THE DEVELOPMENT OF A NONVERBAL TEST OF ACCELERATED LONG-TERM FORGETTING FOR USE WITH PEOPLE WITH EPILEPSY

Joanne Crossley

Submitted in accordance with the requirements for the degree of

Doctor of Clinical Psychology (D. Clin. Psychol.)

The University of Leeds

School of Medicine

Academic Unit of Psychiatry and Behavioural Sciences

August 2016

ACKNOWLEDGMENTS

I would like to thank my supervisors, Dr Richard Allen, Dr Anna Weighall and Dr Steven Kemp for all their encouragement and support. It has been a challenging couple of years and I am particularly grateful for their understanding when things got tough. I have really appreciated all their guidance, suggestions and feedback. Thanks also to Prof Alan Baddeley who offered advice on the earlier versions of the task.

I would like to thank the Max Hamilton Fund for the additional funds to carry out the study and to Epilepsy Action (EA), particularly Amanda Stoneman and Cliff Challenger, for helping to advertise the project on the website and through the local EA groups.

I would like to express my gratitude to Emma Beech, Polly Barker and Alexia Holt for all their hard work in recruiting and assessing their friends and family. I am also indebted to Helen Hill for creating the visual scenes and only charging me 'mate's rates'. Thanks mate!

Thank you to all the participants for giving up their valuable time to participate, for their enthusiasm for the project and for providing detailed feedback, which has helped with the development of the task.

Finally, to Andy, thank you for supporting me on another doctoral journey. You know I could not do it without you. I promise this is the last one! And to Hannah, for all the bedtimes I've missed….you've got your Mummy back.

ABSTRACT

Background

Accelerated long-term forgetting (ALF) is a novel form of memory impairment whereby some people with epilepsy (PWE) demonstrate ‘normal’ patterns of learning and memory over short retention intervals (i.e. 20-30 minutes) but then experience rapid forgetting over longer delays. Currently, there is no consensus on the measures used to assess ALF and little attention has been paid to how clinicians should assess ALF in clinical practice. The aim of this research was to develop and pilot a clinically feasible nonverbal measure of ALF.

Method

Phase 1 comprised the initial development and piloting of the Action-People-Places (APP) test materials and procedure. Six versions were created and each version was piloted on small numbers of healthy adults and/or PWE. Modifications to each version were made in line with participants’ performance and feedback. Phase 2 involved administering the final APP test, as well as a brief battery of neuropsychological tests, to 32 healthy adults and an individual, SK, with a confirmed diagnosis of ALF, to assess its reliability and validity. Comparisons were made to two PWE involved in Phase 1. Telephone follow-ups were undertaken at 24 hours, 1-week and 3-weeks. Phase 3 evaluated its acceptability using a brief structured interview format.

Results

Healthy adults demonstrated forgetting on the APP test. There were no floor effects but some evidence of ceiling effects. The test had modest levels of reliability (.67-.83). Older age was associated with increased forgetting over time. There were some associations with existing memory measures. SK demonstrated ALF on the task. The APP test was considered acceptable to participants.

Discussion

Despite several limitations to the study, including the use of an unmatched healthy adult group, the APP test appears to be a promising measure of ALF, which

is worthy of further development with larger patient groups and a more representative control group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AD	Alzheimer's Disease
AED	Antiepileptic Drug
ALF	Accelerated Long-Term Forgetting
ANOVA	Analysis of Variance
APP	Action-People-Places
DMI	Delayed Memory Index
EA	Epilepsy Action
EEG	Electroencephalogram
EMQ	Everyday Memory Questionnaire
FSIQ	Full Scale IQ
GCSE	General Certificate of Secondary Education
HADS	Hospital Anxiety and Depression Scale
ILAE	International League Against Epilepsy
IMI	Immediate Memory Index
LTM	Long-Term Memory
MCI	Mild Cognitive Impairment
MTL	Medial Temporal Lobe
MTLE	Medial Temporal Lobe Epilepsy
MTS	Mesial Temporal Sclerosis
MTT	Multiple Trace Theory
NFQ	No Formal Qualification
NVQ	National Vocational Qualification
PRI	Perceptual Reasoning Index
PSI	Processing Speed Index
PWE	People with Epilepsy
RBANS	Repeatable Battery for the Assessment of Neuropsychological Status
REM	Rapid Eye Movement
RESUS	Research Experience Scheme for Undergraduate Students
SLI	Specific Language Impairment
SPSS	Statistical Package for Social Sciences
STM	Short-Term Memory
TEA	Transient Epileptic Amnesia
TLE	Temporal Lobe Epilepsy
TOPF UK	Test of Premorbid Functioning UK version
UoL	University of Leeds
VCI	Verbal Comprehension Index
VWMI	Visual Working Memory Index
WAIS	Wechsler Adult Intelligence Scale
WMI	Working Memory Index
WMS	Wechsler Memory Scale

INTRODUCTION

Background to the Project

People with epilepsy (PWE) often report memory problems. This may be due to the underlying aetiology of the epilepsy; the effects of recurrent seizures; the side effects of antiepileptic drug (AED) treatment and/or psychosocial issues (Aldenkamp & Bodde, 2005; Meador, 2002). These memory problems can have a significant impact on an individual's psychological functioning and quality of life (e.g. Baker, Taylor & Hermann, 2009). However, some studies have failed to find a relationship between subjective self-report of memory problems and performance on objective memory measures (e.g. Marino et al., 2009; Thompson & Corcoran, 1992). Several reasons for this lack of association have been suggested but one may be related to the phenomenon of accelerated long-term forgetting (ALF).

ALF is a relatively recently identified memory impairment, whereby individuals demonstrate 'normal' levels of learning and memory over the short retention intervals typically assessed in standardised memory tests (i.e. 20-30 minutes) but then experience 'abnormal' rates of forgetting over longer delays of days and weeks. ALF has largely been associated with temporal lobe epilepsy (TLE). There is a growing interest in ALF, as not only does it help explain some of the memory problems experienced by PWE, but it also adds to the understanding of the processes involved in memory functioning (Leritz, Grande & Bauer, 2006).

Currently, there is no consensus on the measures and procedure to assess ALF, which may have contributed to some of the mixed findings in this area. Researchers have either adapted existing memory tests or developed their own materials. As detecting ALF requires measuring forgetting over multiple intervals, little attention has been paid to how to assess this in clinical practice. Therefore, this thesis aims to develop a test of ALF for use in clinical practice with PWE. This introductory chapter will provide a brief overview of epilepsy and models of memory and forgetting. The research into ALF and the methodological issues in its assessment will be summarised and the clinical implications of this project will be discussed.

Epilepsy

Epilepsy is one of the most common neurological conditions, affecting approximately 600 000 people in the United Kingdom (Joint Epilepsy Council, 2011). Previously, epilepsy was defined as the tendency to have recurrent epileptic seizures (Gastaut, 1973). However, in 2005, it was redefined as '*a disorder of the brain characterised by an enduring predisposition to generate epileptic seizures and by the neurobiological, cognitive, psychological, and social consequences of the condition'* (Fisher et al., 2005, p.471). A more recent definition has been published, which re-conceptualises epilepsy as a 'disease' rather than a 'disorder' to emphasise its seriousness (Fisher et al., 2014). Epilepsy is not a single disease but encompasses a variety of heterogeneous disorders that are symptoms of an underlying neurological disorder (Stokes, Shaw, Juarez-Garcia, Camosso-Stefinovic & Baker, 2004).

An epileptic seizure has been defined as a '*transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain'* (Fisher et al., 2005, p.471). Seizures are divided into generalised (i.e. originate within and rapidly engage bilaterally distributed networks) and focal (i.e. originate within networks limited to one hemisphere), with or without impairments in consciousness or awareness (Berg et al., 2010). Over the last decade, there have been several revisions made to the classification of seizures and terms such as focal seizures have replaced simple and complex partial seizures; although these are still commonly used. The International League Against Epilepsy (ILAE) is currently seeking consultation on further revisions to the classification of seizure types.

There is also a classification system for epilepsies and epilepsy syndromes. Epilepsies have previously been classified as to whether they are localisation-related (i.e. have focal-onset seizures) or generalised and whether the underlying cause is idiopathic (i.e. genetic), symptomatic (i.e. lesion or identifiable pathology) or cryptogenic (i.e. unknown case but presumed symptomatic) (Commission of the ILAE, 1989). Due to advances in understandings and knowledge about epilepsy, the classification of epilepsies is also currently being revised (Scheffer et al., 2014).

Treatment with AEDs is the first approach in the management of PWE (NICE, 2012). Most people with newly diagnosed epilepsy will achieve seizure freedom but approximately 20-40% of patients will never experience seizure remission (Kwan & Sander, 2004). For those with drug resistant (refractory) epilepsy, alternative treatments are epilepsy surgery and vagal nerve stimulation.

Epilepsy is '*more than seizures*' having impacts on multiple aspects of an individual's life (Engel, Jr & Pedley, 2008, p.2), including cognitive functioning. There is a large body of literature showing that PWE experience cognitive problems, in particular, memory problems (e.g. Baker, Jacoby, Buck, Stalgis & Monnet, 1997; Giovagnoli, Mascheroni & Avanzini, 1997; Hermann, Seidenberg, Lee, Chan & Rutecki, 2007).

Memory

Memory is the ability to encode, store, retain and recall information. It is important as it enables us to learn from past experiences, interact with the world and build relationships. It is a complex concept that involves interaction between a number of processes and systems. The three main processes of encoding (receiving and processing information); storage (creating and maintaining a permanent record of the information) and retrieval (capacity to recall the stored information) are thought to underlie memory ability.

Models of Memory

Atkinson and Shiffrin (1968) proposed the modal model of memory with memory consisting of a sensory store, short-term memory (STM) and long-term memory (LTM) store. They suggested that information passes through each store via rehearsal. However, Baddeley and Hitch (1974) offered the working memory model as an alternative, which was later updated (Baddeley, 2000). This suggests that STM is not a unitary store but is made up of different subsystems: the phonological loop, visuo-spatial sketch pad and episodic buffer, controlled by the central executive. These systems enable information to be held in mind and manipulated. Rather than flowing in a single direction, information from these different systems interacts with the LTM systems (Baddeley, 2015b).

Similarly, LTM is unlikely to be a single system and has been subdivided into procedural/nondeclarative memory and declarative memory, which comprises episodic and semantic memory (e.g. Squire, 1987; Tulving, 1972) (see Table 1). The different systems are thought to process different types of information and are mediated by different brain areas, although they are likely to interact with each other (e.g. Baddeley, 2015a; Squire, 2004).

Table 1: Summary of memory systems

Memory system	**Description (length information is held)**
Sensory Memory	Initial storage of sensory stimuli (approx. 2 secs)
Working Memory/ Short-Term Memory	System for holding, processing and manipulating information of limited capacity (< 30 seconds). May comprise multiple components - phonological loop, visuo-spatial scratchpad, central executive and episodic buffer (Baddeley, 1986; Baddeley, 2000)
Long-Term Memory	Longer-term storage (> 30 seconds) of information of virtually unlimited capacity
Declarative Memory	Accessed explicitly/consciously
Episodic Memory	Memory for personally experienced events
Semantic Memory	General knowledge and facts
Procedural/Nondeclarative memory	Accessed implicitly/unconsciously (e.g. motor skills, habits, priming, associative learning)
Prospective Memory	Memory to remember to do things

Consolidation

Consolidation is thought to be the process by which memories are passed to long-term declarative memory and are gradually reorganised or changed to make them more resistant to disruption (Alvarez & Squire, 1994). The medial temporal lobes (MTLs) are thought to be heavily implicated in this process. Early evidence for this comes from the case study of patient HM who had an experimental procedure involving removal of the bilateral MTLs to help control refractory epilepsy, which rendered him with severe amnesia (e.g. Milner, Corkin & Teuber, 1968; Scoville & Milner, 1957). HM was unable to form new memories following the procedure (anterograde amnesia) and had a temporally graded retrograde amnesia (i.e. more recent memories from prior to the surgery were forgotten more easily than memories from further in his past-‘Ribot’s law’, 1881, cited in Wixted, 2004).

Alvarez and Squire (1994) proposed a neural network model known as the ‘standard theory of consolidation’. They argued that declarative memories are stored temporarily in the MTL memory system, which comprises the hippocampus, entorhinal, perirhinal and parahippocampal cortices. Over time, these memories are gradually stored more permanently in the neocortex, as neocortical representations, through reciprocal connections and interaction with the MTL. Damage to the MTL, as occurred with HM, for example, leads to the inability to form new memories as information is not established in LTM and also leads to the loss of more recent memories, as older memories are more likely to be preserved in neocortical areas, which rely less on the MTL structures (Squire & Alvarez, 1995).

Two mechanisms are thought to underlie this process: synaptic and systems consolidation. Synaptic (fast) consolidation takes place over minutes to hours after learning (Dudai, 2004). It is thought to take place within the hippocampus where synaptic connections are strengthened through the process of long term potentiation. In contrast, systems (slow) consolidation is a longer process that can last from days to years (Dudai, 2004). It involves the activation of the connections between the MTL and the diffuse cortical areas (e.g. prefrontal cortex, parietal and temporal cortex), which are involved in storing the memory trace in a large-scale distributed network. Over time, the connections between the different cortical areas are strengthened and may become independent of the MTL (Dickerson & Eichenbaum, 2010; Dudai, 2004; Squire & Alvarez, 1995).

Multiple trace theory (MTT) (Nadel & Moscovitch, 1997), however, suggests that memory traces are not transferred to the neocortex but the hippocampus always remains involved and is re-activated whenever a memory trace is retrieved. The trace is then recoded so the more a memory is recollected, the more traces for that memory are created, making it more retrievable and less vulnerable to MTL damage (Nadel & Moscovitch, 1997).

Winocur, Moscovitch and Bontempi (2010) have built on MTT further with the transformation hypothesis. They propose that there is a difference in the way episodic and semantic memories are processed. A newly formed episodic memory, which is context-specific, is dependent on the hippocampus but the hippocampus is also involved with forming a more schematic ('gist') version of the memory, which is stored in the neocortex. Either of these memories may be retrieved depending on the retrieval cues. This means that MTL damage or disruption will impair episodic memories for recent events, if a 'gist' memory has not yet been able to be formed. If a semantic/schematic version has been formed, individuals are able to use this, which provides an alternative explanation for the temporal gradient found in retrograde amnesia (Winocur et al., 2010).

Sleep is thought to play an important role in the consolidation of new memories (see Diekelmann & Born, 2010 for a review). Several studies have shown that individuals retain more information if they have a period of sleep following learning (e.g. Marshall & Born, 2007). This seems to be more effective if sleep occurs close to learning (Diekelmann & Born, 2010). There is a debate between whether sleep is beneficial due to an active process (i.e. memory traces are preferentially reactivated during sleep, strengthening them and making them less vulnerable to interference) or a passive process (i.e. sleep protects memories from the interference found in everyday life) (Atherton, Nobre, Zeman & Butler, 2014; Diekelmann & Born, 2010; Wixted, 2010). Diekelmann and Born (2010) have also reviewed the evidence suggesting that synaptic consolidation seems to be supported by rapid eye movement (REM) sleep and system consolidation by slow-wave sleep. The synaptic homeostasis hypothesis (i.e. the reduction of synaptic strength during sleep) and active system consolidation hypothesis (i.e. re-activation of memories during sleep) are two current theories which attempt to account for the role of sleep in memory consolidation. A discussion of these theories is beyond the scope of this thesis but for reviews see Diekelmann and Born (2010) and Huijgen and Samson (2015).

Retrieval

Most memory research tests retrieval using either recall or recognition memory paradigms. Recall memory tests can either be free recall (e.g. participants are asked to recall as many items in a list as they can remember) or cued recall (e.g. participant given the category to aid retrieval of a word list). Recognition memory can be tested by presenting a mixture of old and new material after a delay and asking individuals to identify the old material.

Recognition is thought to be based on a dual process model involving recollection and familiarity, which are proposed to be mediated by different brain systems (Huijgen & Samson, 2015). Recollection involves remembering the details of a memory (e.g. 'was this object shown on the left of the right?'). As the hippocampus is thought to be context-specific [cf. transformation hypothesis (Winocur et al., 2010)], the hippocampus and parahippocampal cortex are thought to be important for recollection (Huijgen & Samson, 2015). In contrast, familiarity refers to a sense of knowing (e.g. participant asked to respond ‘yes/no’ to whether a word was presented previously). This relies less on context and is thought to be mediated by the perirhinal cortex (Huijgen & Samson, 2015). This explains why people with hippocampal damage may perform better on recognition tasks based more on familiarity judgments (Huijgen & Samson, 2015; Leritz et al., 2006).

Memory tests assess retention of a memory, which is measured at a single point in time. However, this is different to forgetting, which is measured at two or more points in time and is represented by a decline in performance (MacDonald, Stigsdotter-Neely, Derwinger & Bäckman, 2006). In comparison to research on memory and its underlying processes, research into forgetting has largely been neglected (Della Sala, 2010).

Forgetting

Forgetting is the ‘*inability to recall something now that could be recalled on a previous occasion*’ (Tulving, 1974, p.74) or ‘*the absence of expression of previously properly acquired memory in a situation that normally would cause such expression. This can reflect actual memory loss or a failure to retrieve existing memory’* (Hardt, Nader, & Nadel, 2013, p.111). Therefore, forgetting can refer to whether a memory is inaccessible (i.e. unable to be retrieved) or unavailable (i.e. lost from memory storage) (Tulving & Pearlstone, 1966). Ebbinghaus was the first to demonstrate a forgetting curve for information over time (see Figure 1). He found that recent memories are forgotten at a quicker rate than older memories, where the rate of forgetting slows down. This is related to Jost’s (1897’s) law of forgetting, which states that if two memories are of the same strength but different ages, the older memory will decay more slowly than the younger (cited in Wixted, 2004). Forgetting is thought to be a highly adaptive process, as there are energy and efficiency costs to remembering (Frankland, Köhler & Josselyn, 2013; Hardt et al., 2013; Ludowig et al., 2010; Roediger, Weinstein & Agarwal, 2010).

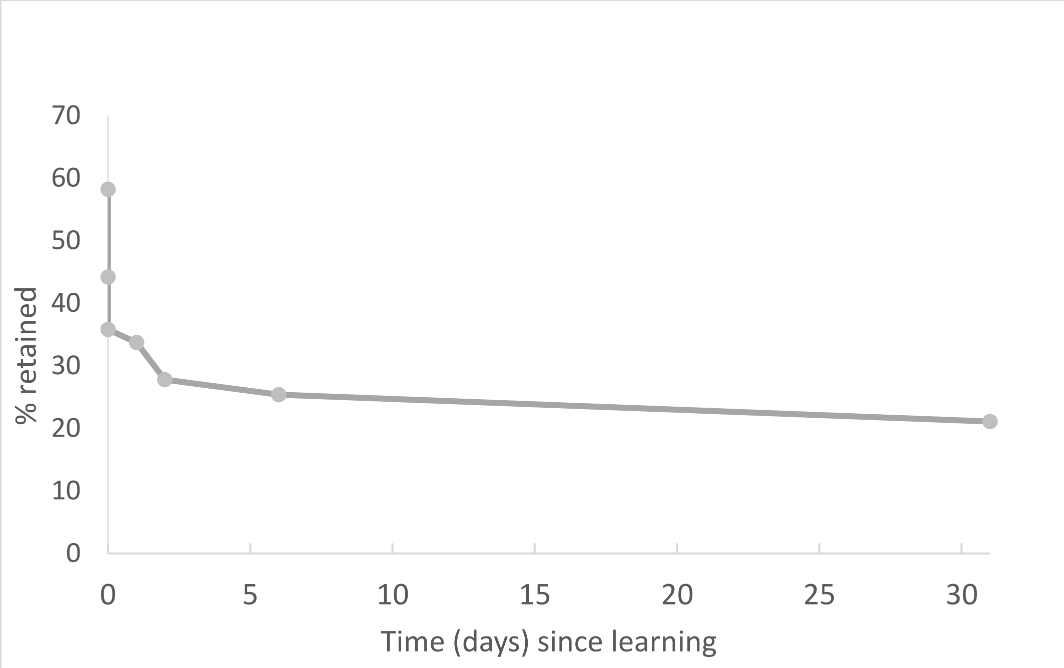


Figure 1: Forgetting curve using data from Ebbinghaus (1885, p.76)

Theories of Forgetting

The underlying mechanisms of forgetting are still unclear. However, there are several different theories. Decay theory (Peterson & Peterson, 1959) suggests that memories weaken over time. This theory has been criticised for re-stating the forgetting curve and not proposing the underlying mechanisms by which decay occurs (Roediger et al., 2010). More recent reviews, however, have hypothesised that hippocampal neurogenesis (i.e. generation of new neurons) may be the neurobiological mechanism by which memories are continually 'cleared' from the hippocampus leading to a 'decay-like' process (Frankland et al., 2013). Frankland and colleagues (2013) propose that these new cells and the synaptic connections they form within the hippocampus disrupt existing memory traces making it less likely that a retrieval cue will be able to reactivate the existing memory trace. Similarly, Hardt and colleagues (2013) also propose a 'decay-like' process, whereby unnecessary memories are actively removed from the hippocampus mainly during sleep. They suggest that it is the hippocampal component of a memory which is lost. As this contains the spatial-temporal context, LTM shifts over time to being more semantic rather than episodic (Hardt et al., 2013) (cf. transformation hypothesis, Winocur et al., 2010).

Interference theory suggests that forgetting is caused by '*the wrong memory being accessed by a particular cue*' (McGeoch, 1942 cited in Roediger et al., 2010, p.10). This suggests that there is competition between different memories. Interference has been divided into proactive interference (i.e. effects of prior learning on to-be-remembered material) and retroactive interference (i.e. effects of new information after material has been encoded). Neurogenesis and the formation of new memories may act as a form of non-specific retroactive interference (Frankland et al., 2013; Wixted, 2004). In support of interference theory, there is evidence that a short period of wakeful rest immediately following learning increases retention (e.g. Dewar, Alber, Butler, Cowan & Della Sala, 2012; Dewar, Alber, Cowan & Della Sala, 2014). Interestingly, this has also been shown to occur for amnesic patients even after seven days (Alber, Della Sala & Dewar, 2014) suggesting that rest may protect compromised memory consolidation systems from interfering information and enhances early consolidation (Alber et al., 2014; Dewar et al., 2014; Hardt et al., 2013). This is consistent with the literature showing the beneficial effects of sleep on memory (e.g. Diekelmann & Born, 2010; Wixted 2004).

Failure of consolidation, therefore, may be an important process in forgetting. Interference may disrupt consolidation (and re-consolidation) processes (Hardt et al., 2013; Nader & Hardt, 2009; Wixted, 2004). As discussed previously, memories are initially fragile until they are consolidated to a more permanent store in the neocortex where they become less vulnerable to disruption (e.g. Nadel & Moscovitch, 1997; Squire & Alvarez, 1995; Winocur et al., 2010). This can help explain Jost's law (Wixted, 2004) and the temporal gradient observed in retrograde amnesia (Dewar, Cowan & Della Sala, 2010). However, Brown and Lewandowsky (2010) argue that consolidation is not as important for forgetting and instead offer a model based on temporal distinctiveness. Similarly, Watkins and Watkins (1975) suggest that forgetting is more likely to occur when a cue is associated with multiple memories (retrieval-cue overload theory).

It may be that multiple processes are involved in forgetting. Frankland et al. (2013) proposes that the outcome of a memory may depend on competition between consolidation and decay processes. For example, during childhood when there are high rates of neurogenesis, children are unable to form long-term memories, leading to infantile amnesia (i.e. the phenomenon whereby adults are unable to remember very early childhood experiences) (Josselyn & Frankland, 2012). Similarly, Sadeh, Ozubko, Winocur and Moscovitch (2014) suggest that the causes of forgetting may vary depending on the nature of the memory. Memories that rely on the hippocampus, such as recollection, are sensitive to decay (cf. Hardt et al., 2013) whereas memories that are less hippocampal-dependent, such as familiarity, are more vulnerable to interference (Sadeh et al., 2014).

In addition to the involvement of multiple processes, there may be multiple factors associated with non-pathological forgetting. MacDonald and colleagues (2006) reviewed the literature on factors affecting forgetting. They found that a slower rate of learning, poorer cognitive performance and older age increased rates of forgetting. Similarly, Mary, Schreiner and Peigneux (2013) found that older healthy adults had a higher forgetting rate after seven days compared with younger adults. The authors suggested that their study provides support for age-related declines in verbal memory consolidation processes, which may be associated with more frequent intra-sleep awakenings in older age. Baddeley, Rawlings and Hayes (2013) also found that older adults had an increased rate of forgetting on a test of verbal memory, which they suggested was indicative of the phenomenon of ALF.

Accelerated Long-term Forgetting

ALF is a novel form of memory impairment whereby individuals demonstrate ‘normal’ patterns of learning and anterograde memory performance over relatively short retention intervals (approx. 20-30 minutes) but then experience rapid forgetting over longer delays of days and weeks (e.g. Blake, Wroe, Breen & McCarthy, 2000; Butler & Zeman, 2008). It seems to be restricted to declarative memory, in particular episodic memory (Tramoni et al., 2011), implicating a role for the MTL. Therefore, it is unsurprising that it has typically been associated with people with TLE and in particular those with transient epileptic amnesia (TEA) (Butler & Zeman, 2008; Butler et al., 2009; Elliott, Isaac & Muhlert, 2014; Fitzgerald, Mohamed, Ricci, Thayer & Miller, 2013a; Manes, Graham, Zeman, de Luján Calcagno & Hodges, 2005; Muhlert et al., 2011).

TLE is the most common form of localisation-related epilepsy (Blair, 2012). It is associated with focal seizures with impairments in awareness or consciousness (previously known as complex partial seizures) arising from the temporal lobes.

People with temporal lobe seizures report experiences such as déjà vu, rising epigastric sensations, feelings of fear, and automatisms such as chewing or lip smacking. It is most commonly caused by mesial temporal sclerosis (MTS) but other causes include central nervous system infections, tumours and head trauma (Blair, 2012).

TEA is a distinctive syndrome but is a subtype of medial temporal lobe epilepsy (MTLE) that is characterised by recurrent, brief episodes of isolated memory loss (Asadi-Pooya, 2014; Butler et al., 2007; Zeman, Butler, Muhlert & Milton, 2013). During the amnesic episodes, other cognitive functions are intact. The episodes often occur on waking and last less than an hour. TEA usually responds well to AED treatment (Asadi-Pooya, 2014). It has been associated with onset of epilepsy in middle to older age and is more common in men (Butler & Zeman, 2008). Many people with TEA complain of memory difficulties between seizures. However, TEA can occur without ALF and ALF can occur in the absence of evidence for TEA (Kemp, Illman, Moulin & Baddeley, 2012).

ALF was initially described in case reports (e.g. de Renzi & Luchelli, 1993; O'Connor, Sieggreen, Ahern, Schomer & Mesulam, 1997). Fitzgerald et al. (2013a) reviewed the 12 single case studies of ALF and found that seven patients had TLE, two had TEA and one had generalised seizures. Only two did not have epilepsy. Of these, one patient had brain hypoxia with damage to the posterior temporal lobes (de Renzi & Luchelli, 1993). The remaining patient, who reported memory loss for events after a night of sleep, was diagnosed with functional amnesia. Her pattern of memory impairment was different to other reports of ALF in the literature and her memories could be retrieved at longer delays under certain conditions (Smith et al., 2010), therefore, should not be included in the review. Elliott and colleagues (2014) reported that that the majority of the case studies had abnormal structural brain abnormalities in the temporal lobe. These findings have led to a growing number of group studies of ALF in people with TLE.

The most recent reviews (Elliott et al., 2014; Fitzgerald et al., 2013a) have indicated that some group studies have not found for evidence for ALF in TLE patients (Bell, Fine, Dow, Seidenberg & Hermann, 2005; Bell, 2006; Giovagnoli, Casazza, Avanzini, 1995) but this may be complicated by these studies showing impaired memory performance at the standard delays; differences in clinical characteristics between the patient groups (e.g. pre and post-surgical patients), and differences in study methodology. The overwhelming majority of studies have found that people with TLE or TEA demonstrate ALF when comparing group means (Elliott et al., 2014; Fitzgerald et al., 2013a). However, epilepsy is heterogeneous and not all TLE patients demonstrate ALF (Elliott et al., 2014; Fitzgerald et al., 2013a). In a recent study by Miller, Flanagan, Mothakunnel, Mohamed and Thayer (2015), 47% of PWE demonstrated ALF, which they defined using cut-scores determined from normative data. All but one of these patients had TLE. Older age and higher IQ was also predictive of ALF on verbal recall tasks. However, it is important to note that 17% of healthy adults were also defined as having ALF in this study (Miller et al., 2015).

ALF has been found for both verbal and nonverbal material (e.g. Cassel, Morris, Koutroumanidis & Kopelman, 2016; Fitzgerald et al., 2013a; Mameniskiene, Jatuzis, Kaubrys & Budrys, 2006; Muhlert et al., 2011) and there is some evidence for laterality effects. Those with left-sided temporal lobe damage showed more rapid forgetting for verbal material (e.g. Blake et al., 2000; Narayanan et al., 2012; Ricci, Mohamed, Savage & Miller, 2015a). Those with right-sided temporal lobe damage showed evidence for rapid forgetting of nonverbal material; although this was not consistently found (Bell & Giovagnoli, 2007; Blake et al., 2000; Butler & Zeman, 2008; Elliott et al., 2014; Miller et al., 2015; Narayanan et al., 2012; Wilkinson et al., 2012). Muhlert and colleagues (2011) found ALF for both verbal and visual material regardless of lateralisation, which may reflect that some visual tasks can often be verbally encoded. This will be discussed further in a later section.

Proposed Mechanisms of ALF

As individuals with ALF are able to demonstrate complete learning and retrieval of information at short delays, ALF is unlikely to be due to an acquisition or a retrieval deficit (e.g. Blake et al., 2000; Lah, Mohamed, Thayer, Miller & Diamond, 2014). ALF has also been found on recognition as well as recall tasks (e.g. Blake et al., 2000; Butler & Zeman, 2008; Tramoni et al., 2011) providing further support against a retrieval deficit explanation. In contrast, some studies have found intact recognition (e.g. Martin et al., 1991; Ricci, Mohamed, Savage, Boserio & Miller, 2015b); although this may reflect the length of delay or the testing procedure (Ricci et al., 2015b). McGibbon and Jansari (2013) in their single case study found that recognition memory appeared intact on testing but inspection of the individual's performance suggested that he was not using recollection but a sense of familiarity to identify the target. This fits with the dual process model of recognition and suggests that the hippocampus may be more implicated in ALF (Huijgen & Samson, 2015; Isaac & Mayes, 1999; Leritz et al., 2006; Ricci et al., 2015b).

The most discussed hypothesis is that ALF is more likely to reflect a failure in memory consolidation (Blake et al., 2000; Butler, Muhlert & Zeman, 2010; Hoefeijzers, Dewar, Della Sala, Zeman & Butler, 2013; Huijgen & Samson, 2015; Isaac & Mayes, 1999; Muhlert, 2012). As memories are retained after shorter intervals but quickly forgotten at longer time intervals, ALF supports the idea of systems (slow) consolidation (Dudai, 2004; Hoefeijzers et al., 2013; Huijgen & Samson, 2015). However, Cassel and colleagues (2016) propose that ALF may reflect disruption to 'early' consolidation processes and this either becomes apparent immediately or becomes more progressive, dependent on other epilepsy characteristics, such as polytherapy (i.e. treatment with two or more AEDs) or presence of MTS.

As discussed earlier, sleep is thought to be important for memory consolidation (e.g. Diekelmann & Born, 2010) so it was initially proposed that ALF may reflect deficits in sleep-dependent consolidation, particularly as TEA is associated with amnesic episodes upon awakening. However, there is emerging evidence that this is not the case (Atherton et al., 2014; Deak, Stickgold, Pietras, Nelson & Bubrick, 2011; Dewar et al., 2010; Hoefeijzers, Dewar, Della Sala, Butler, & Zeman, 2014; McGibbon & Jansari, 2013). Hoefeijzers and colleagues (2014) tracked ALF for word lists over a 24 hour period and 1-week after learning in patients with TEA. They found that ALF could be detected within 3-8 hours of learning and no further ALF was demonstrated following a night of sleep. The authors argued that their findings support the theory that ALF results from failures in memory consolidation that are occurring within a few hours of learning rather than failures in the consolidation processes that occur during sleep. However, as they did not measure forgetting 3-8 hours post-learning after a period of sleep, they could not conclude whether the results were due to impairments that happen during wakefulness specifically.

Atherton and colleagues (2014) compared the performance of 11 patients with TEA and 12 age and IQ matched healthy controls on a word-pair associate task after a sleep and a wake condition. People with TEA showed the same sleep benefit for memory as healthy controls and demonstrated ALF when learning was followed by a day of wakefulness. Their results support the findings from a pilot study by Deak and colleagues (2011). As more rapid forgetting occurred only when learning was followed by a period of wakefulness and not sleep in both these studies, this provides further support for the hypothesis that ALF is caused by impairments to consolidation processes that are not sleep-dependent. Instead, ALF may be due to an undetected encoding abnormality (Atherton et al., 2014) or represents a daytime consolidation deficit (Huijgen & Samson, 2015).

The mechanisms underlying ALF in PWE are unclear but it could result from structural or functional abnormalities in the MTL, which may be caused by the underlying structural brain pathology; seizures or subclinical epileptiform activity; AEDs or psychological mechanisms (Butler et al., 2010; Butler & Zeman, 2008; Huijgen & Samson, 2015).

In line with the standard theory of consolidation (and MTT) (Alvarez & Squire, 1994; Nadel & Moscovitch, 1997), damage to MTL structures, in particular the hippocampus, would mean that memories would be vulnerable to disruption and would predict the observed pattern of abnormal forgetting found in ALF (Mayes et al., 2003). In support of this, ALF was found to be greater for verbal memory in patients with left unilateral hippocampal sclerosis than patients with a normal hippocampus (Narayanan et al., 2012). However, Miller and colleagues (2015) in their small sample did not find that a hippocampal lesion was necessary for ALF. Similarly, Butler, Kapur, Zeman, Weller and Connelly (2012) found no relationship between hippocampal volume and ALF. Lah and colleagues (2014) found that those with and without hippocampal abnormalities demonstrated ALF on a verbal memory task but this became apparent at different lengths of delay. Those with a hippocampal abnormality, usually MTS, demonstrated ALF after 24 hours, whereas those without had ALF at seven days post-learning. This is consistent with the theory that memories become less dependent on the hippocampus over time.

Ricci and colleagues (2015b) compared memory for recent autobiographical experiences in people with TLE with hippocampal lesions (e.g. MTS, temporal lobectomy, dysembryoplastic neuroepithelial tumour); those without hippocampal lesions; extratemporal lobe epilepsy and age and IQ matched controls. Only those with hippocampal lesions demonstrated ALF over the first 24 hours suggesting that the hippocampus is important for consolidation over this period. Only those with extratemporal lobe epilepsy demonstrated ALF for recognition of the autobiographical experience after four days. This supports the standard theory of consolidation (Alvarez & Squire, 1994) (i.e. diffuse cortical areas are involved after a longer delay) and the literature proposing extrahippocampal involvement in recognition tasks (e.g. Huijgen & Samson, 2015). Poorer recognition after four days was also associated with occurrence of seizures and presence of epileptiform discharges during the delay interval, as measured using ambulatory electroencephalogram (EEG) monitoring, suggesting that (subclinical) epileptic activity could also contribute to the consolidation deficit by disrupting transfer between the hippocampus and neocortex (Fitzgerald, Thayer, Mohamed & Miller, 2013b; Ricci et al., 2015b).

Consistent with this, Evans, Elliott, Reynders and Isaac (2014) have investigated the contribution of seizures by investigating whether there was a difference in ALF in seven patients before and after amygdalo-hippocampectomy. The patients exhibited ALF after one week compared to 25 healthy controls pre-surgery but not post-surgery (when they had achieved seizure freedom); although it was harder to match their initial learning post-surgery, presumably due to the effects of the resection. The authors conclude that seizure activity may be playing a role. Their findings are consistent with a single case report (Gallassi et al., 2011) and some group studies, which found that seizure frequency not hippocampal pathology was associated with ALF (Wilkinson et al., 2012).

There is a large literature investigating the detrimental impact of AEDs on cognitive functioning (e.g. Aldenkamp, de Krom & Reijs, 2003; Loring & Meador, 2001; Mula & Trimble, 2009). However, fewer studies have investigated the role of

AEDs in ALF. One study has suggested that patients with high AED serum levels are at higher risk of accelerated forgetting but they only tested patients after 30 minutes and this was a retrospective study (Jokeit, Krämer & Ebner, 2005). AED polytherapy has been associated with ALF at 'early' retention intervals (Cassel et al., 2016); although other studies have failed to find this association (Muhlert et al., 2011). ALF has been demonstrated before a patient with TEA commenced AED treatment (Jansari, Davis, McGibbon, Firminger & Kapur, 2010) and patients with TEA often demonstrate improvements in memory after starting AED treatment (Butler & Zeman, 2008; O'Connor et al., 1997) suggesting AEDs may have a limited causal role in ALF. Similarly, despite previous research suggesting that mood disorders can impact on cognitive functioning (e.g. Mula, Trimble & Sander, 2003; Paradiso, Hermann, Blumer, Davies & Robinson, 2001), there is little evidence that lowered mood leads to ALF (Butler & Zeman, 2008; Fitzgerald et al., 2013b; Witt, Glöckner & Helmstaedter, 2012).

There is an interesting recent case report that highlights that a pattern of memory performance consistent with ALF may occur without any structural brain pathology. Burgess and Chadalavada (2016) described a 38 year old male, WO, who presented with anterograde amnesia following a routine anaesthetic injection and root canal dental procedure. WO was able to learn new material but this was rapidly forgotten if it was not recalled or rehearsed within 90 minutes. Initially, a psychological explanation for WO's memory difficulties was hypothesised due to the absence of any known brain pathology and EEG abnormalities. However, the authors suggested that he may have a breakdown in a biological mechanism that is responsible for protein synthesis affecting late consolidation processes. WO demonstrated ALF after very short intervals and on nondeclarative memory tasks, which has not been found in TLE patients, so his difficulties may represent a different form of amnesic syndrome. Nevertheless, identification of potential biological mechanisms is an interesting avenue for further research of ALF in PWE.

In summary, it is difficult to identify the underlying mechanisms of ALF. It may involve multiple mechanisms, which may operate at different times and may depend on the material to be remembered (Cassel et al., 2016; Hoefeijzers et al., 2013; Ricci et al., 2015b). Alternatively, the mixed findings may reflect differences in methodology and the methodological limitations of studies (Cassel et al., 2016;

Elliott et al., 2014; Fitzgerald et al., 2013a; MacDonald et al., 2006). Elliott and colleagues (2014) in their critical review of ALF studies identified three main methodological issues: selection of appropriate test material and procedures; selection of appropriate control participants, and degree of initial learning and forgetting.

Measurement of ALF

Matching initial learning is thought to be important when assessing ALF, as degree of initial learning is related to the rate of forgetting (Elliott et al., 2014; Isaac & Mayes, 1999). Higher levels of learning are thought to lead to slower rates of forgetting. Therefore, if initial learning is not equated, differences in forgetting curves between patients and controls may reflect differential levels of learning and ALF may instead reflect an acquisition deficit rather than an early consolidation deficit (Cassel et al., 2016; Elliott et al., 2014; Wilkinson et al., 2012). Learning can be equated through extending exposure times, using a multiple presentation procedure or learning to criterion (Elliott et al., 2014). Equating initial learning by utilising a multiple presentation procedure may lead to 'overlearning' which may mask early forgetting (Elliott et al., 2014). Therefore, learning to criterion has been considered a better alternative (MacDonald et al., 2006). This involves individuals learning a task until they reach a specified criterion but the level of criterion has varied between studies

In addition, there is variation in the measurement of ALF (Cassel et al., 2016; Elliott et al., 2014; Fitzgerald et al., 2013a). A recent review has summarised the number of delays used, the timing of the delay and when ALF was observed across 37 ALF studies (Cassel et al., 2016). Some studies have assessed individuals after one long delay (e.g. Blake et al., 2000) and other studies have employed multiple delays (e.g. Hoefeijzers et al., 2013). Fifty-one per cent of studies have used three delays; 24% five delays; 11% two delays and one study tested recall after six delays. After the 'standard delay' (i.e. 20-30 mins), retention intervals have ranged from one hour (e.g. Wilkinson et al., 2012) to 8-weeks (e.g. Blake et al., 2000) with most occurring at 24 hours, 1-week and 3-weeks post-learning (see Figure 2). ALF has most frequently been observed after one week but this may reflect the methodology employed rather than this being when ALF becomes apparent (Cassel et al., 2016).

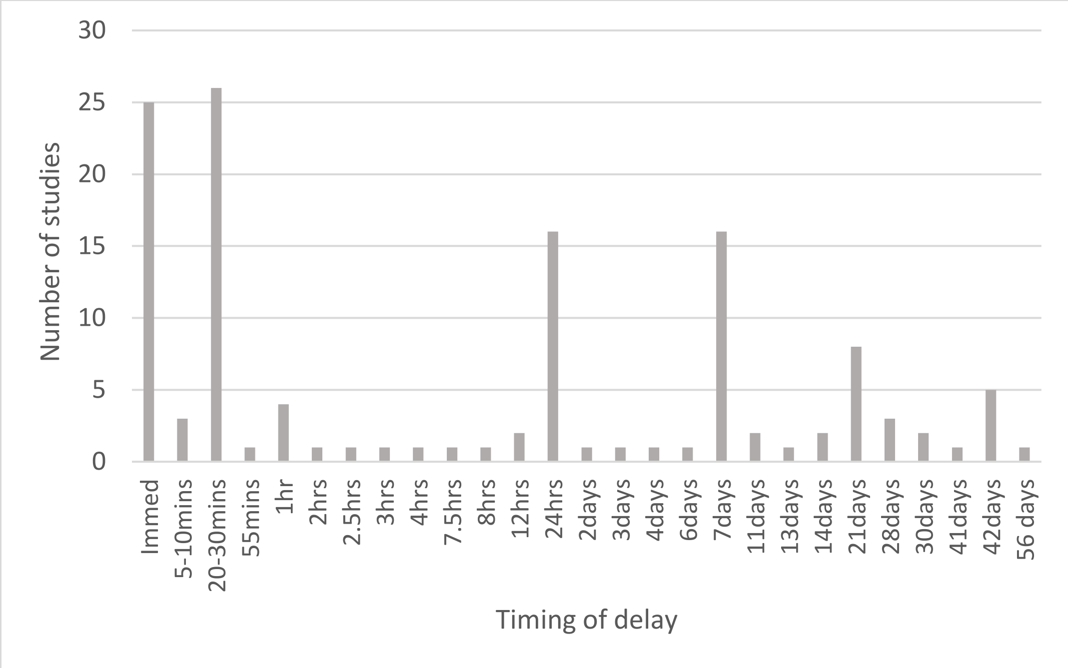


Figure 2: Variation in the timings of delays used in ALF studies (data taken from Cassel et al., 2016).

Similarly, there is no consensus on the measures used, which makes meta-analysis and comparisons between studies challenging (Elliott et al., 2014; Fitzgerald et al., 2013a). In assessing ALF, researchers have either developed their own materials or adapted existing standardised memory tests [e.g. Wechsler Memory Scale (WMS), Rey Auditory Verbal Learning Test, Selective Reminding Test, Rey-Osterreith Complex Figure Recall Test, Graham-Kendall Memory for Designs Test], which is problematic for several reasons. Firstly, there may not be parallel versions of the retrieval questions, which may interact with memory retention (Baddeley et al., 2013; Elliott et al., 2014). Secondly, there is a lack of normative data to determine 'abnormal' forgetting (Zeman et al., 2013). Thirdly, some studies have assessed recognition rather than recall, which is less sensitive to forgetting (MacDonald et al., 2006) and may explain why other clinical populations characterised by temporal lobe damage, such as Alzheimer's Disease (AD), have not consistently demonstrated ALF, which would be predicted based on their underlying pathology (Guerts, van der Werf & Kessels, 2015).

Based on their critical review, Elliott and colleagues (2014) proposed seven recommendations that future ALF studies should incorporate and could be used to guide the development of new tests to assess ALF (see Table 2). These recommendations can be used as a basis to evaluate the methodology of studies in ALF and the extent to which novel tests consider these issues.

Table 2: Recommendations for studies in ALF (Elliott et al., 2014)

Recommendation	
1	Groups are matched for age and IQ
2	Both verbal and nonverbal tests are used
3	Tests include recall and recognition paradigms
4	Distractor tasks are used to ensure that information is retrieved from LTM
5	Initial learning is equated (e.g. by extending exposure times, using a multiple presentation procedure, learning to criterion)
6	Avoid ceiling and floor effects (e.g. by experimental manipulations)
7	Minimise opportunities for rehearsal of test material

Whilst these recommendations are important for research studies, little attention has been paid to assessing ALF in clinical practice, which brings additional considerations. Following up patients face-to-face may be less feasible for clinicians when there are resource implications to follow-ups and patients may already travel long distances to attend specialist epilepsy centres. There is a need, therefore, to develop clinically acceptable and feasible measures for use in clinical practice (Zeman et al., 2013).

One potential option to avoid additional clinic visits is the use of telephone administration. Previous studies have found comparable performance in a variety of populations between in-person and telephone administration of neuropsychological assessments (e.g. Baddeley et al., 2013; Mitsis et al., 2009; Taichman et al., 2005).

The Crimes Test

The Crimes Test (Baddeley et al., 2013) is a recently developed test of verbal ALF. It is a constrained prose recall task that uses a matrix structure, instead of a narrative structure. This means that each story is organised to contain a standard set of features, which are related to each other so that the relationship between the features can be tested. Constrained prose recall tasks were originally developed as a way of measuring the capacity of the episodic buffer in working memory (Baddeley, 2003; Baddeley, Hitch & Allen, 2009). Within the Crimes Test, each story involves a standard set of features, bound together into an episode (Baddeley et al., 2013). The advantages of a matrix-based approach are that a large number of probe recall questions can be generated, which can be divided across testing sessions to reduce the possibility that repeated retrieval at each delay will interfere with memory for the stories (Baddeley et al., 2013; Jansari et al., 2010). It also ‘standardises’ the material in a story. Stories are often difficult to score and are usually based on counting the number of ‘story units’ recalled but the creation of these units is subjective and individuals may vary in their performance, as they may give different weightings to different levels of detail within the story (Baddeley et al., 2013).

The Crimes Test involves people listening to four stories of different crimes, with each crime involving a location, victim and criminal. Participants are then asked 20 cued-recall questions about each crime immediately and after a delay (e.g. ‘what was the location of the hit and run?’). At subsequent delays, different questions can be asked, which test recall for the same features but in a different direction (e.g. ‘what was the crime committed on the bridge?’). The Crimes Test has been piloted for use in young and older healthy adults, with delayed recall tests being carried out by telephone. It is a promising measure of ALF that did not have floor or ceiling effects and telephone follow-ups were considered acceptable by participants (Baddeley et al., 2013). This test has detected ALF after 24 hours in older adults with TLE (Drane, 2012) and is currently being piloted for clinical use in several epilepsy centres in the UK.

Nonverbal Tests of ALF

ALF for nonverbal memory has not been tested as frequently (Fitzgerald et al., 2013a), which may reflect that there are fewer tests of nonverbal memory functioning even at standard delays (Bengner & Malina, 2010).

Muhlert and colleagues (2011) have developed visual scenes, using the Family Pictures sub-test of the WMS-III as their framework, to assess nonverbal ALF. These scenes have been further developed by Elliott (2010) and used by Evans and colleagues (2014). In their test, participants were shown four scenes containing five objects located in each corner of the scene (e.g. a crocodile, bucket and spade, parasol, boat and a golfer presented within the beach scene). Each object was presented individually for three seconds and then the entire scene was shown for 10 seconds. Recall of each individual object and spatial location within the scene was tested at each delay. Recognition of each individual object was tested using a forced choice recognition paradigm.

However, there are potential difficulties with their visual scenes. Firstly, individuals may potentially respond using schematic representations held in LTM (e.g. the inclusion of a bus and a bike in the ‘street’ scene or bucket and spade in the ‘beach’ scene). Secondly, the pictures are not integrated into a coherent scene, which reduces ecological validity. Thirdly, the same questions are asked at each delay potentially leading to better retention, masking forgetting (Karpike & Roediger, 2007). Finally, as the objects within the scene are presented individually before being seen in the entire scene, this may tap into verbal rather than visual encoding.

It has been difficult to create ‘pure’ measures of nonverbal/visual memory, as many tasks can be verbally encoded and recalled so are less sensitive to detecting deficits in people with right MTL lesions (Brown, Roth, Saykin & Beverly-Gibson, 2007; Baddeley, Hitch, Quinlan, Bowes & Stone, 2015; Chapin, Busch, Naugle & Najm, 2009; Narayanan et al., 2012). Those that rely less on verbal encoding and retrieval, however, suffer from additional practical problems, which would make the feasibility of longer-term follow-ups more difficult (Bengner & Malina, 2010). For example, Barkas and colleagues (2012) have designed a virtual Morris water maze task to assess ALF of spatial learning, which participants navigate using a computer. The test was sensitive to non-dominant hippocampal damage but participants had to return to the laboratory to complete the test. Similarly, a measure of memory for faces has been developed (The Alsterdorfer Faces Test), which is thought to be lateralised to the right temporal lobe and offers parallel versions, but this currently can only be completed in the presence of a researcher/clinician (Bengner & Malina, 2010). To try and overcome this problem, a recent study used the Aggie Figures (Majdan, Sziklas & Jones-Gotman, 1996), which involves participants learning a set of 15 abstract line drawings, as a measure of nonverbal ALF (Miller et al., 2015). Participants were contacted by telephone and instructed to draw as many of the figures that they could remember and were asked to post back their drawings. Participants were only followed-up once after seven days. Based on response rates to postal questionnaires, it could be speculated that an increased number of follow-ups would be associated with increased attrition. Additionally, this test was also found not to be sensitive to right-sided lesions and it only detected ALF in two patients with TLE but five healthy controls. This suggests that the task is not sensitive as a measure of longer-term forgetting and/or their normative sample was not representative of the healthy population.

Given the recommendation by Elliott et al. (2014) to include verbal and nonverbal measures of ALF and the potential problems with existing nonverbal tasks, there is a need, therefore, for the development of a *clinically* feasible measure of nonverbal ALF as an analogue to the Crimes Test.

Clinical Relevance

The development of new measures to detect ALF is an important area to explore because PWE, particularly those with TLE, commonly report memory problems (Baker et al., 1997; Giovagnoli et al., 1997; Thompson & Corcoran, 1992). As individuals may be experiencing an abnormal rate of forgetting from long-term memory beyond delays of 20-30 minutes, which are typically assessed in traditional memory assessments, these difficulties may not be detected using current standardised memory measures (Butler et al., 2010; Manes, Serrano, Calcagno, Cardozo & Hodges, 2008). These difficulties may be having a significant impact on their everyday life, psychological well-being and overall quality of life (Fisher et al., 2000; Hall, Isaac & Harris, 2009). Therefore, there is a need to develop valid procedures that can detect this form of impairment in clinical practice (Elliott et al., 2014; Ladowsky-Brooks, 2015), whilst reducing the need for patients to attend additional clinic appointments, which is associated with financial and time implications (Zeman et al., 2013).

Detecting strengths and weakness in verbal and visual memory is important in informing appropriate rehabilitation and support, for example, the encouragement of rehearsal (Blake et al., 2000; Fitzgerald et al., 2013a; Jansari et al., 2010); use of external memory aids, such as SenseCam (e.g. Hodges, Berry, & Wood, 2011) and medication review (Lah et al., 2014).

Having a nonverbal equivalent of The Crimes Test would allow for comparisons of verbal and nonverbal memory functioning within an individual (Djordjevic & Jones-Gotman, 2012). This would also help assess whether an individual was experiencing a general memory consolidation problem or difficulties with memory for particular types of information (Elliott et al., 2014). This comparison is beyond the scope of this research project but would be a consideration in potential future test developments.

In addition, ALF has been found in people with mild cognitive impairment (MCI) (Coen, 2011; Walsh, Wilkins, Bettcher, Butler, Miller & Kramer, 2014) due to the structural and neurophysiological similarities between the memory problems observed in TLE and MCI (Holler & Trinka, 2014). As it is important to detect dementia-related cognitive decline early, and MCI is considered to be a prodrome for AD, the development of novel, clinically feasible tools may have potentially much wider clinical implications (Coen, 2011; Walsh et al., 2014).

Research Aims

Therefore, this project aims to develop a nonverbal test of ALF. This will comprise a series of visual scenes that incorporate a set of features, which allows for a large number of questions to be generated and divided into separate question sets that can be asked at different delays (i.e. similar matrix structure to the Crimes Test). This reduces the effects of repeated retrieval on memory. By administering the test over the telephone, this will reduce the need for face-to-face follow-ups. It is acknowledged that by using a telephone administration procedure the test materials may be verbally encoded and will need to be verbally recalled. Therefore, the task may measure both verbal and visual memory and ultimately may have reduced sensitivity to the side of epilepsy focus (Chapin et al., 2009). However, this reflects a pragmatic balance between creating a nonverbal test, which avoids the resource implications to both patients and clinicians of costs of attendance at follow-up clinics.

This project has three main aims:

1) To develop and pilot a nonverbal test of ALF for potential use in clinical practice with PWE.
2) To assess the psychometric properties (e.g. reliability, validity, floor and ceiling effects) of this test.
3) To evaluate the feasibility and acceptability of the test materials and procedure to PWE.

Hypotheses

Based on the above literature review, several hypotheses have been developed, which will be tested in Phase 2 of the project. It is predicted that healthy adults[1] will score less after a three week delay compared with their immediate recall on the new nonverbal test. Patient SK is expected to show a more rapid rate of forgetting at longer delays [i.e. there will be a significant main effect of time, group (healthy adults vs. SK) and group*time interaction]. Older healthy adults will demonstrate increased rates of forgetting. Performance on the new test is expected to positively correlate with subjective memory complaints and existing memory measures in order to confirm that is has validity. The methodology used to meet these aims and test these hypotheses will be discussed in the next two chapters.

[1] The term ‘healthy adults’ will be used to refer to the people without epilepsy who participated in this study. This is not meant to imply that PWE are not healthy but will be used as a way to distinguish those with and without epilepsy. The term ‘control’ was considered but as the group was not matched to the PWE in this study in a number of important demographic and clinical variables this term was not felt to be appropriate.

PHASE 1: METHOD AND RESULTS

Overview of the Chapter

This chapter will outline the research methods used to develop the Action-People-Places (APP) test, a nonverbal test of ALF, for use with PWE. As shown in Figure 3, this project consisted of three phases. Phase 1 comprised the development of the visual test materials and procedure. The intention was to pilot this initial version on a small number of healthy adults and PWE to assess for floor and ceiling effects. However, as it quickly became apparent that the initial version was too difficult, the test was modified after piloting on two healthy adults. Due to the challenges in developing a test with an acceptable difficulty level, several further modifications were needed based on participants' performance on the test and their feedback. Each revised version, six in total, was tested on small numbers of healthy adults and/or PWE. The process of this pilot work and the modifications made to the test will be described in this chapter.

Phase 2 involved administering the finalised version of the test, alongside other neuropsychological and self-report measures, to a larger number of healthy adults and a person, SK, with a confirmed diagnosis of ALF to assess its validity. Comparisons were also made with the two PWE (P020, P021) who completed the final version of the APP test as part of Phase 1. Phase 3 incorporated participant feedback to evaluate the acceptability of the proposed test. The design, participants and procedure for Phases 2 and 3 will be provided in the following chapters.

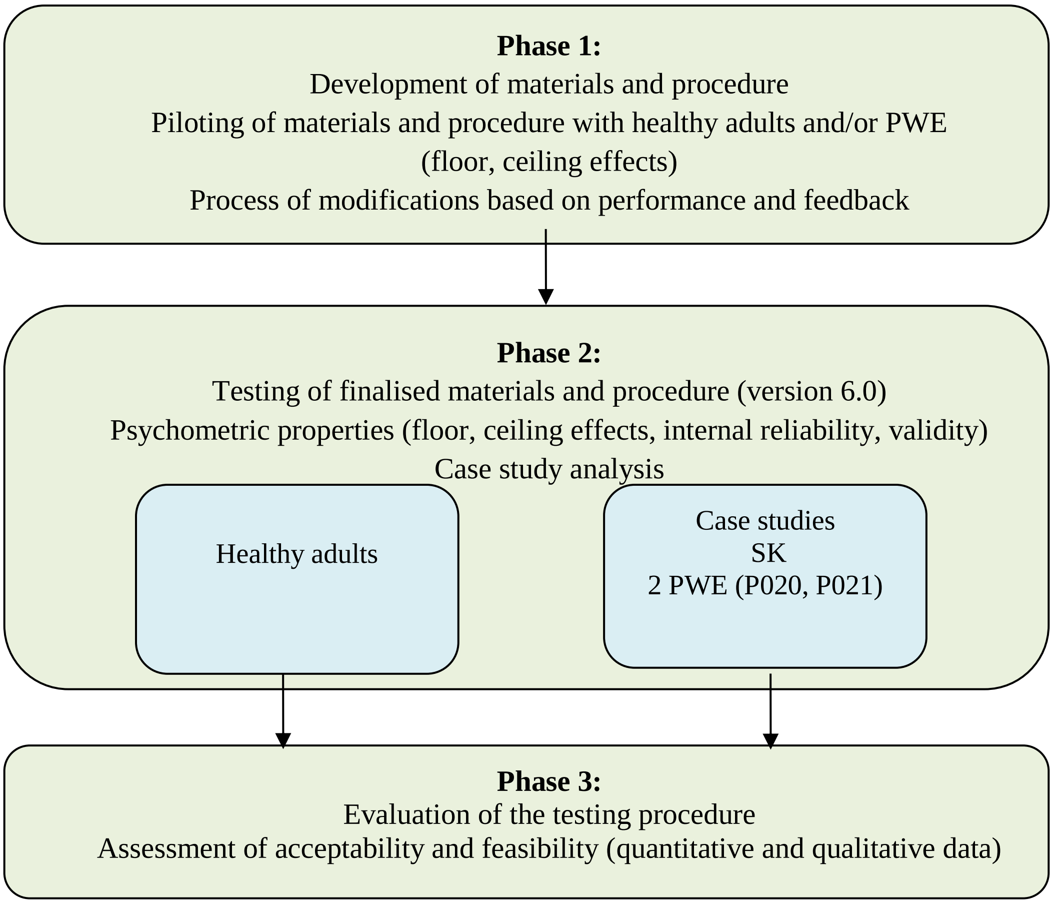


Figure 3: Phases of the project

Development of Materials and Procedure

Development of Visual Materials

The nonverbal materials initially comprised four emotionally neutral visual scenes. Each scene was constructed using a matrix structure, as used in The Crimes Test. Each scene contains a location, an actor, an action and a recipient. For example, the station scene (location) comprises a soldier (actor) pointing at (action) a nurse (recipient) (see Figure 4).

The scenes were constructed so that they were semantically plausible but the features could not easily be guessed from the other features (e.g. the bar scene would not depict an actor buying the recipient a drink). This was to reduce participants relying on semantic representations stored in LTM. The actions needed to be easily recognisable in a two-dimensional scene; emotionally neutral so as to avoid any potential participant distress and involve the characters interacting with each other so that cued-recall questions could be generated by asking either 'who did the pointing?' or 'who was pointed at?'

Figure 4: The station scene

In order to make it easier to refer to the characters during testing, the actors and recipients were referred to by their occupations and roles, which are identifiable by their uniform. They were also chosen so that they could plausibly be found within any scene but not highly associated with a particular scene (for example, using a lollipop lady was rejected as this potentially would be most closely associated with the street scene). Similarly, careful consideration was given to selecting the gender of the characters. As there is evidence to suggest that occupational gender stereotypes are immediately incorporated into verbal representations and are difficult to suppress (e.g. Oakhill, Garnham & Reynolds, 2005), it is unclear whether characters that are congruent or incongruent with gender stereotypes would interfere with visual encoding and make the scenes more or less likely to be remembered. Therefore, some scenes had congruent gender stereotypes (e.g. male builder, female nurse) and some did not (e.g. female priest, female doctor). However, it is important to note that these were based on the research team's opinion and the congruence or incongruence of stereotypes was not formally assessed using ratings.

The scenes were created by a freelance graphic designer. Table 3 summarises the features found in each of the initial scenes. The images of the initial four scenes can be found in Appendix A.

Table 3: Summary of features found in the initial four scenes

Location	**Actor**	**Action**	**Recipient**
Station	Soldier	Pointing	Nurse
Bar	Doctor	Following	Builder
Street	Fireman	Shouting	Priest
Park	Police officer	Waving	Air hostess

Development of Cued Recall Questions

As in The Crimes Test, memory performance on the APP test is assessed by cued-recall (Baddeley et al., 2013). Questions are constructed so that one feature cues another feature, for example, the action cues retrieval of the actor in ‘Who was pointing?’ or actor cues the action in ‘what was the soldier doing?’ This allows a large number of questions to be generated that test memory for the features from different directions. This means several question sets can be compiled so different questions are asked at each time point, therefore, reducing the effects of repeated retrieval and potential re-encoding (Elliott et al., 2014; Jansari et al., 2010).

When cueing using only the four features (i.e. location, actor, recipient, and action), only 12 questions could be generated for each scene generating a total of 48 questions across all scenes. As the Crimes Test asks 20 questions at each delay, further questions needed to be generated in order to keep the tasks comparable. Hair colour of the actor and recipient were incorporated as additional features in order to generate 20 questions per delay. This enabled 30 questions to be generated for each scene (120 in total), which were divided into five question sets (immediate recall, standard delay recall, 24 hour, 1-week, 3-weeks). Hair colour was selected as an additional cueing feature as it could be varied across the eight characters relatively easily and taps into participant’s memory for perceptual details within the scene. The 20 cued recall questions at each testing occasion were balanced across scenes (i.e. five questions from each scene) and type of feature cued.

Development of Test Procedure

The images and instructions were presented using PowerPoint either on an iPad or laptop. The participant viewed the instructions whilst they were read aloud by the researcher:

"You are going to be shown four images of different scenes. Each scene will contain somebody who we will call an 'actor', who will be performing an action to another person who we will call a 'recipient'. You will see each scene for 3 seconds.

For each scene, try to remember, where the scene is taking place. Who the actor is and what they look like. Who the recipient is and what they look like and what action is being performed. Do you have any questions? Let's start the test…"

In the initial version of the task, participants viewed each image for three seconds, with a one second fixation cross between scenes. Immediately after viewing the scenes, they were asked the 20 cued-recall questions. After 20-30 minutes, they were asked a different set of 20 questions. During the delay, the participants were asked not to rehearse the material and the researcher engaged them in conversation or they were given a break and returned after the specified amount of time.

Ethical Approval

Ethical approval for Phase 1 of the study was granted by the University of Leeds (UoL) School of Psychology Research Ethics Committee (14-0214). Amendments were submitted and approved to recruit PWE from local Epilepsy Action (EA) groups into Phase 1 (15-0612) and for the recruitment of PWE from EA into Phase 2 and 3 (16-0064).

Results: Initial Piloting and Modifications

Participants

In total, 12 healthy adults recruited by convenience sampling, were involved in Phase 1. They were largely recruited from friends, family and colleagues of the author. Nine PWE recruited from local EA groups and through advertisements placed on EA's website and social media were also involved in this phase. The number of participants for this phase of the study was not decided in advance but evolved with the iterative process of modifying the test. As several modifications needed to be made to the test due to its initial difficulty level, not all healthy adults and PWE completed each revised version. Each version was piloted on small numbers of healthy adults and/or PWE.

Table 4 illustrates the demographic and clinical characteristics of the 21 participants who were involved in Phase 1 and the version of the test they piloted. Of the healthy adults, 50% were male; they had an average age of 36.67 years (*SD* = 14.54), ranging from 22-61 years. They had an average of 13.25 years of education (*SD* = 2.26), ranging from 11-18 years. The majority had achieved A-levels or equivalent. Of the PWE, there were five females and four males, ranging in age from 27-70 years, with an average age of 49.56 years (*SD* = 14.69). They had an average of 15.11 years of education (*SD* = 4.26), ranging from 11-21 years and the majority had achieved a degree; although a substantial minority (44%) had no formal educational qualifications.

Table 4: Demographic characteristics of the Phase 1 participants

Test version piloted	ID	Group	Gender	Age	Yrs of education	Highest educational level
1.0	C001	HA	F	61	11	GCSEs or equivalent
	C002	HA	M	57	11	GCSEs or equivalent
2.0	C003	HA	M	26	12	NVQ Level 3
	C004	HA	F	25	13	A-levels
	C005	HA	M	29	16	Degree
3.0	C006	HA	M	34	13	NVQ
	C007	HA	M	33	12	NVQ
	C008	HA	M	61	11	GCSE or equivalent
	C009	HA	F	36	15	A-levels
4.0	C010	HA	F	32	18	Postgraduate
	P011	PWE	F	70	11	NFQ
	P012	PWE	M	68	11	NFQ
	P013	PWE	F	50	11	NFQ
5.0	P014	PWE	F	58	11	NFQ
	P015	PWE	M	41	17	Degree
	P016	PWE	F	43	21	Postgraduate
6.0	C017	HA	F	24	12	A-levels
	C018	HA	F	22	15	A-levels
	P019	PWE	M	27	20	Postgraduate
	P020	PWE	M	34	15	Degree
	P021	PWE	F	55	19	Degree

Note. F = Female; M = Male; HA = Healthy adults; PWE = People with epilepsy; GCSE = General Certificate of Secondary Education; NVQ = National Vocational Qualification; NFQ = No Formal Qualification

Procedure

Potential healthy adult participants were approached by the author and given information sheets. If they gave written informed consent to take part, they were asked to complete the APP test and recall was assessed immediately and after the standard delay of 20-30 minutes. They were then asked for detailed feedback on the task.

For PWE, the author attended a local EA branch meeting and informed the group of the research study and gave out information sheets. Interested participants left their contact details and after 24 hours, they were contacted by the researcher and a mutually convenient time was arranged for assessment. Three participants also contacted the author to participate after seeing the research study advertised on EA's social media pages. All participants in the Phase 1 pilot study were offered the opportunity to enter a prize draw for a £15 high street voucher (or equivalent donation could be made to EA on their behalf).

Due to the challenges in creating a test that was acceptable to both healthy adults and PWE, small numbers of participants (healthy adults and/or PWE) piloted different versions of the task and modifications were made at each stage based on their performance and feedback. This iterative process will be described in more detail below. A summary of the key features of each version of the APP test is shown in Table 5.

Table 5: Key features of each version of the APP test

Key features	**v1.0**	**v2.0**	**v3.0**	**v4.0**	**v5.0**	**v6.0**
No of scenes	4	4	4	4	3	3
Actions	Pointing Shouting Waving Following	Pointing Shouting Waving Giving a gift	Pointing Shouting Waving Giving a gift	Pointing Shouting Waving Giving a gift	Pointing Shouting Waving	Pointing Shouting Waving
People (actor, recipient)	Soldier, Nurse Fireman, Priest Policeman, Air hostess Doctor, Builder	Soldier, Nurse Fireman, Priest Policeman, Air hostess Doctor, Builder	Soldier, Nurse Fireman, Priest Policeman, Air hostess Doctor, Builder	Soldier, Nurse Fireman, Priest Policeman, Air hostess Doctor, Builder	Soldier, Nurse Fireman, Priest Policeman, Air hostess	Soldier, Nurse Fireman, Priest Policeman, Air hostess
Places	Station Street Park Bar	Station Street Park Bar	Station Street Park Bar	Station Bar Street Park	Station Street Park	Station Street Park
No of questions (per delay)	20	20	20	20	15	15
Exposure period	3s	10s	10s	10s	10s	10s
Example item	No	No	No	No	No	Yes
Learning to criterion	No	No	Yes-60% criterion (or 3 learning trials)	Yes-60% criterion (or 3 learning trials)	Yes-80% criterion (or 3 learning trials)	Yes-80% criterion (or 3 learning trials)

Version 1.0

The first two (healthy adult) participants were tested using the materials and procedure described above. The scores and feedback for these two participants are presented in Table 6.

Table 6: Recall scores and feedback from version 1.0 (*n* = 2)

ID	Immediate recall (%)	Delayed recall (%)	Feedback/observations
C001	30	20	Some questions were complicated to understand/suggest re-phrasing Action in the bar was unclear Gave answer of 'pulling trolley' for action in park scene
C002	15	35	Time to view scenes was too short Would not have thought to look at hair colour-suggested prompting participants Guessed a lot of answers Gave answer of 'pulling trolley' for action in park scene 'Following' action not clear
Mean (*SD*)	22.50 (10.61)	27.50 (10.61)	

As the scores of both participants were low (on average only recalling 23% and 28% immediately and after a short delay), and the participants reported guessing, several changes were made to the task materials, procedure and cued-recall questions.

As both participants could not identify the 'following' action, this was changed to a more easily recognisable action of 'giving a gift'. Both participants had responded that the action in the park scene was the air hostess 'pulling a trolley' rather than the police officer 'waving', which was the intended correct response. Therefore, the air hostess' suitcase trolley was removed from this scene. Both participants reported that if this change was made they would still be able to identify her occupation from her clothing.

As participants reported that three seconds exposure to the scenes was very quick, this was increased to 10 seconds per scene, which is in line with the WMS-III Family Pictures subtest. As one participant indicated that they did not consider looking at hair colour, the instructions were modified to encourage participants to pay attention to the character's features, such as eye, hair and skin colour. Some of the problematic questions were re-phrased to make them easier to comprehend, for example, 'What colour hair did the person who was with the person with red hair have?' was changed to 'Someone had red hair. What colour hair did the other person in that scene have?'

In summary, the key changes made version 1.0 were:

- 'Following' action in the bar scene changed to 'giving a gift'
- Changes made to the air hostess character (i.e. removal of her trolley suitcase)
- Exposure period increased to 10 seconds
- Re-phrasing of some cued recall questions
- Instructions added to explicitly direct participants to pay attention to the character's features, such as eye, hair, skin colour.

Version 2.0

The scores and feedback for the three healthy adult participants who were assessed using the version 2.0 materials and procedure are shown in Table 7.

Table 7: Recall scores and feedback from version 2.0 (n = 3)

ID	**Immediate recall (%)**	**Delayed recall (%)**	**Feedback/observations**
C003	45	50	Wished could have seen again Example as part of instructions would have helped, unsure what to expect
C004	60	85	Clear questions and instructions Not too easy or too hard
C005	35	15	Difficult test Example or practice might have helped prepare self for test
Mean (*SD*)	46.67 (12.58)	50.00 (35.00)	

On average, participants scored 47% immediately and 50% after a 20-30 minute delay. However, there was variability in the participant's performance, ranging from 35-60% immediately and 15-85% after a standard delay. Participant C003 said that despite understanding the instructions, they had forgotten to pay attention to the action in each scene, which affected their responses to the cued-recall questions. Participant C005 reported finding the test very difficult.

Given the variable performance and the low scores obtained by two of the participants, further modifications to the test were necessary. There was no negative feedback given about the scenes. Participants correctly identified all the characters, action and locations in their answers, therefore, no modifications were made to the materials.

All participants reported that the instructions were clear but two said that they would have appreciated an example/practice item to help orient them to the task and manage expectations. However, as this would have required development and construction of a new location, action and characters that met the criteria and would have required further funds to pay the designer, an alternative option was considered at this stage. As learning to criterion is one potential way to match initial learning, particularly when comparing healthy controls with clinical groups (Elliott et al., 2014), the next stage was to pilot a learning to criterion procedure. As discussed in the introductory chapter, learning to criterion involves presenting the material until a pre-defined score is reached.

A criterion of 60% was chosen, as high criterion levels (e.g. 100% on two successive trials) have been associated with ‘overlearning’, which can mask ALF (Elliott, 2010; Elliott et al., 2014). In the subsequent versions of the task, participants were shown the scenes until they scored 12/20 (60%) or completed three initial learning trials, whichever was the sooner, to minimise any potential participant distress.

The key changes made to version 2.0:

- Introduction of 60% learning to criterion

Version 3.0

Four healthy adult participants were assessed using version 3.0 materials and procedure. Their scores and feedback are presented in Table 8.

Table 8: Recall scores and feedback from version 3.0 (*n* = 4)

ID	**Trial I (%)**	**Trial II (%)**	**Delayed recall (%)**	**Feedback/observations**
C006	35	65	55	Practice would be helpful Spent time focusing on small details in first trial
C007	35	85	85	First trial very difficult Confusion re female priest
C008	20	60	45	Worked in a hospital so felt the nurse could be a doctor due to her uniform Too many questions about hair colour
C009	30	65	60	Example would be helpful as easier the second time Focused on eye colour and no questions asked
Mean (*SD*)	30.00 (7.07)	68.75 (11.09)	61.25 (17.02)	

After the first trial, participant's scores remained low ranging between 20-35%. However, all participants reached learning to criterion after the second trial and did not require a third learning trial. The average recall score after two trials was 68.75%. After a 20-30 minute delay, recall performance was still variable, ranging from 45%-85%, but the percentage retention of material (i.e. delayed recall relative to immediate recall) was better (ranging from 75-100%).

Only minor modifications to the instructions were made at this stage. In order to provide reassurance when people were finding the task difficult, participants were explicitly told that 'You are not expected to get everything right'. As some participants had reported that they had paid attention to irrelevant details (e.g. counting the number of bottles in the bar scene), they were instructed to pay attention 'in general to where the scene is taking place'. The focus was also removed from directing participants to pay attention to hair, eye and skin colour to just hair colour. After these modifications, it was felt that it would be useful to pilot it on a small number of PWE.

In summary, the changes made to version 3.0 were:

- Modifications to instructions- before answering questions, participants explicitly told that they are not expected to get everything right to provide reassurance.
- Participants directed to pay attention to the general features of where the scene is taking place (i.e. what kind of location it is).
- Participants directed to pay attention to actor/recipients hair colour rather than eye, skin and hair colour.

Version 4.0

One healthy adult (C010) and three PWE piloted version 4.0.

Table 9: Recall scores and feedback from version 4.0 (*n* = 4)

ID	**Trial I (%)**	**Trial II (%)**	**Trial III (%)**	**Delayed recall (%)**	**Feedback/observations**
C010	45	85	NA- Reached criterion	55	Difficult task but easier on second trial Hair colour and style needs to be more distinctive
P011	30	40	35	40	Clear instructions Questions were easy to understand
P012	25	25	45	40	Unsure about identity of some of the characters from their clothing (e.g. priest, air hostess). Bar scene does not make real-world sense
P013	10	15	20	30	Not stressful Useful to have the instruction presented as read aloud- helps support people with memory problems Air hostess was not clear
Mean (*SD*)	27.50 (14.43)	41.25 (30.92)	33.33 (12.58)	41.25 (10.31)	

Note. NA = Not administered

As shown in Table 9, the PWE found the APP test difficult. None met the 60% learning to criterion after the third trial. After a 20-30 minute delay, all scored less than 50%.

Due to their poorer performance relative to the healthy adults and the importance of equating initial learning, further modifications were made to the task. Two of the participants had reported finding it difficult to identify the characters from their clothing. As it would not have been feasible for the graphic designer to alter the characters at this stage, no further modifications were made to the characters. However, participants would be scored as answering correctly, if they could demonstrate that they had a visual representation of the scene e.g. if they said ‘the lady in the red suit’ rather than explicitly saying ‘air stewardess’.

As none of the participants with epilepsy had reached the 60% learning to criterion, the number of scenes presented as part of the task was reduced to three scenes to try and reduce the initial memory load. The bar scene was chosen as the scene to be removed from the task for several reasons: i) it was the only scene that introduced an object (a gift) into the action ii) it involved a doctor, which some participants had confused with the nurse iii) some participants had reported being distracted by additional features in the scene (e.g. bottles in the background).

As there were now only three scenes, the questions at each delay were reduced to 15 and were balanced to include five questions per scene. The remaining questions across the other delays were checked to ensure that they were still balanced across the types of questions asked and features cued.

Even though PWE were struggling to reach the 60% learning to criterion, it was decided to increase the learning to criterion to 80% to make the task consistent with other tests and ALF studies (e.g. Kemp et al., 2012; Ricci et al., 2015a; Wilkinson et al., 2012). However, due to the potential poorer learning in the epilepsy group, a decision was made not to exclude from this study those who may not reach criterion from the telephone follow-ups.

In summary, the key changes made version 4.0 were:

- Test reduced to three scenes.
- Questions reduced to 15 at each delay.

- Increased learning to criterion to 80% consistent with other studies in the ALF literature.

Version 5.0

Three PWE were assessed using version 5.0. Their scores are shown in Table 10. As this version has a maximum score of 15, percentage recall scores are rounded to the nearest whole number. Despite the reduction in material presented, PWE were still struggling with learning the three scenes. None of the participants reached the 80% learning to criterion and only one participant (P015) would have reached the previous 60% learning to criterion on the third trial. After a standard delay, performance across all three participants was low, ranging from 20-53%. Based on previous feedback by several previous participants, an example item (the bar scene) was introduced into the test procedure.

Table 10: Recall scores and feedback from version 5.0 (*n* = 3)

ID	**Trial I (%)**	**Trial II (%)**	**Trial III (%)**	**Delayed recall (%)**	**Feedback/observations**
P014	7	20	27	20	Not stressful Instructions made sense
P015	33	60	67	53	Found being asked about finer details, such as hair colour, more difficult Priest and air stewardess more difficult to identify
P016	20	53	27	40	Expected a video due to reference to ‘actions’ and ‘actors’ Practice/example may have helped manage expectations
Mean (*SD*)	20.00 (13.00)	44.33 (21.36)	40.33 (23.09)	37.67 (16.62)	

Key changes made to version 5.0:

- Introduction of the bar scene as an example

Version 6.0

In order to ensure that the introduction of an example did not lead to ceiling effects in healthy adult participants, the task was piloted on a further two healthy adults and an additional person with epilepsy (P019). As shown in Table 11, one of the healthy adults (C018) performed at ceiling at the second trial and after a standard delay. The person with epilepsy (P019) scored highly on this version, scoring 93% after a third trial and after a 20 minute delay. Participants continued to feedback concerns about the distinctiveness of the characters, in particular the priest and air hostess, but due to limited financial and time resources, changes to these characters could not be made at this stage. However, participants were not penalised for not providing the correct occupational identity and providing visual details instead (e.g. lady in the green cloak, woman in business dress) so this was felt not to affect their recall performance.

Table 11: Recall scores and feedback from version 6.0 (*n* = 3)

ID	**Trial I (%)**	**Trial II (%)**	**Trial III (%)**	**Delayed recall (%)**	**Feedback/observations**
C017	47	80	NA- Reached criterion	67	Example was helpful to know what going to be asked Depending on religious background, priest may be unclear
C018	67	100	NA- Reached criterion	100	Practice item was useful, as expected to see video from instructions Actions were clear Learning trials enabled to focus on hair colour 2nd time of viewing scenes
P019	53	47	93	93	Air stewardess identity unclear Practice item helpful Waving may be too subtle to identify
Mean (*SD*)	55.67 (10.26)	75.67 (26.76)	-	86.67 (17.39)	

Note. NA = Not administered

Over the various versions of the task, the mean performance of the small numbers of healthy adults and/or PWE who completed each version, had improved in line with modifications (see Figure 5). Therefore, version 6.0 was considered to be the finalised version of the task for use in Phase 2 of the project.

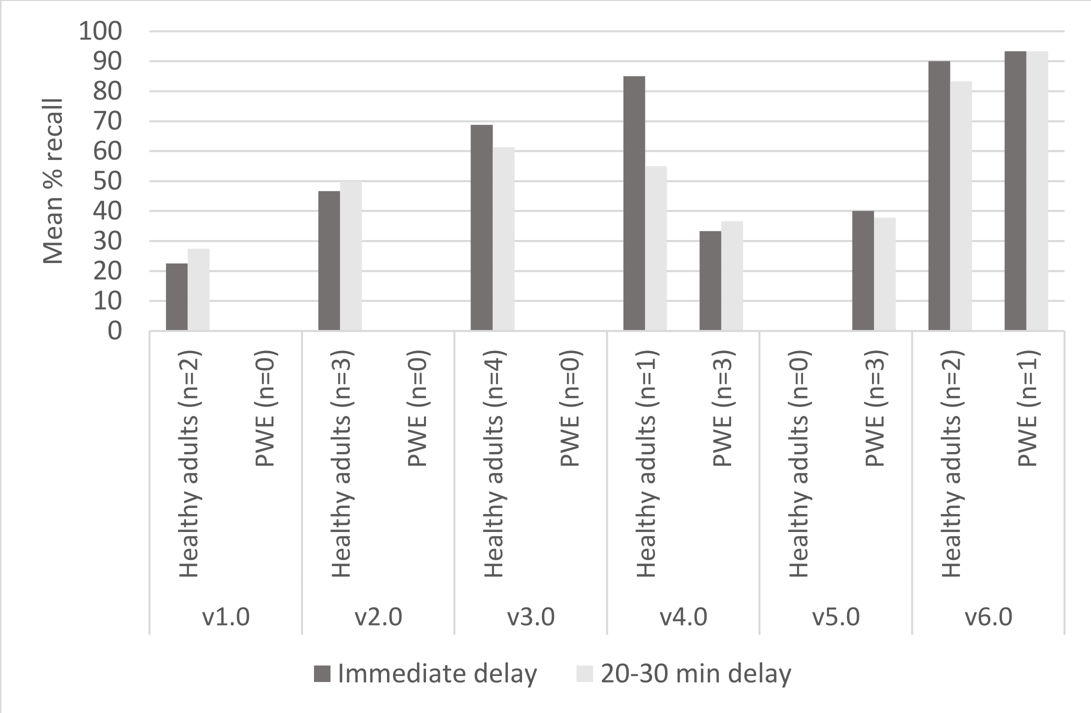


Figure 5: Summary of mean performance over modified versions of the APP test

Final APP Test Procedure

The finalised testing procedure is illustrated in Figure 6 and the final three scenes can be found in Appendix B. The APP test is administered using PowerPoint. The participant views the instructions whilst they are read aloud by the researcher:

"You are going to be shown three images of different scenes. Each scene will contain somebody who we will call an 'actor', who will be performing an action to another person who we will call a 'recipient'. You will see each scene for 10 seconds.

For each scene, try to remember, where the scene is taking place. In general, the kind of location it is. Who the actor is and what they look like. You might want to pay attention to features such as their hair colour. Who the recipient is and what they look like, again paying attention to features such as their hair colour. And what action is being performed.

You will see an example scene first. This is the bar scene. This is the doctor. We call her the 'actor' because she is giving a gift to the builder. We call him the 'recipient' in this scene. So once you've seen all the scenes, I'll ask you a series of questions, for example:

'Who was the recipient in the bar?' 'You would say the builder';

'Who was giving a gift to somebody?' 'You would say the doctor'

'What colour was the doctor's hair?' 'You would say blonde'.

Do you have any questions? Let's start the test…"

Participants view each scene for 10 seconds, with a one second fixation cross between scenes. Immediately after viewing the scenes, they are asked 15 cued-recall questions. If the participant scores less than 12 (i.e. < 80%), they are shown the scenes again and asked the same cued-recall questions. If they score less than 12 on this second learning trial, they see the scenes for a final time. After the third learning trial (or if they score 12 or more on any learning trial), they are asked a different set of 15 cued-recall questions after a 20-30 minute delay (see Appendix C). One point is given for a correct response. Points are summed to give a total recall score at each delay. Total scores range from 0-15. Percentage recall scores are rounded to the nearest whole number. Participants' recall is tested at longer delays over the telephone.

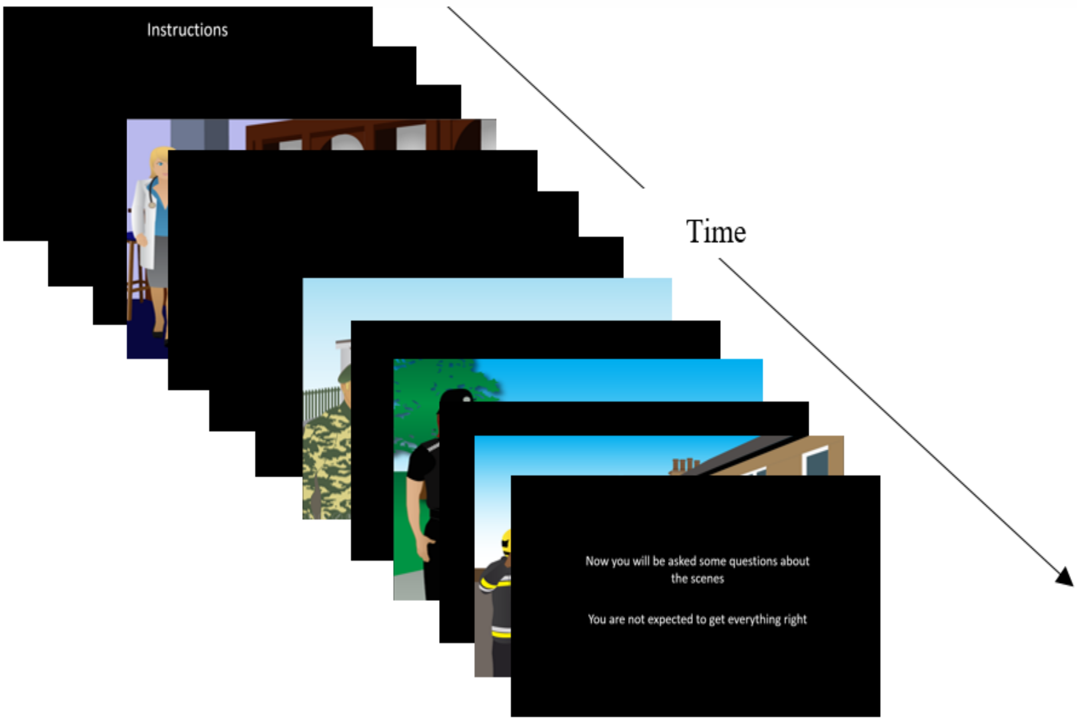


Figure 6: Finalised APP test procedure

Telephone Follow-ups

Due to the difficulties in modifying the test to avoid floor and ceiling effects over the shorter delays (immediate recall and after 30 minutes), the telephone follow-ups had not been piloted at this stage. Two PWE (P020, P021), recruited via EA's social media, completed the finalised version of the task, including the three telephone follow-ups at 24 hours, 1-week and 3 weeks post-learning. They were offered a £10 high street voucher for completing the initial assessment and taking part in the study over this longer time period. Different questions were asked at each delay (see Appendix C). To minimise rehearsal of information, participants were not informed when the follow-ups would occur. Their feedback on the task was also gathered and is presented as part of the Phase 3 results.

Piloting of Full Task

The scores of the two participants are illustrated in Table 12. The two participants both performed very differently on the task. However, this may reflect their different medical histories. Participant 021 reported that she had right TLE following a subarachnoid haemorrhage 18 years ago. She had undergone radiotherapy and stereotactic radiosurgery for an arteriovenous malformation 14 years ago and had noticed significant cognitive changes following this surgery. She reported that neuropsychological assessment had revealed significant deficits in her visual memory functioning[2]. She had never achieved seizure freedom and experiences mainly nocturnal complex partial seizures four to five times a month. In line with her self-reported visual memory difficulties, P021 found the task challenging and did not reach criterion after three learning trials. She consequently struggled to remember the information after 24 hours and over the longer delays.

However, P020, who also had TLE, met learning to criterion on the second trial. He was diagnosed with epilepsy 12 years ago and had not achieved seizure freedom, experiencing frequent complex partial seizures approximately two per fortnight. He had not undergone epilepsy surgery and was treated with a combination of AEDs. He

---

[2] As these participants had been recruited via a charity organisation and not through the NHS, access to medical records could not be obtained to confirm clinical details. However, as this phase of the study was interested in obtaining feedback on the experience of completing the task from PWE regardless of their diagnosis, this was not felt to be detrimental to the pilot findings.

retained all the information after a standard delay and 24 hours. After three weeks, he obtained a recall score of 73% and retained 92% of the information originally learnt.

Table 12: Piloting of the APP test across longer delays with two PWE

	P020		P021	
Delay	**% recall**	**% retention**	**% recall**	**% retention**
Trial I	47	-	0	-
Trial II	80	-	7	-
Trial III	NA-reached criterion	-	27	-
Standard	80	100	20	75
24 hours	80	100	0	0
1 week	87	100[ǂ]	7	25
3 week	73	92	0	0

Note. NA = Not administered. [ǂ] Percentage retention capped at 100%

Despite the variability in individual performance and the floor effects observed in Participant 021, which could be understood in light of her reported neurological history, no further modifications were felt to be necessary and the task was ready to be piloted on a larger number of healthy adults and an individual, SK, with confirmed ALF. Given that these two participants completed the finalised version of the test at each delay, their data will also be presented as part of Phase 2 and compared with the healthy adult data and that of SK.

Summary of the Chapter

Phase 1 of the study aimed to develop the materials and test procedure of the APP test, a nonverbal test of ALF, for use with PWE. The test was constructed using a matrix structure so that each scene contains a set of features (action; location; actor; recipient; actor and recipient hair colour), which enables a large number of cued-recall questions to be generated that test memory for the features from other features within the scene.

A total of 21 participants (12 healthy adults and 9 PWE) were involved in Phase 1 of this study. Due to the challenges in creating materials that avoided floor effects, the test needed to be continually refined with changes made to the number of scenes; phrasing of the questions; exposure period; the inclusion of a learning criterion and an example item. Six versions of the APP test were created and piloted with small numbers of participants (healthy adults and/or PWE) piloting the different versions of the task and the modifications were made based on their test performance and feedback. These iterative modifications over the six versions appeared to increase participants' recall performance helping to avoid floor effects, at least in the healthy adult participants.

The finalised APP test (version 6.0) contained three emotionally neutral scenes, which participants view for 10 seconds each. Participants are asked cued-recall questions immediately after viewing all three scenes. A learning criterion is set at 80% (i.e. 12 of 15) or three learning trials, whichever is the sooner. Participants are tested using a different set of questions after 20-30 minutes (i.e. standard delay), 24 hours, 1-week and 3-weeks. Two PWE piloted the telephone follow-ups at these longer delays.

The finalised APP test was then further tested in Phase 2 with a larger group of healthy adults and a person, SK, with a confirmed diagnosis of ALF. The performance of the two PWE (P020, P021) who piloted version 6.0, as part of Phase 1, were also compared to the healthy adult group and SK to provide additional support for its validity. The next chapter presents the methodology and results of this second phase of the study.

PHASE 2: METHOD AND RESULTS

Overview of the Chapter

This chapter will report on the methodology and results of Phase 2 of the study, which involves the collection of data for the APP test from a larger number of healthy adults and from SK (an individual with confirmed ALF). Comparisons will also be made between the two PWE (P020, P021) who participated in the development of the final version in Phase 1. Firstly, the participants and procedure will be described. The demographic characteristics of the healthy adult sample will be provided and their performance on the APP test will be investigated. Relationships between demographic, neuropsychological and psychological variables will be explored. Floor and ceiling effects will be examined. Finally, the performance of the three PWE (SK, P020, P021) will be compared with the healthy adult sample using a single case methodology (Corballis, 2009; Crawford & Howell, 1998).

Method

Recruitment

The healthy adults were recruited by a convenience sample through friends and family of the researchers[3] (n = 16) and through the UoL School of Psychology participant mailing list (n = 16).

Participants were eligible for this study if they:

- Were aged between 18-65 years
- English was their first language
- Were able to give informed consent.

Participants were not approached to participate or excluded from the analysis if they:

- Had a previous or current neurological or psychiatric history, determined during a semi-structured interview prior to assessment

[3] Three UoL undergraduate Psychology students involved in the Research Experience Scheme for Undergraduate Students (RESUS) assisted the author with recruitment and data collection of the healthy adult group.

- Had an estimated IQ <80, as assessed using the Test of Premorbid Functioning UK (TOPF UK).

These exclusion criteria were chosen as they may have influenced the cognitive functioning of individuals in the healthy adult group and the aim was to recruit a sample of healthy people from the general population. No participants were excluded.

As the aim was to recruit normative data representative of the general population, efforts were made to recruit a wide range of individuals in terms of age, sex and educational background into the study. A screening phase was considered for those participants recruited via the School of Psychology mailing list. Interested participants were asked for some brief demographic details (sex, age and educational history), via a Bristol Online survey (https://www.onlinesurveys.ac.uk/), prior to initial assessment so that a wide range of participants could be selected to take part. Unfortunately, there was insufficient response to the advert to enable screening to take place; therefore, all interested participants were offered an initial assessment. This meant that due to the recruitment strategy the healthy adult group comprises a mainly female, more highly educated sample and is not matched in age, gender and IQ to patient SK. Therefore, caution should be made when considering the normative data and interpreting the results of the case studies.

Case Study of a Person with Confirmed ALF

To contribute to the assessment of the known-groups validity of the test, a patient, SK, with confirmed ALF was identified by the field supervisor, a Consultant Clinical Neuropsychologist. SK has previously been the subject of a case study into ALF (Kemp et al., 2012). SK underwent neuropsychological assessments prior to possible epilepsy surgery in 2005. He then underwent repeated testing in 2006, 2007 and 2009. SK's best neuropsychological test performance fell within the 13^{th}-50^{th} percentile, which was consistent with his general intellectual functioning. However, he reported rapid rates of forgetting that impact on his daily life. On tests adapted for assessment of SK's long-term forgetting rates (i.e. list learning, two stories from the Crimes Test and to scenes from the Family Pictures subtest of the WMS-III), SK demonstrated accelerated forgetting compared with 10 age, gender and IQ matched controls (Kemp et al., 2012).

Procedure

Healthy adults

Healthy adult volunteers either responded to an advertisement placed on the School of Psychology mailing list or were approached to see if they wanted to take part in the study by the researchers. All were provided with information sheets and given at least 24 hours to consider if they wanted to take part. If participants wanted to take part, a mutually convenient time and place to complete the initial assessment was arranged. Initial assessments took place at either the School of Psychology at the UoL or at participant's homes. Forty-four per cent of the healthy adults were recruited and assessed by the author and 66% by the three Psychology undergraduate students who were trained by the author to administer the assessments. The author scored all the neuropsychological assessments to ensure consistency.

Case study

SK was approached by his clinician and verbal consent was obtained for the author to contact him to explain the study. The initial assessment took place at St James's University Hospital.

Initial assessment

At the initial assessment, written informed consent was taken. For the healthy adults, a brief semi-structured interview was conducted to obtain demographic information. This included their sex, age, handedness and educational history (years of formal full-time education, highest qualification level). For SK, additional clinical information was obtained on his duration of epilepsy; age at diagnosis; current antiepileptic medication and current seizure frequency. Written informed consent was also given to access his medical records to confirm these details and provide additional information about his epilepsy and neuropsychological functioning.

In addition to the finalised version of the APP test, all participants were assessed using a brief battery of neuropsychological tests involving a measure of premorbid intellectual functioning (TOPF UK); a brief measure of cognitive status [(Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)], as well as a measure of mood [Hospital Anxiety and Depression Scale (HADS)] and everyday subjective memory complaints [Everyday Memory Questionnaire (EMQ)] to assess convergent validity.

Measures

Test of Premorbid Functioning UK (TOPF UK)

The TOPF UK (Wechsler, 2011) was administered to estimate intellectual and memory ability. The task involves participants reading aloud 70 irregularly spelled words. It takes approximately 10 minutes to complete. Wechsler Adult Intelligence Scale (WAIS-IV) [(Full Scale IQ (FSIQ), Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI) and Processing Speed Index (PSI)] and WMS-IV [(Immediate Memory Index (IMI), Delayed Memory Index (DMI) and Visual Working Memory Index (VWMI)] scores are estimated based on an individual's TOPF UK raw scores. WAIS-IV and WMS-IV scores have a mean score of 100 (*SD* = 15). The test has good reliability (Cronbach's alpha = .95) and validity (Wechsler, 2011).

Repeatable Battery for the Assessment of Neuropsychological Status (RBANS)

The RBANS (Randolph, Tierney, Mohr, & Chase, 1998) is a brief measure of cognitive function for use in adults assessing immediate and delayed memory, visuospatial/constructional, language and attention abilities. Index scores are obtained for each of those domains, with a mean of 100 (*SD* = 15). The RBANS takes approximately 30 minutes to administer. The test has good reliability and validity for use as a screening tool (Strauss, Sherman, & Spreen, 2006) and its psychometric properties have been evaluated in several clinical populations, including traumatic brain injury (McKay, Casey, Wertheimer & Fichtenberg, 2007) and multiple sclerosis (Beatty, 2004).

Everyday Memory Questionnaire (EMQ)

As ALF has been associated with subjective memory complaints, participants were asked to complete the EMQ (Sunderland, Harris, & Baddeley, 1983), which is a self-report measure, assessing the subjective frequency of memory failures in everyday life. It comprises 28-items which participants respond to using a five-point scale of frequency from 'never' to 'once or more a day'. Example items include 'forgetting where you have put something. Losing things around the house' and 'telling someone a story or joke that you have already told them once already'. It takes approximately 10 minutes to complete. Scores range from 0-112. Previous studies have considered scores of 24 and above as 'high' levels of memory problems (e.g. Das Nair et al., 2015). It has good reliability [(Cronbach's alpha .91) (Royle & Lincoln, 2008)] and has been used with PWE (Goldstein & Polkey, 1992; Ricci et al., 2015b).

Hospital Anxiety and Depression Scale (HADS)

As there is evidence that memory functioning can be affected by mood in PWE (e.g. Paradiso et al., 2001), the HADS (Zigmond & Snaith, 1983) was used to measure current mood state. It is a short 14-item tool, which was designed as an instrument to screen for clinically significant anxiety and depression. Seven items assess severity of depression and seven assess anxiety. Patients are asked to rate on a four-point scale to what extent they agree with each item. It takes approximately 10 minutes to complete. Total scores for anxiety and depression range from 0-21. It has good reliability [(Cronbach's alpha = .78-.90) (Panelli et al., 2007)] and has been used in several studies with PWE (e.g. Jacoby, Gamble, Doughty, Marson & Chadwick, 2007; Panelli et al., 2007).

Administration

The initial assessment took approximately 45-60 minutes to complete. Three versions of the APP test were created (A, B and C) with the order of scenes counter-balanced across the three versions to control for primary and recency effects of the scenes. The remaining tests were administered in a fixed order to ensure adequate time passed to test delayed recall after 20-30 minutes on the APP test. The participants completed other study-related tasks during this delay period. Whilst completing cognitively demanding tasks during a delay has been associated with reduced memory for the information (Alber et al., 2014; Craig, Della Sala & Dewar, 2016), a balance needs to be struck with the time burden placed on participants. Completion of other aspects of the study procedure during the short delay has been used in other similar studies (e.g. Dewar, Hoefeijzers, Zeman, Butler & Della Sala, 2015; Hoefeijzers et al., 2014).

The ALF telephone follow-ups were completed at 24 hours, 1-week and 3-weeks post-learning. These time points were chosen as they were the most commonly used delay intervals in ALF studies (see Figure 2). To minimise rehearsal of information, participants were told to not rehearse the material or write anything down and they were not informed when the follow-ups would occur. However, they were aware that they would have three calls over a three week period and convenient times to contact each participant was ascertained at the initial assessment to reduce the risk of loss to follow-up. As seizures may impact on cognitive functioning (e.g. Dodrill, 1986), SK was asked to complete a seizure diary (see Appendix D) over the three week period. This information on any seizures experienced was requested at each telephone interview.

Participants were also asked whether they would like to receive a lay summary of the results at the end of the study. All participants in Phase 2 were offered a £10 high street voucher (or equivalent donation could be made to EA on their behalf) for completing the initial assessment.

Ethics

Approval for the recruitment of healthy volunteers into Phase 2 and 3 was granted by the UoL School of Psychology Research Ethics Committee (14-0214). Amendments were submitted and approved (15-0394) to introduce the screening questionnaire into the recruitment of healthy adults. Approval to recruit and assess people with confirmed ALF (i.e. SK) (and other people with TLE) from NHS clinics was granted by the East Midlands-Nottingham 1 REC (15/EM/0428) and Leeds Teaching Hospitals NHS Trust R&D department (PY15/297).

Statistical Analysis

Once all the assessments were completed and scored, the data was entered onto a Statistical Package for Social Science (SPSS) database. Percentage retention scores were capped at 100%. Quantitative data was analysed using SPSS version 23.0. JASP (version 0.7.5.5; JASP Team, 2016) was used to create graphical displays for some data.

Descriptive statistics were used to assess the baseline demographic and neuropsychological characteristics of the healthy adult group. The primary analysis of forgetting on the APP test in the healthy adult group was undertaken using a one-way repeated measures analysis of variance (ANOVA). This was carried out with both percentage recall scores on the APP test and percentage retention scores (relative to the last learning trial) as the dependent variable and delay as the within subjects variable. Post hoc tests were carried out using Bonferroni tests. Effect sizes were calculated using partial eta-squared.

Floor and ceiling effects were examined using frequency counts. Internal consistency of the delays and the overall APP test was assessed using the Kuder-Richardson reliability coefficient, which is used for items with dichotomous responses (Sherman, Brooks, Iverson, Slick & Strauss, 2011). Reliability was classified as 'adequate' if internal consistency coefficient fell between .80-.99, 'modest' if it fell between .70-.80 and 'low' internal consistency if it fell below .70 (Iverson, 2001; Lance, Butts & Michels, 2006).

Secondary analyses explored relationships between recall scores and rates of forgetting on the APP test and demographic characteristic such as age, years of formal education, estimated intellectual and memory ability and scores on the EMQ and RBANS memory indices, using Pearson correlations.

Performances on the APP test by the three PWE (P020, P021, SK) were plotted in comparison with the healthy adult group and descriptive statistics examined. To compare the performance of SK, P020 and P021 to the healthy adult group, single-case analysis as outlined by Corballis (2009) was carried out. The three PWE were each compared separately to the healthy adult group due to their differing clinical characteristics. As described in Phase 1, P020 and P021 have different clinical histories, which would preclude them being considered together as a 'TLE group' (e.g. P021 has undergone radiotherapy and radiosurgery in addition to a subarachnoid haemorrhage and arteriovenous malformation; P021 has right TLE whereas site of epileptic focus is unknown for P020). They are also different from SK who is considered to have ALF. Therefore, this analysis involved three separate two-way mixed ANOVAs with recall scores on the APP test as the dependent variable, group (healthy adults vs. SK; healthy adults vs. P020; healthy adults vs. P021) as the between subjects factor and delay as the within subjects factor. Effect sizes were calculated using partial eta-squared.

To examine significant interaction effects, comparisons with the normative data at each delay used the modified *t*-test procedure outlined in Crawford and Howell (1998). This methodology was developed for use when the sample size for the normative data is small. It involves using the *t* distribution rather than the normal distribution curve to estimate the probabilities of the participant differing from the normative sample (Crawford & Garthwaite, 2002; Crawford & Howell, 1998). The *t*-

statistic, point estimate of the probability and the 95% confidence limits on the estimate was calculated using the program SINGLIMS.EXE, which accompanied the publication by Crawford and Garthwaite (2002). This modified *t*-test procedure was also used to compare SK's demographic and neuropsychological characteristics to that of the healthy adult group.

The significance level for all analyses was set at $p<.05$. A Bonferroni correction could have been applied to reduce the likelihood of making a Type 1 error due to the number of multiple comparisons being made. However, this would have been too conservative a value (e.g. for correlation analyses p = (.05/171) = .0003) and would have increased the likelihood of making a Type II error. Therefore, given the exploratory nature of the study, uncorrected *p*-values are reported but these should be interpreted with caution.

Power

The primary analysis was to detect forgetting in the healthy adult group. As the APP test had not been piloted on healthy controls at longer delays, estimates of effect sizes were calculated using Cohen's (1988) convention. In order to achieve a power of .80 with an alpha set at .05, with the correlation between repeated measures estimated at .80, it was calculated that six healthy adults were required to detect a large effect (partial $\eta^2 = .14$) and 12 healthy adults were required to detect a medium effect (partial $\eta^2 = .06$). The sample size calculation was undertaken using G*Power (version 3.1.9.2) (Faul, Erdfelder, Lang & Buchner, 2007).

However, given that a secondary analysis was to explore differences on the APP test between the healthy adult group and the patients (SK, P020, P021) using the Crawford and Howell (1998) single case methodology, the sample needed to be larger than this. Crawford and Garthwaite (2006) have considered the power needed to detect differences in single case studies and found that it is modest but suggested that deficits are routinely detected in these studies because the effect sizes are often large due to the effects of neurological damage on cognitive functioning. Using SK's previous performance on the similar Family Pictures subtest, after 30 days, he performed -2.2 *SD* below the healthy control mean (Kemp et al., 2012). Crawford and Garthwaite (2006) suggest that using their single case methodology, the statistical power to detect a large (2 *SD*) deficit is 54.63% when the sample size is 20 and 59.46% when the sample size is 50. As Crawford & Garthwaite (2006) suggest that it may not be worth the additional expenditure of effort to recruit sample sizes greater than 30, this study will aim to recruit a sample size of approximately 35 participants to account for the effects of possible loss to follow up over the longer delays.

Results: Normative Data

Participants

A total of 32 healthy adults from the general population were recruited and assessed in this phase of the study. Half were recruited from friends and family of the researchers and half via the UoL School of Psychology participant mailing list.

Demographics of the Healthy Adults

Table 13 summarises the demographic characteristics of the participants at the initial assessment. The majority of the group were female with an average age of 35.97 years (*SD* = 14.37). They were highly educated with an average of 16.22 years of education (*SD* = 2.71) and the majority had achieved at least an undergraduate degree. They had no history of neurological disorders.

Table 13: Demographic characteristics of healthy adults in Phase 2

Characteristics	***N* = 32**
Mean age, years (*SD*, range)	35.97 (14.37, 18-65)
Sex (*n*, %)	
Male	7 (22%)
Female	25 (78%)
Mean years of education (*SD*, range)	16.22 (2.71, 11-21)
Highest qualification level (*n*, %)	
GCSE or equivalent	2 (6%)
A-levels of equivalent	13 (41%)
Degree	12 (38%)
Postgraduate	5 (16%)

Note. GCSE = General Certificate of Secondary Education

Neuropsychological and Psychological Test Data

As shown in Table 14, the group had an estimated mean premorbid FSIQ of 107.81 (*SD* = 9.21), which falls within the ‘average’ range. Their estimated mean premorbid memory scores also fell within the ‘average’ range. Mean RBANS scores also fell within the ‘average’ range, with the exception of immediate memory, which fell within the ‘high average’ range. The group had an average score of 17.67 (*SD = 11*.17) on the EMQ. The majority (78%) scored below the cut-off of 24, suggesting that most participants were not experiencing significant memory failures in everyday life. Similarly, the group were not experiencing psychological distress with mean scores for anxiety and depression falling within the ‘normal’ ranges.

Table 14: Summary of neuropsychological and psychological scores of the healthy adults in Phase 2

Measure	***N***	**Mean**	***SD***	**Range**
TOPF UK[ǂ]				
Full Scale IQ (FSIQ)	31	107.81	9.21	86-127
Verbal Comprehension Index (VCI)	31	107.00	9.41	87-127
Perceptual Reasoning Index (PRI)	31	107.26	6.90	90-123
Working Memory Index (WMI)	31	107.48	7.26	86-120
Processing Speed Index (PSI)	31	102.84	5.39	89-114
Immediate Memory Index (IMI)	31	106.71	5.03	91-113
Delayed Memory Index (DMI)	31	106.81	5.08	91-114
Visual Working Memory Index (VWMI)	31	109.81	4.93	95-116
RBANS Index Scores				
Immediate Memory	32	114.28	16.33	81-152
Visuospatial	32	103.88	17.12	64-126
Language	32	104.53	9.46	87-120
Attention	32	103.91	14.25	75-132
Delayed Memory	32	103.84	11.72	78-128
Total	32	108.66	13.08	83-129
HADS				
Anxiety	32	7.13	3.77	1-16
Depression	32	3.25	2.51	0-10
EMQ	32	17.67	11.17	2-50

Note.· [ǂ] missing data (*n* = 1) due to administrator error; TOPF UK=Test of Premorbid Functioning-UK version; RBANS=Repeatable Battery for the Assessment of Neuropsychological Status; HADS=Hospital Anxiety and Depression Scale; EMQ=Everyday Memory Questionnaire

APP Test Results

All 32 participants completed the initial assessment. The test was counterbalanced with three versions, A, B and C, which were completed by 38%; 34% and 28% of participants respectively. The follow-up tests were conducted on average after 25 minutes (range 18-30 minutes), 24 hours (range 22-69 hours); 7 days (range 7-14 days) and 21 days (19-30 days). One participant did not complete the 1-week assessment as they were unable to be contacted by the researcher. One participant did not complete the 3-week assessment due to administrator error.

Eight (25%) participants reached learning to criterion in one trial and 18 (56%) reached criterion within two trials. Six (19%) participants failed to reach criterion within three trials but their recall at subsequent testing occasions was still assessed.

Table 15 illustrates the mean percentage recall and retention of participants at each delay. As shown in Figure 7, there was evidence of forgetting across the delays, with participants recalling on average 86.30% (*SD* = 16.75) of the information after a 3-week delay.

Table 15: Performance of healthy adults on the APP test

Delay	***N***	**Mean % recall (*SD*, range)**	**% retention (*SD*, range)**
Immediate delay	32	83.54 (17.4, 33-100)	-
Standard delay	32	84.38 (14.72, 47-100)	96.15 (5.39, 78.57-100)
24 hour	31	81.94 (18.63, 33-100)	94.22 (10.84, 45.45-100)
1 week	32	80.63 (19.33, 33-100)	92.44 (11.88, 45.45-100)
3 week	31	74.19 (22.36, 20-100)	86.30 (16.75, 27.27-100)

A one-way repeated measures ANOVA was carried out. Mauchley's test indicated that the assumption of sphericity had been violated, $\chi^2(9) = 32.30$, $p<.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.64$). The results show that there was a significant effect of delay, F $(2.54, 73.72) = 6.65$, $p = .001$, partial $\eta^2 = .19$, suggesting that the participants were experiencing forgetting over time. Bonferroni post hoc tests revealed that the significant differences fell between the 3-week delay and all the other delays. The significant main effect for delay and forgetting after 3-weeks remained when percentage retention scores were used as the dependent variable, $F(1.87, 54.34) = 9.50$, $p<.001$, partial $\eta^2 = .25$.

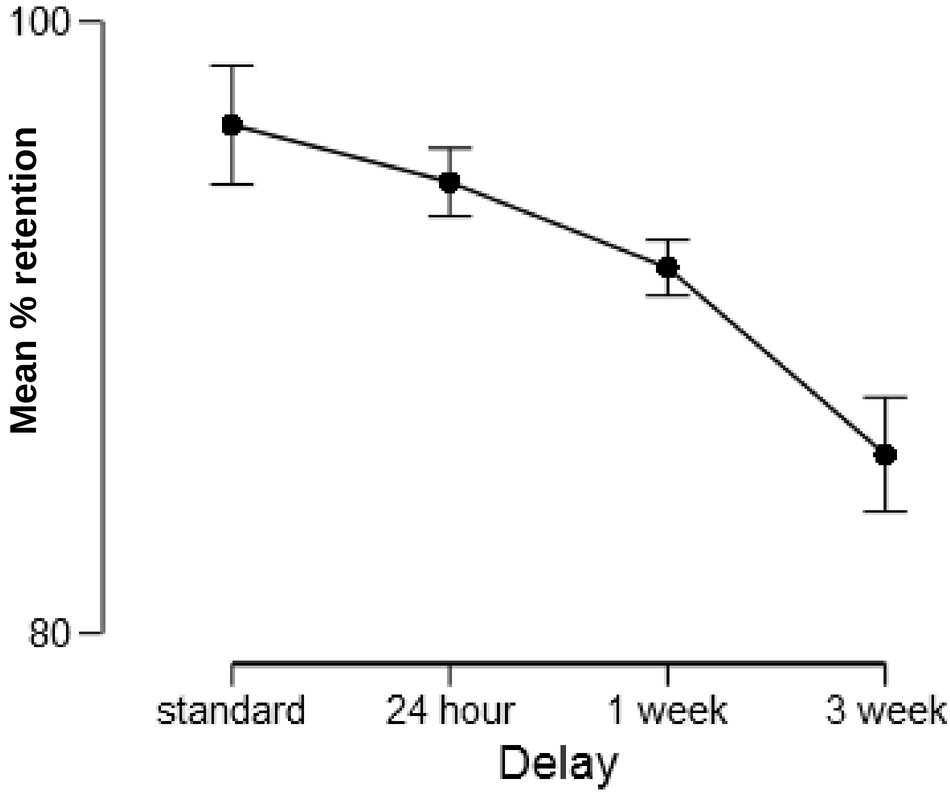


Figure 7: Forgetting curve on the APP test, mean percentage retention and standard error bars of the healthy adults.

Internal Consistency

Internal consistencies for each delay ranged between .67-.83, as shown in Table 16. The most internally consistent delay was 3 week (.83), which had adequate levels of reliability along with the 1 week delay (.82). The least internally consistent delay was the standard delay (.67). The immediate trial and 24 hours had modest levels of reliability, with reliability coefficients falling between .70 and .80.

Table 16: Internal consistency of the APP test

Delay	**Reliability coefficient**
Immediate delay (trial 1)	.75
Standard delay	.67
24 hour	.80
1 week	.82
3 week	.83

Individual analysis of correct responses is shown in Appendix E. As shown in Tables E1-E5, there was variability in the proportion of participants giving correct responses to the cued-recall questions (36-100%). For example, all participants gave the correct response to 'who was the recipient in the park?' asked at 1-week delay, whereas only 36% of people responded correctly to 'what colour is the police officer's hair?', which was asked at 3-weeks delay, suggesting that there may be some variability in the difficulty level of some questions. The proportions of participants giving correct responses to each of the 15 items were 38%-81%; 63-97%; 52%-97%; 41-100%, and 36%-90% for trial I, standard delay, 24 hour delay, 1-week and 3-week delay respectively.

Table 17 highlights the questions where more than 50% of respondents gave incorrect responses. Of these six questions, all involved either actor or recipient hair colour as either the cued feature or the response. Four of the questions related to the police offer or air stewardess in the park scene and two related to the fireman.

Table 17: Questions with > 50% of participants providing incorrect responses

Delay	**Question**	**Correct response**	**Proportion responding incorrectly**
3 week	What colour is the police officer's hair?	Red	65%
Immediate	Who had short black hair?	Fireman	63%
1 week	Someone was waving. What colour hair did they have?	Red	59%
Immediate	Someone had red hair. What colour hair did the other person in that scene have?	(Light) brown	56%
Immediate	What colour hair did the actor in the street have?	(Short) black	56%
3 week	What colour is the air stewardess' hair?	(Light) brown	52%

Correlations with Demographic Variables

As shown in Table 19, there was no relationship between performance on the APP test and years of education. However, there was a significant relationship between age and 3-week delay. Age was negatively correlated with APP test recall and retention scores and this relationship strengthened over the length of delay.

Comparisons with Self-report Questionnaires

As shown in Table 19, there was no relationship between either recall or retention scores at any delay and scores on the EMQ or anxiety or depression scores, as measured by the HADS.

Correlations with RBANS and Premorbid IQ

There was no relationship between FSIQ and recall scores at any delay. Percentage retention at 24 hours was positively correlated with all the estimated WAIS-IV indices (FSIQ, VCI, PRI, WMI, PSI) and all the estimated WMS-IV indices (IMI, DMI, VWMI), suggesting that those who had higher estimated IQ and memory abilities retained more information on the APP test after one day (see Table 20).

Similarly, there was no relationship between RBANS scores and recall scores at any delay. But percentage retention at 24 hours was positively correlated with the Visuospatial Index and Delayed Memory Index, again suggesting that those who had higher memory and visuospatial abilities retained more information on the APP test after one day (see Table 21).

Floor and Ceiling Effects

Floor and ceiling effects were examined. There were no floor effects, with none of the healthy adult participants obtaining a score of 0 at any of the delays. As shown in Table 18, there were some ceiling effects, with between 16-26% of participants correctly answering all the questions at each delay. However, only one (3%) participant remained at ceiling across all the follow-up time points.

Table 18: Floor and ceiling effects on the APP test

Delay	***N***	***N* (%) scoring at ceiling**
Immediate delay	32	5 (16%)
Standard delay	32	7 (22%)
24 hour	31	8 (26%)
1 week	32	6 (19%)
3 week	31	5 (16%)

Table 19: Correlation matrix with demographic and psychological variables

APP test		**Age**	**Education**	**EMQ**	**Anxiety**	**Depression**
Immediate % recall	Pearson's *r*	-.21	-.12	.17	.12	.17
	p-value	.25	.51	.35	.50	.37
Standard % recall	Pearson's *r*	-.32	-.11	.08	.23	.26
	p-value	.08	.55	.67	.20	.15
24 hours % recall	Pearson's *r*	-.30	.03	.25	.29	.25
	p-value	.10	.87	.18	.12	.17
1 week % recall	Pearson's *r*	-.33	-.06	.15	.27	.28
	p-value	.07	.74	.42	.13	.12
3 week % recall	Pearson's *r*	-.39	-.09	.19	.17	.10
	p-value	.03*	.62	.31	.38	.60
% retention standard delay	Pearson's *r*	-.11	-.01	-.06	.21	-.01
	p-value	.56	.97	.75	.25	.95
% retention 24 hours	Pearson's *r*	-.18	.31	.26	.15	.05
	p-value	.33	.09	.15	.44	.78
% retention 1 week	Pearson's *r*	-.29	.03	.12	.13	.08
	p-value	.10	.86	.50	.49	.68
% retention 3 week	Pearson's *r*	-.35	.05	.13	.04	-.02
	p-value	.05*	.79	.48	.82	.91

Note. *$p<.05$, **$p<.01$, ***$p<.001$; EMQ = Everyday Memory Questionnaire

Table 20: Correlation matrix with TOPF UK estimated indices

		TOPF UK							
APP test		**FSIQ**	**VCI**	**PRI**	**WMI**	**PSI**	**IMI**	**DMI**	**VWMI**
Immediate % recall	Pearson's *r*	-.05	-.08	-.09	.00	-.05	.04	.05	.04
	p-value	.78	.67	.63	.99	.80	.82	.79	.81
Standard % recall	Pearson's *r*	-.04	-.06	-.07	-.01	-.04	.03	.03	.03
	p-value	.82	.76	.71	.97	.83	.88	.86	.88
24 hours % recall	Pearson's *r*	.15	.12	.13	.23	.17	.28	.29	.27
	p-value	.42	.54	.48	.21	.38	.14	.13	.14
1 week % recall	Pearson's *r*	.07	.03	.05	.15	.09	.20	.21	.20
	p-value	.70	.86	.78	.41	.64	.27	.25	.29
3 week % recall	Pearson's *r*	-.10	-.14	-.10	-.02	-.08	.02	.04	.01
	p-value	.60	.45	.59	-.93	.68	.90	.86	.95
% retention std delay	Pearson's *r*	.05	.07	.06	.02	.05	.00	.01	.01
	p-value	.79	.70	.75	.93	.81	.99	.97	.97
% retention 24 hours	Pearson's *r*	.44	.41	.48	.51	.47	.52	.52	.50
	p-value	.02*	.02*	<.01**	<0.01**	<.01**	<.01**	<.01**	<.01**
% retention 1 week	Pearson's *r*	.13	.09	.16	.22	.17	.25	.26	.24
	p-value	.49	.63	.40	.23	.38	.17	.16	.20
% retention 3 week	Pearson's *r*	-.06	-.10	-.03	.03	-.03	.07	.08	.05
	p-value	.75	.60	.88	.87	.87	.72	.69	.80

Note. **p*<.05, ***p*<.01, ****p*<.001; TOPF UK = Test of Premorbid Functioning-UK version; FSIQ = Full Scale IQ; VCI = Verbal Comprehension Index; PRI = Perceptual Reasoning Index; WMI = Working Memory Index; PSI = Processing Speed Index; IMI = Immediate Memory Index; DMI = Delayed Memory Index; VWMI = Visual Working Memory Index

Table 21: Correlation matrix with RBANS indices

APP test		RBANS Immediate Memory	Visuospatial	Language	Attention	Delayed Memory	Total
Immediate % recall	Pearson's *r*	-.13	-.03	-.08	.21	-.31	-.10
	p-value	.47	.85	.66	.25	.09	.59
Standard % recall	Pearson's *r*	.04	.04	.05	.04	-.21	-.00
	p-value	.81	.83	.81	.82	.26	.98
24 hours % recall	Pearson's *r*	.03	.15	-.02	.23	-.04	.12
	p-value	.89	.43	.94	.21	.82	.52
1 week % recall	Pearson's *r*	.05	.02	-.09	.18	-.02	.05
	p-value	.80	.91	.63	.34	.90	.79
3 week % recall	Pearson's *r*	-.20	.14	-.28	.13	-.14	-.09
	p-value	.29	.47	.12	.48	.47	.63
% retention std delay	Pearson's *r*	.13	.21	.11	-.03	.08	.16
	p-value	.50	.24	.54	.86	.65	.38
% retention 24 hours	Pearson's *r*	.12	.36	-.08	.24	.38	.33
	p-value	.53	.04*	.65	.20	.04*	.07
% retention 1 week	Pearson's *r*	.06	.11	-.15	.11	.19	.11
	p-value	.77	.55	.42	.55	.30	.54
% retention 3 week	Pearson's *r*	-.15	.21	-.33	.02	.07	-.03
	p-value	.42	.27	.07	.92	.69	.86

Note. *$p<.05$, **$p<.01$, ***$p<.001$. RBANS=Repeatable Battery for the Assessment of Neuropsychological Status

Results: Case Study

In order to provide further support for the validity of the APP test, it was undertaken by patient SK, an individual with confirmed ALF. His performance was compared with the healthy adult group and the two patients who completed the test at all delays as part of Phase 1.

Description of the Case

SK is a 40 year old male who was diagnosed with bilateral TLE approximately 10 years ago. He has severe and uncontrollable epilepsy with severe and unpredictable seizures. He reported that his seizures are mainly nocturnal and occur in clusters, approximately every 4-5 weeks. His current AEDs are carbamazepine (500mg); zonisamide (500mg) and perampanel (8mg). SK has had a vagal nerve stimulator inserted. He has no formal qualifications and reported having six years of formal education. SK did not differ significantly in age from the mean of the healthy adult group, $t(31) = 0.28$, $p = .39$, but he had significantly fewer years of education, $t(31) = -3.71$, $p<.001$.

Neuropsychological Testing

SK's results across the neuropsychological test measures compared with the healthy adult group are illustrated in Table 22. SK had a premorbid FSIQ of 87, which falls within the 'low average' range and is significantly lower than the mean estimated IQ of the healthy adult group ($t(31) = -2.22$, $p = .017$). His premorbid memory scores are estimated to fall within the 'average' range but are significantly lower than the estimated mean scores of the healthy adult group. The majority of his RBANS indices fell within the 'low average' range; although his Total RBANS index fell within the 'borderline' range. The majority of his index scores are also significantly lower than those of the healthy adult group with the exception of the Visuospatial and Attention indices. SK has received neuropsychological input for his cognitive difficulties and consistent with this, his Delayed Memory index fell within the 'extremely low' range. Unsurprisingly, given his diagnosis of ALF, he reported a high level of memory failures in his everyday life on the EMQ, which was significantly higher than the mean of the healthy adult group ($t(31)= 3.10$, $p = .002$). He did not report experiencing psychological distress, with levels of anxiety and depression, as measured by the HADS, similar to the healthy adult group and falling within the 'normal' range.

Table 22: Neuropsychological test scores of SK compared to healthy adult group

Measure	**SK**	**Healthy adults (Mean, *SD*)**	***p*-value (95%CI)**
TOPF UK			
Full Scale IQ	87	107.81 (9.21)	.017 (.002-.057)*
Verbal Comprehension Index	88	107.00 (9.41)	.028 (.004-.082)*
Perceptual Reasoning Index	92	107.26 (6.90)	.019 (.002-.061)*
Working Memory Index	88	107.48 (7.26)	.006 (<.001-.028)**
Processing Speed Index	90	102.84 (5.39)	.013 (.001-.046)*
Immediate Memory Index	92	106.71 (5.03)	.004 (<.001-.018)**
Delayed Memory Index	92	106.81 (5.08)	.004 (<.001-.018)**
Visual Working Memory Index	95	109.81 (4.93)	.003 (<.001-.015)**
RBANS			
Immediate Memory	81	114.28 (16.33)	.027 (.004-.078)*
Visuospatial	89	103.88 (17.12)	.199 (.102-.324)
Language	83	104.53 (9.46)	.016 (.002-.054)*
Attention	88	103.91 (14.25)	.140 (.060-.252)
Delayed Memory	64	103.84 (11.72)	.001 (.000-.007)***
Total	76	108.66 (13.08)	.010 (.001-.037)**
HADS			
Anxiety	5	7.1 (3.8)	.295 (.178-.430)
Depression	2	3.3 (2.5)	.306 (.188-442)
EMQ	53	17.7 (11.2)	.002 (.000-.011)**

Note. **p*<.05, ***p*<.01, ****p*<.001; TOPF UK = Test of Premorbid Functioning-UK version; RBANS = Repeatable Battery for the Assessment of Neuropsychological Status; HADS = Hospital Anxiety and Depression Scale; EMQ = Everyday Memory Questionnaire

APP Test Results

SK's scores on the APP test were compared to the healthy adult data and to the two participants (P020, P021) who participated in the development of the final version of the APP test in Phase 1. As illustrated in Figure 8, SK was able to learn the information and reached criterion in two trials. He was able to retain the information after a 20-30 minute delay but then forgot the information at a rapid rate over the longer intervals of days and weeks, despite his initial learning. His performance was at floor after seven days. He admitted that his correct response at three weeks was a 'guess'. In contrast, P020, who was matched in age ($t(31) = -0.14$, $p = .45$) and years of education ($t(31) = -0.44$, $p = .33$), to the healthy adult group performed in a similar manner to this group across the delays. Interestingly, P021 who was also matched in age ($t(31)=1.30$, $p = .10$) and years of education

($t(31)=1.01$, $p = .16$) to the healthy adult group but had right TLE and reported significant visual memory problems following surgery did not reach learning to criterion and was performing at floor after 24 hours.

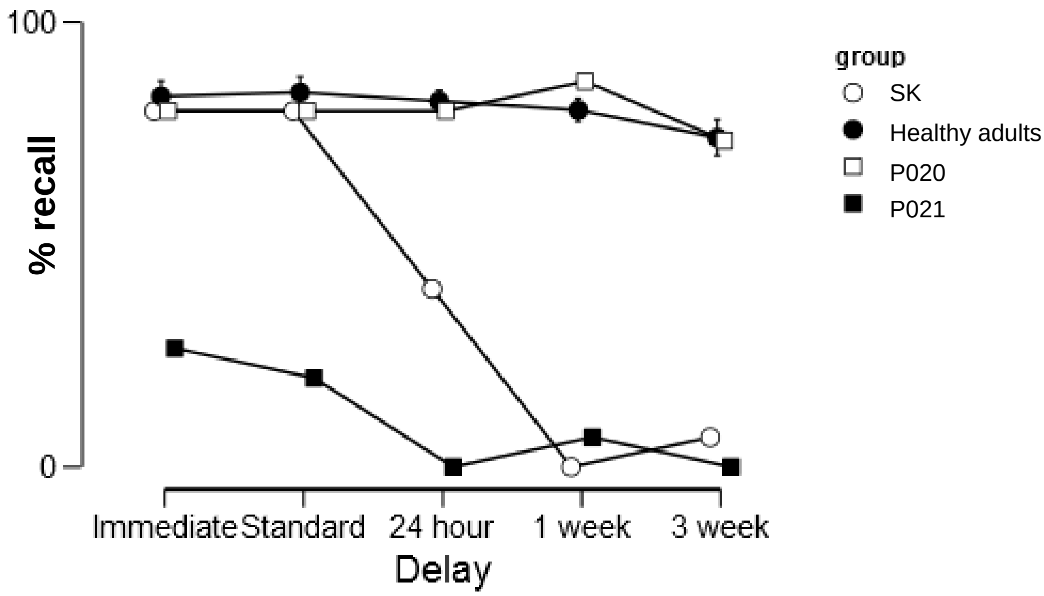


Figure 8: Performance on the APP test by PWE and healthy adults

In order to explore whether there were statistically significant differences in the performance on the APP test between SK and the healthy adult group, a two-way 2 (group: healthy adults, SK) x 5 (delay: immediate, standard, 24 hour, 1 week, 3 week) mixed ANOVA with repeated measures on delay was carried out (Corballis, 2009). Mauchley's test indicated that the assumption of sphericity had been violated, χ^2 (9) = 32.30, $p<.001$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.64$). The results show that there was a significant effect of delay, $F(2.54, 73.73) = 22.31$, $p<.001$, a main effect of group, $F(1, 29) = 5.01$, $p = .03$, and a delay and group interaction, $F(2.54, 73.73) = 16.04$, $p<.001$, suggesting that SK performed significantly differently from the healthy adult group across the delays. To establish where this interaction lies, comparisons were carried out between the two groups (Crawford & Howell, 1998). SK performed similarly to the healthy adults after an immediate and 20-30 minute delay. He then experienced a significantly more rapid rate of forgetting after 24 hours, 1-week and 3-weeks, indicative of ALF (see Table 23).

Table 23: Comparison between SK and healthy adults on the APP test

APP test	**SK**	**Healthy adults (*N*=30) Mean (*SD)***	***p*-value (95%CI)**
Trial I	47	-	-
Trial II	80	-	-
Trial III	NA-reached criterion	-	-
Immediate	80	83.33 (17.92)	.43 (.29-.57)
Standard	80	84.22 (14.80)	.39 (.26-.53)
24 hour	40	82.22 (18.88)	.018 (.002-.059)*
1 week	0	80.22 (19.84)	<.001 (.00-.002)***
3 week	7	74.00 (22.72)	.004 (.001-.018)**

Note. **p*<.05, ***p*<.01, ****p*<.001. NA= Not administered.

To compare the performance of P020 to the healthy adult group, a two-way 2 (group: healthy adults, P020) x 5 (delay: immediate, standard, 24 hour, 1 week, 3 week) mixed ANOVA with repeated measures on delay was carried out. Mauchley's test indicated that the assumption of sphericity had been violated, χ^2 (9) = 32.30, *p*<.001, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = 0.64). The results show that there was no significant effect of delay, *F*(2.54, 73.73) = 0.766, *p* = .50, no main effect of group, *F*(1, 29) = 0.002, *p* = .96, and no delay and group interaction, *F*(2.54, 73.73) = 0.233, *p* = .84, suggesting that P020 performed similarly to the healthy adult group and did not demonstrate ALF.

A further two-way 2 (group: healthy adults, P021) x 5 (delay: immediate, standard, 24 hour, 1 week, 3 week) mixed ANOVA with repeated measures on delay was undertaken to compare P021 to the healthy adult group. Mauchley's test indicated that the assumption of sphericity had been violated, χ^2 (9) = 32.30, *p*<.001, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε = 0.64). The results show that there was a significant effect of delay, *F*(2.54, 73.73) = 2.95, *p*<.05, a main effect of group, *F*(1, 29) = 15.81, *p*<.001, but no delay and group interaction, *F*(2.54, 73.73) = 1.25, *p* = .30, suggesting that while P021 performed significantly lower than the healthy adults there was no difference between their pattern of performance across the delays. However, it is important to note that while P021 demonstrates some forgetting, they cannot decline further as they are at floor after 24 hours. It is possible that ALF may emerge had they managed to encode and retain more information initially.

Summary of the Chapter

A total of 32 healthy adults were recruited and assessed as part of this second phase of the APP test's development. They completed the APP test as well as a brief battery of neuropsychological tests and self-reported measures. The majority of this group were females with an average age of 36 years, ranging from 18-65 years. The majority were educated to at least undergraduate degree level and had an average of 16 years of formal education. The mean group predicted FSIQ was within the average range and the mean Total RBANS score was within the average range.

The APP test demonstrated forgetting in this group of healthy adults, with participants recalling less of the information after a 3-week delay. The test has modest levels of reliability, with internal consistencies for each of the delays, ranging between .67 and .83. Some items were found to be more difficult than others, particularly those that involved hair colour as a cued feature or response. There was some evidence of ceiling effects but not floor effects. Age was negatively correlated with APP performance after three weeks. Percentage retention after 24 hours was positively correlated with estimated WAIS-IV and WMS-IV scores, and the Visuospatial and Delayed Memory indices on the RBANS. There was no relationship between performance on the APP test and mood or subjective report of memory problems. However, these exploratory analyses should be interpreted with caution due to the small sample size and number of multiple comparisons made.

An individual with confirmed ALF, SK, completed the APP test and his performance was compared with the healthy adult data using single case methodology. His performance on the APP task was consistent with what would be predicted for someone with ALF. He was able to learn the material initially and to the same level as the healthy adult group but then forgot the information at a rapid rate over longer intervals. ALF was apparent after 24 hours. This suggests that the task is a promising measure of ALF. However, this finding should be interpreted with caution due to SK not being matched in terms of education and IQ to the healthy adult group. However, his performance can be contrasted with two PWE (who are matched in age and education to the healthy adult group) who participated in the development of the task in Phase 1. One participant, P020, performed similar to the healthy adult group and did not demonstrate ALF. P021, on the other hand, performed significantly worse than healthy adults. She struggled to acquire the information initially and was performing at floor after 24 hours, which is consistent with her self-reported clinical history of right TLE following a subarachnoid haemorrhage and severe visual memory impairments following radiotherapy and stereotactic radiosurgery for an arteriovenous malformation. As she reached floor and had poorer learning, it is possible that ALF may have emerged if she had been able to retain more of the information during the learning stage.

This study also aimed to establish whether the task would be acceptable to service users, in particular, the use of telephone follow-ups. This will be examined in the following chapter.

PHASE 3: ACCEPTABILITY OF THE TASK

Overview of the Chapter

This chapter will detail the methodology and results of Phase 3 of the study, which aimed to evaluate the acceptability of the task to participants. The design and procedure will be described. The results from the quantitative and qualitative feedback will be presented.

Method

Design

In order to assess the acceptability of the test to patients, participants were asked for their feedback at their 3-week telephone follow-up, using a brief structured interview format. This contained both quantitative and qualitative questions to elicit participants experience and attitudes towards completing the test and the telephone follow-ups.

Questionnaire Development

The items for the brief interview were developed from a similar study investigating patient's attitudes towards computerised neuropsychological testing (Parikh et al., 2013). The quantitative questions asked participants how much they agreed with a series of statements using a five point Likert scale from 'strongly disagree' to 'strongly agree'. They were asked to rate their agreement with:

- Overall, I was satisfied with the new memory test
- I felt uncomfortable completing the new memory test
- I was satisfied with completing the follow-ups by telephone
- The instructions for the test were easy to understand
- I found the test stressful
- I was concerned about my privacy during the telephone follow-ups.

They were asked, if given the choice, whether they would have preferred to have met face-to-face or conduct the follow-ups over the telephone and the reasons for their choice. Participants were also asked open-ended questions to elicit their feedback about the task and the research study more generally.

Statistical Analysis

For Phase 3, both quantitative and qualitative data was collected. Descriptive statistics were used to examine attitudes by healthy adults and the small numbers of

PWE to the task. The qualitative data generated from the open-ended questions was analysed using thematic analysis (Braun & Clarke, 2006).

Results

Thirty (94%) of the healthy adults provided feedback on the task. Two participants were not asked for feedback due to administrator error. Feedback was also gathered from SK and the two people recruited from EA who completed the task at all testing occasions in Phase 1 of the study.

Quantitative Feedback

Figure 9 illustrates the responses to each statement by healthy adults (n = 30) and PWE (n = 3). As shown in Figure 9, participants gave positive feedback towards the task. None of the healthy adults expressed any dissatisfaction with the APP test or with completing follow-ups by telephone. None felt concerned about their privacy. All felt that the instructions were easy to understand. Only a small minority (10%) indicated that they felt uncomfortable completing the task and five (16.6%) felt that it was stressful. Similarly, PWE also expressed that they had no concerns about their privacy. They found the instructions easy to understand and did not find the task stressful. One of the participants recruited via EA felt uncomfortable completing the task due to their severe visual memory problems, which impacted on their overall satisfaction with the task. All were satisfied with completing the follow-ups by telephone.

When asked whether they would have preferred to meet over the telephone or face-to-face, nearly three quarters of the healthy adult participants reported that they prefer telephone administration because of the convenience, ease and time savings associated with telephone follow-ups compared with travelling to an appointment. Two expressed no preference and six participants stated they would prefer face-to-face because of their own personal preference for a more personal style. Only one person mentioned that they felt they may make more errors over the telephone, as they may be distracted by whatever task they may be completing at the time of the call. Of note, all the participants with epilepsy indicated that they would prefer telephone follow-ups because of the convenience.

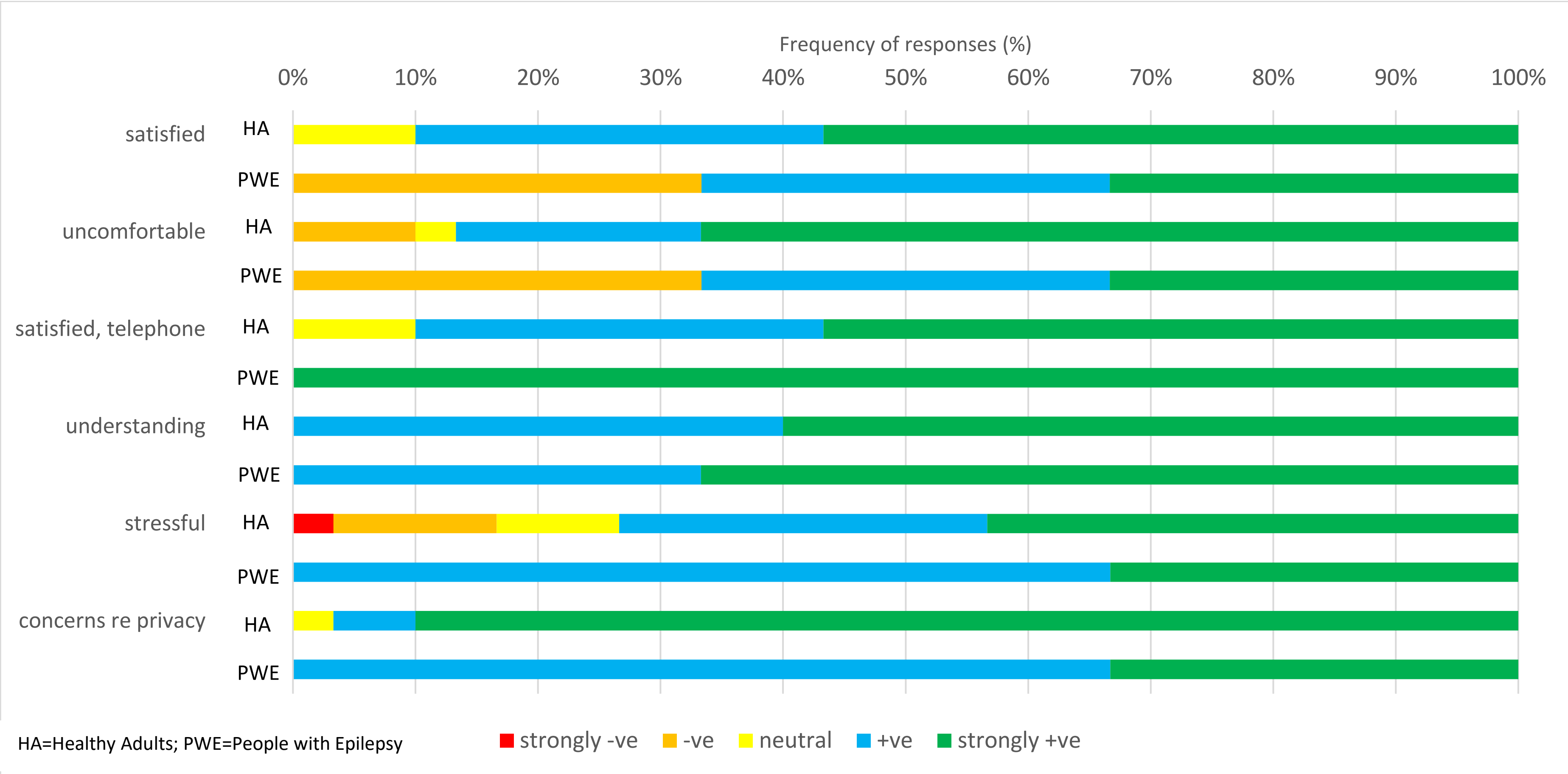


Figure 9: Quantitative feedback on the APP test

Qualitative Feedback

Participants were asked for their open-ended feedback on the APP test. As shown in Figure 10, feedback focused on the scenes; instructions; administration of the task and suggestions for future versions of the task.

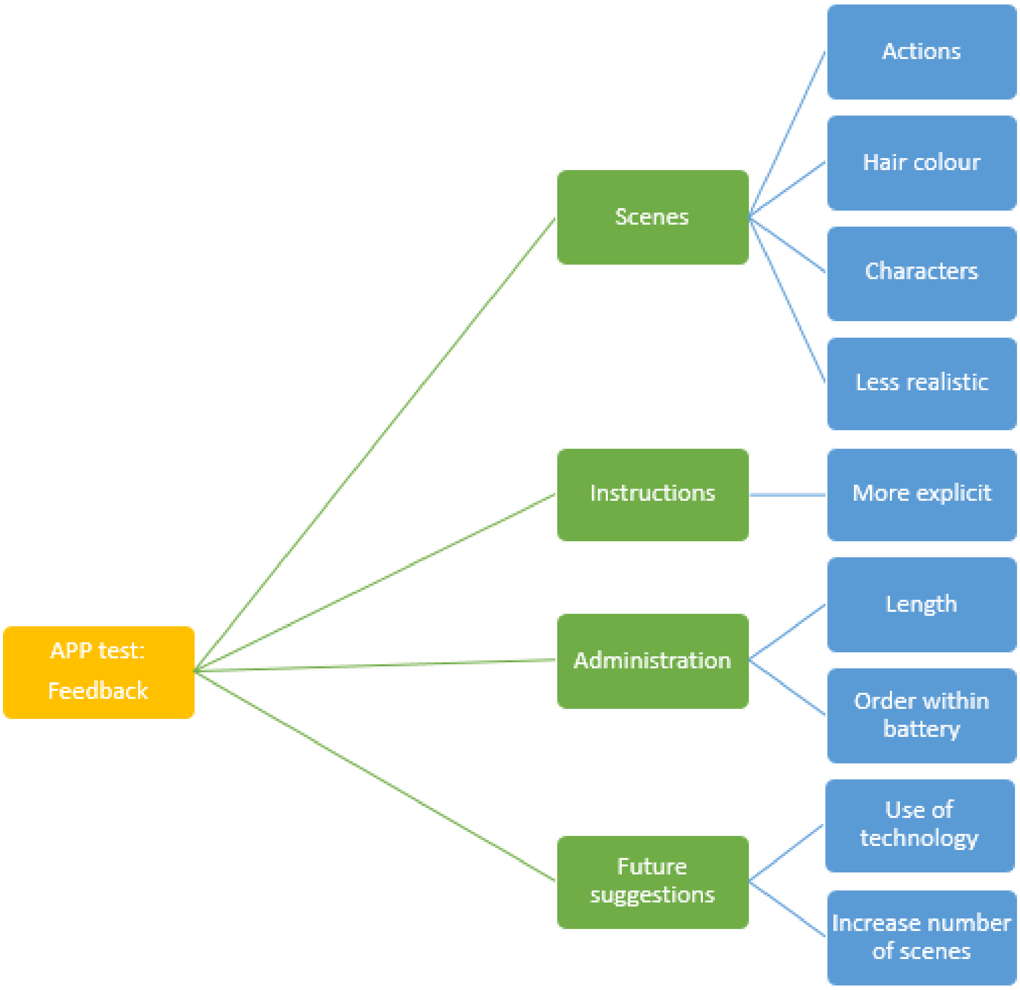


Figure 10: Qualitative feedback on the APP test

Several participants commented that some of the actions were too subtle and that some of the characters' professions were unclear, in particular, the air stewardess. Some participant's commented that the hair colour of the characters needs to be made more distinctive and that it was difficult to identify the hair colour of the fireman, as he was wearing a helmet. Two participants commented that they felt the task was difficult because the scenes are cartoons and 'less realistic'.

Many participants reported that the task and the instructions were clear and straightforward and the example helped, as they expected to view video clips from the initial instructions. However, one participant felt that the instructions need to direct participants to pay attention to hair colour more explicitly and another felt that it would be helpful to be directed whether to guess or state they do not know an answer when they are not sure.

Two participants commented on the length of the task; although this was unclear whether they were referring to the APP test specifically or the length of the initial assessment, which included other neuropsychological tasks. One person commented on the order of the test battery and felt that they may have performed better if the APP test was administered later in the battery. Two participants commented on the use of technology in the future to help with the administration of the telephone follow-ups and one felt that increasing the number of scenes would have been an enjoyable challenge for them.

Summary of the chapter

The task was seen as being acceptable to all participants and participants coped well with completing the follow-ups by telephone. The majority indicated that they preferred this method of administration due to the ease and convenience. Participants suggested several ways the task could be improved. In the next chapter, the results will be discussed in relation to the literature and ways the task can be developed further.

DISCUSSION

Overview of the Chapter

This chapter will discuss the results of this study, which aimed to develop a nonverbal test of ALF for use in clinical practice with PWE. There will be a discussion of the findings from the process of the development of the task, which will be considered within the context of the previous literature. Finally, the limitations, strengths and clinical implications of the research will be reviewed and recommendations for future work will be proposed.

Findings from this Research

ALF is a novel memory impairment, whereby, individuals demonstrate normal levels of learning and memory over short retention intervals but then demonstrate more rapid forgetting over longer intervals (Blake et al., 2000; Butler & Zeman, 2008; Elliott et al., 2014; Fitzgerald et al., 2013a). ALF may result from disruption to a slow memory consolidation process that takes place over days and weeks (Dudai, 2004). As neuropsychologists typically only assess retention after 20-30 minutes, these memory impairments may not be detected on standardised memory assessments. Therefore, it has been argued that memory should be tested at longer delays in those who report experiencing memory problems, in particular, those with TLE or TEA (Ladowsky-Brooks, 2015). As there are currently no standardised measures for assessing ALF, this study aimed to develop a clinically feasible nonverbal measure of ALF.

Initial Development and Piloting

A recently developed verbal measure of ALF, The Crimes Test (Baddeley et al., 2013), employed a matrix structure instead of a narrative structure within a constrained prose recall task. This means that each story contains a standard number of features and the relationship between features can be tested. This is important as it allows a large number of questions to be generated, which can be asked at the different delays, to limit the effects of repeated retrieval (cf. MTT, Nadel & Moscovitch, 1997). Jansari and colleagues (2010) investigated the effects of repeated retrieval of a story in a single patient with TEA compared with matched controls. Both the TEA patient and controls retained more information over a four week delay when the same story was retrieved at each time point, suggesting that repeated retrieval can be protective for people with ALF as well as healthy controls (Jansari et al., 2010).

As the aim was to develop and pilot a nonverbal analogue of the Crimes Test, a similar matrix structure was employed in the APP test. Several visual scenes were created that contained a standard number of features (location, action, actor, recipient, hair colour), which were integrated into a coherent scene. Different cued-recall questions were then asked at each of the five delays (immediate, 30 minutes, 24 hours, 1-week, and 3-week). The APP test was also designed to be a clinically feasible measure so a telephone follow-up administration procedure was conducted to reduce the resource implications for both patients and clinicians associated with face-to-face testing.

A total of 12 healthy adults and nine PWE recruited from local EA groups were involved in the first phase of the study. This phase of the study became an iterative process due to the challenges involved in developing materials that were of an acceptable difficulty level. In total, six versions were developed and piloted. Each version was piloted on a small number of healthy adults and/or PWE. Consistent with previous research that has found reduced memory and learning in PWE (e.g. Baxendale et al., 1998; Hermann, Seidenberg, Schoenfeld & Davies, 1997), the PWE in this study struggled with learning the visual material in the initial versions of the task. The patients were recruited from EA so no clinical information about their seizures or epilepsy syndrome was known. However, as initial learning can affect rate of forgetting (Elliott et al., 2014; Isaac & Mayes, 1999), it was important that learning was equated to ensure the detection of ALF and not just an acquisition deficit. Therefore, amendments were made to the task to facilitate initial learning, such as increasing the exposure time and introducing a learning to criterion (Elliott et al., 2014). Whilst attempts were made to modify the test in line with participant's performance and feedback, due to resource limitations, some changes could not be made (e.g. changes to the characters) so it is acknowledged that the modifications were limited in their scope.

Finalised Version of the APP Test

The final APP test was administered to 32 healthy adults and patient SK (an individual with confirmed ALF) to assess its psychometric properties (i.e. reliability, validity, floor and ceiling effects). Two PWE also completed the final version as part of the initial development in Phase 1, therefore, their APP data was used for further comparison in Phase 2.

The healthy adult group demonstrated forgetting on the APP test using both percentage recall and percentage retention scores, which is consistent with the literature on non-pathological forgetting and the forgetting curve for information over time. Participants recalled significantly less after three weeks compared to all the other delays. As participants were tested with different question sets at each delay, it is difficult to evaluate whether this represents a 'genuine' forgetting curve or whether it may reflect differences in the difficulty level of the different questions sets. However, on inspection of individual cued-recall questions, half of the most difficult questions (i.e. ones where >50% of participants failed to answer correctly) were from the immediate delay. Therefore, it is less likely that the observed forgetting was due to the later delays being more difficult and instead more likely that it represents 'normal' forgetting. To test this hypothesis, the question sets asked at each delay could have been counterbalanced across delays and this could be piloted in future development of the test.

Reliability and validity

Reliability refers to the '*consistency of measurement of a given score*' (Sherman et al., 2011, p.847). Reliability coefficients range between 0-1.00, with higher coefficients meaning that the test is more reliable. In Phase 2, internal consistency (i.e. extent to which individual items within a test measure the same cognitive ability/domain) was assessed using the Kuder-Richardson reliability coefficient. Internal consistencies of the APP test varied between .67 and .83 across the multiple delays. The delay with the lowest internal consistency coefficient was the standard delay and the highest coefficients were found at the longer delays (1-week = .82, 3-week = .83). The majority of delays were above >.70, suggesting that the test had modest levels of reliability and measured a unified construct of memory (Iverson, 2001; Lance et al., 2006). There are additional forms of reliability evidence (test-retest; alternate form; inter-rater reliability) and it will be important that future development of the APP test assesses these different forms of reliability in larger sample sizes and in clinical groups (Sherman et al., 2011).

Similarly, there are different types of validity evidence. Validity refers to '*the degree to which a test actually measures what it is intended to measure'* (Sherman et al., 2011, p.847). The assessment of the validity of a test is an on-going process, which is determined by collecting evidence from both healthy individuals and clinical populations (Sherman et al., 2011; Strauss & Smith, 2009). In Phase 2 of this research, evidence for convergent (i.e. associations with measures of the same construct) and known groups validity (i.e. picks up expected differences between groups) was tested. Evidence supporting the validity of the APP test will be discussed below.

Ageing

Several studies have found that older age is associated with ALF in non-clinical populations (Baddeley et al., 2013; MacDonald et al., 2006; Mary et al., 2013). Baddeley and colleagues (2013) compared healthy adults aged 18-25 years with those aged 49-76 years. Mary and colleagues (2013) compared younger adults (aged 18-30 years) with older adults (aged 65-75 years). Both studies found evidence for ALF in the older adult group. Therefore, it was hypothesised that forgetting on the APP test would correlate with age. Older age was associated with lower percentage recall and retention scores on the APP test and this strengthened across the length of retention interval with significant associations found at 3-weeks. This is consistent with the results of the study by Baddeley and colleagues (2013) using the Crimes Test, which also failed to find a difference between the younger and older adult group after an immediate delay but which became significant at longer delays.

The failure to find an association with forgetting at the earlier delays possibly may relate to the age of the sample. In this study, participants ranged from 18-65 years, with an average age of 36 years (*SD* = 14.37 years). Therefore, none of the healthy controls fell within the older adult range defined by Mary and colleagues (2013) in their study. Similarly, Miller and colleagues (2015) failed to find an effect of age on ALF in their healthy control group. They divided their participants into those aged 18-39 years and those aged 40-60 years (Miller et al., 2015). It may be that ALF is not as apparent at these relatively younger ages.

Older age has been associated with reduced episodic memory possibly due to reduced effectiveness in encoding (Craik & Rose, 2012). Trahan (1992) investigated forgetting in 255 healthy adults using the Visual Reproduction subtest of the WMS. They found that older adults performed more poorly than younger adults but the authors argued that this reflected a learning deficit rather than increased forgetting; although participants were only tested after 30 minutes. In contrast, Mary and colleagues (2013) found that their older participants did not differ in their quality of encoding in learning and suggested that their results support a hypothesis of age-related declines in consolidation processes. Unfortunately, the sample size in this study was not large enough to divide participants into groups of younger and older adults and assess rates of learning on the APP test or assess group*time interaction effects for forgetting. Nevertheless, the association with age and poorer recall after 3-weeks provides some support for the validity of the APP test as a measure of longer-term forgetting.

Subjective self-report of memory problems

The APP test did not correlate with subjective measures of memory complaints, which was initially unexpected. Butler and colleagues (2009) found that measures of ALF and mood predicted scores on the EMQ. However, several other studies have also reported a lack of association between forgetting and subjective memory complaints (Blake et al., 2000; Evans et al., 2014; Fitzgerald et al., 2013b; Muhlert et al., 2011; van der Werf, Guerts & de Werd, 2016). Blake and colleagues (2000) found that those who rated their memory problems as being a significant nuisance in their everyday life did not have reduced forgetting after an eight week interval. They argued that this may have been because they only assessed longer-term verbal memory and if they had measured longer-term nonverbal memory they may have found an association. A similar argument could be made for this study where only longer-term memory for nonverbal material was assessed. Inspection of individual items on the EMQ shows that it asks people about a range of memory problems involving both verbal (e.g. telling someone a story or joke that you have already told them once already) and nonverbal memory (e.g. failing to recognise, by sight, close relatives or friends that you meet frequently) and different processes of recall and recognition. Therefore, it could be speculated that only particular items would be related to performance on the APP test. This could be examined in future work.

In addition, other factors may account for higher levels of self-reported memory complaints, such as anxiety and depression. For example, Witt and colleagues (2012) only found correlations between objective memory performance and subjective memory after four weeks and not at earlier delays and mood was more strongly related to subjective memory complaints. The relationship between mood and EMQ scores was not examined in this study as this was beyond the scope of this thesis. However, mood was not associated with forgetting on the APP test, which is consistent with other studies that have not found a causal role for psychological factors in ALF (Butler & Zeman, 2008; Fitzgerald et al., 2013b).

Finally, in this study, there was very little variation in subjective memory complaints in this healthy adult sample. Most healthy adult participants reported experiencing few memory problems in everyday life and performed well on the APP test. In contrast, SK reported experiencing a high number of memory complaints on the EMQ and performed at floor after seven days. Therefore, it would be important to consider this relationship further in a larger clinical sample.

Objective memory measures

Forgetting on the APP test did correlate with some of the standardised memory measures as predicted. Increased forgetting after 24 hours was significantly associated with the Delayed Memory Index of the RBANS and the estimated WMS-IV indices. This provides some support for convergent validity in this healthy adult sample; although it less clear why forgetting at other delays was not associated with these measures.

There was also an unpredicted but interesting association between increased forgetting at 24 hours and the Visuospatial Index of the RBANS. This could imply that APP test performance is associated with visuospatial skills, providing some support for it being a nonverbal test, but this will be discussed further in a later section. However, it is important to acknowledge the increased probability of making a Type I error due to the large number of associations explored and the small sample size.

Person with confirmed ALF

One of the important sources of evidence for the validity of the test is the finding that SK demonstrated ALF on the APP test. SK has previously been published as a case report of ALF (Kemp et al., 2012). He has bilateral TLE and experiences frequent and severe seizures. He has undergone repeated neuropsychological testing, which showed that he has anterograde memory performance broadly in line with his general intellectual functioning. However, he experiences severe memory problems that impact on his daily functioning and reports several instances of rapid forgetting in his daily life, such as forgetting that he had attended a funeral several weeks later (Kemp et al., 2012). When his performance on tests of long-term forgetting were compared with an age, gender and IQ matched control group five times over a period of 28 days, he was found to have accelerated forgetting for visual and verbal material regardless of initial learning (Kemp et al., 2012).

On the APP test, he was able to learn the material and reached criterion within two trials, which is consistent with the healthy adult data. After a 30 minute delay, there was no difference between his performance and that of the healthy adult group. However, he then experienced rapid forgetting of the information after 24 hours. After one week, SK performed at floor. He reported 'guessing' an answer after three weeks, suggesting that he had some memory of the scenes, but could not recall the specific context of the information. This indicates that SK is experiencing accelerated forgetting of information after a day, which is consistent with a failure of a slow memory consolidation process (Atherton et al., 2014; Hoefeijzers et al., 2013; Muhlert et al., 2011). It is important to note that SK only reported experiencing two seizures during the three week period, which both occurred during sleep. These occurred at day 15 and 16. As SK performed at floor after seven days, his poor performance is unlikely to be due to the direct effects of seizures. However, it is important to note that the healthy adult group was not matched to SK in terms of estimated IQ, education and general cognitive abilities so caution should be made when interpreting his lowered performance compared with the healthy adult group, as these factors may impact on forgetting (MacDonald et al., 2006).

SK's performance was consistent with his self-reported memory difficulties on the EMQ and those he reported experiencing in daily life. Anecdotally, SK reported that he found it very difficult to recall details from his birthday which was only a few weeks prior to testing as part of this study. He also struggled to remember completing the initial testing at the 3-week follow-up and needed to be reminded at each telephone follow-up call the nature and purpose of the study. His performance on the APP test was also consistent with his previous performance on tests of long-term forgetting (Kemp et al., 2012). In the previously published case study, SK's nonverbal memory was assessed using two scenes from the Family Pictures subtest of the WMS-III. This is a similar task to the APP test, as it requires participants to remember characters within visual scenes, although the same questions were asked at each delay (e.g. recall of who was in the scene, where the characters were and what they were doing). This is potentially problematic as each testing occasion may result in being a learning trial, which can lead to increased retention, masking forgetting (Karpicke & Roediger, 2007).

On this adapted Family Pictures subtest, SK reached criterion within three trials, which was significantly longer than matched controls. However, this was less than on a list learning task, where he was unable to reach the criterion of 80% and he requested to discontinue the task due to the distress it caused him. His pattern of performance on the Family Pictures subtest was similar to his performance on the APP test. On the Family Pictures test, he retained 50% after four days; 30% after 11 days, and 20% after 28 days in comparison to the controls who were, on average, still retaining 80% at the longest delay. This consistency in performance with another nonverbal task suggests the APP test is a promising measure of ALF. Whilst, there are differences between the two tasks (e.g. three scenes versus two scenes; issues of repeated retrieval; timing of delay) and the length of time (approximately seven years) that has passed between the two assessments, it is positive that the APP test detects evidence of ALF using a different measure at a different time in SK's life.

Additional support for the ability of the APP test to detect ALF also comes from the comparison of the two PWE who completed the final version of the test as part of Phase 1. These participants who were matched in age and education performed very differently on the test from each other and SK. As these patients were recruited from EA and not from NHS clinics, their medical records could not be accessed to confirm clinical information. P020 had TLE but the site of his epileptic focus was unknown. He experienced frequent seizures but had not undergone epilepsy surgery. His scores on the APP test did not differ from those of the healthy adult group. This is consistent with previous reports that suggest there is variability in ALF in TLE and not all people with TLE demonstrate ALF (Elliott et al., 2014; Fitzgerald et al., 2013a, Milller et al., 2015). In contrast, P021 found the test challenging and struggled to learn the information to criterion initially. She performed at floor after

24 hours. Both these factors mean that rates of forgetting can be underestimated. While P021 demonstrated some forgetting, ALF may have possibly emerged if she had retained more information initially. P021 has right TLE and she reported undertaking neuropsychological assessments previously that detailed visual memory impairments. In addition, she has undergone radiotherapy and stereotactic radiosurgery for an arteriovenous malformation. While it is difficult to speculate on her neuropsychological performance due to a lack of detailed clinical information about her neurological history, her pattern of performance on the APP test is not surprising. Whilst the test requires further validation in a larger clinical sample, the findings from this small case-series analysis suggest that the test has some sensitivity to detecting memory impairments and is worthy of further testing and development, particularly with more well matched healthy adults.

Acceptability of the Test

An important component of this study was to develop a task that was clinically feasible and acceptable to PWE. Previous studies have suggested that telephone testing is a reliable method of administration (e.g. Baddeley et al., 2013; Taichman et al., 2005; Unverzagt et al., 2007). In a study of healthy elderly individuals where individuals received telephone cognitive testing and in-person testing, using a two-way cross-over design, telephone administration was found to be comparable (Mitsis et al., 2010). Similarly, an as yet unpublished study using the Crimes Test has also show no difference between phone and in-person testing after one week (Allen et al., in prep). However, generally little attention has been paid to participant's experience and preference.

Using quantitative and qualitative questions to elicit attitudes towards the test, the APP test was considered acceptable to participants. The majority of healthy adults and all PWE felt that telephone follow-ups were acceptable and preferable to face-to-face testing due to their ease and convenience.

There are some concerns about administering neuropsychological assessments over the telephone, particularly around the lack of nonverbal communication, which may make examiners less sensitive to any potential participant distress. Despite some of the PWE experiencing memory problems (SK and P021), neither provided feedback that they found it more challenging to answer the questions over the telephone. All were given the opportunity to debrief and discuss the experience as part of the research study. None expressed experiencing any frustration during the telephone calls. Only one participant, from the healthy adult group, responded that they felt that they were more distractible during the telephone follow-up, as they were completing another task when the call was made. However, in this situation, it would be sensible to re-arrange the call for a more convenient time when the person is in a quiet room with no interruptions or distractions. Finally, none of the participants, in this study, reported any health problems, such as hearing loss, which would impact on telephone assessment. In this situation, modifications to the task would need to be considered but with an awareness that this would impact on reliability and validity, as with any neuropsychological assessment (Hill-Briggs, Dial, Morere & Joyce, 2007).

Evaluation of Methodology

Elliott and colleagues (2014) critically reviewed the methodology of ALF studies and the lack of appropriate methods to study this memory phenomenon. On the basis of their review, they outlined seven recommendations (see Table 2 in the introductory chapter) that future studies should follow when designing new tests. These will now be considered in turn in relation to the APP test.

1. *Groups are matched for age and IQ*

Unfortunately, due to the recruitment strategy employed in this research, which involved advertising through the UoL School of Psychology participant mailing list and via friends and family of undergraduate research students, the healthy adult group were highly educated. They had a mean 16 years of education (*SD* = 2.71) and the majority had achieved at least an undergraduate degree. They also had a higher predicted FSIQ (mean estimated IQ = 107.81, *SD* = 9.21) than SK (estimated IQ = 87). While SK performed in line with the healthy adult group on the APP test immediately and after a 30 minute delay, as discussed previously, caution must be made in interpreting whether his more rapid rate of forgetting is due to ALF.

Efforts were made to try and recruit a wide range of people in terms of age, sex and educational background through a number of methods. Firstly, a screening questionnaire was devised which interested participants were asked to complete prior to selection but due to a limited number of people responding to the advertisement all eligible participants were included for an initial assessment. Secondly, ethical approval was also granted to recruit people from the local community through, for example, posters in local church groups or libraries; although this recruitment strategy was not employed due to the time constraints involved in the research. However, future development of the test should involve an age and IQ matched control group when comparisons are made with a larger clinical group.

2. Both verbal and nonverbal tests are used

ALF has been found for both verbal and nonverbal information (Cassel et al., 2016; Fitzgerald et al., 2013a; Muhlert et al., 2011) and there is some evidence for material specificity. People with left TLE are more likely to experience more rapid forgetting of verbal information and those with right TLE are more likely to experience accelerated forgetting of nonverbal information (e.g. Blake et al., 2000; Narayanan et al., 2012); although this is less consistently found for right-sided lesions (Wilkinson et al., 2012). By assessing memory for both kinds of information, strengths and weaknesses can be identified and appropriate memory strategies can be tailored for the individual.

The APP test has been designed as a nonverbal analogue of the recently developed Crimes Test. A comparison of the two tests was considered to be beyond the scope of this thesis, as it would have required larger sample sizes to detect differences between the two tests and would potentially have made the assessment too lengthy for participants, if it was also included alongside the other neuropsychological and self-report measures. Therefore the Crimes Test was not administered in this study. However, this would be a useful avenue for further studies and would fulfil the recommendations of Elliott and colleagues (2014). Comparisons of the performance of this healthy adult group on the APP test to the performance of healthy adults on the Crimes Test collected by Baddeley et al., (2013) was considered. However, this would be difficult due to differences in the methodology (e.g. age differences in the two samples, length of delays used). The Crimes Test is currently being piloted in several epilepsy centres in the UK and data is being gathered which will be used to assess its potential for use in clinical practice. Therefore, there is future scope for the APP test to be used alongside the Crimes Test in future larger clinical samples.

Whilst participants are asked to remember visual scenes on the APP test, it is acknowledged that the APP test may be verbally encoded and is verbally retrieved through the use of cued-recall questions. Previous studies have found that similar tasks to the APP test, such as the Family Pictures subtest of the WMS-III, have been found to measure verbal memory and be less sensitive to right medial temporal lobectomy (Chapin et al., 2009). A review of declarative memory in children with specific language impairment (SLI) found that children with SLI performed more poorly that their peers (average effect size 0.53, 95%CI 0.29-0.77) on Picture Tasks (memory for visual scenes). The authors suggested that this was because the pictures can be verbally encoded and argued that the children's observed visual memory difficulties reflected their underlying verbal working memory deficits (Lum & Conti-Ramsden, 2013).

In addition, as the features within the scenes have been integrated into a coherent scene, they could potentially be incorporated into a narrative representation, which could also make them more meaningful (Ladowsky-Brooks, 2015). Craik and Lockhart's (1972) level of processing model suggests that deeper processing leads to better recall. In the APP test, the interaction of the characters within the scenes may lead to a richer semantic elaboration and deeper encoding and retention than pictures of more concrete objects (Baddeley & Hitch, 2017). Similarly, it has been suggested that there are differences in how semantic and episodic memories are processed and that over time 'gist' versions of memories are created, which are less hippocampal-dependent (Winocur et al., 2010). Therefore, people with TLE and hippocampal sclerosis, in particular, may not be as susceptible to ALF for these types of scenes. This could be investigated in future studies by comparing the performance of those with different underlying pathologies and epilepsy syndromes (e.g. TLE vs. idiopathic generalised epilepsy). However, questions were asked about perceptual details, such as the character's hair colour, to test individual's memory for more visual elements of the scene, which may be more context and hippocampal-dependent.

However, the APP test does have advantages over previous nonverbal ALF tasks. For example, the test developed by Muhlert and colleagues (2011) presented elements of their scenes in a list-like fashion by showing objects individually in each quadrant of the scene for three seconds. As these objects had verbal labels (e.g.

barrels, clown, cat, lady, mop), this may lead to more verbal encoding and appears more similar to a verbal list learning task. In addition, in their visual scenes test, answers could potentially be guessed based on semantic representations held in LTM (e.g. clown in the stage scene). In the APP test, characters were chosen so that they could plausibly be found within any scene to reduce any benefits of distinctiveness but the characters were not closely associated with a particular scene so that they relied more on episodic memory (Schmidt, 2008).

Generally, there are fewer measures of nonverbal memory, as it is difficult to create 'pure' visual memory tasks using stimuli that cannot be verbally encoded (Baddeley et al., 2015; Brown et al., 2007). The WMS-IV has attempted to more accurately assess visual memory function through the new visual subtests (Design Memory, Spatial Addition and Symbol Span) (Wechsler, 2010). Similarly, virtual water maze tasks, faces tests and memory for real-world events have been created (Barkas et al., 2012; Bengner & Malina, 2010; Muhlert, Milton, Butler, Kapur & Zeman, 2010; Narayanan et al., 2012). However, all these tasks pose a challenge to measurements of ALF where there are practical problems to repeated testing and how to assess longer term memory function in a way that does not pose a burden to participants and clinicians. The aim of this study was to create a task that is feasible in clinical practice.

The Doors and People test (Baddeley, Emslie & Nimmo-Smith, 1994) has been designed to assess verbal and visual memory and includes recall and recognition. The features involved in the Doors task would lend itself to a similar matrix structure to the APP test (e.g. colour of door, age, condition, shape, function of door) (Baddeley et al., 2015) and so would be a possible candidate for development into a nonverbal test of ALF (Baddeley, personal communication). However, it is unclear how longer-term follow-ups with this task could also be conducted without avoiding the problem of verbal retrieval.

This study chose to use telephone administration, which as discussed earlier, was considered to be acceptable to participants. Recent ALF studies, have used postal return of abstract designs (e.g. Miller et al., 2015) or incorporated the use of technology (e.g. Gascoigne et al., 2014). For example, a recent study of ALF assessed nonverbal memory using a Design Locations task in children, where they were asked to remember the locations of abstract designs. After seven days, they were emailed a link and asked to complete the task online (Gascoigne et al., 2014). Cassel and colleagues (2016) also used a telephone administration procedure but asked participants to open a computer file containing pictures during the telephone call. The use of technology, such as using smartphone apps, email, or videoconferencing, is something that could be incorporated into the future development of the APP test. Possible ideas could involve participants drawing aspects of the scene so that it is retrieved in a less verbal manner or incorporating recognition items (see below).

3. Tests include recall and recognition paradigms

As discussed earlier, an important feature of the APP test is the use of a large number of cued-recall questions to create parallel question sets. However, it is also important to assess recognition (Elliott et al., 2014). Recognition is based on the two processes of familiarity and recollection, which are thought to be mediated by different systems (Huijgen & Samson, 2015). Familiarity involves the perirhinal cortex, whereas recollection relies on the hippocampus and parahippocampal cortex (Huijgen & Samson, 2015). There are mixed findings regarding whether individuals demonstrate ALF on recognition tasks. Some studies have found ALF for recognition, which suggests that ALF may represent disruption to consolidation processes (Blake et al., 2000; Dewar et al., 2015; Tramoni et al., 2011). However, others have suggested that recognition is intact (e.g. Martin et al., 1991; Ricci et al., 2015b), which suggests that ALF may in fact be a retrieval deficit. Therefore, the inclusion of recognition paradigms would potentially help identify the underlying mechanisms of ALF but it also has important clinical implications. Memory strategies would be more appropriately tailored if ALF represented a failure of retrieval rather than consolidation.

However, as discussed in the section on nonverbal tasks above, there are practical problems to assessing visual recognition memory, especially using the telephone administration methodology employed in this study. Therefore, a recognition component was not included at this stage of the test's development. This could be incorporated into future tasks verbally either by asking forced choice recognition questions (e.g. 'was the fireman in the park scene?') or by using email/video-conferencing to visually present recognition items. As part of the test development, several alternative scenes were created (e.g. a supermarket scene) so these could be included as foil items. Characters can also be removed from the scenes and superimposed onto different scenes or have their features changed relatively easily to create a number of distractor items. Future versions of the test and future validation studies could be undertaken to compare whether there are differences between recognition that is assessed verbally compared with selecting a target image from a foil items which could be presented visually.

4. Distractor tasks are used

Distractor tasks involve participants repeating sequences of words or numbers to prevent rehearsal in working memory and ensure that retrieval is from LTM. In visual memory tasks, this could also prevent participants verbally encoding the stimuli or using subvocal rehearsal (Baddeley, 2003). Elliott and colleagues (2014) recommended that participants complete a 10 second distractor task after presentation of the material. A distractor task was considered in this study, however, as PWE were having difficulty learning and recalling the material in the initial piloting, this was not employed in case it added another level of complexity and led to floor effects in PWE. However, this could be incorporated into future developments of the task and comparisons made between those who have or have not completed a distractor task.

5. Initial learning is equated

During the initial piloting, attempts were made to equate initial learning. Exposure times were increased from three seconds to 10 seconds to improve initial learning and a learning to criterion was introduced (Elliott et al., 2014; MacDonald et al, 2006). This is in contrast to the Crimes Test, which only used a single presentation procedure, as the piloting for the Crimes Test suggested that the material was relatively easy to learn (Baddeley et al., 2013). However, the authors accepted this might not be ideal for an epilepsy group, who may find the task more challenging and a learning to criterion procedure may be incorporated into the Crimes Test in the future (Baddeley et al., 2013).

On the APP test, the learning to criterion was set at 80% rather than requiring, for example, perfect recall on successive tasks, to reduce the effects of over-learning, which may mask ALF. In Phase 2 of the task, initial learning was equated between the healthy adult group and SK, which provides further support that SK was demonstrating ALF and that his reduced performance after the longer delays is unlikely to reflect an acquisition deficit.

6. Avoid ceiling and floor effects

Through initial piloting, modifications were made to the APP test to try and avoid floor and ceiling effects. In Phase 2, no floor effects were noted for the healthy adult participants. SK reached floor after one week. However, SK was able to reach criterion and scored 50% after 24 hours, suggesting that his poor performance may be capturing his severe memory difficulties rather than the task being too difficult for people with TLE more generally. However, P021 also reached floor after 24 hours, which means that forgetting rates may be underestimated (Elliott et al., 2014) Testing on a larger clinical sample is needed to demonstrate the proportion of PWE who demonstrate floor effects at these longer delays.

There was some evidence of ceiling effects in healthy participants; although the test was still able to demonstrate evidence of forgetting over time in this group. Only one participant scored at ceiling across all the delays. Again, a larger sample is needed to assess whether further modifications are needed to reduce potential ceiling effects without leading to floor effects.

7. Minimise opportunities for rehearsal of test material

During Phase 2 of the study, participants completed other neuropsychological tasks during the 30 minute delay. This reduced the opportunities for rehearsal as participants were involved in other cognitively demanding tasks. As discussed in the introduction, there is an interesting literature around interference and the impact of a wakeful rest on retention and forgetting (Alber et al., 2014; Dewar et al., 2012; Dewar et al., 2014). Therefore, it is important when collecting further normative data and in the standardisation of the task that all participants are subject to the same activity during the delay, as this may potentially affect retention at longer delays.

For the telephone follow-ups, participants were informed that they would be contacted via telephone at various points over the three week delay in order to reduce attrition from the study. However, they were not told when the calls would occur in order to minimise rehearsal and they were asked explicitly not to rehearse the scenes or write any information down. This study did not ask whether people explicitly rehearsed the material, for example as in the study by Alber and colleagues (2014), but this could be asked in future development. As the follow-ups were conducted over the telephone, it cannot be verified whether people used written prompts; although this seems unlikely and there is no apparent motivation or gain to be made from engaging in this behaviour.

In summary, the test aimed to develop a nonverbal analogue to The Crimes Test, which fulfilled Elliott and colleagues (2014) recommendations for verbal and nonverbal tests of ALF. Efforts were made to equate initial learning, to avoid floor and ceiling effects in healthy adult participants and to minimise the opportunities for rehearsal of test material. The APP test requires further development and validation to ensure that groups are matched for age and IQ, to incorporate a recognition paradigm and to use a distractor task. Further limitations of the study will be discussed below.

Limitations of this Research

There are several limitations of this study. Firstly, as discussed in the earlier section, the test was deigned to be a nonverbal test, but it may not be a ‘pure’ measure of nonverbal memory due to it being verbally encodable and verbally retrieved. The extent to which it relies on verbal memory is unclear but could be examined in future research using methods detailed in the Recommendations section below. Given that the characters, scenes and actions have such clear verbal labels, its reliance on verbal memory is likely to be high. However, little attention has been paid to how to assess ALF in clinical practice and it was important to balance creating a valid test with one that is clinically feasible, practical to administer over the telephone, and acceptable to patients. As discussed, there are several potential ways that this task could be adapted to make it less reliant on verbal memory (e.g. use of email links, smartphone apps, videoconferencing).

Secondly, there are limitations in the representativeness of the healthy adult sample. It is important that normative samples are representative and matched on important variables (Brooks, Sherman, Iverson, Slick & Strauss, 2011). As discussed earlier, the healthy adult group were significantly more educated, majority were female and had a higher estimated IQ and levels of cognitive functioning than SK. Despite SK having comparable levels of learning and retention over the standard delays on this task, these factors may have contributed to his poorer performance on the APP test rather than ALF. Therefore, caution should be made when drawing conclusions in comparison to this healthy adult sample until a larger, more representative sample is obtained. To provide further support for the use of this task in clinical practice, it will be important to compare SK to an age, and IQ matched control group.

Additionally, little attention has been paid to calculating power for single case studies using the methodology employed in this study but it is thought to be modest (Crawford & Garthwaite, 2006). Crawford and Garthwaite (2006) calculated the size of the deficit required for 80% power given a range of sample sizes and measurement errors. Based on a task reliability of .85, the size of the deficit required to achieve 80% power to detect a deficit ranged from 2.76SD (for a control sample size of 50) to 2.87SD (for a control sample size of 20) meaning that this method is only powered to detect relatively large deficits. The rate of forgetting that determines ALF is still unclear due to differences in methodology so further work is needed to determine the size of deficit required to detect ALF.

It would have been useful to have an age and IQ matched TLE group who would be compared with the healthy adult group to provide further support for the validity of the task. Having people with a confirmed diagnosis of left TLE and right TLE would also have been potentially useful to make predictions about their performance and test for potential known-groups validity; although laterality is not always found for nonverbal tests. Ethical approval has been granted to undertake this work and this would be a useful for further validation of the task.

Thirdly, there are limitations in terms of the APP test and the features comprising the scenes, which was provided via feedback in Phase 3 of the study. Participants commented that some of the actions were too subtle and that some of the characters were not easily recognisable. In addition, there may be further problems in terms of their cultural specificity, for example, the inclusion of the female priest. It is acknowledged that some of these limitations were identified by participants in Phase 1 but unfortunately due to resource limitations, the materials were limited in the ways that they could be modified. For example, due to budget constraints, there was not enough funds to request further changes by the graphic designer. Therefore, the pilot phase was limited in its scope and potential for modification and had to rely on ways the task could be delivered to reduce the potential for floor effects. In addition, there are several ways that the test materials could have been constructed, for example, they could have been constructed using photographs of real-life scenes and could have been comprised using a number of standard features that did not involve people, places or actions. With an increased budget, more time could have been spent piloting different versions of the stimulus materials and gathering more formal data. For example, the APP test tried to create scenes that avoided participants using their schematic representations held in LTM to 'guess' the answers and tried to avoid gender stereotypes. However, mainly due to the time constraints, the scenes and characters were decided on informally within the research team. On reflection, ratings of gender stereotypes and distinctiveness in scenes could have been more formally gathered, using a team of raters and comparing agreement ratings. In addition, more extensive piloting could have been undertaken; although this would have impacted on the ability to carry out Phase 2 and 3. As Phase 1 was an iterative process but only involved small numbers of healthy adults and/or PWE at each stage, decisions were often made to alter the materials based on the feedback of this small sample.

Strengths of the Research

However, despite these limitations there are a number of strengths to this research project. Firstly, given that this test was designed for use with PWE in clinical practice, PWE were involved in all phases of the project and their feedback was solicited and responded to in Phase 1. Patient and public involvement is important to ensure that the research is relevant and meaningful to the population it is designed to help (INVOLVE, 2012). In this study, directly asking participants about their experience in Phase 3, ensured that the APP test, including the telephone follow-ups are acceptable to participants. Additionally, modifications in line with feedback has meant that the APP test has been developed to be clear, accessible and the instructions are easy to understand for both clinicians and patients. The test is easy to administer and requires little training. The three undergraduate psychology students who assisted with recruitment and assessment in this project were trained to administer it within a short training session.

Secondly, the extensive piloting phase helped to avoid floor effects, at least in the healthy adult group, and ensure that the test was sensitive to the experience of participants (e.g. addition of instructions to reassure individuals that they are not expected to get everything right).

Thirdly, the design of the APP test allows for parallel cued-recall questions to be generated, which avoids the problem of repeated retrieval in longer-term memory studies, which may mask accelerated forgetting.

Finally, an individual with confirmed ALF performed in line with predictions on the APP test. His performance was contrasted with the healthy adult group and two patients who were involved in the development of the APP test in Phase 1. Whilst caution needs to be made regarding these comparisons due to the healthy adult group not being matched on a number of potentially confounding variables and lack of confirmed clinical information on the two PWE, the differences in their patterns of performance, suggests that the APP test is a promising measure of nonverbal ALF worthy of further development.

Clinical Implications

This study is important because memory problems can have a significant impact on an individual's day–to-day functioning, psychological well-being and quality of life (e.g. Baker et al., 2009). Many PWE report memory problems over longer time intervals, which are not detected on current standardised measures. For example, if SK was only assessed after the standard 30 minute delay, his performance would have been in line with the healthy adult group and his rapid rate of forgetting would not have been detected. However, detecting his longer term memory problems may help to provide an explanation for the difficulties he reported experiencing in his daily life. This may be important in validating his experience of memory problems. The assessment of ALF may also help to identify the nature of an individual's memory problems and ascertain whether they are due to difficulties in learning and encoding, for example, or due to faster forgetting. This may then help to identify appropriate memory strategies such as repeated rehearsal or the use of external memory aids (Jansari et al., 2010; Lee-Donaldson, 2011).

These preliminary results from the APP test suggest that the assessment of long-term forgetting of nonverbal information can be completed relatively quickly over the telephone and at little cost to patients and clinicians. Currently, it is difficult to estimate the prevalence of ALF in people with TLE. Approximately half of people with TEA may demonstrate ALF and estimates range between 30-55% for people with TLE (Zeman et al., 2013). While not all people with TLE who report memory problems may require assessment at these longer intervals, it could easily be incorporated into neuropsychological assessments, particularly for those who complain of rapid forgetting of recently acquired memories. However, it is important to acknowledge that the assessment of ALF would require the collection of normative data with a more representative population. In addition, the results of this study also suggest that there is little relationship between subjective report of memory problems and assessments of accelerated forgetting such as the APP test. Future research could consider whether it is possible to create more sensitive subjective report measures that could be used as potential screening tools to identify when more detailed assessment using accelerated forgetting measures may be appropriate.

Recommendations for Future Research

As discussed, the APP test appears to be a promising measure of ALF, however, neuropsychological test development is a long process. Therefore, there are lots of areas for development for this test and for future research. Some suggestions for future research have already been highlighted so for brevity they will only be mentioned briefly here.

Modifications to the task

Some modifications need to be made to some of the characters (e.g. distinctiveness of the hair colour, make some of the characters more recognisable). The cued-recall questions that involved these features were the ones that the majority of participants found most difficult to answer. Following these changes, it would be important to reassess reliability and evaluate whether this improves internal consistency coefficients. There is also some validation work to be undertaken around its cultural sensitivity. In addition, further work could consider methods to make the test less reliant on verbal encoding and recall and to include a recognition paradigm. This may incorporate advances in technology, such as the use of smartphone apps or email, as described earlier.

Further assessment of its psychometric properties

The APP test requires further assessment of its reliability and validity with a larger clinical sample. Predictions could be made about differences between those with right and left TLE and those with extratemporal lobe epilepsy. The test could be compared with existing tests, including the Crimes Test (thus providing assessment of convergent validity). Dual-task methodology could also be used to specify the degree of verbal involvement. Similarly, functional imaging studies could help determine the involvement of the left and right hippocampus in the task. A study could also be designed to compare the reliability and validity of face-to-face versus telephone testing, which could be combined with cost-effectiveness data. This could also consider the impact of retrieving information in the same or different context (context-dependent cues- Godden & Baddeley, 1975; Tulving, 1974).

Further testing of its use as a clinical tool

An interesting and important area for future development would involve the collection of a large, more representative normative sample. This normative data could then be used to determine clinically significant change and define ALF (cf. Miller et al., 2015). This may help to determine the length of delay that is necessary to detect ALF. Additionally, as it has been suggested that ALF has been associated with MCI, which is an early indicator for Alzheimer's Disease, further investigation with this clinical population may have important clinical implications for its earlier detection.

Further research in ALF

This study was not designed to explore the cognitive mechanisms underlying ALF but to develop a clinically feasible measure. However, studies of ALF are important not only as a way of understanding the memory problems experienced by PWE but they also provide important information about theories of memory functioning (Dickerson & Eichenbaum, 2010). The APP test, particularly with its practical value of telephone follow-ups, could be incorporated into future research studies.

Conclusions

In conclusion, the APP test appears to be a promising measure of ALF. While many of the findings need to be interpreted with caution due to the small sample size, lack of a matched comparison reference group, limited materials and the number of multiple comparisons, there are some interesting findings. Whilst there are some ceiling effects, the APP test is able to demonstrate forgetting in a healthy adult sample. It has modest levels of reliability and there is evidence to support its validity. In particular, SK, a patient with confirmed ALF, demonstrated ‘normal’ levels of learning and retention after a standard delay of 30 minutes but then experienced rapid forgetting of the information over a three week delay indicating that the test is able to detect nonverbal ALF using a telephone administration procedure. Importantly, participants found the telephone follow-ups acceptable and preferable to face-to-face testing, which suggest that the APP test shows some promise as a potential test of ALF for use in experimental exploration and clinical contexts. Based on this early work, the test is definitely worthy of further research and development.

REFERENCES

Alber, J., Della Sala, S. & Dewar, M. (2014). Minimizing interference with early consolidation boosts 7-day retention in amnesic patients. *Neuropsychology 28*(5), 667-675. doi:10.1037.neu0000091

Aldenkamp, A.P. & Bodde, N. (2005). Behaviour, cognition and epilepsy. *Acta Neurologica Scandinavica 112*, 19-25.

Aldenkamp, A.P., de Krom, M. & Reijs, R. (2003). Newer antiepileptic drugs and cognitive issues. *Epilepsia 44*, 21-29.

Alvarez, P. & Squire, L.R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Science USA 91*, 7041-7045.

Asadi-Pooya, A.A. (2014). Transient epileptic amnesia: A concise review. *Epilepsy & Behaviour 31*, 243-245.

Atherton, K. E., Nobre, A. C., Zeman, A. Z. & Butler, C. R. (2014). Sleep-dependent memory consolidation and accelerated forgetting. *Cortex 54(March)*, 92–105. doi:10.1016/j.cortex.2014.02.009

Atkinson, R. C. & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of Learning and Motivation - Advances in Research and Theory 2*, 89–195. doi:10.1016/S0079-7421(08)60422-3

Baddeley, A. (1986). *Working memory*. New York: Clarendon Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences 4(11)*, 417-432 doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders* 36, 189-208.

Baddeley, A. (2015a). What is memory? In Baddeley, A., Eysenck, M.W. & Anderson, M.C (Eds.), *Memory*. London: Psychology Press pp.3-22.

Baddeley, A. (2015b). Working Memory. In Baddeley, A., Eysenck, M.W. & Anderson, M.C (Eds.), *Memory*. London: Psychology Press pp.67-106.

Baddeley, A.D., Emslie, H. & Nimmo-Smith, I. (1994). *Doors and people: A test of visual and verbal recall and recognition.* Bury St Edmunds, Suffolk Thames Valley Test Company.

Baddeley, A. D. & Hitch, G. (1974). Working memory. In G.A. Bower (Ed) *Recent advances in learning and motivation.* Vol 8, Academic Press, New York, pp. 47–89. doi:10.1016/S0079-7421(08)60452-1

Baddeley, A.D. & Hitch, G.J. (2017). Is the levels of processing effect language-limited? *Journal of Memory and Language 92*, 1-13.

Baddeley, A. D., Hitch, G. & Allen, R.J. (2009). Working memory and binding in sentence recall. *Journal of Memory & Language 61,* 438-456.

Baddeley, A.D., Hitch, G.J., Quinlan, P.T., Bowes, L. & Stone, R. (2015). Doors for memory: A searchable database. *The Quarterly Journal of Experimental Psychology,* doi. 10.108D/1470218.2015.1087582

Baddeley, A., Rawlings, B. & Hayes, A. (2013). Constrained prose recall and the assessment of long-term forgetting: The case of ageing and the Crimes Test. *Memory* doi:10.1080/09658211.2013.865753

Baker, G., Jacoby, A., Buck, D., Stalgis, C. & Monnet, D. (1997). Quality of life of people with epilepsy: A European study. *Epilepsia 38*(3), 353–62.

Baker, G.A., Taylor, J. & Hermann, B. (2009). How can cognitive status predispose to psychological impairment? *Epilepsy & Behaviour 15*(*2 Suppl 1*), S31-S35.

Barkas, L., Redhead, E., Taylor, M., Shtaya, A., Hamilton, D.A. & Gray, W.P. (2012). Fluoxetine restores spatial learning but not accelerated forgetting in mesial temporal lobe epilepsy. *Brain 135,* 2358-2374.

Baxendale, S. A., Van Paesschen, W., Thompson, P. J., Connelly, A., Duncan, J. S., Harkness, W. F. & Shorvon, S. D. (1998). The relationship between quantitative MRI and neuropsychological functioning in temporal lobe epilepsy. *Epilepsia 39*, 158–166. doi:10.1111/j.1528-1157.1998.tb01353.x

Beatty, W.W. (2004). RBANS analysis of verbal memory in multiple sclerosis. *Archives of Clinical Neuropsychology 19(6),* 825-834.

Bell, B.D. (2006). WMS-III logical memory performance after a two-week delay in temporal lobe epilepsy and control groups. *Journal of Clinical and Experimental Neuropsychology 28,* 1435-1443.

Bell, B.D., Fine, J., Dow, C., Seidenberg, M. & Hermann, B.P. (2005). Temporal lobe epilepsy and the selective reminding test: The conventional 30-minute delay suffices. *Psychological Assessment 17,* 103-109.

Bell, B.D. & Giovagnoli, A.R. (2007). Recent innovative studies of memory in temporal lobe epilepsy. *Neuropsychology Review 17,* 455-476. doi:10.1007/s11065-007-9049-3

Bengner, T. & Malina, T. (2010). Long-term face memory as a measure of right temporal lobe function in TLE: The Alsterdorfer Faces Test. *Epilepsy Research 89,* 142-147.

Berg, A.T., Berkovic, S.F., Brodie, M.J., Buchhalter, J., Cross, H., van Emde Boas, W., Engel, J., French, J., Glauser, T.A., Mathern, G.W., Moshé, S.L., Nordli, D., Plouin, P. & Scheffer, I.E. (2010). Revised terminology and concepts for organization of seizures and epilepsies: Report of the ILAE Commission on classification and terminology, 2005-2009. *Epilepsia 51*, 676-685.

Blair, R.D.G. (2012). Temporal lobe epilepsy semiology. *Epilepsy Research and Treatment,* 1-10. doi:10.1155/2012/751510

Blake, R. V, Wroe, S. J., Breen, E. K. & McCarthy, R. A. (2000). Accelerated forgetting in patients with epilepsy: Evidence for an impairment in memory consolidation. *Brain: A Journal of Neurology 123 Pt 3*, 472–483. doi:10.1093/brain/123.3.472

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Quantitative Research in Psychology 3,* 77-101.

Brooks, B.L., Sherman, E.M.S., Iverson, G.L., Slick. D.J. & Strauss, E. (2011). Psychometric foundations for the interpretation of neuropsychological test results. In Schoenberg, M.R. & Scott, J.G. (Eds). *The little black book of neuropsychology: A syndrome-based approach*. Springer: New York pp. 893-922.

Brown, G.D.A. & Lewandowsky, S. (2010). Forgetting in memory models: Arguments against trace decay and consolidation failure. In Della Sala, S. (Ed). *Forgetting*. Psychology Press, New York, pp.49-76.

Brown, F.C., Roth, R.M., Saykin, A.J. & Beverly-Gibson, G. (2007). A new measure of visual location learning and memory: Development and psychometric properties for the Brown Location Test (BLT). *Clinical Neuropsychology 1(5),* 811-825.

Burgess, G.H. & Chadalavada (2016). Profound anterograde amnesia following routine anesthetic and dental procedure: A new classification of amnesia characterised by intermediate-to-late-stage consolidation failure? *Neurocase 22(1),* 84-94.

Butler, C.R., Bhaduri, A., Acosta-Cabronero, J., Nestor, P.J., Kapur, N., Graham, K.S., Hodges, J.R. & Zeman. A.Z. (2009). Transient epileptic amnesia: Regional brain atrophy and its relationship to memory deficits. *Brain 132,* 357-368.

Butler, C. R., Graham, K. S., Hodges, J. R., Kapur, N., Wardlaw, J. M. & Zeman, A. Z. J. (2007). The syndrome of transient epileptic amnesia. *Annals of Neurology 61*, 587–598. doi:10.1002/ana.21111

Butler, C., Kapur, N., Zeman, A., Weller, R. & Connelly, A. (2012). Epilepsy-related long-term amnesia: Anatomical perspectives. *Neuropsychologia 50,* 2973-2980.

Butler, C., Muhlert, N. & Zeman, A. (2010). Accelerated long term forgetting. In Della Sala, S. (2010). *Forgetting*. Psychology Press, New York, pp.211-238.

Butler, C. R., & Zeman, A. (2008). Recent insights into the impairment of memory in epilepsy: Transient epileptic amnesia, accelerated long-term forgetting and remote memory impairment. *Brain 131,* 2243-2263.

Cassel, A., Morris, R., Koutroumanidis, M. & Kopelman, M. (2016). Forgetting in temporal lobe epilepsy: When does it become accelerated? *Cortex 78,* 70-84.

Chapin, J.S., Busch, R.M., Naugle, R.I. & Najm, I.M. (2009). The Family Pictures of the WMS-III: Relationship to verbal and visual memory following temporal lobectomy for intractable epilepsy. *Journal of Clinical and Experimental Neuropsychology 31,* 498-504.

Coen, R. F. (2011). Neuropsychological assessment of mild cognitive impairment/prodromal Alzheimer's disease. *Aging Health 7(1),* 155–162. doi:10.2217/ahe.10.85

Cohen J (1988). *Statistical power analysis for the behavioural sciences*. (2nd edition). Hillsdale, NJ: Erlbaum.

Commission of the ILAE (1989). Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy. *Epilepsia 30*, 389-399.

Corballis, M.C. (2009). Comparing a single case with a control sample: Refinements and extensions. *Neuropsychologia 47*, 2687-2689.

Craig, M., Della Sala, S. & Dewar, M. (2014). Autobiographical thinking interferes with episodic memory consolidation. *PLOS One 9(4)*, 1-9.

Craik, F.I.M. & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behaviour 11*, 671-684.

Craik, F.I.M. & Rose, N.S, (2012). Memory encoding and aging: A neurocognitive perspective. *Neuroscience and Biobehavioural Reviews 36(7),* 1729-1739.

Crawford, J.R. & Garthwaite, P.H. (2002). Investigation of the single case in neuropsychology: Confident limits on the abnormality of test scores and test score differences. *Neuropsychologia 40*, 1196-1208. Accompanying program SINGLIMS.EXE. Retrieved from http://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodsComputerPrograms.HTM

Crawford & Garthwaite, P.H. (2006). Methods of testing for a deficit in single-case studies: Evaluation of statistical power by Monte Carlo simulation. *Cognitive Neuropsychology 23(6)*, 877-904.

Crawford, J.R. & Howell, D.C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist 12(4)*, 482-486.

Das Nair, R., Lincoln, N.B., Fitzsimmons, D., Brain, N., Montgomery, A., Bradshaw, L., Drummond, A., Sackley, C., Newby, G., Thornton, J., Stapleton, S. & Pink, A. (2015). Rehabilitation of memory following brain injury (ReMemBrIn): Study protocol for a randomised controlled trial. *Trials 16.*

De Renzi, E. & Luchelli, F. (1993). Dense retrograde amnesia, intact learning capability and abnormal forgetting rate: A consolidation deficit? *Cortex 29*, 449-466.

Deak, M.C., Stickgold, R., Pietras, A.C., Nelson, A.P. & Bubrick, E.J. (2011). The role of sleep in forgetting in temporal lobe epilepsy: A pilot study. *Epilepsy & Behaviour 21*, 462-466.

Della Sala, S. (2010). Preface. In Della Sala, S. (Ed). *Forgetting*. Psychology Press, New York, pp. xii-xiv.

Dewar, M., Alber, J., Butler, C., Cowan, N. & Della Sala, S. (2012). Brief wakeful resting boosts new memories over the long term. *Psychological Science 23(9),* 955-960.

Dewar, M., Alber, J., Cowan, N. & Della Sala, S. (2014). Boosting long-term memory via wakeful rest: Intentional rehearsal is not necessary, consolidation is sufficient. *PLOS One 9,* 1-10.

Dewar, M., Cowan, N. & Della Sala, S. (2010). Forgetting due to retroactive interference in amnesia: Findings and implications. In Della Sala, S. (2010). *Forgetting*. Psychology Press, New York, pp.185-211.

Dewar, M., Hoefeijzers, S., Zeman, A., Butler, C. & Della Sala, S. (2015). Impaired recognition in transient epileptic amnesia. *Epilepsy & Behaviour 42*, 107-116.

Dickerson, B.C., & Eichenbaum, H. (2010). The episodic memory system: Neurocircuitry and disorders. *Neuropsychopharmacology 35,* 86-104.

Diekelmann, S. & Born, J. (2010). The memory function of sleep. *Nature Reviews: Neuroscience 11*, 114-126.

Djordjevic, J. & Jones-Gotman, M. (2012). Neuropsychological assessment of memory in patients with epilepsy. In A. Zeman, N. Kapur, & M. Jones-Gotman (Eds.) *Epilepsy and memory*. Oxford: OUP, pp. 177–188.

Dodrill, C. (1986). Correlates of generalised tonic-clonic seizures with intellectual, neuropsychological, emotional and social function in patients with epilepsy. *Epilepsia 27*, 399-411.

Drane, E.S.M. (2012). *An exploration of the experience of living with epilepsy in later life*. Unpublished doctoral dissertation, University of Oxford, Oxford, UK.

Dudai, Y. (2004). The neurobiology of consolidation, or, how stable is the engram? *Annual Review of Psychology*, 55, 51-86.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. Oxford, Dover.

Elliott, G. (2010). *Accelerated long-term forgetting in temporal lobe epilepsy: Test development, a group comparison and case series analysis*. Unpublished doctoral dissertation, University of Sheffield, Sheffield, UK.

Elliott, G., Isaac, C. L., & Muhlert, N. (2014). Measuring forgetting: A critical review of accelerated long-term forgetting studies. *Cortex 54*, 16–32. doi:10.1016/j.cortex.2014.02.001

Engel, J. J. & Pedley, T. (2008). *Epilepsy: A comprehensive textbook*. Philadelphia: Lippincott, Williams & Wilkins.

Evans, S. J., Elliott, G., Reynders, H. & Isaac, C. L. (2014). Can temporal lobe epilepsy surgery ameliorate accelerated long-term forgetting? *Neuropsychologia 53*, 64–74. doi:10.1016/j.neuropsychologia.2013.11.007

Faul, F., Erdfelder, E., Lang, A.G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural and medical sciences. *Behaviour Research Methods 39,* 175-191.

Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., Forsgren, L., French, J., Glynn, M., Hesdorffer, D.C., Lee, B. I., Mathern, G.W., Moshé, S.L., Perucca, E., Scheffer, I.E., Tomson, T., Watanabe, M. & Wiebe, S. (2014). ILAE Official Report: A practical clinical definition of epilepsy. *Epilepsia 55,* 475–482. doi:10.1111/epi.12550

Fisher, R. S., van Emde, B. W., Blume, W., Elger, C., Genton, P., Lee, P. & Engel Jr, J. (2005). Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia 46*, 470-472.

Fisher, R. S., Vickrey, B. G., Gibson, P., Hermann, B., Penovich, P., Scherer, A. & Walker, S. (2000). The impact of epilepsy from the patient's perspective I: Descriptions and subjective perceptions. *Epilepsy Research 41*, 39–51. doi:10.1016/S0920-1211(00)00126-1

Fitzgerald, Z., Mohamed, A., Ricci, M., Thayer, Z. & Miller, L. (2013a). Accelerated long-term forgetting: A newly identified memory impairment in epilepsy. *Journal of Clinical Neuroscience : Official Journal of the Neurosurgical Society of Australasia 20(11),* 1486–91. doi:10.1016/j.jocn.2013.04.037

Fitzgerald., Z., Thayer, Z., Mohamed, A. & Miller, L.A. (2013b). Examining factors related to accelerated long-term forgetting in epilepsy using ambulatory EEG monitoring. *Epilepsia 54,* 819-827.

Frankland, P.W., Köhler, S. & Josselyn, S.A. (2013). Hippocampal neurogenesis and forgetting. *Trends in Neurosciences 36(9),* 497-503.

Gallassi, R., Sambati, L., Poda, R., Maserati, M.S., Oppi, F., Guilioni, M. & Tinuper, P. (2011). Accelerated long-term forgetting in temporal lobe epilepsy: Evidence of improvement after left temporal pole lobectomy. *Epilepsy & Behaviour 22,* 793-796.

Gascoigne, M.B., Smith, M.L., Barton, B., Webster, R., Gill, D. & Lah, S. (2014). Accelerated long-term forgetting in children with temporal lobe epilepsy. *Neuropsychologia 59*, 93-102.

Gastaut, H. (1973). *Dictionary of epilepsy. Part I: Definitions.* World Health Organisation.

Giovagnoli, A.R., Casazza, M. & Avanzini, G. (1995). Visual learning on a selective reminding procedure and delayed recall in patients with temporal lobe epilepsy. *Epilepsia 36*, 704-711.

Giovagnoli, A. R., Mascheroni, S. & Avanzini, G. (1997). Self-reporting of everyday memory in patients with epilepsy: Relation to neuropsychological, clinical, pathological and treatment factors. *Epilepsy Research 28*, 119–128. doi:10.1016/S0920-1211(97)00036-3

Godden, D.R. & Baddeley, A.D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology 66(3),* 325-331.

Goldstein, L. H. & Polkey, C. E. (1992). Behavioural memory after temporal lobectomy or amygdalo-hippocampectomy. *The British Journal of Clinical Psychology 31(Pt 1)*, 75–81.

Guerts, S., van der Werf, S.P. & Kessels, R.P.C. (2015). Accelerated forgetting? An evaluation on the use of long-term forgetting rates in patients with memory problems. *Frontiers in Psychology 6*, 1-9.

Hall, K. E., Isaac, C. L. & Harris, P. (2009). Memory complaints in epilepsy: An accurate reflection of memory impairment or an indicator of poor adjustment? A review of the literature. *Clinical Psychology Review 29(4)*, 354–67. doi:10.1016/j.cpr.2009.03.001

Hardt, O., Nader, K. & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences 17(3),* 111-120.

Hermann, B.P, Seidenberg, M., Lee, E. J., Chan, F. & Rutecki, P. (2007). Cognitive phenotypes in temporal lobe epilepsy. *Journal of the International Neuropsychological Society 13*, 12-20.

Hermann, B. P., Seidenberg, M., Schoenfeld, J. & Davies, K. (1997). Neuropsychological characteristics of the syndrome of mesial temporal lobe epilepsy. *Archives of Neurology 54*, 369–376. doi:10.1001/archneur.1997.00550160019010

Hill-Briggs, F., Dial, J.G., Morere, D.A. & Joyce, A. (2007). Neuropsychological assessment of persons with physical disability, visual impairment or blindness, and hearing impairment or deafness. *Archives of Clinical Neuropsychology 22,* 389-404.

Hodges, S., Berry, E. & Wood, K. (2011). SenseCam: A wearable camera that stimulates and rehabilitates autobiographical memory. *Memory*. doi:10.1080/09658211.2011.605591

Hoefeijzers, S., Dewar, M., Della Sala, S., Butler, C. & Zeman, A. (2014). Accelerated long-term forgetting can become apparent within 3-8 hours of wakefulness in patients with transient epileptic amnesia. *Neuropsychology, 29*, 117-125.

Hoefeijzers, S., Dewar, M., Della Sala, S., Zeman, A. & Butler, C. (2013). Accelerated long-term forgetting in transient epileptic amnesia: An acquisition or consolidation deficit? *Neuropsychologia, 51*, 1549-1555.

Holler, Y. & Trinka, E. (2014). What do temporal lobe epilepsy and progressive mild cognitive impairment have in common? *Frontiers in Systems Neuroscience 8,* 1-8.

Huijgen, J. & Samson, S. (2015). The hippocampus: A central node in a large-scale brain network for memory. *Revue Neurologique 171*, 204-216.

INVOLVE (2012). *Briefing notes for researchers: Involving the public in NHS, public health and social care research*. INVOLVE, Eastleigh.

Isaac, C.L. & Mayes, A.R. (1999). Rate of forgetting in amnesia: I. Recall and recognition of prose. *Journal of Experimental Psychology: Learning, Memory and Cognition 25,* 942-962.

Iverson, G.L. (2001). Interpreting change on the WAIS-III/WMS-III in clinical samples. *Archives of Clinical Neuropsychology 16,* 183-191.

Jacoby, A, Gamble, C., Doughty, J., Marson, A. & Chadwick, D. (2007). Quality of life outcomes of immediate or delayed treatment of early epilepsy and single seizures. *Neurology*, *68*(15), 1188–96. doi:10.1212/01.wnl.0000259411.78423.50

Jansari, A.S., Davis, K., McGibbon, T., Firminger, S. & Kapur, N. (2010). When ‘long-term memory’ no longer means ‘forever’: Analysis of accelerated long-term forgetting in a patient with temporal lobe epilepsy. *Neuropsychologia 48*, 1707-1715.

JASP Team (2016). JASP (Version 0.7.5.5)[Computer software]. Retrieved from https://.jasp-stats.org.

Joint Epilepsy Council (2011). *Epilepsy prevalence, incidence and other statistics* (pp. 1–13). Leeds, UK.

Jokeit, H., Krämer, G. & Ebner, A. (2005). Do antiepileptic drugs accelerate forgetting? *Epilepsy & Behaviour 6*, 430-432.

Josselyn, S.A. & Frankland, P.W. (2012). Infantile amnesia: A neurogenic hypothesis. *Learning & Memory 19*, 423-433.

Karpike, J.D. & Roediger III, H.L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language 57*, 151-162.

Kemp, S., Illman, N.A., Moulin, C.J.A. & Baddeley, A.D. (2012). Accelerated long-term forgetting (ALF) and transient epileptic amnesia (TEA): Two cases of epilepsy-related memory disorder. *Epilepsy & Behaviour 24*, 382-388.

Kwan, P. & Sander, J. W. (2004). The natural history of epilepsy: An epidemiological view. *Journal of Neurology, Neurosurgery, and Psychiatry 75*, 1376–1381. doi:10.1136/jnnp.2004.045690

Ladowsky-Brooks, R.L. (2015). Four-hour delayed memory recall for stories: Theoretical and clinical implications of measuring accelerated long-term forgetting. *Applied Neuropsychology: Adult* doi.10.1080/23279095.2015.1030670

Lah, S., Mohamed, A., Thayer, Z., Miller, L. & Diamond, K. (2014). Accelerated long-term forgetting of verbal information in unilateral temporal lobe epilepsy: Is it related to structural hippocampal abnormalities and/or incomplete learning? *Journal of Clinical and Experimental Neuropsychology 36*, 158-169.

Lance, C.E., Butts, M.M. & Michels, L.C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organisational Research Methods 9*, 202-220.

Lee-Donaldson, K. (2011). *Rehabilitation of accelerated long-term forgetting using external memory aids: An investigation of a diary and a wearable automatic camera.* Unpublished doctoral dissertation, University of Sheffield, Sheffield, UK.

Leritz, E.C., Grande, L.J. & Bauer, R.M. (2006). Temporal lobe epilepsy as a model to understand human memory: The distinction between explicit and implicit memory. *Epilepsy & Behaviour 9*, 1-13. doi:10.1016/j.yebeh.2006.04.012

Loring, D.W. & Meador, K.J. (2001). Cognitive and behavioural effects of epilepsy treatment. *Epilepsia 42*, 24-32.

Ludowig, E., Möller, J., Bien, C.G., Münte, T.F., Elger, C.E. & Rosburg, T. (2010). Active suppression in the mediotemporal lobe during directed forgetting. *Neurobiology of Learning and Memory 93*, 352-361.

Lum, J.A.G. & Conti-Ramsden (2013). Long-term memory: A review and meta-analysis of studies of declarative and procedural memory in specific language impairment. *Topics in Language Disorders 33(4)*, 282-297.

MacDonald, S.W.S., Stigsdotter-Neely, A., Derwinger, A. & Bäckman, L. (2006). Rate of acquisition, adult age, and basic cognitive abilities predict forgetting: New views on a classic problem. *Journal of Experimental Psychology: General 135(3)*, 368-390.

Majdan, A., Sziklas, V. & Jones-Gotman, M. (1996). Performance of healthy subjects and patients with resection from the anterior temporal lobe on matched tests of verbal and visuoperceptual learning. *Journal of Clinical and Experimental Neuropsychology 18*, 416-460.

Mameniskiene, R., Jatuzis, D., Kaubrys, G. & Budrys, V. (2006). The decay of memory between delayed and long-term recall in patients with temporal lobe epilepsy. *Epilepsy & Behaviour 8*, 278-288.

Manes, F., Graham, K.S., Zeman, A., de Lújan Calcagno, M. & Hodges, J.R. (2005). Autobiographical amnesia and accelerated forgetting in transient epileptic amnesia. *Journal of Neurology, Neurosurgery and Psychiatry 76,* 1387-1391.

Manes, F., Serrano, C., Calcagno, M.L., Cardozo, J. & Hodges, J. (2008). Accelerated forgetting in subjects with memory complaints: A new form of Mild Cognitive Impairment? *Journal of Neurology 255*, 1067-1070.

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

Marshall, L. & Born, J. (2007). The contribution of sleep to hippocampus dependent memory consolidation. *Trends in Cognitive Science 11*, 442-450.

Martin, R.C., Loring, D.W., Meador, K.J., Lee. G.P., Thrash, N. & Arena, J.G. (1991). Impaired long-term retention despite normal verbal learning in patients with temporal lobe dysfunction. *Neuropsychology 5*, 3-12.

Mary, A., Schreiner, S. & Peigneux, P. (2013). Accelerated long-term forgetting in aging and intra-sleep awakenings. *Frontiers in Psychology 4,* 1-11.

Mayes, A.R., Isaac, C.L., Holdstock, J.S., Cargia, P., Gummer, A. & Roberts, N. (2003). Long-term amnesia: A review and detailed illustrative case study. *Cortex, 39,* 567-603.

McGibbon, T. & Jansari, A.S. (2013). Detecting the onset of accelerated long-term forgetting: Evidence from temporal lobe epilepsy. *Neuropsychologia 51*, 114-122.

McKay, C., Casey, J.E., Wertheimer, J. & Fichtenberg, N.L. (2007). Reliability and validity of the RBANS in a traumatic brain injured sample. *Archives of Clinical Neuropsychology 22(1)*, 91-98.

Meador, K. J. (2002). Cognitive outcomes and predictive factors in epilepsy. *Neurology 58,* S21-S26.

Miller, L.A., Flanagan, E., Mothakunnel, A., Mohamed, A. & Thayer, Z. (2015). Old dogs with new tricks: Detecting accelerated long-term forgetting by extending traditional measures. *Epilepsy & Behaviour 45,* 205-211.

Milner, B., Corkin, S., & Teuber, H. L. (1968). Further analysis of the hippocampal amnesic syndrome: 14 year follow-up study of HM. *Neuropsychologia 6*, 215-234.

Mitsis, E.M., Jacobs, D., Luo, X., Andrews, H., Andrews, K. & Sano, M. (2009). Evaluating cognition in an elderly cohort via telephone assessment. *International Journal of Geriatric Psychiatry 25,* 1099-1166.

Muhlert, N. (2012). The enigma of long-term forgetting. *Seizure 21,* 77-78.

Muhlert, N., Grünewald, R., Hunkin, N. M., Reuber, M., Howell, S., Reynders, H. & Isaac, C. L. (2011). Accelerated long-term forgetting in temporal lobe but not idiopathic generalised epilepsy. *Neuropsychologia*, *49*(9), 2417–26. doi:10.1016/j.neuropsychologia.2011.04.018

Muhlert, N., Milton, F., Butler, C.R., Kapur, N. & Zeman, A.Z. (2010). Accelerated forgetting of real-life events in transient epileptic amnesia. *Neuropsychologia, 48,* 3235-3244.

Mula, M. & Trimble, M.R. (2009). Antiepileptic drug-induced cognitive adverse effects: Potential mechanisms and contributing factors. *CNS Drugs 23,* 121-137.

Mula, M., Trimble, M.R. & Sander, J.W. (2003). The role of hippocampal sclerosis in topiramate-related depression and cognitive deficits in people with epilepsy. *Epilepsia 44,* 1573-1577.

Nadel, L. & Moscovitch, M. (1997). Memory consolidation, retrograde amnesia and the hippocampal complex. *Current Opinion in Neurobiology 7(2),* 217-227.

Nader, K. & Hardt, O. (2009). A single standard for memory: The case for reconsolidation. *Nature Reviews: Neuroscience 10(3)*, 224-234.

Narayanan, J., Duncan, R., Greene, J., Leach, J.P., Razvi, S., McLean, J. & Evans, J.J. (2012). Accelerated long-term forgetting in temporal lobe epilepsy: Verbal, nonverbal and autobiographical memory. *Epilepsy & Behaviour 25,* 622-630.

NICE (2012). *Epilepsies: The diagnosis and management of the epilepsies in adults and children in primary and secondary care*. Retrieved from https://www.nice.org.uk/Guidance/CG137.

Oakhill, J., Garnham, A. & Reynolds, D. (2005). Immediate activation of stereotypical gender information. *Memory and Cognition 33,* 972-983.

O'Connor, M., Sieggreen, M.A., Ahern, G., Schomer, D. & Mesulam, M. (1997). Accelerated forgetting in association with temporal lobe epilepsy and paraneoplastic encephalitis. *Brain and Cognition 35*, 71-84.

Panelli, R. J., Kilpatrick, C., Moore, S. M., Matkovic, Z., D'Souza, W. J. & O'Brien, T. J. (2007). The Liverpool Adverse Events Profile: Relation to AED use and mood. *Epilepsia 48*(3), 456–63. doi:10.1111/j.1528-1167.2006.00956.x

Paradiso, S., Hermann, B. P., Blumer, D., Davies, K., & Robinson, R. G. (2001). Impact of depressed mood on neuropsychological status in temporal lobe epilepsy. *Journal of Neurology, Neurosurgery, and Psychiatry, 70*, 180–185. doi:10.1136/jnnp.70.2.180

Parikh, M., Grosch, M.C., Graham, L.L., Hynan, L.S., Weiner, M., Shore, J.H. & Cullum, M. (2013). Consumer acceptability of brief videoconference-based neuropsychological assessment in older individuals with and without cognitive impairment. *The Clinical Neuropsychologist 27(5)*, 808-817.

Peterson, L.R. & Peterson, J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology 58(3)*, 193-198.

Randolph, C., Tierney, M. C., Mohr, E., & Chase, T. N. (1998). The Repeatable Battery for the Assessment of Neuropsychological Status (RBANS): Preliminary clinical validity. *Journal of Clinical and Experimental Neuropsychology 20*, 310–319. doi:10.1076/jcen.20.3.310.823

Ricci, M., Mohamed, A., Savage, G., Boserio, J. & Miller, L.A. (2015b). The impact of epileptiform abnormalities and hippocampal lesions on retention of recent autobiographical experiences: Adding insult to injury? *Neuropsychologia 66*, 259-266.

Ricci, M., Mohamed, A., Savage, G. & Miller, L.A. (2015a). Disruption of learning and long-term retention of prose passages in patients with focal epilepsy. *Epilepsy & Behaviour 51*, 104-111.

Roediger, H.L., Weinstein, Y. & Agarwal, P.K. (2010). Forgetting: Preliminary considerations. In Della Sala, S. (Ed). *Forgetting*. Psychology Press, New York, pp.23-34.

Royle, J. & Lincoln, N. B. (2008). The Everyday Memory Questionnaire-revised: Development of a 13-item scale. *Disability and Rehabilitation 30*, 114–121. doi:10.1080/09638280701223876

Sadeh, T., Ozubko, J.D., Winocur, G. & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Science 18(1)*, 26-36.

Salas-Puig, J., Gil-Nagel, A, Serratosa, J. M., Sánchez-Alvarez, J. C., Elices, E., Villanueva, V. & Porcel, J. (2009). Self-reported memory problems in everyday activities in patients with epilepsy treated with antiepileptic drugs. *Epilepsy & Behaviour 14*(4), 622–7. doi:10.1016/j.yebeh.2009.01.015

Scheffer, I.E., Berkovic, S.F., Capovilla, G., Connolly, M.B., Guilhoto, L., Hirsch, E., Moshé, S.L., Nordli, D., Zhang, Y. & Zuberi, S.M. (2014). *The Organisation of the Epilepsies: Report of the ILAE Commission on Classification and Terminology*. Retrieved from ttp://www.ilae.org/Visitors/Centre/Organization.cfm

Schmidt, S.R. (2008). Distinctiveness and memory: A theoretical and empirical review. *Learning and Memory: A Comprehensive Reference 2*, 125-144.

Scoville, W. B. & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry 20*, 11-21.

Sherman, E.M.S., Brooks, B.L., Iverson, G.L., Slick, D.J. & Strauss, E. (2011). Reliability and validity in neuropsychology. In Schoenberg, M.R. & Scott, J.G. (Eds). *The little black book of neuropsychology: A syndrome-based approach* Springer Science, pp. 873-892. doi 10.1007/978-0-387-76978-3_30.

Smith, C.N., Frascino, J.C., Kripke, D.L., McHugh, P.R., Treisman, G.J. & Squire, L.R. (2010). Losing memories overnight: A unique form of human amnesia. *Neuropsychologia 48*, 2833-2840.

Squire, L.R. (1987). *Memory and brain*. New York: Oxford University

Squire, L.R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory 8*, 171-177.

Squire, L.R. & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: A neurobiological perspective. *Current Opinion in Neurobiology 5*, 169-177.

Stokes, T., Shaw, E. J., Juarez-Garcia, A., Camosso-Stefinovic, J. & Baker, R. (2004). *Clinical guidelines and evidence review for the epilepsies: Diagnosis and management in adults and children in primary and secondary care.* London: Royal College of General Practitioners.

Strauss, E. H., Sherman, E. M. S. & Spreen, O. (2006). *A compendium of neuropsychological tests*. Oxford University Press, USA.

Strauss, M.E. & Smith, G.T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology 5*, 1-25.

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behaviour*. doi:10.1016/S0022-5371(83)90229-3

Taichman, D., Christie, J., Biester, R., Mortensen, J., White, J., Kaplan, S., Hansen-Flaschen, J., Palevsky, H.I., Elliott, C.G. & Hopkins, R.O. (2005). Validation of a brief telephone battery for neurocognitive assessment of patients with pulmonary arterial hypertension. *Respiratory Research 6(1)*, 39-45.

Thompson, P. J. & Corcoran, R. (1992). Everyday memory failures in people with epilepsy. *Epilepsia 33 Suppl 6*, S18–S20.

Trahan, D.E. (1992). Analysis of learning and rate of forgetting in age-associated memory differences. *The Clinical Neuropsychologist 6(2)*, 241-246.

Tramoni, E., Felician, O., Barbeau, E.J., Guedj, E., Guye, M., Bartolomei, F. & Ceccaldi, M. (2011). Long-term consolidation of declarative memory: Insight from temporal lobe epilepsy. *Brain 134*, 816-831.

Tulving, E. (1972). Episodic and semantic memory. *Organization of Memory, 1* 381–403. doi:10.1017/S0140525X00047257

Tulving, E. (1974). Cue-dependent forgetting. *American Scientist, 62*, 74-82.

Tulving, E. & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behaviour 5(4),* 381-391.

Unverzagt, F.W., Monahan, P.O., Moser, L.R., Zhao, Q., Carpenter, J.S., Sledge, G.W. & Champion, V.L. (2007). The Indiana University telephone-based assessment of neuropsychological status: A new method for large-scale neuropsychological assessment. *Journal of the International Neuropsychological Society 13(5),* 799-806.

Van der Werf, S.P., Guerts, S. & de Werd, M.M.E. (2016). Subjective memory ability and long-term forgetting in patients referred for neuropsychological assessment. *Frontiers in Psychology 7,* 1-8.

Walsh, C.M., Wilkins, S., Bettcher, B.M., Butler, C.R., Miller, B.L., & Kramer, J.H. (2014). Memory consolidation in aging and MCI after 1 week. *Neuropsychology, 28,* 273-280.

Watkins, O.C. & Watkins, M.J. (1975). Build-up of proactive inhibition as a cue overload effect. *Journal of Experimental Psychology: Human Learning and Memory 104,* 442-452.

Wechsler, D. (2010). *Wechsler Memory Scale-Fourth UK Edition (WMS-IV UK).* Pearson, UK.

Wechsler, D. (2011). *Test of Premorbid Functioning UK Edition (TOPF UK): Administration, Scoring and Techniques Manual*. Pearson, UK.

Wilkinson, H., Holdstock., J.S., Baker, G., Herbert, A., Clague, F. & Downes, J.J. (2012). Long-term accelerated forgetting of verbal and non-verbal information in temporal lobe epilepsy. *Cortex 48,* 317-332.

Winocur, G., Moscovitch, M. & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal-neocortical interactions. *Neuropsychologia* 48, 2239-2356.

Witt, J., Glöckner, C. & Helmstaedter, C. (2012). Extended retention intervals can help to bridge the gap between subjective and objective memory impairment. *Seizure 21*, 134-140.

Wixted, J.T. (2004). On common ground: Jost's (1897) law of forgetting and Ribot's (1881) law of retrograde amnesia. *Psychological Review 111(4)*, 864-879.

Wixted, J.T. (2010). The role of retroactive interference and consolidation in everyday forgetting. In Della Sala, S. (2010). *Forgetting*. Psychology Press, New York, pp.285-313.

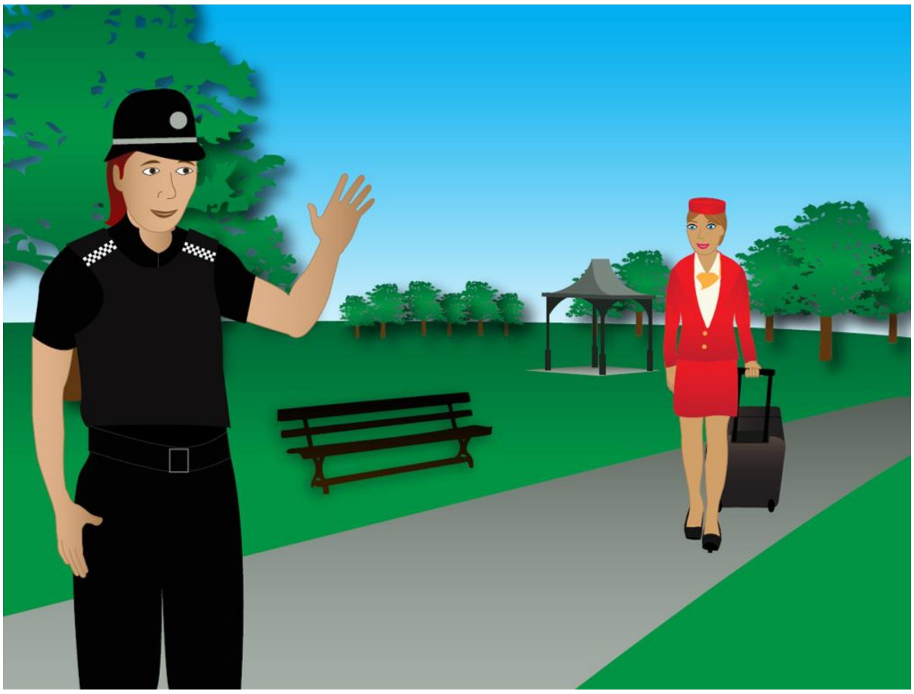
Zeman, A., Butler, C., Muhlert, N., & Milton, F. (2013). Novel forms of forgetting in temporal lobe epilepsy. *Epilepsy & Behaviour 26*, 335-342.

Zigmond, A S. & Snaith, R. P. (1983). The hospital anxiety and depression scale (HADS). *Acta Psychiatrica Scandinavica 67,* 361–370. doi:10.1016/S0016-5085(01)83173-5

APPENDIX A INITIAL SCENES OF THE ACTION-PEOPLE-PLACES (APP) TEST

Park scene

Police officer (actor) waving (action) at the air hostess (recipient)



Station scene

Soldier (actor) pointing (action) at the nurse (recipient)

Bar scene

Doctor (actor) following (action) the builder (recipient)

Street scene

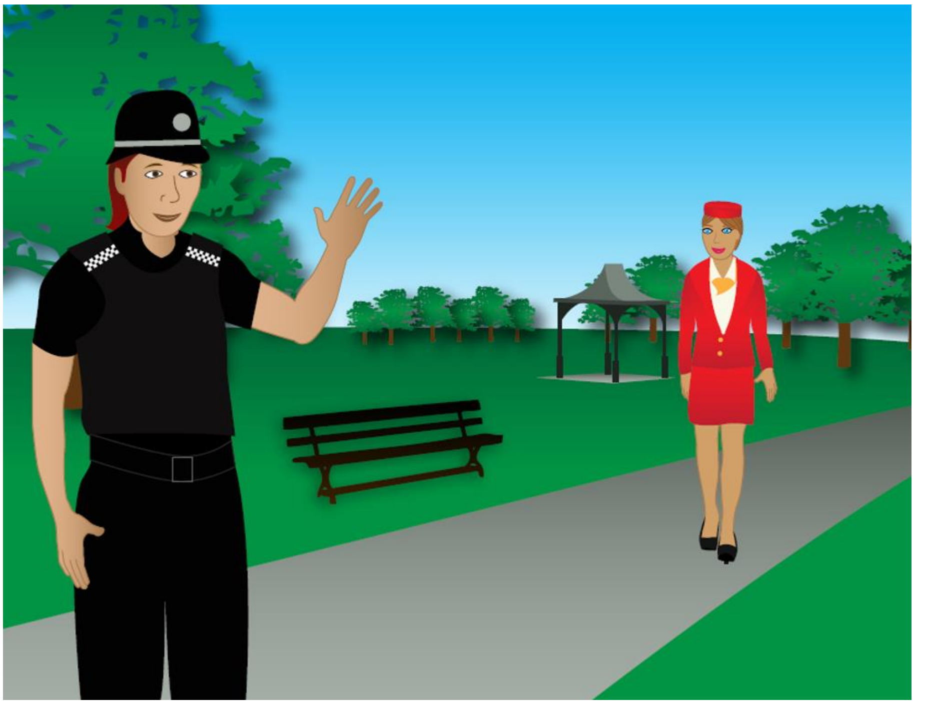
Fireman (actor) shouting (action) to the priest (recipient)

APPENDIX B FINAL SCENES OF THE APP TEST

Example item

Park scene

Police officer (actor) waving (action) at the air hostess (recipient)



Station scene

Soldier (actor) pointing (action) at the nurse (recipient)

Street scene

Fireman (actor) shouting (action) to the priest (recipient)

APPENDIX C FINAL APP TEST RECALL QUESTIONS

Immediate delay

	Question	Answer	Response	Scoring
1	Where was the fireman?	*Street*		
2	Who was the actor at the station?	*Soldier*		
3	Who was with the police officer?	*Air stewardess*		
4	Who had short black hair?	*Fireman*		
5	What colour is the soldier's hair?	*Blonde*		
6	What action was being performed in the park?	*Waving*		
7	Who had grey hair?	*Priest*		
8	Someone was with the air stewardess. What colour hair did they have?	*Red*		
9	What was happening to the priest?	*Shouted at*		
10	Who was being pointed at?	*Nurse*		
11	Someone had red hair. What colour hair did the other person in that scene have?	*(Light) brown*		
12	What colour is the nurse's hair?	*Black*		
13	Who was waving?	*Police officer*		
14	What colour hair did the actor in the street have?	*Black*		
15	What was happening to the person with long black hair?	*Pointed at*		

Total ________

Standard delay: 20-30 minutes

	Question	Answer	Response	Scoring
1	Who was the actor in the street?	*Fireman*		
2	Who was with the nurse?	*Soldier*		
3	What colour is the priest's hair?	*Grey*		
4	Where was the police officer?	*Park*		
5	What colour is the fireman's hair?	*Black*		
6	Someone had light brown hair. What colour hair did the other person in that scene have?	*Red*		
7	What was happening to the person with grey hair?	*Shouted at*		
8	What was the soldier doing?	*Pointing*		
9	What was happening to the air stewardess?	*Waved at*		
10	Someone had short blonde hair. What colour hair did the other person in that scene have?	*Black*		
11	Someone was being waved at. What colour hair did they have?	*(Light) brown*		
12	Who was the recipient at the station?	*Nurse*		
13	Where was the person with short black hair?	*Street*		
14	Who was with the person with red hair?	*Air stewardess*		
15	In which scene was someone pointing?	*Station*		

Total

24 hour: Telephone follow-up

	Question	Answer	Response	Scoring
1	Who was with the air stewardess?	*Police officer*		
2	Who was being shouted at?	*Priest*		
3	Who was pointing?	*Soldier*		
4	What was the police officer doing?	*Waving*		
5	Who was with the person with grey hair?	*Fireman*		
6	Someone was with the nurse. What colour hair did they have?	*Blonde*		
7	Who was the recipient in the street?	*Priest*		
8	What action was being performed at the station?	*Pointing*		
9	Where was someone waving?	*Park*		
10	Who was with the soldier?	*Nurse*		
11	What was the person with short black hair doing?	*Shouting*		
12	Where was the person with long black hair?	*Station*		
13	What colour hair did the recipient in the park have?	*Light brown*		
14	Where was the priest?	*Street*		
15	Who had light brown hair?	*Air stewardess*		

Total ____________

1 week: Telephone follow-up

	Question	Answer	Response	Scoring
1	Who was shouting?	*Fireman*		
2	Where was the air stewardess?	*Park*		
3	Someone was with the priest. What colour hair did they have?	*Black*		
4	Who had long black hair?	*Nurse*		
5	Someone was waving. What colour hair did they have?	*Red*		
6	Who had short blonde hair?	*Soldier*		
7	What action was being performed in the street?	*Shouting*		
8	Who was the recipient in the park?	*Air stewardess*		
9	Someone was being pointed at. What colour hair did they have?	*Black*		
10	Who was with the fireman?	*Priest*		
11	What was happening to the nurse?	*Pointed at*		
12	What was happening to the person with light brown hair?	*Being waved at*		
13	Where was the soldier?	*Station*		
14	Where was the person with grey hair?	*Street*		
15	Someone was with the police officer. What colour hair did they have?	*(Light) brown*		

Total ________

3-week: Telephone follow-up

	Question	Answer	Response	Scoring
1	Where was the nurse?	*Station*		
2	Where was someone shouting?	*Street*		
3	Who was the actor in the park?	*Police officer*		
4	What was the person with short blonde hair doing?	*Pointing*		
5	Someone had grey hair. What colour hair did the other person in that scene have?	*Black*		
6	Who was with the person with short blonde hair?	*Nurse*		
7	What was the fireman doing?	*Shouting*		
8	Where was the person with red hair?	*Park*		
9	What colour hair did the recipient in the street have?	*Grey*		
10	Who was being waved at?	*Air stewardess*		
11	Someone was with the soldier. What colour hair did they have?	*Black*		
12	What colour is the police officer’s hair?	*Red*		
13	Who was with the person with long black hair?	*Soldier*		
14	Who was with the priest?	*Fireman*		
15	What colour is the air stewardess’ hair?	*(Light) brown*		

Total ________

APPENDIX D SEIZURE DIARY

The Leeds Teaching Hospitals NHS
NHS Trust

UNIVERSITY OF LEEDS

The Development of a Non-Verbal Test of Long-Term Memory for Use with People with Epilepsy

Investigators: Joanne Crossley, Richard Allen, Steven Kemp

Seizure diary

As part of the research study, it would be helpful if you could record the details of your seizures since completing the initial assessment**.**

How to use this diary

Pages 2-3 are for you to record the details of your seizures. Use one line per day. The days are numbered from the initial assessment.

If you have lots of seizures in a day, there are some additional lines on page 4 if you need more space. But please remember to add the date.

You will not be asked to hand this diary back into the research team but Joanne Crossley (researcher) will ask you for this information when she carries out the telephone follow-ups. We would be really grateful, if you could have this information available when she contacts you.

Thank you for taking the time to complete this diary

Seizure diary

Day/Date	No of seizures	Time of seizure(s)	Awake or asleep	Type of seizure(s) (if known)
0 (day of assessment)				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

15				
16				
17				
18				
19				
20				
21				

Please continue on the next page if you need more space

APPENDIX E SUPPLEMENTARY DATA

Table A.1: Proportion of correct responses at immediate delay (Trial I)

Question		**% of correct responses**
1	Where was the fireman?	78
2	Who was the actor at the station?	81
3	Who was with the police officer?	75
4	Who had short black hair?	38
5	What colour is the soldier's hair?	66
6	What action was being performed in the park?	72
7	Who had grey hair?	78
8	Someone was with the air stewardess. What colour hair did they have?	56
9	What was happening to the priest?	59
10	Who was being pointed at?	69
11	Someone had red hair. What colour hair did the other person in that scene have?	44
12	What colour is the nurse's hair?	53
13	Who was waving?	59
14	What colour hair did the actor in the street have?	44
15	What was happening to the person with long black hair?	59

Table A.2: Proportion of correct responses at standard delay

Question		**% of correct responses**
1	Who was the actor in the street?	97
2	Who was with the nurse?	91
3	What colour is the priest's hair?	94
4	Where was the police officer?	97
5	What colour is the fireman's hair?	75
6	Someone had light brown hair. What colour hair did the other person in that scene have?	63
7	What was happening to the person with grey hair?	88
8	What was the soldier doing?	91
9	What was happening to the air stewardess?	97
10	Someone had short blonde hair. What colour hair did the other person in that scene have?	63
11	Someone was being waved at. What colour hair did they have?	69
12	Who was the recipient at the station?	97
13	Where was the person with short black hair?	81
14	Who was with the person with red hair?	78
15	In which scene was someone pointing?	88

Table A.3: Proportion of correct responses at 24 hour delay

Question		% of correct responses
1	Who was with the air stewardess?	87
2	Who was being shouted at?	97
3	Who was pointing?	84
4	What was the police officer doing?	90
5	Who was with the person with grey hair?	90
6	Someone was with the nurse. What colour hair did they have?	58
7	Who was the recipient in the street?	97
8	What action was being performed at the station?	84
9	Where was someone waving?	94
10	Who was with the soldier?	84
11	What was the person with short black hair doing?	74
12	Where was the person with long black hair?	87
13	What colour hair did the recipient in the park have?	52
14	Where was the priest?	90
15	Who had light brown hair?	61

Table A.4: Proportion of correct responses at 1 week delay

Question		% of correct responses
1	Who was shouting?	88
2	Where was the air stewardess?	97
3	Someone was with the priest. What colour hair did they have?	72
4	Who had long black hair?	84
5	Someone was waving. What colour hair did they have?	41
6	Who had short blonde hair?	81
7	What action was being performed in the street?	91
8	Who was the recipient in the park?	100
9	Someone was being pointed at. What colour hair did they have?	59
10	Who was with the fireman?	91
11	What was happening to the nurse?	91
12	What was happening to the person with light brown hair?	78
13	Where was the soldier?	91
14	Where was the person with grey hair?	94
15	Someone was with the police officer. What colour hair did they have?	53

Table E.5: Proportion of correct responses at 3 week delay

	Question	**% of correct responses**
1	Where was the nurse?	90
2	Where was someone shouting?	90
3	Who was the actor in the park?	77
4	What was the person with short blonde hair doing?	58
5	Someone had grey hair. What colour hair did the other person in that scene have?	74
6	Who was with the person with short blonde hair?	65
7	What was the fireman doing?	87
8	Where was the person with red hair?	90
9	What colour hair did the recipient in the street have?	87
10	Who was being waved at?	90
11	Someone was with the soldier. What colour hair did they have?	61
12	What colour is the police officer’s hair?	36
13	Who was with the person with long black hair?	71
14	Who was with the priest?	87
15	What colour is the air stewardess’ hair?	48