Peer Support Coordinator. Each volunteer undergoes training to learn how to work with and support other service-users. The service works in conjunction with voluntary organisations, including the Share Project and Mind, amongst others.

Figure 3.1. Structure of the Lewisham Peer Support Service



The Peer Support Service offers a range of informal arts and skills-based groups, with open- membership and often unstructured to allow for different skill levels and paces of working. Groups offered at Lewisham include art, creative writing, gardening, social

events (including sports groups, community visits to museums or events and walking groups). A Women's Group is being set up, and suggestions for other groups are being considered. The groups are formed in a way that promotes safety and acceptance, allowing service-users the opportunity to interact with peers and professionals on an equal level.

2.2 Leadership

Drawing upon BPS guidelines (2010), the researcher identified a gap in the peer support services being offered at Lewisham IPTT. Some of the key areas outlined in the guidelines, including inclusivity and evaluation, had not been addressed by the service. Through consultation with the research supervisor and Peer Support Coordinator, a service need was identified, whereby the idea for the current project was developed – the peer support groups had never been formally evaluated, and it was also unclear whether the service was accessible to all service-users attending Lewisham IPTT, particularly very young or very old clients, clients from Black and Minority Ethnic (BAME) backgrounds and all genders. Further, the involvement of peer support volunteers in the data collection process was included following the findings of Clark, Scott, Boydell and Goering (1999), which indicated that service-users were more likely to provide honest feedback, when interviewed by a peer, and as per the NHS England (2015) recommendations on involving service-users in the evaluation of services.

2.3 Aims of the study

The aims of the evaluation were:

1. To develop and pilot an evaluation questionnaire by gaining qualitative feedback from service-users and peer support volunteers, via a semi-structured focus group
2. To gain quantitative feedback on service-users' experience of the peer support groups, via the use of an evaluation questionnaire
3. To gain qualitative feedback from service-users on their experience of the peer support groups and any suggestions for improvement of the groups, via the use of an evaluation questionnaire
4. To explore whether BAME groups, all sexes and all age groups were appropriately represented within the peer support group caseload, in comparison to the wider Lewisham IPTT caseload.

3. Method

3.1 Ethical approval

A proposal for the project was drafted in collaboration with the Peer Support Coordinator and research supervisor, and submitted to the Psychological Medicine and Integrated Care Clinical Academic Group (PMIC CAG) Clinical Audit and Effectiveness Team for consideration. Approval was gained on 15th February 2017.

3.2 Design and procedure

A draft evaluation questionnaire was developed by the researcher through consultation with the Peer Support Coordinator and a Peer Support Volunteer. This consisted of

demographic questions (i.e. age, ethnicity, gender), 20 questions to be rated along a five-point Likert scale and six open-ended questions giving participants the opportunity to provide qualitative information. The 20 qualitative questions included 10 questions under the theme of “My Daily Skills” and 10 under the theme of “My Mental Health and Well-Being”. Following this, a focus group was set up, inviting all service-users attending the peer support groups. The focus group lasted two hours, with two breaks, and was facilitated by the Peer Support Coordinator and two Peer Support Volunteers. Tea and biscuits were provided to create an informal atmosphere. Facilitators employed a semi-structured approach, offering prompts in the form of questions, whilst encouraging participants to contribute openly to the discussion (Hayes, 1997). The researcher attended the focus group as an observer and scribe. The purpose of the focus group was to discuss and adapt the evaluation questionnaire, and to pilot it as well.

Suggestions made within the focus group were incorporated into the final version of the evaluation questionnaire. This was then circulated among the Peer Support Coordinator and Peer Support Volunteers for final approval. The final version of the evaluation questionnaire consisted of demographic questions (i.e. age, ethnicity, and gender), 19 questions rated along a five-point Likert scale and six open-ended questions. One of the questions under the “My Daily Skills” sub-section was not deemed suitable by the focus group and was removed, leaving nine questions under this sub-section and 10 under the “My Mental Health and Well-Being” sub-section. The language used for the rating scale was amended as per suggestions made by the focus group, to make it more accessible. Data from evaluation questionnaires were collected over three months, giving every service-user attending the peer support groups the opportunity to complete one. The Peer Support team developed a checklist system to monitor who had

completed the questionnaire. Evaluation questionnaires were given to participants at the end of each peer support group, by the Peer Support Coordinator and Peer Support Volunteers. These were completed anonymously and returned to the Peer Support Coordinator or Peer Support Volunteers, who passed these on to the researcher. All clients offered the evaluation questionnaire agreed to complete it.

3.3 Participants

For the focus group, peer support volunteers developed a flyer, which stated the purpose, date and time the group was to be held. Flyers were handed out to service-users at each peer support group held in Lewisham. The peer support coordinator and volunteers also verbally informed service-users of the purpose of the focus group, asking for volunteers.

All service-users attending at least one of the Lewisham peer support groups were asked to complete the evaluation questionnaire. They were asked to report on their overall experience of the peer support groups, regardless of the one they were attending at the time.

3.4 Thematic Analysis

Thematic analysis, as outlined by Braun and Clarke (2006), was conducted to analyse the qualitative data gathered from the focus group and the qualitative items on the evaluation questionnaire. An inductive approach was employed, as per the guidance.

Qualitative data gathered from the focus group was in writing. This was recorded verbatim where possible and collated in an Excel spread sheet. Qualitative data gathered

from each evaluation questionnaire was collated within an Excel spread sheet. Written codes were used to extract ideas and concepts from each piece of data provided, which allowed for broader themes to be identified. These were further refined through consultation with the research supervisor and main themes of interest were agreed and reported on.

3.5 Demographic data

Demographic data was collected as part of the evaluation questionnaire. This included information on age, gender and ethnic background. Research has suggested that people from BAME backgrounds and older age clients may not access peer support groups, due to a lack of information or knowledge (Bignall, Butt & Pagarani, 2002), stigma (Pathways RTC, 2013), anxiety, or mistrust (World Health Organisation, WHO; 2017). To explore whether the peer support groups were accessible to clients of all ages, genders and ethnicities, demographic data gathered from the evaluation questionnaires were compared against the entire Lewisham IPTT caseload, as on August 2017.

4. Results

4.1 Focus group

One focus group was held on 14th March 2017, led by two Peer Support Volunteers and the Peer Support Coordinator. Eight service-users and peer support volunteers attended the focus group to talk about their experiences of the peer support groups – things they found helpful, things that were less helpful and suggestions for improvement of groups. The draft evaluation questionnaire was piloted, and suggestions were made to improve

this. Qualitative questions from the survey were used to guide the focus group. A total of 12 themes were identified:

Helpful

Structure of the groups

Service-users found the structure of the groups to be particularly helpful. They commented on the "informality" and liked that there was "no agenda" and "no expectations". For those who preferred a more structured approach, group facilitators provided a "structure and directive". Focus group participants agreed that the variety of groups offered meant that "different groups meet different goals and suit different needs" and the freedom to "choose" was valued. The timings and location of the groups allowed for "good transport links and taxi access". Focus group participants also commented on the peer support groups being "more about the process rather than the skills", identifying this as a strength.

Atmosphere of the groups

A large part of the focus group discussion centred on the atmosphere within the peer support groups. Participants consistently described the atmosphere as "relaxed", and "an escape", "liberating" and "therapeutic". Service-users valued the "collaborative" and "non- judgemental" nature of the groups, which helped to facilitate a "supportive" atmosphere, within which "everybody feels included". A major theme identified was around "acceptance" and "empathy", with the groups fostering an environment where people "don't get blamed" and can openly "share feelings".

Socialising in the groups

Focus group participants particularly valued the "social aspect" of the groups, and the opportunity to "meet peers who understand". The groups were found to be helpful in "reducing isolation" and "encouraging to get out of the house more". The "positive conversations" and "sharing ideas on what helps" (with regards to overcoming difficulties) were valued. Being able to talk about mental health difficulties openly was "normalising", through finding "things in common with peers". As a result, several focus group participants felt that they had made "new friends" and found "companionship" through the groups.

Meeting goals

Participants felt that the groups encouraged them to "do things differently" and meet their goals. Some felt the groups helped them to "meet practical goals, like timekeeping" and "exercise", and "learning new skills for the future or improve on existing skills" including "learning English", "learning to take care of plants" and "writing songs". Others found them helpful to "gain confidence", "learning to say 'no'" and to "be more helpful towards others". The groups offered participants the opportunity to "explore new activities and challenges" and "do something we wouldn't usually do", which "helped recovery". Achieving these goals made participants feel "proud".

Overcoming stigma

Another major theme identified was around overcoming the stigma of mental health difficulties, which overlaps with previous themes. One participant described the peer support groups as having "no stigma, no labelling and no stereotyping". This was

achieved by the group facilitator's "knowledge of mental health problems" and their "lived experience", which helped participants to "overcome boundaries". Focus group participants felt that the groups were a place where "mental health is not a taboo". Overall, focus group participants felt that this was the most important aspect of the peer support groups, and the one thing that was most helpful.

Not helpful

Practical aspects of the groups

In general, participants felt that some of the venues for the group were not ideal. For example, the venue where the focus group was held, the Crypt at the local Parish Church in Lewisham, was where some of the peer support groups were held. This venue was found to be "cold" with "heavy doors", which made entry to the venue difficult, particularly for service-users with physical health issues. Further, the toilet on site was "inaccessible" for wheelchair users or those with mobility issues, resulting in people being excluded. One participant felt that the open membership of the groups was unhelpful, as "people come and go so we are not able to keep a consistent group", which they felt impacted on their anxiety in group situations. Others felt that the timings of the group (i.e. before 17:00pm) were also not ideal, as "people are lonely in the evening".

Frequency and length of the groups

Whilst participants enjoyed the groups, it was felt they were too infrequent. All focus group participants agreed that the "fortnightly pattern of groups is not helpful". Further, the length of groups was reported as being "not long enough". Participants felt that the two-hour slot was often not enough "to set up and set down". Further, concerns were

raised regarding “clash of timings” of certain groups, which limited service-user’s accessibility to all groups.

Remembering dates and times of the groups

One of the biggest challenges reported by focus group participants was "remembering which week the group is on". Although a paper timetable for the year was shared with service-users, this was not found to be adequate as reminders were not provided.

Suggestions for improvement of the groups

More variety

Although focus group participants reported enjoying the groups offered (i.e. art, creative writing, gardening, and social group), it was felt that there was room to expand and diversify further. Suggestions were made for a "coffee morning and informal chat", a "reading and discussion group", and for more social events like "local music festivals". Participants also felt that they would benefit from psychoeducation or workshop-based groups to learn about mental health. Suggestions were made for a "well-being group" or a "medical group", which giving service-users the opportunity to learn from "visiting professionals", in an informal way, about living well with a mental health diagnosis. Participants also identified a need for "befriending", to support people to attend groups when initially referred.

Greater frequency and length

Drawing from the reported benefits of the peer support groups, focus group participants felt that "weekly sessions" for "longer than two hours" would be ideal for some of the groups that were currently running (i.e. art, creative writing).

Setting up reminders for participants

Suggestions were made for "calendar-style dates for groups" to be circulated by "setting up reminders through email alerts. Having a group email could ensure all are informed". "Text reminders" were also suggested, "for those who may not have access to email".

Participants also made suggestions to the draft evaluation questionnaire. Minor amendments were suggested regarding language, including the ratings provided for the rating scales. One question relating to acquiring a job was removed from the questionnaire as it was not considered relevant in this context.

4.2 Evaluation questionnaire (by questions)

A total of 41 evaluation questionnaires were returned. This sample was highly representative of the Lewisham IPTT clients regularly attending the Peer Support groups between March and June 2017, owing to the monitoring system employed by the Peer Support Coordinator and Volunteers. There was missing qualitative data in three surveys.

Overall responses to each quantitative evaluation questionnaire item are presented below in the form of pie charts. This was found to be the best way to present the data gathered, as a reflection of the group responses. Data has also been provided on the average ratings for each question, to highlight areas of strength and those that may require further improvement.

4.2.1 My mental health and well-being

Question 1: The group helped me to... Improve my self-esteem/feel better about myself

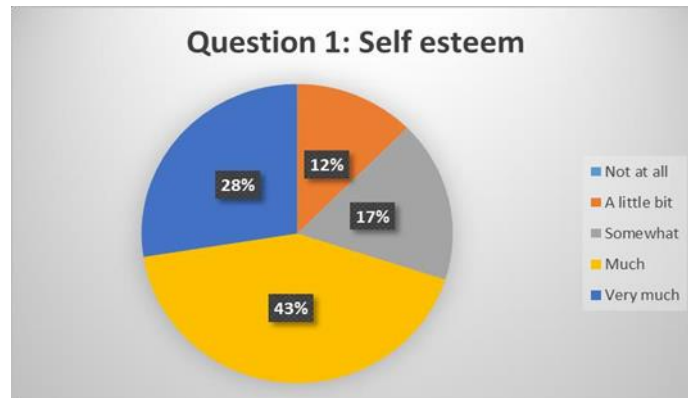


Figure 3.2. Service-user responses to improved self-esteem

Most clients (71%) felt that the groups had helped them feel better about themselves and improve their self-esteem to a large extent, and 100% of respondents reported improvement in this area to some extent. Respondents identified that the groups had met their expectations “to feel happier within myself” and “to lessen anxiety socially”. On average, respondents rated this area of improvement as 4 (i.e. experiencing it “much”).

Question 2: The group helped me to... Improve my mood/emotional well-being

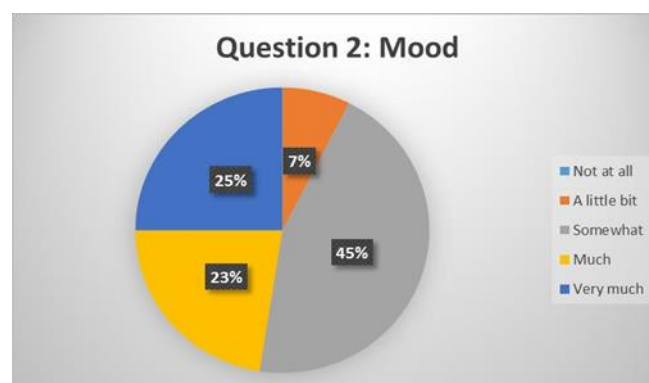


Figure 3.3. Service-user responses to improved mood

48% of respondents felt that the groups have helped them to improve their mood and their emotional well-being, with 100% of respondents noting an improvement in this area to some extent (52% reported “somewhat” and “a little bit”). Respondents reported choosing to attend the peer support groups “for emotional support”, “to understand and overcome emotional problems” and “to develop my ability to support others”.

Question 3: The group helped me to... Make new friends/meet new people

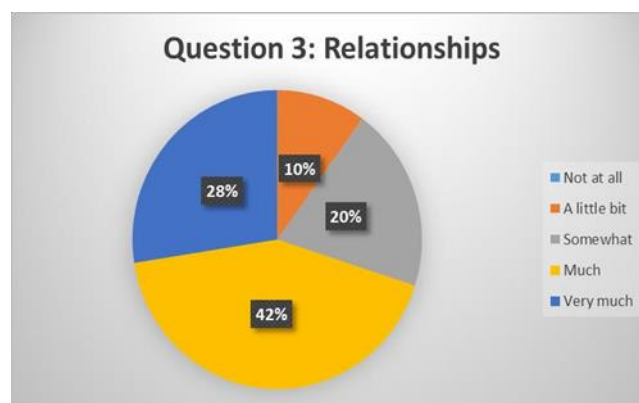


Figure 3.4. Service-user responses to forming relationships

100% of respondents felt the groups had helped them to make new friends and meet new people, with 70% reporting this to a large extent (i.e. “much” or “very much”). An overwhelming number of respondents identified this as one of the main reasons for them choosing to attend the groups, and most respondents also reported this need being met. The average rating for this area was 4 (“much”), indicating that improvement was noted to a large extent.

Question 4: The group helped me to... Share my feelings with others/talk to others

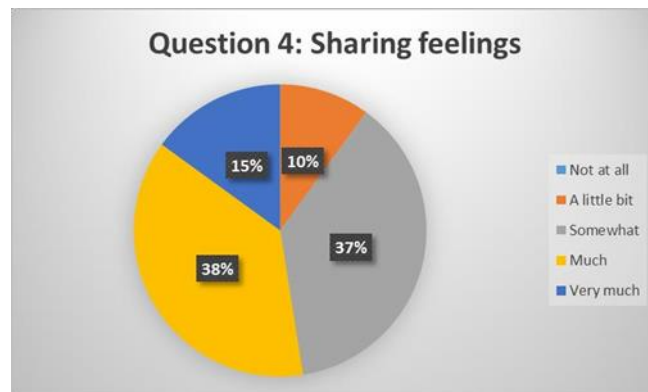


Figure 3.5. Service-user responses to sharing feelings

Overall, all respondents were satisfied with improvement in this area, with 52% reporting feeling more able to share their feelings with others. Qualitative responses included feeling able to “share the story” and “talking with other people”. The average rating for this area was 4, indicating respondents experienced improvement to a large extent.

Question 5: The group helped me to... Feel accepted by others

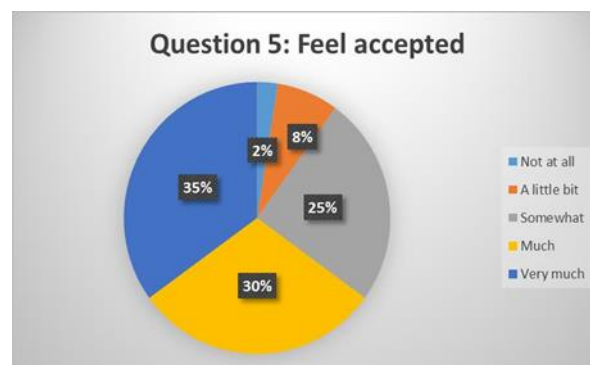


Figure 3.6. Service-user responses to feeling accepted

An overwhelming 65% of respondents reported feeling more accepted by others to a greater extent upon attending the peer support groups. Several respondents reported

feeling “safe” and that the groups helped combat the “stigma of mental ill health”, with a “high level of acceptance” and “sense of belonging”. One respondent reported feeling “not at all” accepted by the group, however, the reasons for this were not reflected in the qualitative data provided. On average, respondents experience improvement in this area, with an average rating of 4.

Question 6: The group helped me to... Feel understood by others

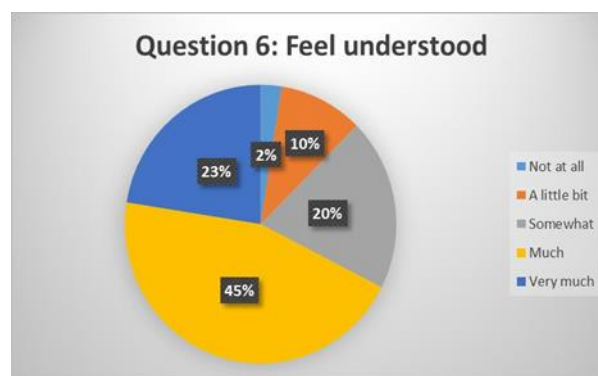


Figure 3.7. Service-user responses to feeling understood

68% of respondents reported feeling more understood by others, following attending the groups and meeting “non-judgemental peers”. 30% of respondents reported improvement in this area to some extent (“somewhat” and “a little”), with one respondent rating no improvement at all. Qualitative responses did not indicate reasons for this. The average rating for improvement in this area was 4 (“much”).

Question 7: The group helped me to... Feel included/less isolated or alone

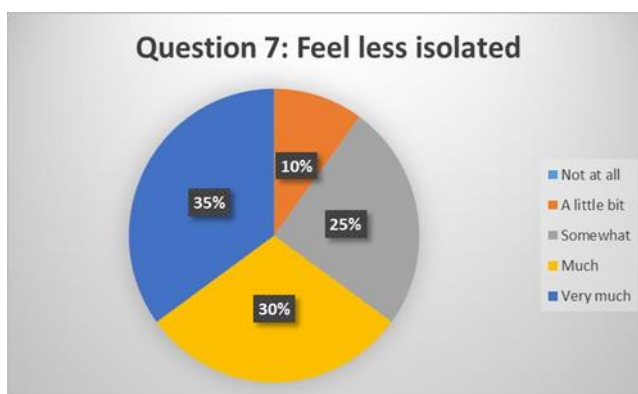


Figure 3.8. Service-user responses to feeling less isolated

100% of respondents reported feeling less isolated and more included, to some extent, with 65% reporting this to a greater extent, and 35% to a lesser extent. The qualitative data reflected this as another important reason respondents chose to attend the peer support groups, with one respondent reporting that they chose to attend the groups “because I found myself isolated since I got unwell, my friends disappeared” and that the groups had helped them to “find new friends”. An average rating of 4 (“much”) was noted, reflecting this area was much improved for those attending the groups.

Question 8: The group helped me to... Feel comfortable

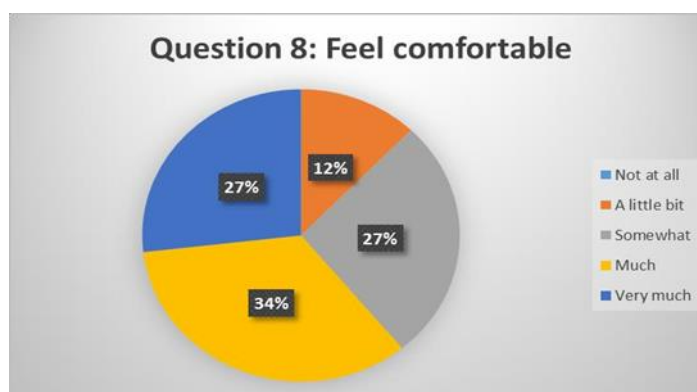


Figure 3.9. Service-user responses to feeling comfortable

All respondents reported feeling comfortable in the groups they attended to some extent, with 61% reporting this to a greater extent. Respondents found it “easy to fit in with everyone”, found the groups had a “relaxed atmosphere” and felt “comfortable around others”. This was also noted to be an important area for respondents, with an average rating of 4 (“much”).

Question 9: The group helped me to... Get support with my difficulties/learn ways of coping

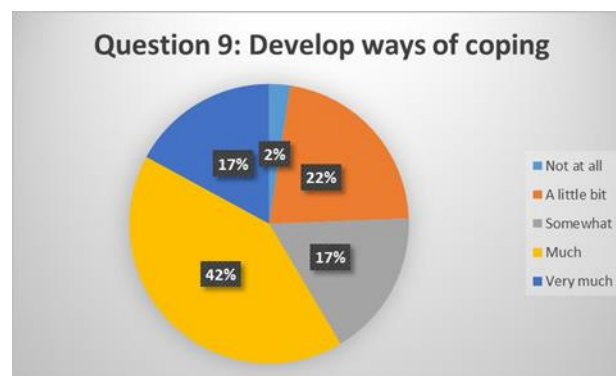


Figure 3.10. Service-user responses to developing ways of coping

A majority (59%) of respondents reported finding the groups helpful in learning new ways of coping and getting support. Respondents reported that the most helpful things about the groups were developing “creative ways to help with my emotional problems” and that they chose to attend “to help my development in my recovery”. One respondent did not report any improvement in this area, rating 1 (“not at all”), however, did not provide reasons for this. On average, respondents rated improvement in this area as 3 (“somewhat”), with suggestions for a “wellbeing group” to build on this.

Question 10: The group helped me to... Develop my self-confidence

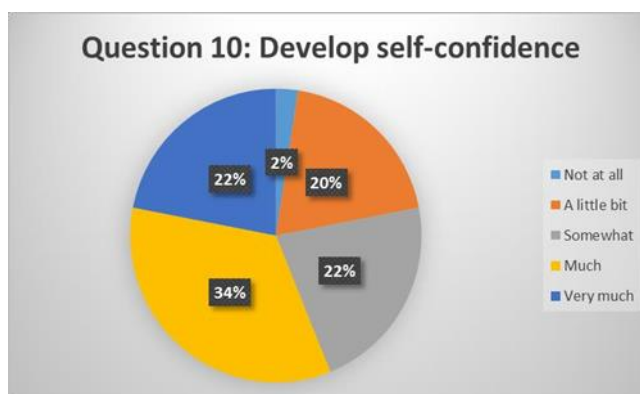


Figure 3.11. Service-user responses to developing self-confidence

56% of respondents reported finding the groups helpful in developing their self-confidence, with only one respondent rating improvement in this area as 1 (“not at all”). Reasons for this low rating were not provided, whilst other respondents felt the groups helped them to “gain confidence”, “integrate” and “talk”. Respondents rated this area as 4 (“much” improved), on average.

Overall ratings for “My Mental Health and Well-Being” questions

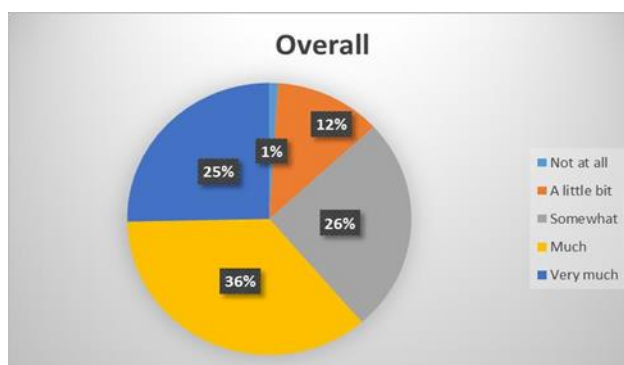


Figure 3.12. Overall service-user responses to questions on “mental health and well-being”

Across all responses within the “My Mental Health and Well-Being” sub-section, 99% indicated the groups were helpful in these areas (61% to a large extent, 38% to a small extent). 1% of responses indicated the groups had not been helpful at all, although qualitative data did not reflect this. The most frequent rating provided was 4 (“much” improvement), indicating, on average, the needs of respondents in this area had been met by the groups.

4.2.2 My daily skills

Question 1: The group helped me to... Gain new skills required for work/jobs

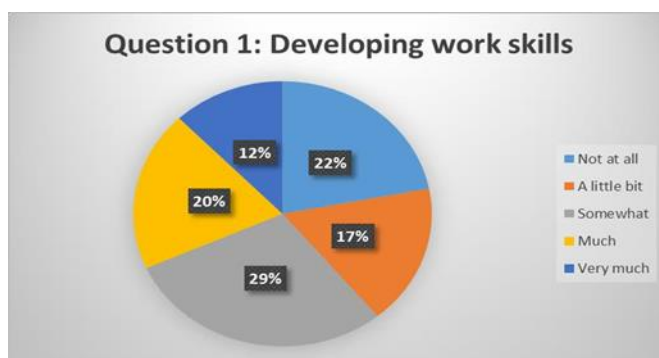


Figure 3.13. Service-user responses to developing work skills

78% of respondents reported the peer support groups helped them develop new skills for work or jobs, at least to some extent. 22% of respondents did not feel the groups helped them, with one respondent reporting that they found it “frustrating that it doesn’t lead to work sooner”. Suggestions for improvement included “short courses on budgeting and money management” and to “have learning segments”.

Question 2: The group helped me to... Develop my interpersonal/social skills

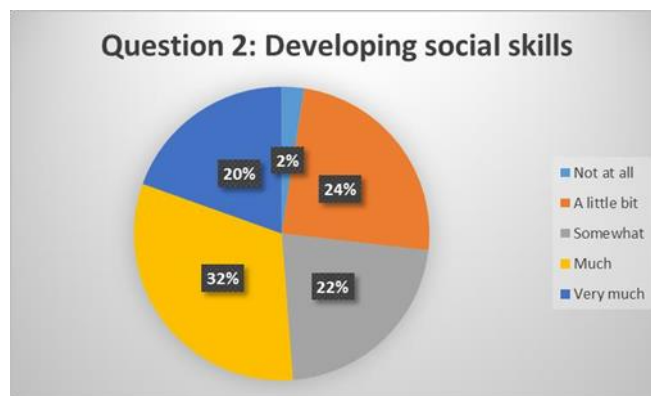


Figure 3.14. Service-user responses to developing social skills

A majority of 52% of respondents felt that the groups had helped them to develop their interpersonal and social skills to a great extent, and 46% reported that this was the case to a lesser extent. “Developing communication skills” and “talking and communicating with others” was one of the main things respondents reported finding helpful about the groups. Only 2% (i.e. one respondent) did not feel the groups helped him develop in this area at all. No qualitative data was recorded to suggest reasons for this. On average, this area was rated a 4 (“much”), indicating overall satisfaction with this area.

Question 3: The group helped me to...Carry out daily activities

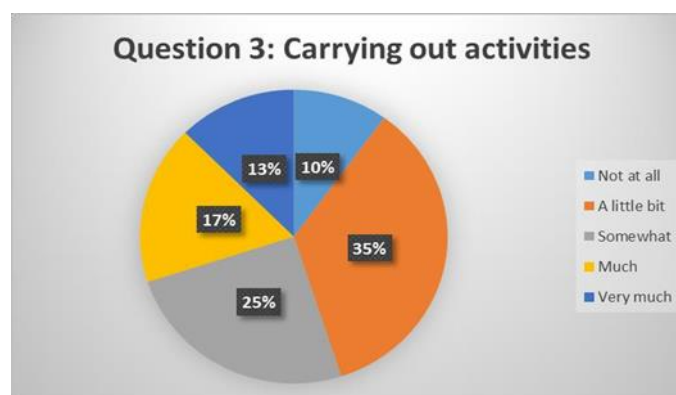


Figure 3.15. Service-user responses to carrying out activities

30% of respondents reported that the groups helped them to carry out activities of daily living, like shopping, cooking and cleaning, to a great extent, whilst a larger number of respondents (60%) felt the groups helped them do this to a small extent. Respondents reported that the groups met their expectations of “getting out of the house” and developing “better mental health and attitude for living”. 10% of respondents did not feel the groups helped them in this area at all, making suggestions like “exercise class (mindfulness for beginners)” as possible areas for improvement. An average rating of 3 (“somewhat”) was given for this area, across all respondents.

Question 4: The group helped me to... Develop my hobbies and interests

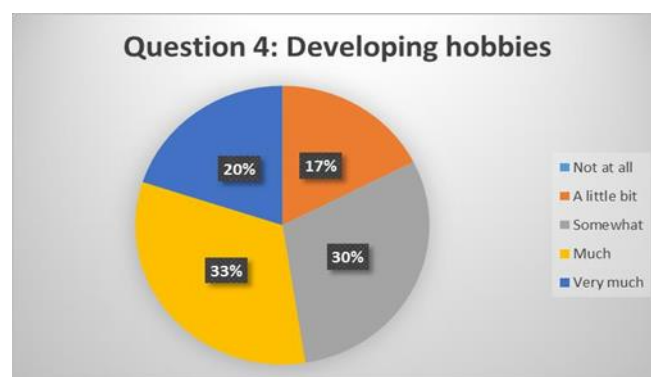
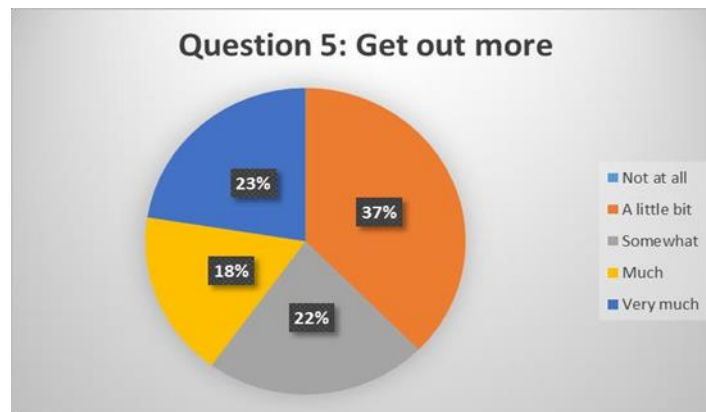


Figure 3.16: Service-user responses to developing hobbies

An overwhelming 100% of respondents felt that the peer support groups helped them to develop their hobbies and interest, at least to some extent. Respondents reported valuing the opportunity to “develop my creative skill”, “to develop writing skills in a safe environment” and receiving “help to be creative”. On average, a rating of 4 (“much”) was given to this area.

Question 5: The group helped me to... Get out more**Figure 3.17.** Service-user responses to getting out more

All respondents reported finding the groups helpful in getting out more, with 41% rating this as “very much” or “much”, and 59% rating this as “somewhat” or “a little”.

Respondents reported that the groups had encouraged them to “come out of home” and that the groups “somewhat” met their expectations in this area. An average rating of 3 (“somewhat”) was recorded for this area.

Question 6: The group helped me to... Try new and different things**Figure 3.18.** Service-user responses to trying new things

95% of respondents felt that the peer support groups helped them to try new things, at least to some extent, with only 5% reporting that they did not feel the groups achieved this at all. Some respondents reported that “I didn’t know what to expect but I’m glad I made the effort, the creative writing is very helpful”, whilst others felt there was room for improvement, making suggestions for different groups they would like to experience, including “yoga, reading group”, “performance group, alternative/complimentary therapies”. Overall, respondents felt the groups achieved this area, with an average rating of 4 (“much”).

Question 7: The group helped me to... Socialise more

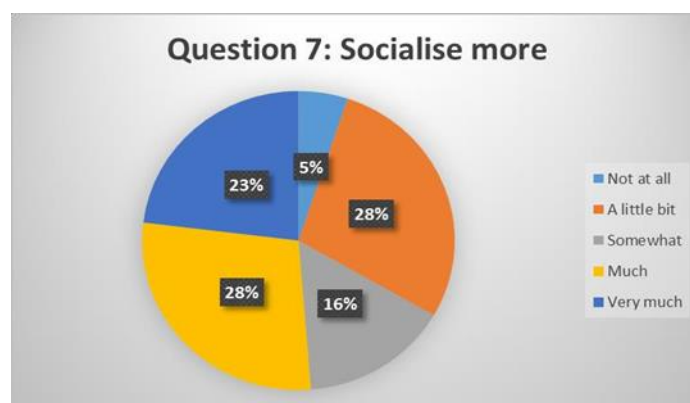


Figure 3.19. Service-user responses to socialising more

Most respondents (95%) felt the groups helped them to socialise more, at least to some extent. The social aspect of the groups was one of the most positively comments on areas across all respondents. A small number of respondents (5%) did not feel the groups helped them to socialise more at all, but no qualitative responses were recorded for reasons behind this. However, an average rating of 3 (“somewhat”) appears to be representative of this.

Question 8: The group helped me to... Ask for help when I needed it/access support

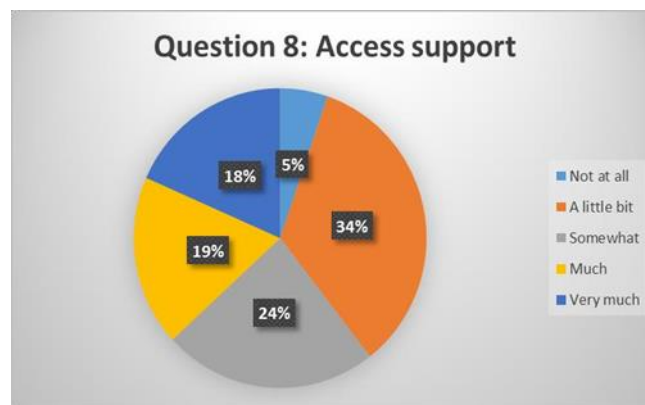


Figure 3.20. Service-user responses to accessing support

Most responses recorded for this area were positive, with 95% reporting that they had found the groups helpful in accessing support and being able to seek help when required. One respondent reported that they chose to attend the groups because “I wanted others to know that you do get better and you do learn to manage it”, whilst another felt the groups helped them “to join some activities which help in my recovery process”. A small number (5%) of respondents did not feel the groups helped them in this area at all. One respondent reported that they had chosen to attend due to being “told to go to work on doing things for my own self-development... until can get therapy”. An average rating of 3 (“somewhat”) was given for this area, across all respondents.

Question 9: The group helped me to... Take care of my needs better

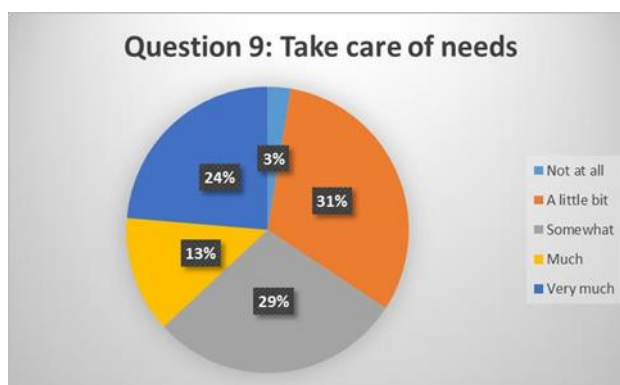


Figure 3.21. Service-user responses to taking care of needs

37% of respondents felt the peer support groups helped them to take care of their own needs to a large extent, whilst 60% of respondents felt the groups helped them to this to a small extent. Respondents reported that the groups met their expectations around gaining “self- respect”, “to feel happier within myself”, and encouraging “self-expression”. Only 3% (two respondents) rated this area as 1 (“not at all”), but reasons for this were not indicated. On average, a rating of 3 (“somewhat”) was given for this area.

Overall ratings for “My Daily Skills” questions

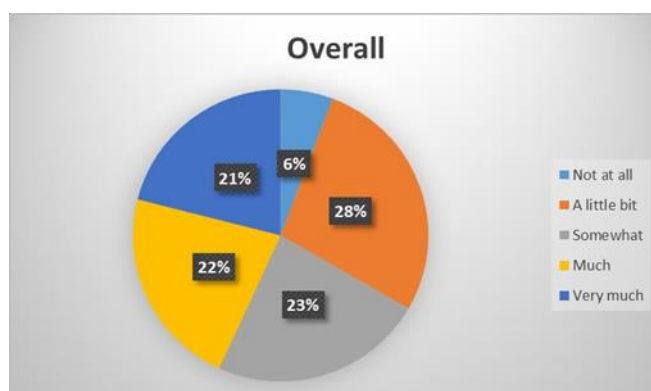


Figure 3.22. Overall service-user responses to questions on “daily skills”

Across all responses for all questions within the “My Daily Skills” sub-section of the evaluation questionnaire, 94% of responses indicated that the groups had been helpful in these areas, at least to some extent (43% to a large extent, 51% to a small extent). Only 6% of responses indicated that the groups had not been helpful at all, which was reflected in the qualitative data which recorded respondent’s suggestions for groups that enhanced their skills for work and suggestions for a greater variety of groups to meet needs in this area. The most frequent rating provided was 2 (“a little” helpful), indicating, on average, the needs of respondents in this area had been met to a small extent by the groups, but that there were areas for improvement.

4.3 Themes from qualitative data (by questions)

Question 1: What did you find most helpful about the group?

An overwhelming number of responses indicated that respondents valued the opportunity to "meet new people" in a "safe environment" as the most helpful thing about the peer support groups. Respondents found the groups "supportive" and "friendly", and helpful in "reducing isolation". Several respondents reported that the presence and "energy" of the Peer Support Coordinator fostered this safe environment.

Question 2: What did you find least helpful about the group?

Most respondents reported finding the "times of groups" unsuitable, providing reasons like "work" and "school hours limitation". Some respondents finding it difficult when participants "might be upset and then aggressive" or "too 'unwell' to participate" fully.

Question 3: Why did you choose to attend peer support groups?

In general, respondents chose to attend the peer support groups to “gain support”,

"reduce social isolation" and "develop my confidence". Others chose to attend to "develop my skills" and "to support others. Reasons to attend across all respondents appeared to complement each other.

Question 4: What did you expect to gain from the group when you first attended?

Some respondents did not know what they expected from the groups, whilst others reported expecting "normality", to "lessen anxiety socially" and to "gain creative skills". Others expected the groups to "help in my recovery process" and gain "better mental health and attitude for living".

Question 5: Did the group meet your expectations?

All respondents reported that they felt that some of the groups met their expectations, at least some of the time. One participant reported preferring the creative writing group to the art group, whilst others reported that the groups had "exceeded their expectations" as they had none to begin with.

Question 6: Do you have any suggestions on how we can improve our groups? Are there any other groups you would like?

Most suggestions made were around "having more sessions and materials" in the groups. "Reliable budget and venues" was also noted as a suggestion. Overall, respondents were satisfied with the groups, and suggestions were made for other groups they would like, including "yoga, reading group, coffee morning", "learning segments" to develop practical skills, "short course on budgeting and money" and "alternative therapies", including "mindfulness and meditation".

4.4 Demographic data

Demographic data was collected for each participant, as part of the evaluation questionnaire, including age, gender and ethnicity. This data was compared to the demographic data for the Lewisham IPTT caseload as of August 2017, to assess accessibility of the peer support groups for service-users from minority groups. Data is presented in bar graphs below.

4.4.1 Ethnicity data

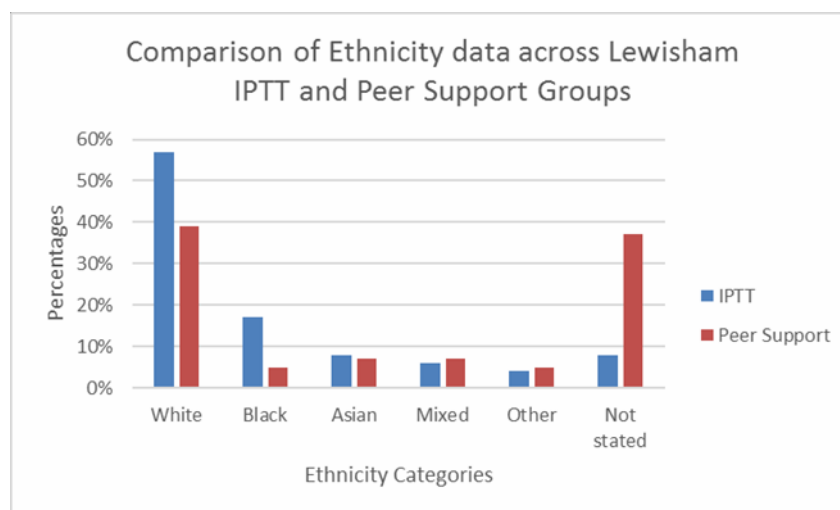


Figure 3.23. Comparison of ethnicity data across Lewisham IPTT and Peer Support groups

From the data presented in the graph, it may be said that service-users from a White ethnic background (this includes White English/Welsh/Northern Irish/British/Scottish, Irish, Gypsy/Irish Traveller and Any other) and Black ethnic background (this included Black African, Caribbean and Any other) are under-represented within the Peer Support Groups, as compared to the wider Lewisham IPTT caseload. However, there is a lot of missing data ("not stated" = 37%) for this item on the evaluation questionnaire. Hence, the data presented may not be an accurate representation of the ethnic backgrounds of

service-users accessing the peer support groups, and no conclusive ideas may be formulated.

4.4.2 Gender data

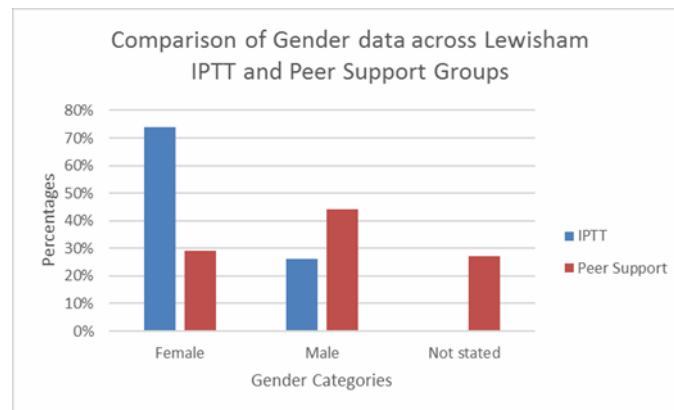


Figure 3.24. Comparison of gender data across Lewisham IPTT and Peer Support groups

From the data presented in the graph, it may be said that females are slightly under-represented, whilst males slightly over-represented within the Peer Support Groups, as compared to the Lewisham IPTT caseload. However, due to a large amount of missing data ("not stated"= 27%) this may not be an accurate representation.

4.4.3 Age data

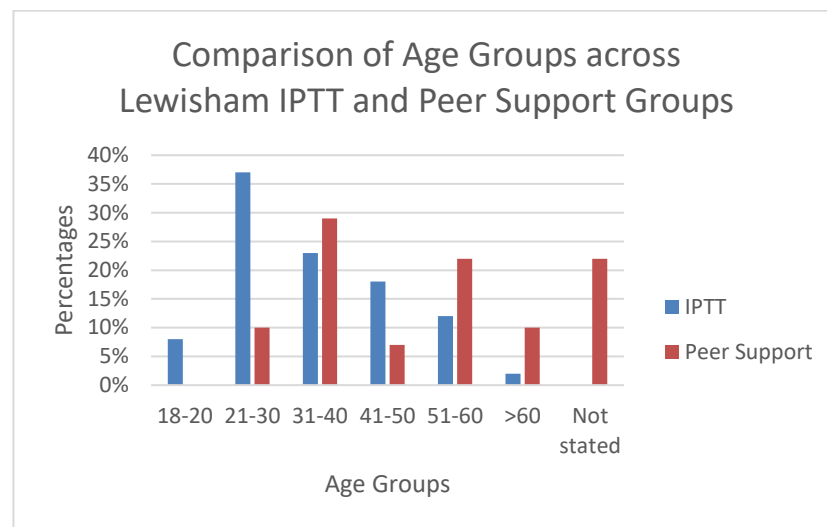


Figure 3.25. Comparison of age groups across Lewisham IPTT and Peer Support groups

The graph displays that younger and mid age groups (from 18 to 30 and 41 to 50) are significantly under-represented within the peer support groups, whilst the older age groups (>51) appear to be well represented. However, results may be non-conclusive due to a high volume of missing data from the evaluation questionnaires ("not stated" = 22%).

5. Discussion

5.1 Summary of results

Overall, the results from the evaluation questionnaire indicate a high level of satisfaction with the peer support groups. All areas within the evaluation questionnaire, including those related to daily living skills and mental health recovery, were found to be beneficial, at least to some extent. Only a small proportion of respondents reported

dissatisfaction with some aspects of the groups, particularly related to development of skills related to work or carrying out daily living activities.

Qualitative data gathered via the evaluation questionnaire and the focus group was also found to be overwhelmingly positive, with respondents identifying the benefits of the peer support groups, particularly in the areas of "reducing isolation", "overcoming stigma" and "feeling included". Further, respondents reported enjoying the opportunity to "try something new" and to "nudge my brain" into "developing new skills or enhancing existing skills". Qualitative data helped to gain a better understanding of aspects of the groups that respondents were less satisfied with, particularly around the practical issues in relation to funding and venues. The opportunity to share suggestions for improvements and new groups for the future generated helpful responses, which will be taken into consideration in the future development of the service.

The aim of this evaluation was to address a gap in the service, by developing and piloting an evaluation questionnaire, for use in the future, as an outcome measure. In the past, the Lewisham IPTT Peer Support Groups have been informally evaluated, by gaining sporadic feedback from service-users. However, this data has never been formally analysed and disseminated, for service development and improvement. As outlined by the BPS guidelines (2010), evaluation is an essential part of the process of developing and facilitating service- user involvement services. This study successfully met this aim, in collaboration with peer support volunteers and service-users, promoting involvement throughout. The evaluation questionnaire successfully gained meaningful feedback of the peer support service. The results were disseminated by sharing the

results with the Lewisham IPTT clinical team at a team meeting, where all core clinical staff and Peer Support staff were present.

Findings from the evaluation questionnaire and focus group were in line with previous research, which identified the benefits of peer support groups as being inclusive, assimilating knowledge of mental health issues, sharing feelings with peers experiencing similar difficulties and meeting new people (Wann, 1995). Of note was the reported benefit of "overcoming professional boundaries and relationships", which has been found to be a barrier to service-users engaging with mental health services (Morrison, Doucet & LeBlanc, 2008). To further address this barrier to engagement, data collection for this study (disseminating surveys and leading the focus group) was carried out by Peer Support Volunteers, with support from the Peer Support Coordinator and the researcher. Previous research has indicated that service-users are more likely to report the negative aspects of their experience when interviewed by a peer, due to an "interviewer effect" (Clark et al., 1999), as a peer-to-peer relationship fosters greater trust and confidence.

Further, feedback from peer support volunteers who facilitate some of the groups evaluated, but attend others, was also included within this study. Peer support volunteers often benefit as much, if not more than, the service-users they are supporting (Ontario Centre of Excellence for Child and Youth Mental Health, 2016), by gaining confidence and a sense of well-being. The process of sharing personal experiences of recovery to support this appears to be helpful, as one respondent reported: "I wanted others to know that you do get better and you do learn to manage it".

Finally, this study aimed to address the barrier of accessibility and inclusivity of minority groups (age, gender, ethnicity), who are often under-represented within mental health services (BPS, 2010). However, due to a large volume of missing demographic data from the evaluation questionnaire, this aim of the study was not achieved, as conclusive results could not be established.

5.2 Possible areas for improvement of Peer Support groups

The evaluation questionnaires and focus group generated several areas that could be improved upon, to further enhance the experience of the peer support groups. Each of these has been discussed below.

Budgeting and venues

Service-users commented on the lack of consistent and suitable venues for their groups. Specific issues around the accessibility of some venues, as well as the poor heating, were identified as areas of concern. Further, issues around a lack of material for some groups (i.e. art) were raised. Suggestions were made around ensuring accessible venues are offered, so as not to exclude service-users with mobility difficulties from the groups, and for funds to be allocated to purchase new materials for the groups. This may be a difficult area to address due to current lack of funding within the wider NHS, but will be raised with managers within Lewisham IPTT, for consideration.

Variety of groups

Although, overall, service-users were satisfied with the variety of creative groups offered by the peer support team, it was felt that some structured groups around skills training (for example, budgeting, DIY, well-being and/or meditation) would be

beneficial. Suggestions were also made for expansion of the existing creative groups, for example, holding an art exhibition of work completed in the art group, or a reading group as an extension of the creative writing group. These suggestions may be facilitated by the peer support team.

Reminders

Suggestions were made for reminders of group dates and times to be sent out via email or text message to service-users keen to attend. It was felt that a paper timetable, which was being used currently, was easily misplaced, and reminders could help to reduce rates of attrition. A group email, based on those who opt-in and consent to the service, may be set up, and reminders sent out regularly in a calendar-style format.

Frequency and length of groups

Many respondents felt that the frequency of the groups was too low and the length too short. Service-users considered the time it takes to set up and clear up after each group, which they felt could take up time from the activity. At the time of the focus group, the Peer Support Coordinator was working on increasing the frequency of some groups (for example, art) from fortnightly to weekly. Further, the peer support team were also considering increasing the time of the group, to incorporate set up and clear up. This will need to be considered in the context of available resources (i.e. volunteer time and venue availability).

6. Limitations of the evaluation and future recommendations

Several limitations to the current study were identified, which will be discussed here, along with recommendations for future research in the area.

Firstly, due to time constraints, only one focus group was facilitated, in which the new evaluation questionnaire was piloted. To establish a stronger measure of outcome, a second focus group may have been helpful in ensuring that all ground had been covered with regards to feedback on the draft questionnaire and any changes to be made (MacDougall & Fudge, 2001; Stewart, Shamdasani & Rook, 2007).

Secondly, the questionnaire was rolled out only for a short period, between March and June 2017, due to time constraints relating to project completion. Although the peer support team made every effort to ensure all group attendees during this period were given the opportunity to complete the questionnaire, it is unclear whether the sample is fully representative. Ideally, the questionnaire would be given to each attendee midway through their attendance to a group, or prior to leaving. Stages of engagement were not monitored in this study, which may have had an impact on the data gathered, if some respondents were new to the groups.

Thirdly, responses from peer support volunteers and service-users attending the groups were analysed collectively in this study. Although this provided an overview of the experience of the groups, it may be beneficial to look at the responses separately, to gain a better understanding of individual experiences of the groups. This is in conjunction with research that suggests peer support volunteers may benefit more than those who receive the support themselves (Bracke, Christiaens & Verhaeghe, 2008), which may have implications for the results of the current study. For the future, the experiences of these two groups should be looked at individually.

Finally, the demographic data gathered was insufficient to study the issues of accessibility and inclusivity. Demographic questionnaires were presented at the back of the questionnaire, which possibly led to them not being completed. Prompting respondents to complete this information may be helpful, or ensuring respondents are given enough information as to why this data is being carried, may also help address the issue of missing data. Another suggestion may be to gain consent from respondents to access this data from service records. However, for this to work, confidentiality and ethical issues will need to be seriously considered.

7. Conclusions

This evaluation study provided good preliminary data on the experience of service-users attending peer support groups. It also led to the establishment of a satisfactory evaluation survey which may be used continually and amended as deemed suitable to the service requirements. Areas for improvement of the groups have also been identified, along with suggestions of how these may be carried out. Further work is required to address the issue of accessibility.

8. References

- Bignall, T., Butt, J., & Pagarani, D. (2002). *Something to do: The development of peer support groups for young black and minority ethnic disabled people*, Bristol: Joseph Rowntree Foundation/Policy Press.
- Bracke, P., Christiaens, W., & Verhaeghe, M. (2008). Self-esteem, self-efficacy, and the balance of peer support among persons with chronic mental health problems. *Journal of Applied Social Psychology, 38*, 436-459.
- Braun, V. & Clarke, V. (2006). *Using thematic analysis in psychology. Qualitative Research in Psychology, 3*(2), 77-101.
- British Psychological Society (2010). *Good Practice Guidelines to Support the Involvement of Service-users and Carers in Clinical Psychology Services*. Leicester: British Psychological Society.
- Clark, C. C., Scott, E. A., Boydell, K. M., & Goering, P. (1999). Effects of client interviewers on client-reported satisfaction with mental health services. *Psychiatric Services, 50*(7), 961-63.
- Davidson, L., Chinman, M., Kloos, B., Weingarten, R., Stayner, T., & Tebes, J. (1999). Peer support among individuals with severe mental illness: A review of the evidence. *Clinical Psychology: Science and Practice, 6*, 165-187.
- Davidson, L., Bellamy, C., Guy, K., & Miller, R. (2012). Peer support among persons with severe mental illnesses: a review of evidence and experience. *World Psychiatry, 11*(2), 123-28.
- Hayes, D. (1997). Helping teachers to cope with large classes. *ELT Journal, 51*(2), 106-116.
- MacDougall, C., & Fudge, E. (2001). Planning and recruiting the sample for focus groups and in-depth interviews. *Qualitative Health Research, 11*, 117-126.

- Mead, S., Hilton, D., & Curtis, L. (2001). Peer support: A theoretical perspective. *Psychiatric Rehabilitation Journal*, 25(2), 134-41.
- Morrison, W., Doucet, C., & LeBlanc, M. (2008). *Best Practices: Early Intervention, Outreach and Community Linkages for Youth with Substance Use Problems*. Ottawa: Health Canada.
- NHS England (2015). *Liaison and diversion manager and practitioner resources: Service-user involvement*. Offender Health Collaborative.
- Omeni, E., Barnes, M., MacDonald, D., Crawford, M., & Rose, D. (2014). Service-user involvement: a survey of service-user and staff perspectives. *BMC Health Service Research*, 14, 491-504.
- Ontario Centre of Excellence for Child and Youth Mental Health (2016). *Youth peer support in a mental health context. Evidence In-Sight*. Retrieved from <http://www.excellenceforchildandyouth.ca/resource-hub/evidence-in-sightdatabase>.
- Pathways RTC (2013). *Implementing the peer support specialist role: providing direct, individualized support in a local program*. Portland, OR: Research and Training Center for Pathways to Positive Futures, Portland State University.
- Repper, J., Aldridge, B., Gilfoyle, S., Gillard, S., Perkins, R., & Rennison, J. (2013). *Peer support workers: Theory and practice. Implementing Recovery through Organisational Change Briefing (ImROC)*. London: Centre for Mental Health and Mental Health Network.
- Sherman, P., & Porter (1991). Mental health consumers as case management aids. *Hospital and Community Psychiatry*, 42, 494-498.
- Solomon, P. (2004). Peer support/peer provided services underlying processes, benefits, and critical ingredients. *Psychiatric Rehabilitation Journal*, 27(4), 392-401.

- Stewart, D. W., Shamdasani, P. N., & Rook, D. W. (2007). *Focus Groups: Theory and Practice (2nd ed.)*. Thousand Oaks, CA: Sage.
- Together Charity, & National Survivor User Network (2014). *Service-user Involvement in the Delivery of Mental Health Service Briefing*. London: Together UK.
- Wann, M. (1995). *The development of 'self-help' activities. Findings: Social Care Summary 5*. York: Joseph Rowntree Foundation.
- World Health Organisation (2017). *Creating peer support groups in mental health and related areas – WHO QualityRights training to act, unite and empower for mental health (pilot version)*. Geneva: World Health Organisation.

9. Appendices

Appendix A: Peer Support group evaluation questionnaire

Peer Support Groups - Evaluation Questionnaire

Group:

My Mental Health and Well-being

The group helped me to:

1. Improve my self-esteem/feel better about myself

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

2. Improve my mood/emotional well-being

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

3. Make new friends/meet new people

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

4. Share my feelings with others/talk to others

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

5. Feel accepted by others

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

6. Feel understood by others

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

7. Feel included/less isolated or alone

1	2	3	4	5
----------	----------	----------	----------	----------

Not at all	A little bit	Somewhat	Much	Very Much
------------	--------------	----------	------	-----------

8. Feel comfortable

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

9. Get support with my difficulties/learn ways of coping

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

10. Develop my self-confidence

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

My Daily Skills

The group helped me to:

1. Gain new skills required for work/jobs

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

2. Develop my interpersonal/social skills

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

3. Carry out daily activities (like shopping, cooking, cleaning)

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

4. Develop my hobbies and interests

1	2	3	4	5
Not at all	A little bit	Somewhat	Much	Very Much

5. Get out more

1	2	3	4	5
----------	----------	----------	----------	----------

Not at all	A little bit	Somewhat	Much	Very Much
------------	--------------	----------	------	-----------

6. Try new and different things

1 Not at all	2 A little bit	3 Somewhat	4 Much	5 Very Much
------------------------	--------------------------	----------------------	------------------	-----------------------

7. Socialise more

1 Not at all	2 A little bit	3 Somewhat	4 Much	5 Very Much
------------------------	--------------------------	----------------------	------------------	-----------------------

8. Ask for help when I needed it/access support

1 Not at all	2 A little bit	3 Somewhat	4 Much	5 Very Much
------------------------	--------------------------	----------------------	------------------	-----------------------

9. Take care of my needs better

1 Not at all	2 A little bit	3 Somewhat	4 Much	5 Very Much
------------------------	--------------------------	----------------------	------------------	-----------------------

What did you find most helpful about the group?

What did you find least helpful about the group?

Why did you choose to attend Peer-support groups?

What did you expect to gain from the group when you first attended?

Did the group meet your expectations?

Do you have any suggestions on how we can improve our groups? Are there any other groups you would like?

Demographics

Age:

Gender:

Ethnicity:

Current IPTT engagement:

Completed
On going
Yet to start
N/A

Have you attended any other Peer-support Groups?

Art

Gardening

Writing

Yes No
Women's Support Group

Appendix B: Focus group questions

Peer Led focus group 14/03/2017

1. What did you find most helpful about the groups?
2. What did you find least helpful about the groups?
3. Why did you choose to attend Peer-support groups?
4. What did you expect to gain from the group when you first attended?
5. Did the group meet your expectations?
6. Do you have any suggestions on how we can improve our groups? Are there any other groups you would like?
7. Are there any suggestions/changes you would like to make to the evaluation questionnaire?

Prompts – What do you think of the wording?

What do you think of the rating scale?

What do you think of the questions? Are they relevant?