An Investigation of the Reliability of the Addenbrooke's Cognitive Examination – Revised (ACE-R)

and

Clinical Research Portfolio

VOLUME I

(Volume II Bound Separately)

Stephanie Crawford

July 2010

Section of Psychological Medicine
Division of Community Based Sciences

Submitted in part fulfilment of the requirements for the Degree of

Doctor in Clinical Psychology

Acknowledgements

First and foremost I would like to thank my research supervisor Professor Jonathan Evans for his unfailing patience, wisdom and enthusiasm throughout my training. My sincerest thanks go also to my field supervisors, Dr Joanne Robertson and Dr Leigh Whitnall for their advice, guidance and reassuring words. A particular thank you goes also to the University of Glasgow Media Services for their involvement in developing my research vignettes, their professional expertise was invaluable. My thanks too to the three actors, Elsie, Ray and Pene, who took part in my vignettes; their performances made such a significant contribution to the quality of the vignettes. I would like to thank all of the older adult community mental health teams in Greater Glasgow for their cooperation throughout my study. My thanks also go to the nurses and trainee clinical psychologists who took the time to participate in my research. And finally my biggest and most grateful thank you, for their ongoing support throughout my training, goes to my wonderful family and friends.

Volume I

Table of Contents **Page**

Chapter 1: Systematic literature review

A systematic review of the accuracy and clinical utility of the Addenbrooke's Cognitive Examination and the Addenbrooke's Cognitive Examination - Revised

Stephanie Crawford[1*]

[1]Section of Psychological Medicine, University of Glasgow

[*]Address for Correspondence
Section of Psychological Medicine
Division of Community Based Sciences
University of Glasgow
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XY
E-mail: scrawford@nhs.net

Key words: Dementia, Screening, Diagnosis, Mild Cognitive Impairment

Word count: 6317

Prepared in accordance with requirements for submission to the International Journal of Geriatric Psychiatry (Appendix 1.1)

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology (D.Clin.Psy)

Abstract

Objectives

This review examines the diagnostic accuracy and clinical utility of the Addenbrooke's Cognitive Examination (ACE) and its updated version, the Addenbrooke's Cognitive Examination – Revised (ACE-R), in relation to the diagnosis of dementia.

Design

A systematic search of relevant databases was conducted covering the period 2000 to April 2010 and specific journals and reference lists were hand searched. Identified studies that fulfilled the inclusion criteria were reviewed using a tailored, methodological quality rating checklist.

Results

The systematic search process identified 9 studies for review (7 relating to the ACE, 2 on the ACE-R). Each study is described individually before strengths and weaknesses across studies are considered. Diagnostic accuracy measures are presented for 6 out of the 9 studies.

Conclusion

The studies included in this review convey the ACE and ACE-R as tools capable of providing information on a range of cognitive domains and differentiating well between those with and those without cognitive impairment. Further research examining how the tools distinguish between dementia subtypes and Mild Cognitive Impairment would further the evidence base.

Introduction

Dementia is a disorder involving progressive, global, cognitive decline (Hannay et al., 2004). In 2005 it was estimated that 24.3 million people worldwide had dementia, with an anticipated worldwide increase of 4.6 million new cases per year (Ferri et al., 2005). Dementia diagnosis and the subsequent care pathway derived from the diagnosis is currently an area of high interest. It has been estimated that only a third of those with dementia receive a diagnosis (National Audit Office, 2007), thus national priority benchmarks for NHS service providers have been implemented to increase rates of early diagnosis (for an example of this see the Scottish Government's, Delivering for Mental Health, 2006). Whilst under-diagnosis is described as 'the current norm' it has been recognised that early diagnosis can provide individuals with *"...the chance to prevent future problems and crises and to benefit more from positive interventions"* (Department of Health, 2009, p.7).

There is an increasing evidence base to support this position. For example, research has found that intervention following early diagnosis can improve quality of life and delay institutionalisation (Banarjee et al., 2009; Gaugler et al., 2005). A prerequisite to effective post-diagnosis support must be an accurate assessment, capable of identifying individuals in the early stages of dementia. Such an assessment would be maximally useful if it could identify the subtype of dementia and detail the cognitive domains showing impairment. This information is essential to enable appropriate early intervention and management (e.g. correct medication and symptom specific compensation strategies).

Dementia is commonly used as an umbrella term for a set of progressive neurological conditions that cause cognitive impairment, including: Alzheimer's disease (AD),

Vascular dementia (VD), Fronto-temporal dementia (FTD), Lewy Body dementia (LB) and others. Each subtype of dementia has its own symptom profile and disease course. However, commonalities exist between the different subtypes and between them and other disorders (e.g. depression and Vitamin B12 deficiency), this complicates the diagnostic process. Before a dementia diagnosis can be established other possible causes of cognitive impairment should be ruled out; this includes confirming that the presenting symptoms are not due to normal age related changes. The term 'Mild Cognitive Impairment' (MCI) is used to describe individuals who present with a degree of memory impairment, but do not meet full diagnostic criteria for dementia. A significant number of people with MCI will go on to develop dementia, though some will not: this group therefore requires accurate detection and monitoring.

Screening tools in dementia assessment

Potentially, cognitive screening tools offer a time-efficient, objective initial assessment of cognitive functioning. The results from cognitive screening tools are not sufficient to make a diagnosis of dementia and should only be used as part of a comprehensive assessment (Smith et al., 2008). Nonetheless, screening tools are typically judged by their ability to accurately distinguish between those with and those without dementia, based on cut off scores. It has been noted that in several screening tool studies the dementia participants have severe cognitive impairment (Boustani et al., 2003). These participants almost inevitably score poorly compared to healthy controls, who tend to perform near ceiling. Studies which only compare these two participant groups fail to capture and represent the clinical utility required from a screening tool. Evidence from such studies may aid detection of moderate or severe dementia, but be less accurate in detecting mild dementia (Boustani et al., 2003). It has been stated that screening tools should be able to provide:

"...good sensitivity and specificity for all dementia types in unselected populations, and ... elicit information about key cognitive abilities, which can then be compared with neuropsychological profiles in different types of dementia" (Cullen et al., 2007, p.9).

In 2000, the ACE was published as a screening tool to detect mild dementia and distinguish between Alzheimer's disease and Fronto-temporal dementia (Mathuranath et al., 2000). The ACE incorporates the Mini Mental State Examination (MMSE), and includes items assessing the cognitive domains of: Attention and orientation, Fluency, Language, Visuospatial and Memory. Subscale scores are provided for each cognitive domain. In 2006, a revised version of the ACE, the Addenbrooke's Cognitive Examination – Revised (ACE-R) was published (Mioshi et al., 2006). Within the last decade the ACE has been cited as a potentially useful screening tool in guideline documents by the National Institute for Health and Clinical Excellence (NICE, 2006) and the Scottish Intercollegiate Guidelines Network (SIGN, 2006).

Statistical assessment of screening tools

Gifford and Cummings (1999) outlined guidelines on the methodological standards which should be reported for dementia screening tools. They identified three measures of reliability (inter-rater reliability, intra-rater reliability and internal consistency) and three types of validity (content validity, construct validity and criterion validity - notably sensitivity and specificity), which *"...should be evaluated and reported for a screening test"* (Gifford and Cummings, 1999, p.224). They also highlighted three potential sources of bias in screening tool studies. These were: i) Spectrum bias (when the test varies across population subgroups e.g. age, gender); ii) Verification bias (when only patients scoring below the cut off point receive further gold standard assessment); iii) Review bias

(when the test under review and the gold standard are administered and/or interpreted by the same person).

There is a recognised distinction between psychometric properties which are used to evaluate tests and those which contribute to interpretation of an individual's test results. Group comparison statistics (such as intra-rater reliability and construct validity) should not be used to support the clinician seeking to select the best tests to aid in diagnosing individual patients: criterion validity methods are a more helpful aid to such diagnostic decision making (Smith et al., 2008). Table 1 defines the recommended measures for aiding clinical decision making (Smith et al., 2008). These measures include: sensitivity, specificity, likelihood ratios (LR), positive predictive values (PPV), negative predictive values (NPV) and post test probabilities (PTP). The PPV, NPV and PTP calculations require base rate information. The base rate is the proportion of people in a larger reference sample (e.g. in the general population, or in people attending a memory clinic) who have the condition of interest. Since the usefulness of dementia screening tools is primarily assessed using measures of diagnostic accuracy, a screening tool review will necessarily have to explore diagnostic accuracy, even though dementia screening tools should never be the sole means of diagnostic decision making.

INSERT TABLE 1 HERE

Systematic review objectives

This review evaluates the literature available on the ACE and ACE-R with the following objectives:

i) To review the diagnostic accuracy of the ACE and ACE-R in diagnosing dementia.

ii) To examine the evidence for utilising the ACE and ACE-R as cognitive profiling

tools for differential diagnosis.

Method

Search strategy

The following electronic bibliographic databases were searched: All Evidence Based Medicine reviews, EMBASE, Medline, PsychINFO. The search was limited to the time period 2000 to April 2010 because the original ACE paper was first published in 2000. The databases were searched using various search terms, including: "Addenbrooke*", "dement*", "screen*" and "cognitive impair*" (see Appendix 1.2 for full strategy). Titles and abstracts of citations identified were examined to identify articles featuring either the ACE or the ACE-R. The following journals were hand searched for the time period 2000 to April 2010: Dementia and Geriatric Cognitive Disorders; International Journal of Geriatric Psychiatry. Reference lists of the included studies were also checked to identify further relevant papers.

Inclusion and exclusion criteria

The titles and abstracts of papers identified as featuring the ACE or ACE-R were screened against the following inclusion and exclusion criteria.

Inclusion Criteria:

- Studies investigating the diagnostic accuracy of the ACE or ACE-R.
- Studies considering the use of the ACE or ACE-R as tools for identifying and/or differentiating between types of dementia or between dementia and other disorders.
- If more than one study used the same participants only the most up to date study was included.

Exclusion Criteria:

- Studies that were not in English.
- Studies that investigated translated versions of the ACE/ACE-R.
- Studies that used the ACE/ACE-R to track changes in cognition over time.
- Studies in which the tools featured only as part of a wider assessment; or the tools were not the primary focus of the study.
- Response letters or guides were excluded and one study was also excluded because it was exclusively exploring the use of the ACE-R as a tool for identifying cognitive impairment in individuals who had a brain injury.

Some of the excluded studies, where appropriate, will be commented on in the discussion.

Assessment of methodological quality

To rate the methodological quality of the included studies a rating checklist was devised based on the SIGN Methodology Checklist 5 for diagnostic studies (Scottish Intercollegiate Guidelines Network, SIGN, 2007) and the Standards for the Reporting of Diagnostic accuracy studies (STARD) statement (Bossuyt et al., 2003). The rating checklist had a maximum score of 34 points (see Appendix 1.3 for a copy of the checklist). All papers were rated by the author. A second rater assessed 50% of the studies as a means of examining the inter-rater reliability of the checklist. Across all the checklist items that were subject to inter-rater assessment, there was 84% agreement between the raters; where discrepancies occurred, these were resolved through discussion.

Results

Outcome of search process

Forty-five ACE or ACE-R studies were initially identified; of which nine papers met the

inclusion criteria. Figure 1 provides a flow diagram outlining the systematic process of identifying the nine studies discussed in this review from the initial forty-five ACE or ACE-R studies. A reference list of the excluded studies is provided in Appendix 1.4.

INSERT FIGURE 1 HERE

Whilst the studies included have in common a focus on either the ACE or ACE-R as a tool to distinguish between those with and those without cognitive impairment or between different types of cognitive impairment, there is nonetheless, substantial heterogeneity between them. Each study will be reported individually before commonalities and differences between the studies are highlighted. ACE and ACE-R studies are reported separately. A summary table of demographic information across studies is provided in Table 2.

INSERT TABLE 2 HERE

Table 3 provides a summary of the key strengths and limitations identified by the rating checklist for each study. Studies are described below in order of their score on the methodological checklist.

INSERT TABLE 3 HERE

Review of findings for ACE papers

Mathuranath et al. (2000): 27/34 on the rating checklist.

This was the original article introducing the ACE. It aimed to validate the ACE as a screening tool to detect mild dementia and differentiate between Alzheimer's disease

(AD) and Fronto-temporal dementia (FTD). The study consisted of a clinic group referred to a memory clinic (diagnosed independently of the ACE) and controls. The control and clinic groups excluded individuals if they had major depression, mixed dementia or cognitive impairment of non neurodegenerative aetiology (e.g. alcoholism).

Two cut off points were identified for distinguishing between participants with dementia and those without. The 88/100 cut off point was developed via calculation of two standard deviations below the mean composite score for the control group. The other cut off (83/100) was determined by estimating the probability of diagnosing dementia in the clinic group, at varying potential cut offs and selecting the optimal one.

Different performance profiles were identified across the subscale scores for AD and FTD groups. The FTD patients performed better on tests of orientation and delayed memory, whilst the AD group did relatively better on language and fluency items; these results are consistent with established neuropsychological profiles of the two dementia subtypes. The 'VLOM ratio' was introduced as a proposed method for distinguishing between AD and FTD, by comparing scores on verbal fluency and language items to those on orientation and delayed memory recall. Maximal sensitivity and specificity details for the VLOM ratio were provided. Mathuranath et al. (2000) investigated the internal consistency, criterion validity and construct validity of the ACE. It was concluded that the ACE maintained good sensitivity across different subtypes and severity of dementia as defined by Clinical Dementia Ratings (CDR). The CDR provides an indicator of an individual's stage in the disease process.

Dudas et al. (2005): 25/34 on the rating checklist.

The aim of this study was to investigate the ability of the ACE to distinguish between

cognitive deficits resulting from dementia, versus those from affective disorders. In addition to dementia groups the study included a group of participants with Major Depressive Disorder (MDD); a group with affective symptoms and a mixed group (in which it was unclear if their impairment was due to a dementia or an affective aetiology). A significant difference in ACE performance was found between controls compared to the dementia and mixed groups. There was no significant difference between the MDD, affective and control groups. When cognitive domains were examined, a combined affective and MDD group was significantly impaired in memory and verbal fluency compared to controls. The mixed group was indistinguishable from AD and FTD groups on the category and verbal fluency items. In the mixed group, 15 out of 16 patients scoring lower than 88/100 on the ACE at initial assessment had a confirmed dementia within 2 years.

Davies et al. (2008): 24/ 34 on the rating checklist.

This study aimed to assess the ACE's ability to differentiate between Alzheimer's disease (AD) and Semantic dementia (SD, a subtype of Fronto-temporal dementia). ACE items were grouped into 12 sub scores and performances on these sub scores were compared between the SD and AD groups. SD participants performed significantly poorer on naming and reading items, whereas AD participants were significantly poorer on orientation items. A 'semantic index' was developed which consisted of the naming and reading scores, serial 7s, orientation in time and drawing scores. Values of less than zero on the index were found to be predictive of SD rather than AD, with 88% sensitivity and 90% specificity. The study did not detail how the AD participants were recruited and how and when their diagnosis of AD was made. It is also unclear when the AD participants completed the ACE. The discussion does not provide guidance about when in clinical practice this proposed semantic index should be utilised and whether it should be used in

addition to, or instead of the VLOM, which was introduced by Mathuranath et al. (2000).

Galton et al. (2005): 21/34 on the rating checklist.

The aim of this study was to address the relative value of the ACE in comparison with more detailed, neuropsychological tests and evaluation of the medial temporal lobe (via magnetic resonance imaging, MRI) in predicting participant conversion from questionable dementia (QD, a concept similar to MCI) to Alzheimer's disease (AD). The ACE had the best combination of Positive Predictive Value (PPV) and sensitivity compared to other neuropsychological tests, for predicting participant conversion from QD to AD. An ACE cut off of 80 provided best separation between converters and non converters and was the single best predictor of progression to AD; more so than neuropsychological tests evaluating one cognitive domain and MRI imaging results. The study reports sensitivity and specificity values for the ACE; however, it does not explicitly report which ACE cut off point these values are for, presumably they are for a cut off score of 80.

Reyes et al. (2009): 19/34 on the rating checklist.

This study investigated the validity of the ACE as a means of assessing cognitive function in patients with Parkinson's disease (PD). Participants from a PD outpatient clinic were recruited (n=44). Participants were excluded if they had depression, dopamine dysregulation syndrome, a history of drug abuse, cognitive decline secondary to a systemic or other degenerative disease. The Mattis Dementia Rating Scale (MDRS) was used as the gold standard reference method. The study used a variety of tools to measure the disease progression and the symptoms of Parkinson's. The ACE and the MDRS were found to correlate well ($r = 0.91$, $p<0.0001$), thus it was concluded that the ACE was a valid tool for dementia evaluation in PD.

Larner (2007a): 19/34 on the rating checklist.

Larner's (2007a) study reports ACE data from 285 consecutive patients referred with cognitive complaints to a cognitive function clinic, there was no participant exclusion criteria. All participants received a full assessment (including 1 year follow-up) and were subsequently classified, independently of their ACE performance, as individuals with either dementia or non-dementia. The study did not provide any demographic information on the participants recruited nor on what conditions were present in the group defined as 'non dementia'. The 88 and 83 cut offs reportedly had good sensitivity but poorer specificity; an alternative cut off of 75 improved specificity. The VLOM ratio was found to have reasonable sensitivity and specificity for the recommended cut off proposed by Mathuranath et al. (2000) to indicate AD; however, sensitivity for the FTD cut off was poor.

Bak et al. (2005): 14/34 on the rating checklist.

The aim of this study was to examine the cognitive profile of three subcortical dementia disorders associated with parkinsonism (Progressive supranuclear palsy (PSP), Corticobasal degeneration (CBD) and Multiple systems atrophy (MSA)) compared with a group of participants with AD and a group of healthy controls. The ACE and the Dementia Rating Scale (DRS) were used to compare the different participant groups. The AD participants were recruited through a memory clinic and the parkinsonism participants were recruited from other studies (PSP, Litvan et al., 1996; MSA, Gilman et al., 1998; CBD, Riley and Lang, 2000). Two of the subcortical dementia groups (PSP and CBD) along with the AD group were significantly impaired on the ACE total score, compared to controls; however, the ACE did not detect cognitive impairment in the other subcortical group (MSA); the verbal fluency sub test was the only subscale score which

distinguished between this group and the controls.

Review of findings for ACE-R papers

Mioshi et al. (2006): 23/34 on the rating checklist.

This article introduced the ACE-R and aimed to validate the clinical utility of the revisions of the ACE. Three groups of participants were included in the study: dementia (AD, FTD, LB), healthy controls and MCI. Participants were excluded if they had a psychiatric disorder, mixed dementia or cognitive impairment caused by something other than a neurodegenerative disease. Two cut off scores were defined (88 and 82) based on sensitivity, specificity and positive predictive value calculations at different prevalence rates. Participant performance on the ACE compared to the ACE-R was investigated; however, it is unclear when the ACE and ACE-R were undertaken in relation to each other. A subgroup of age and education matched participants was elicited from control, AD, and MCI groups; from these subgroups the MCI group performance was found to be between the AD and control groups, although a suggested cut off point for the MCI group was not given and the small numbers in each group limits the conclusions that can be drawn.

The study reported that spectrum bias was avoided by: *"including participants with different dementia syndromes and with a broad range of impairment (MMSE scores ranging from 9 to 30)"* (2006, p.1084). It is unclear if reported MMSE scores are extracted from ACE-R results or if they were completed separately. Concurrent and convergent validity and reliability were explicitly assessed and reported to be 'significant' and 'good'. Mioshi et al. (2006) provide a table reporting the positive likelihood ratios of dementia at a range of cut off points.

Larner (2007b): 13/34 on the rating checklist.

This article is a brief research letter which describes a prospective study of the ACE-R conducted for 100 consecutive patients referred to a cognitive function clinic. The breakdown of the subsequent diagnoses that these participants received is included. The sensitivity and specificity of the ACE-R for identifying dementia at the cut off scores of 88 and 82 are provided. Results for a cut off score of 75 are also provided. The study concludes that the ACE-R has good sensitivity and that in clinical practice test specificity and positive predictive value may be improved by a lower cut off.

Critical review of studies

The studies which rated highest on the rating checklist were: Mathuranath et al. (2000), Dudas et al. (2005), Davies et al. (2008) and Mioshi et al. (2006) (scoring 27, 25, 24, and 23 respectively). The studies which had the lowest scores on the rating checklist were: Bak et al. (2005) and Larner (2007b) (14 and 13 respectively).

Most studies used the reference standard, the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's disease and Related Disorders Association criteria (NINCDS-ADRDA, McKhann et al., 1984), with the exception of Reyes et al. (2009) who used the Mattis Dementia Rating Scale. Most studies (except the two Larner studies 2007, a+b) provide demographic information in terms of participant sex and average age. Four studies also provide information on years of education (Mathuranath et al., 2000; Dudas et al., 2005; Mioshi et al., 2006; Reyes et al., 2009) and four on severity of symptoms using the CDR (Mathuranath et al., 2000; Mioshi et al., 2006; Galton et al., 2005; Dudas et al., 2005).

In 2007 a study reported that years of education had a significant impact on performance

on the Malaysian version of the ACE (Mathuranath et al., 2007). Whilst this does not necessarily mean the same is true for the English version, it does highlight the need for educational experience to be routinely reported and for future research to explore whether or not education is a potential moderating factor for ACE/ACE-R performance. Since the impact of education on ACE/ACE-R performance requires further investigation it is unclear whether or not it is a potential source of spectrum bias. Furthermore, since a criticism of screening tool studies has been that their dementia participants often have a more severe level of dementia (Boustani et al., 2003) it would be beneficial if dementia symptom severity was routinely reported.

Under reported elements across most of the studies, (with the exception of Mathuranath et al., 2000 and Dudas et al., 2005) were the numbers, training and expertise of the persons administering and interpreting the ACE/ACE-R and the reference standard. With the exception of Larner (2007a) all of the studies failed to explicitly state if the ACE/ACE-R results were interpreted without knowledge of the reference standard results. With these two factors not being consistently reported it is not possible to know if there was a review bias present in most of the studies. Included studies also lacked information on how they assessed the reference standard (e.g. what specific tests, interview schedules and scans were used). Without this information it is difficult to judge how detailed these assessment were.

Four of the studies (Larner 2007 a+b; Reyes et al., 2009; Bak et al., 2005) did not report the period of time between administration of the reference standard and ACE/ACE-R. Whether or not the reference standard was independent of the ACE/ACE-R (i.e. the ACE/ACE-R did not form part of the reference standard) was not explicitly mentioned in three of the articles. In addition Mioshi et al. (2006) and Galton et al. (2005) report using

the MMSE as part of the assessment for the reference standard. Since the MMSE is incorporated within the ACE/ACE-R this indicates that there was not complete independence between the reference standard and the ACE/ACE-R, this is an area that both STARD and SIGN diagnostic checklists highlight, because overlap here may introduce bias (Bossuyt et al., 2003; SIGN, 2007). Furthermore, the majority of studies, with the exception of Reyes et al. (2009) and Galton et al. (2005), do not mention if there were any drop outs/deaths in the studies.

In several studies the ACE/ACE-R was administered at initial assessment to a group of participants with unknown aetiologies and reviewed once subsequent diagnostic decisions had been made (Mathuranath et al., 2000; Larner, 2007a+b; Reyes et al., 2009). This methodology minimises verification bias because it means all those in the clinical group receive both the ACE/ACE-R and reference standard. However in studies involving a control group, although not explicitly reported, it appears to the reader that none of the participants in the control groups received the reference standard; which means verification bias cannot be ruled out. In addition it is unclear from the studies how many participants required follow up assessment to establish diagnosis. If follow up was required it seems possible (depending on the time elapsed) that presentation could have altered from when the ACE/ACE-R was originally completed.

Mathuranath et al. (2000) and Mioshi et al. (2006) reported internal consistency and convergent validity. None of the studies included in this review assessed inter-rater or intra-rater reliability. In terms of the criterion validity, most studies reported sensitivity and specificity information. However, sensitivity and specificity values cannot assist clinicians in making decisions about the probability of disease in individual patients (Akobeng, 2006; Grimes and Schulz, 2002). Only Larner (2007a+b) and Mioshi et al.

(2006) report likelihood ratios. Furthermore, only Mathuranath et al. (2000), Mioshi et al. (2006), Galton et al. (2005) and Larner (2007a+b) report predictive values; however, the Larner studies (2007a+b) do not report the base rates used to calculate these values.

This review calculated the positive and negative likelihood ratios for all included studies which had provided relevant sensitivity and specificity information for ACE/ACE-R cut offs. Furthermore, using estimated base rate information, predictive values and post test probability values were also calculated. These values are available in Table 4.

INSERT TABLE 4 HERE

The base rates used are based on information regarding the prevalence of dementia at memory clinics in England and the prevalence of dementia in the general population of people aged between 60-69, because the average age of participants in the majority of studies is within this age range. The memory clinic base rate for dementia was taken as 54% (NICE, 2010), whilst the base rate for dementia in the general population aged 60-69 was estimated to be 1.3% (Knapp and Prince, 2007).

Discussion

The ACE and ACE-R are screening tools designed to aid in the detection of dementia. In the research covered in this review, they have been applied to a range of populations; different cut off scores have been identified and some consideration has been given to symptom profiles. Table 4 demonstrates that across the range of studies the ACE/ACE-R remain statistically robust, although the PPV, NPV and PTP results for the population and memory clinic base rates emphasise that underlying base rate has a significant influence

on the probability of a patient in clinical practice with a given result actually having dementia. It is essential that if these values are reported in the literature readers know the base rates they are derived from, so that they may compare them to the base rates in their own clinical context.

Across the majority of reviewed studies there was a lack of information on: number, training and expertise of the persons executing and interpreting the ACE/ACE-R and reference standard; withdrawals; if the ACE/ACE-R results were interpreted without knowledge of the reference standard results and how the reference standard assessment was conducted. These are items that SIGN and/or STARD diagnostic accuracy guidelines consider important and their absence means that it is not possible to know the potential for verification and review bias present in these studies. Future ACE/ACE-R studies would benefit from including this information.

In the studies included in this review there was also a lack of information on how those with Vascular dementia and Lewy Body dementia perform on the ACE/ACE-R. Vascular dementia is the second most common type of dementia (after AD), thus how patients with Vascular dementia perform on the ACE/ACE-R deserves further investigation.

Diagnostic accuracy

To be fit for use clinically, screening tools need to have statistically robust cut off scores. This allows distinction between individuals who are in the early stages of a dementia and those who may have a mild cognitive impairment of a different aetiology or no impairment at all.

The original articles by Mathuranath et al. (2000) and Mioshi et al. (2006) scored highly

on the rating tool and provide two cut off points for differentiating between those with and those without dementia. The Larner studies (2007a+b) did not score as highly on the rating tool, due mainly to missing information, nonetheless they offer an alternative cut off based on an unselected participant group. Therefore, there are currently three potential cut offs identified in the literature for the ACE (88, 83, 75) and the ACE-R (88, 82, 75).

Galton et al. (2005) reported that a cut off score of 80 on the ACE was best at distinguishing between those with questionable dementia who converted to dementia and those who did not, two years post assessment. Dudas et al. (2005) reported that in their 'mixed' group (affective and dementia symptoms) of the participants who scored below 88 on the ACE, 15/16 went on to develop dementia. A study not included in this review, followed up groups of participants diagnosed with MCI at 2 years post diagnosis (Mitchell et al., 2009). The study found that a combination of the ACE and the Paired Associates Learning (PAL) test was predictive of status after 2 years, concluding that those scoring >88 on the ACE and <14 errors on the PAL were at low risk of dementia. Therefore it would seem that in clinical practice those scoring below 88 are at an increased risk of dementia, and thus require further assessment and monitoring.

Taking into account the reviewed literature and the information from Table 4 it would seem that the 88 cut off is able to distinguish well between those with some degree of cognitive impairment and those without. Conversely a cut off score of 75 seems to capture those who are highly likely to have dementia. The 83 cut off and indeed any score between 75-88 may be suggestive of an early dementia or mild cognitive impairment. Due to the different participant groups across studies it is not possible to provide more specific information on these cut off scores. In Mathuranath et al. (2000)

the 83 cut off had good sensitivity, specificity, likelihood ratios and post test probabilities (in memory clinic base rate), for distinguishing between dementia and non-dementia in a clinic group. However, it is not clear if the clinic group participants classified as 'non dementia' had some level of mild cognitive impairment. Furthermore, exclusion criteria had been applied to the clinic group, omitting those with psychiatric disorders, mixed dementias and cognitive impairments of non neurodegenerative aetiology. Clinicians using the ACE/ACE-R in clinical practice should refer to the information in Table 4 as a guide to aid assessment when considering individual scores. It is essential that when Table 4 results are interpreted the participant groups involved in each study are taken into account.

Grimes and Schulz (2002) note that information gained from the measurements available in Table 4 are:

"Predicated on an assumption that is often clinically unrealistic i.e. that all people can be dichotomized as ill or well. Often those tested simply do not fit neatly into these designations: they might be possibly ill, early ill, probably well or some other variety" (p.882).

This statement is consistent with the results for the current review, because it is not clear whether or not the ACE/ACE-R is able to distinguish between MCI and dementia, and if it is what the cut offs for MCI should be. This highlights why screening tools, in isolation, are not sufficient means of diagnosis.

Utility of the ACE/ACE-R for differential diagnosis

The original article presenting the ACE suggested that AD and FTD groups could be

distinguished by their subscale scores. It was hypothesised that those with AD would be more impaired on the orientation and delayed memory recall items while those with FTD would be more impaired on verbal fluency and language items. The VLOM ratio was introduced as a means of objectively measuring this contrasting pattern in performance. The VLOM ratio has been investigated in three of the nine studies in this review. Mathuranath et al. (2000) and Mioshi et al. (2006) conclude that it is a useful ratio to calculate, whereas Larner (2007a) questions its specificity. Interestingly a paper not included in this current review explored the VLOM in the French version of the ACE and found that only 1 out of 9 patients with Fronto-temporal dementia were identified using the recommended VLOM cut offs (Bier et al., 2004). However, the FTD patients in this study all presented with 'the pure frontal form' of the disease. In another study by Mioshi et al. (2007) the relationship between the ability to perform activities of daily living and cognitive dysfunction was investigated across three variants of Fronto-temporal dementia. The ACE-R was used in the study as a measure of cognition. The study concluded that Fronto-temporal dementia has a significant impact on activities of daily living but that this impairment is not captured by the ACE-R. Mioshi et al. (2007) suggested that the ACE-R may have limited ability to detect FTD. It therefore remains uncertain whether the VLOM is a clinically useful measure.

Davies et al. (2008) reported an index for distinguishing Semantic dementia from Alzheimer's disease. However, no other study has examined this index and so it is not possible to conclude from the current review what its clinical utility is. The subtests implicated in this index as being more impaired in SD participants were naming and reading, whereas orientation items were reported as more impaired for AD participants. In Dudas et al.'s (2005) study there was no significant difference between the major depressive disorder, affective and control groups. However, when the cognitive domains

were examined a combined affective and MDD group was significantly impaired in memory and verbal fluency compared to the controls.

Based on the articles in this review, it would seem that there is not, as yet, a well established evidence base for certain subscale profile patterns on the ACE/ACE-R being indicative of certain disorders. However, this does not detract from the fact that the subscale information available in the ACE/ACE-R is a key strength of these tools. This information enables clinicians with knowledge of the neuropsychological profiles of different dementia subtypes, to obtain qualitatively rich information from patient subscale performance. Such information can assist in guiding further assessment. Therefore the subscales offer an important source of information to clinicians who have the expertise to extract qualitative information from them.

Limitations of the review

This review did not include studies investigating translated versions of the ACE/ACE-R. During the initial search stage of this review sixteen such studies were identified, indicating that these studies form a significant part of the ACE/ACE-R literature base which this review is unable to comment on. Thus, a future review of the translated studies would be desirable.

A further limitation of this review concerns the adapted methodological checklist used to assess quality in the articles included. Since screening tools should never be used in isolation to make diagnostic judgements it may be that the rating checklist had limited applicability because it was based on diagnostic accuracy guidelines. The guidelines for diagnostic tests are perhaps more suited to medical tests rather than assessments of cognitive function. This review adapted diagnostic guidelines and took into account

guidelines for dementia screening tools (Gifford and Cummings, 1999), in an attempt to minimise any rating checklist difficulties.

Areas for future research

None of the included studies in this review assessed the intra-rater or inter-rater reliability of the ACE/ACE-R. Without this information it is not possible to know if the ACE/ACE-R are reliably scored across different raters and by the same raters at different time points. These issues should be addressed in future research. This review also identified that whilst sensitivity and specificity were widely reported other measures of diagnostic accuracy, which are a greater aid to clinical decision making, LRs, PPV, NPV and PTP were not routinely reported. It would be beneficial for future understanding and comparison among ACE/ACE-R studies if this information could be routinely reported; this would include explicitly reporting the base rates used in calculations.

Additionally, in this review none of the articles provided any indicator about whether or not conclusions drawn in studies about the ACE are applicable to the ACE-R. It would be helpful if this could be clarified in future as the majority of studies in this review use the ACE (7vs.2), however it has now been widely replaced by the ACE-R in clinical practice.

Conclusions

This review suggests that statistically the ACE and ACE-R are robust tools for detecting cognitive impairment and has provided readers with information on diagnostic accuracy, which they may refer to when reviewing individual patient performance on the ACE/ACE-R. The ability of the tool to distinguish between MCI and dementia and also between dementia subtypes is an issue which requires further research. Currently there is

not enough evidence to advocate the use of certain subscale symptom profiles; nonetheless domain subscale information offers a rich source of qualitative information to aid the assessment process.

References

Akobeng AK. 2006. Understanding diagnostic tests 2: likelihood ratios, pre and post test probabilities and their use in clinical practice. *Acta Paediatr* **96**: 487-491.

Bak TH, Crawford LM, Hearn VC, et al. 2005. Subcortical dementia revisited: similarities and differences in cognitive function between progressive supranuclear palsy (PSP), corticobasal degeneration (CBD) and multiple system atrophy (MSA). *Neurocase* **11**: 268-273.

Banarjee S, Wittenberg R. 2009. Clinical and cost effectiveness of services for early diagnosis and intervention in dementia. *Int J of Geriatr Psychiatry* **24**(7): 748-754.

Bier JC, Ventura M, Donckels V, et al. 2004. Is the Addenbrooke's Cognitive Examination effective to detect frontotemporal dementia? *J Neurol* **251**(4): 428-431.

Bossuyt PM, Reitsma JB, Bruns DE, et al. 2003. Standards for reporting of Diagnostic Accuracy. Toward complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* **326**: 41-44.

Boustani M, Peterson B, Hanson L, et al. 2003. Screening for Dementia in Primary Care: a summary of the evidence for the U.S. preventive services task force. *Ann Intern Med* **138**: 927-937.

Cullen B, O'Neill B, Evans J.J, et al. 2007. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry* **78**: 790-799.

Davies RR, Dawson K, Mioshi E, et al. 2008. Differentiation of semantic dementia and Alzheimer's disease using the Addenbrooke's Cognitive Examination (ACE). *Int J of Geriatr Psychiatry* **23**: 370-375.

Department of Health. 2009. *Living well with dementia: a national dementia strategy*. [Internet] DH Publications: London. Available at: http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_094051.pdf [Accessed on 5th June 2010].

Dudas RB, Berrios GE, Hodges JR. 2005. The Addenbrooke's Cognitive Examination (ACE) in the differential diagnosis of early dementias versus affective disorder. *Am J Geriatr Psychiatry* **13**(3): 218-226.

Ferri CP, Prince M, Brayne C, et al. 2005. Global prevalence of dementia: a Delphi consensus study. *Lancet* **366**: 2112-2117.

Galton CJ, Erzinclioglu S, Sahakian BJ, et al. 2005. A comparison of the Addenbrooke's Cognitive Examination (ACE), conventional neuropsychological assessment and simple MRI-based medial temporal lobe evaluation in the early diagnosis of Alzheimer's disease. *Cogn Behav Neurol* **18**(3): 144-150.

Gaugler JE, Kane RL, Kane RA, Newcomer R, 2005. Early community-based services utilization and its effects on institutionalization in dementia caregiving. *Gerontologist* **45**: 177-185.

Gifford DR, Cummings JL. 1999. Evaluating Dementia Screening Tests: Methodologic standards to rate their performance. *Neurology* **52**(2): 224-227.

Gilman S, Low PA, Quinn N, et al. 1998. Consensus statement on the diagnosis of multiple systems atrophy. *Auton Nerv Syst* **74**: 189-192.

Grimes DA, Schulz KF. 2002. Epidemiology series uses and abuses of screening tests. *Lancet* **359**: 881-84.

Hannay HJ, Howieson DB, Loring DW, et al. 2004. Neuropathology for Neuropsychologists. In *Neuropsychological Assessment 4th edition*, Lezak MD, Howieson B, Loring DW (eds). Oxford University Press Inc.: New York; 157-285.

Knapp M, Prince M. 2007. *Dementia UK Summary of Key Findings*. [Internet] London: Alzheimer's Society. Available at: http://www.psige.org/psige-pdfs/Dementia_UK_Summary.pdf [Accessed 7th June 2010].

Larner AJ. 2007a. Addenbrooke's Cognitive Examination (ACE) for the diagnosis and differential diagnosis of dementia. *Clin Neurol Neurosurg* **109**: 491-494.

Larner AJ. 2007b. Addenbrooke's Cognitive Examination – Revised (ACE-R) in day-to-day clinical practice, Research Letter. *Age Ageing* **36**(6): 685-686.

Litvan I, Agid Y, Calne D, et al. 1996. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome); report of the NINDS-SPSP international workshop. *Neurology* **47**: 1-9.

Mathuranath PS, Cherian JP, Mathew R, et al. 2007. Mini mental state examination and the Addenbrooke's Cognitive Examination: Effect of education and norms for a multicultural population. *Neurol India* **55**(2): 106-110.

Mathuranath PS, Nestor PJ, Berrios GE, et al. 2000. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* **55**(11): 1613-1620.

McGhee S. 2002. Simplifying Likelihood Ratios. *J Gen Intern Med* **17**: 646-649.

McKhann G, Drachman D, Folstein M, et al. 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* **34**(7): 939-944.

Mitchell J, Arnold R, Dawson K, et al. 2009. Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. *J Neurol* **256**(9): 1500-1509.

Mioshi E, Kipps CM, Dawson K, et al. 2007. Activities of daily living in frontotemporal dementia and Alzheimer disease. *Neurology* **68**: 2077-2084.

Mioshi E, Dawson K, Mitchell J, et al. 2006. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J of Geriatc Psychiatry* **21**: 1078-1085.

National Audit Office. 2007. *Improving services and support for people with dementia*. [Internet] The Stationery Office: London. Available at: http://www.nao.org.uk/publications/0607/support_for_people_with_dement.aspx [Accessed 20th June 2010].

National Institute for Health and Clinical Excellence (NICE). 2010. *Assumptions used in estimating a population benchmark.* [Internet] National Institute for Health and Clinical Excellence. Available at: http://www.nice.org.uk/usingguidance/commisioningguides/memoryassessmentservice/assumptions.jsp [Accessed 20th June 2010].

National Institute for Health and Clinical Excellence (NICE). 2006. *Clinical Guideline 42, Dementia: Supporting people with dementia and their carers in health and social care.* [Internet] NICE: London. Available at: http://www.nice.org.uk/nicemedia/pdf/CG042NICEGuideline.pdf [Accessed 5th December 2009].

Reyes MA, Lloret SP, Gerscovich ER, et al. 2009. Addenbrooke's Cognitive Examination validation in Parkinson's disease. *Eur J Neurol* **16**: 142-147.

Riley DE, Lang AE. 2000. Clinical diagnostic criteria. In *Advances in neurology, corticobasal degeneration and related disorders,* Litvan I, Goetz CG, Lang AE (eds). Lippincott Williams and Wilkins: Philadelphia. **82**: 29-34.

Scottish Government. 2006. *Delivering for Mental Health.* [Internet] Scottish Executive:

Edinburgh. Available at: http://www.scotland.gov.uk/Resource/Doc/157157/0042281.pdf [Accessed 10th February 2010].

Scottish Intercollegiate Guidelines Network (SIGN). 2007. Methodology checklist 5: studies of diagnostic accuracy. In *A guideline developers handbook.* SIGN: Edinburgh; Annex B.

Scottish Intercollegiate Guidelines Network (SIGN). 2006. *Management of patients with dementia, a National Clinical Guideline*, *(SIGN publication 86).* [Internet] SIGN: Edinburgh. Available at: http://www.sign.ac.uk/pdf/sign86.pdf [Accessed 5th December 2008].

Smith GE, Ivnik RJ, Lucas J. 2008. 4 Assessment techniques: Tests, test batteries, norms and methodological approaches. In *Textbook of Clinical Neuropsychology*, Morgan JE, Ricker JH (eds). Taylor and Francis: New York; 38-58.

Table 1: Definitions

Measure	Definition*	Interpretation*
Sensitivity (Sn)	The proportion of people with the target disorder who have a positive test result.	The following mneumonic is used to understand sensitivity: High **S**ensitivity means a **N**egative result rules **out** the diagnosis (SNOUT).
Specificity (Sp)	The proportion of people without the target disorder who have a negative test result.	The following mneumonic is used to understand specificity: High **Sp**ecificity means a **p**ositive result rules **in** the diagnosis (SpPIN).
Likelihood Ratio for a positive test (LR +)	The probability an individual with the disease will have a positive test, divided by the probability of an individual without disease having a positive test result.	An LR+ greater than 1 means a positive test result is more likely to occur in those with the disease than in those without. For disease prevalence of 10-90% a LR+ of 2 increases the probability of disease by 15%, LR+ of 5 by 30%, LR+ of 10 by 45%.
Likelihood Ratios for a negative test (LR-)	The probability of an individual with disease having a negative test, divided by the probability of an individual without disease having a negative result.	An LR- of less than 1 means a negative result is less likely to occur in people with the disease compared to those without. For disease prevalence of 10-90% a LR- of 0.5 reduces the probability by about 15%, LR- of 0.2 by 30% and LR- of 0.1 by 45%.
Positive Predictive Value (PPV)	The proportion of people with positive test results who actually have the disease	To judge the probability an individual patient has/does not have the target disorder based on the PPV and NPV values of their score one must ensure the patient meets the inclusion/exclusion criteria applied in the study and that the prevalence rate in the study is approximately the same as that in the area the patient is from. If the prevalence rate of the disease is low then the PPV will not be close to 1, even when the sensitivity and specificity are high.
Negative Predictive Value (NPV)	The proportion of people with negative test results who do not have the disease	
Post test probability	The proportion of patients with that particular test result who have the condition of interest.	This result will vary depending on the underlying base rate. It allows an individual patient's probability of having the disease after the test result is known to be reported. It allows statements like "Based on the patient having earned a score of y on test z the probability that this patient has the condition of interest is x" (where x is the post test probability).

* Information in table taken from Akobeng (2006), McGhee (2002) and Smith et al. (2008).

Figure 1: Flow diagram

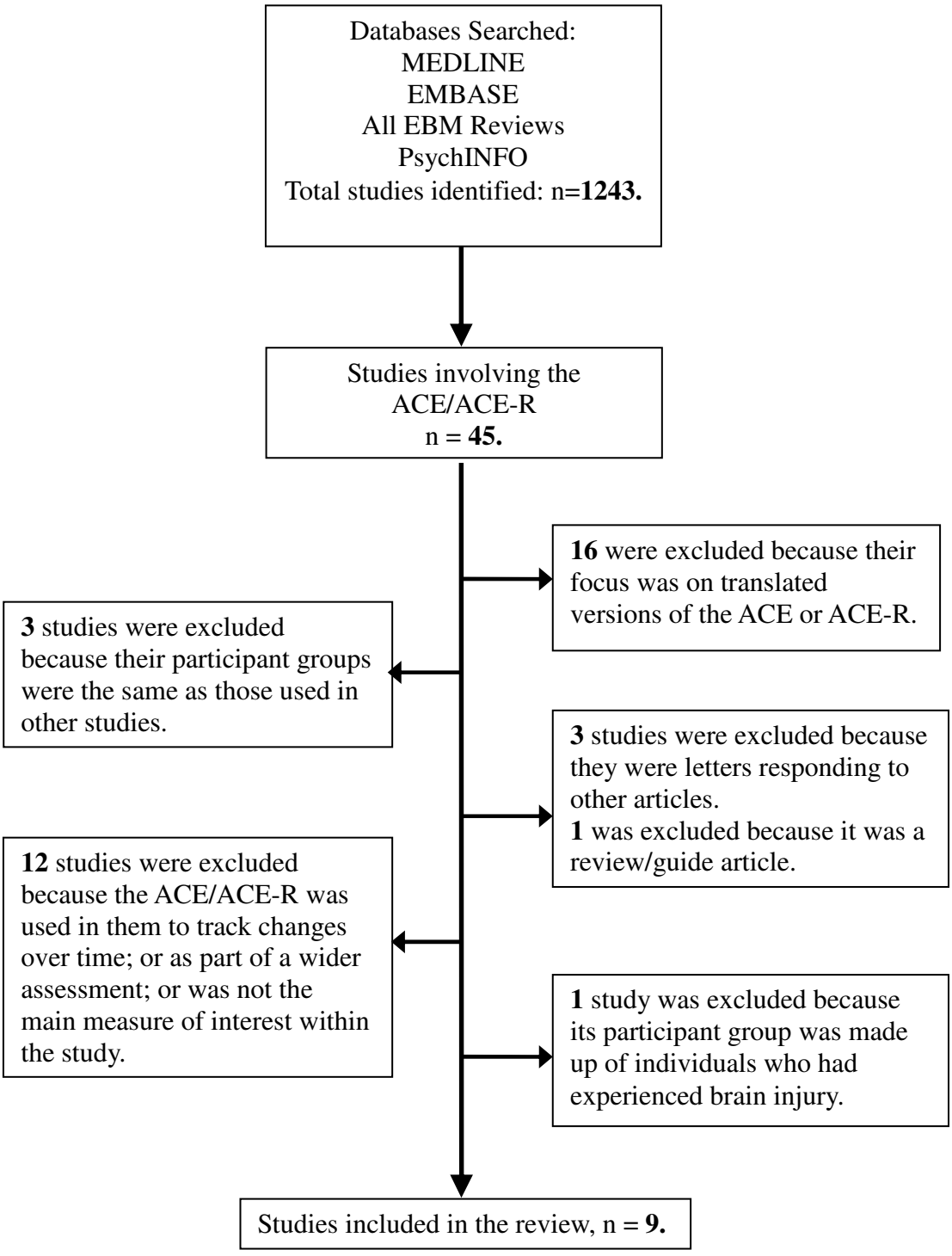

Table 2: Demographic Information*

Study	ACE/ ACE-R	Gold standard	Participant type	Participant number	No. of Males	Age in years (SD)	Education in years (SD)	Mean total score (SD)
Mathuranath et al. 2000	ACE	DSM-IV, NINCDS/ADRDA, NINDS-AIREN and the FTD consensus criteria	Dementia	115	61	66.6 (8.9)	11.1 (2.6)	64.8(18.9)
			Non dementia	24	19	63.8 (7.0)	12.7 (3.6)	88.0 (8.2)
			Controls	127	63	64.4 (9.3)	11.3 (2.6)	93.8 (3.5)
Dudas et al. 2005	ACE	NINCDS/ADRDA, FTD Consensus criteria	AD	63	27	68.9(8.0)	10.6(2.3)	61.9(18.3)
			FTD	27	18	61.3(7.3)	11.3(2.7)	74.2(15.4)
			MD	23	11	59.1(7.9)	11.2(1.7)	89.2(9.2)
			Affective	37	21	54.4(10.2)	12.4(3.1)	89.0(9.2)
			Mixed	22	12	64.1(8.1)	11.1(2.9)	71.0(16.6)
			Control	127	63	64.4 (9.4)	11.3(2.6)	93.9(3.5)
Davies et al. 2008	ACE	NINCDS-ADRDA, FTD Consensus criteria	SD	40	25	62.9(7.0)	NS	56.7(20.3)
			AD matched MMSE	40	25	66.7 (8.8)	NS	68.3 (19.9)
			AD matched ace	40	25	67.2 (7.9)	NS	56.7 (20.3)
Galton et al. 2005	ACE	NINCDS/ADRDA,	Early AD	19	7	66.9 (8.4)	NS	64.0 (7.4)
			QD convert	11	6	70.9(8.9)	NS	78.6 (7.5)
			QD non convert	18	8	9.3(7.8)	NS	91.4 (3.6)
Reyes et al. 2009	ACE	MDRS	PD dementia	13	9	68.6 (11.5)	11.9 (5)	71.7 (8.9)
			PD non dementia	31	18	71.9 (10.5)	14.3 (3.4)	89.8 (6.1)
Larner 2007a	ACE	NINCDS/ADRDA, DSM-IV, Consortium on DLB, FTD consensus criteria, NINDS-AIREN	Non dementia	145	NS	NS	NS	NS
			Dementia	140	NS	NS	NS	NS
Bak et al. 2005	ACE	NINCDS	AD	30	NS	69.3 (8.3)	NS	71.7 (14.0)
			PSP	39	NS	69.2 (5.8)	NS	78.8 (10.2)
			CBD	25	NS	67.1 (7.5)	NS	63.9 (21.7)
			MSA	26	NS	65 (7.2)	NS	85.5 (8.5)
			Controls	30	NS	71.3 (5.5)	NS	94.4 (3.0)
Mioshi et al. 2006	ACE-R	NINCDS/ADRDA, FTD consensus, Consortium on LB	Dementia	142	99	68.8(9.0)	12.8(3.4)	65.4(15.9)
			MCI	36	17	64.4(5.7)	12.7(2.1)	84.2(7.3)
			Control	63	28	60.9(11.6)	NS	93.7(4.3)
Larner 2007b	ACE-R	NS	Dementia AD	33	NS	NS	NS	NS

			Dementia FTD	8	NS	NS	NS	NS
			Dementia VD	2	NS	NS	NS	NS
			Dementia LB	1	NS	NS	NS	NS
			Non dementia MCI	18	NS	NS	NS	NS
			Non dementia Affective	10	NS	NS	NS	NS
			Non dementia PSMI	25	NS	NS	NS	NS

***Table 2 Abbreviations Key**					
AD	Alzheimer's disease	Affective	Affective disorder	CBD	Corticobasal degeneration
Convert	Converter from QD to dementia	FTD	Fronto-temporal dementia	LB	Lewy body dementia
MCI	Mild cognitive impairment	MD	Major Depressive disorder	Mixed	Mixed presentation of dementia & depression symptoms
MSA	Multiple systems atrophy	NS	Not specified	Non convert	Non converter from original QD diagnosis
PD	Parkinson's disease	PSMI	Primarily subjective memory impairment	PSP	Progressive supranuclear palsy
QD	Questionable dementia	VD	Vascular dementia		
Consortium on LB	Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies	DSM-IV	Diagnostic and statistical manual of mental disorders – Fourth edition	FTD consensus criteria	Fronto-temporal lobar degeneration: a consensus on clinical diagnostic criteria
MDRS	Mattis Dementia Rating Scale	NINCDS/ ADRDA	National Institute of Neurological and Communicative disorders and stroke and the Alzheimer's disease and Related Disorders Association	NINDS -AIREN	National Institute of Neurological Disorders and Stroke-Association Internationale pour la Recherche et l'Enseignement on Neurosciences criteria

Table 3: Summary of rating checklist results

Study	Rating	Strengths	Limitations
Mathuranath et al., 2000	27/34	- Clear selection criteria and participant recruitment information. - Adequate information about the ACE and the reference standard in terms of their execution and the period of time between them. - Reference standard independent of the index test and likely to classify the condition correctly. - Demographic information provided and diagnostic accuracy statistics reported.	- No report of whether or not the results were interpreted without knowledge of the results of the reference standard. - Unknown if there were any withdrawals from the study. - More information on researchers qualifications and roles within the study; details about the timescale of the study and a less specific participant group would have been helpful.
Dudas et al., 2005	25/34	- Clear selection criteria and participant recruitment information. - Range of participant groups included. - Reference standard likely to classify the condition and independent of the ACE. - Time period between ACE and reference standard reported. - Study timescale information and demographic information reported.	- No report of whether the ACE was interpreted with or without knowledge of the reference standard. - No information on any study withdrawals. - More information on the researchers involved; the execution of the ACE and the execution of the reference standard would have been helpful.
Davies et al., 2008	24/34	- Clearly explained participant recruitment. - Reference standard likely to classify the condition and independent of the ACE.	- No information on the researchers involved. - No report of whether the ACE was interpreted with or without knowledge of the reference standard.

		- Execution of the ACE explained. - Time period between reference standard and the ACE seemed reasonable. - Demographic information provided and timescale of the study reported.	- No mention of whether or not there were any study withdrawals. - Quite selective participant group. - More information on the selection criteria; execution of the reference standard and the statistical analyses would have been helpful.
Galton et al., 2005	21/34	- Clear selection criteria and participant recruitment information. - Considers ACE application in an MCI group. - Reference standard likely to classify the condition. - Details on the execution of the ACE provided. - Information on withdrawals from the study provided. - Details of the study timescale reported.	- No information on when the ACE was administered during the study and by whom. - The MMSE was used as part of the assessment of the reference standard. - Lack of information on the execution of the reference standard. - No report of whether the ACE was interpreted with or without knowledge of the reference standard. - More demographic information and statistical reporting would have been helpful.
Reyes et al. 2009	19/34	- Clear selection criteria and participant recruitment information. - Reference standard was independent of the ACE. - Demographic information provided. - Statistical analyses seemed appropriate.	- Lack of information on the researchers involved in the study. - Period of time between the reference standard and ACE not reported. - No report of whether the ACE was interpreted with or without knowledge of the reference standard. - No information on study withdrawals. - Study timescale not reported - More information on the execution of the index and reference

			standards and whether the reference standard selected was likely to classify the condition would have been helpful.
Larner, 2007a	19/34	- Good representative sample of participants and selection criteria are mentioned. - Reference standard independent of the index test and likely to classify the condition correctly. - Explicitly reported that the ACE was independent of the reference standard and interpreted without knowledge of it. - Appropriate statistics used (Sensitivity, specificity, LRs, PPV and NPV).	- No detail on participant recruitment or on the expertise of the researchers undertaking the research. - The period between the reference standard and ACE is unknown. - Demographic information on the participants and an explanation for any withdrawals from the study not provided. - More information on selection criteria; execution of the reference standard and ACE; and the timescale of the study would have been helpful.
Bak et al. 2005	14/34	- Selection criteria and participant recruitment mentioned. - Reference standard likely to classify the condition. - Execution of the ACE described. - Some demographic information available. - Participants from a population which required consideration in the literature.	- No information on the researchers involved. - It is unknown whether the reference standard was independent of the ACE. - No report of whether the ACE was interpreted with or without knowledge of the reference standard. - No information on the period of time between the reference standard and the ACE and on the study timescale. - No report of study withdrawals. - Further information would have been helpful on: the selection criteria; participant recruitment; execution of the reference standard; and demographic information.
Mioshi et al. 2006	23/34	- A range of different participant groups were included. - Clear selection criteria and participant recruitment information.	- No report of the researchers involved. - No report of whether the ACE was interpreted with or without knowledge of the reference standard.

		- Reference standard likely to classify the condition and period between it and the ACE-R seems reasonably short. - Timescale of study reported. - Demographic information provided. - Statistics used seem appropriate.	- Study withdrawals are not mentioned. - More information on the execution of the reference standard and ACE-R and demographic information would have been helpful. - The MMSE was used as part of the reference standard assessment.
Larner 2007b	13/34	- Involves a spectrum of participants representative of clinical practice. - Details of different diagnoses given post assessment provided. - Provides selection criteria and participant recruitment information. - Statistics used seem appropriate.	- No information on the researchers involved. - No report given of the execution, independence and blinding of the ACE from the reference standard and vice versa. - Withdrawals from the study are not mentioned. - The study time scale was not reported. - No report of whether the ACE was interpreted with or without knowledge of the reference standard.

Table 4: Diagnostic Accuracy Information

Study	ACE or ACE-R	Cut off score	Sensitivity	Specificity	Likelihood Ratio +	Likelihood Ratio -	Positive Predictive value for base rate of 1.3%	Negative Predictive value for base rate of 1.3%	Post test probability with base rate of 1.3%	Positive Predictive value for base rate of 54%	Negative Predictive value for base rate of 54%	Post test probability with base rate of 54%
Mathuranath et al. 2000	ACE	88	0.93	0.71	3.21	0.10	0.04	1.00	0.04	0.79	0.90	0.79
Larner 2007a	ACE	88	1.00	0.43	1.75	0	0.02	1.00	0.02	0.67	1.00	0.67
Dudas et al. 2005	ACE	88	0.93	0.82	5.17	0.09	0.06	1.00	0.06	0.86	0.91	0.856
Mioshi et al. 2006	ACE-R	88	0.94	0.89	8.55	0.07	0.10	1.00	0.10	0.91	0.93	0.91
Larner 2007b	ACE-R	88	1.00	0.48	1.92	0	0.69	1.00	0.02	0.02	1	0.69
Mathuranath et al. 2000	ACE	83	0.82	0.96	20.50	0.19	0.21	1.00	0.21	0.96	0.82	0.96
Larner 2007a	ACE	83	0.96	0.63	2.59	0.06	0.03	1.00	0.03	0.75	0.93	0.75
Reyes et al. 2009	ACE	83	0.92	0.91	10.22	0.09	0.12	1.00	0.12	0.92	0.91	0.92
Mioshi et al. 2006	ACE-R	82	0.84	1.00	N/A*	0.16	1	1.00	N/A*	1	0.84	N/A*
Larner 2007b	ACE-R	82	0.96	0.72	3.43	0.06	0.04	1.00	0.04	0.80	0.94	0.80
Larner 2007a	ACE	75	0.85	0.83	5.00	0.18	0.06	1.00	0.06	0.85	0.82	0.85
Larner 2007b	ACE-R	75	0.91	0.91	10.11	0.10	0.12	1.00	0.12	0.92	0.90	0.92

* Unable to calculate value because equation involved a denominator of 0

Chapter 2: Major research project

An Investigation of the Reliability of the Addenbrooke's Cognitive Examination – Revised (ACE-R)

Stephanie Crawford[1*]

[1]Section of Psychological Medicine, University of Glasgow

[*]Address for Correspondence
Section of Psychological Medicine
Division of Community Based Sciences
University of Glasgow
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XY
E-mail: scrawford@nhs.net

Key words: Dementia Screening Tool, Reliability, Multiple Raters

Word count: 6221

Prepared in accordance with requirements for submission to the International Journal of Geriatric Psychiatry (Appendix 1.1)

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology (D.Clin.Psy)

Abstract

Objectives

The Addenbrooke's Cognitive Examination – Revised (ACE-R) is a dementia screening tool. The objectives of this study were to investigate rater accuracy in scoring the ACE-R in terms of its total and subscale scores and to examine whether scoring accuracy is affected by participant experience of using the ACE-R.

Methods

Three filmed vignettes of the ACE-R being administered to older adult actors (mock patients) were used to assess scoring accuracy across multiple raters. The vignettes had a pre-determined 'true score'. Study participants were required to complete ACE-R scoring sheets for each vignette. Participants were Community Nurses and Trainee Clinical Psychologists.

Results

Participant scores were compared with the pre-determined true scores as a means of measuring scoring accuracy. The results indicated that the majority of participant scores were either the same as or within a few points of the true scores. However, when compared to the true scores, participant total scores differed significantly on two out of the three vignettes. Scoring accuracy was lowest for the Memory subscale of the ACE-R. Scoring accuracy issues were also identified for the Visuospatial and Attention and orientation subscales. Individual items which had low scoring accuracy were identified.

Discussion

The majority of participants scored the same as or within a few points of the true scores, such deviation is likely to be clinically acceptable, providing over-emphasis is not placed on cut off scores. Professionals using the ACE-R should ensure they are familiar with the scoring guidelines for the items highlighted in this study as having low scoring accuracy.

Introduction

Dementia refers to global cognitive decline, characteristic of progressive conditions such as Alzheimer's disease (Hannay et al., 2004). There is a general consensus that early dementia diagnosis is desirable because it enables earlier intervention and gives *"people the opportunity to make choices and plan for the future while they are still relatively well"* (The National Audit Office, 2007, p.43). Early diagnosis necessarily requires accurate assessment.

The role of cognitive screening in dementia diagnosis

The National Institute for Health and Clinical Excellence (NICE, 2006) states that a dementia diagnosis should only be made following a comprehensive assessment involving history taking, cognitive, mental and physical examination and medication review. Cognitive assessment tools are well established aids in this diagnostic process. In an older adult population cognitive assessment is typically used for three reasons: i) screening for cognitive impairment; ii) differential diagnosis; iii) rating disease severity or monitoring disease progression (Woodford and George, 2007). In clinical practice, where time constraints are omnipresent, the possibility of achieving an objective, quantitative measure of cognitive functioning following a short administration period is highly desirable. This may explain why, in the arena of cognitive assessment, recent decades have witnessed the development of various dementia screening tools to aid the diagnostic process.

The fundamental aim of any cognitive screening tool is to infer from the patient's score, compared to reference norms, the likelihood that genuine cognitive impairment is present. The success of such instruments in achieving this aim depends upon their psychometric properties (Cullen et al., 2007). The time pressures associated with everyday clinical

practice mean that screening tools must be brief and easy to administer. The ideal screen should also provide rich qualitative data; enabling a symptom orientated approach to assessment, which may aid differential diagnosis of different forms of dementia and facilitate better, targeted rehabilitation efforts (Cullen et al., 2007). A central feature of most screening tools is the 'cut-off ': the point at which scores move from being regarded as normal to abnormal (impaired), thus indicating the presence of the condition of interest (i.e. dementia). It is unlikely there will ever be a cognitive screening tool which will always, accurately make this distinction (Hannay and Lezak, 2004). A recent review concluded that useful dementia screening tools should:

"...have good sensitivity and specificity for all dementia types in unselected populations, and... elicit information about key cognitive abilities, which can then be compared with neuropsychological profiles in different types of dementia" (Cullen et al., 2007, p.9).

Currently a universally accepted, standardized screening tool for dementia remains elusive (Gifford and Cummings, 1999).

Measurement issues in cognitive screening tools

It is generally agreed that the measurement obtained from any instrument is a combination of the 'true score' and a degree of measurement error (McDowell, 2006). The fundamental difficulty across all cognitive measures is that the true score is never precisely known; it can only be inferred from the value obtained (Trochim, 2006). It is essential that cognitive tools minimise the degree of measurement error they create: thus making the value they obtain as accurate a reflection of the underlying true score as possible. Measurement errors can be categorised into random errors and systematic errors. Random errors increase group variability but do not affect the average

performance (overall mean) of the group. Systematic errors tend to be consistently either positive or negative and thus are considered a source of bias (Trochim, 2006).

There are numerous sources of potential measurement error, which can affect the reliability of a cognitive tool. These may be broadly conceptualised into three categories: i) client factors, e.g. client motivation, attention, and mood; ii) contextual factors, e.g. setting in time and place; iii) rater factors, e.g. administration and scoring issues.

Reliability refers to a measurements consistency. Inter-rater reliability assesses whether the tool is administered and scored consistently between raters (Trochim, 2006). Traditionally inter-rater reliability has been investigated by assessing agreement between a few (typically expert) raters across numerous trials. Such designs assume that high agreement indicates strong inter-rater accuracy for both administration and scoring. However, these designs fail to investigate administration and scoring accuracy across multiple raters in a clinical setting. Furthermore, by considering administration and scoring collectively these designs may fail to detect important aspects of each.

Inter-rater reliability studies may benefit from using greater numbers of raters; investigating administration and scoring separately; and considering scoring accuracy rather than agreement between raters, since raters may show high agreement, but both have low accuracy. If a tool is not consistently administered and scored by different raters then its clinical utility is limited. Therefore the consistency with which a tool is administered and scored by a range of professionals is a relevant, yet infrequently reported, part of assessing a tool's generalizability. A recent study involving multiple raters scoring filmed vignettes of two brief screening tools reported that the percentage of correct total scores obtained was lower than anticipated (Queally et al., 2010). These

results raise concerns about the lack of attention given to this issue in the literature and highlights that test developers need to consider the implications of scoring inaccuracy and make recommendations in light of these (Queally et al., 2010).

The ACE-R as a dementia screening tool

The Addenbrooke's Cognitive Examination (ACE) was originally designed to detect mild dementia and differentiate Alzheimer's disease from Fronto-temporal dementia (Mathuranath et al., 2000). It was revised in 2006, to produce the ACE-R (Mioshi et al., 2006). The ACE-R has five subscales: Attention and orientation, Language, Fluency, Memory and Visuospatial; each representing a cognitive domain. The subscale scores are summed to produce an overall total score (maximum 100 points). The Mini Mental State Examination (MMSE) is incorporated within the ACE-R. There are three versions of the ACE-R available, to reduce practice effects with repeat testing.

In 2006 the tool was recommended by the Scottish Intercollegiate Guidelines Network, (SIGN 86), as a means of improving initial cognitive testing for dementia. The ACE-R has been translated into 15 languages (Bak and Mioshi, 2007) indicating that its clinical use internationally is becoming increasingly pervasive.

ACE-R reliability

The original ACE study suggested that inter-rater reliability of the ACE was likely to be high since the tool measured cognition in an objective way; however, this study did not formally assess inter-rater reliability and acknowledged that this would be a useful focus for future research (Mathuranath et al., 2000). The ACE-R was reported to have undergone *"design changes to make the test easier to administer"* (Mioshi et al., 2006, p.1078). Nonetheless there are no published data examining the ACE-R's rater reliability.

The current study

A procedure was developed to investigate rater scoring accuracy on the ACE-R, whilst administration was kept constant. Several items in the ACE-R require a degree of rater interpretation. Based on anecdotal evidence from Older Adult Clinical Psychologists, it was hypothesised that professionals may sometimes give clients "the benefit of the doubt", when scoring screening tools. This can inflate the 'test score' from the 'true score', potentially generating conclusions that do not accurately reflect the underlying true score. Such over-scoring may lead to systematic measurement errors. In terms of the ACE-R it was hypothesised that the subscales which involve more subjective scoring judgements (namely the Visuospatial and Language subscales) would be the ones which would be over-scored by professionals.

Furthermore, it may be hypothesised that if individuals do not frequently refer to the ACE-R administration and scoring guide they may develop their own unique scoring methods. Thus scoring accuracy may also be affected by rater experience of the ACE-R.

Objectives

The primary objective was to investigate how accurately the ACE-R is scored by multiple raters with a focus on examining whether participants over-score the ACE-R. Variation of scoring accuracy across ACE-R subscales and degree of ACE-R experience were investigated as secondary objectives.

Primary hypothesis:

Participant ACE-R total scores will differ significantly from the ACE-R true scores because participants will over score the ACE-R.

Secondary hypotheses:

a) Participant scores on the ACE-R subscales that involve more subjective judgements (Visuospatial and Language) will significantly differ from the 'true' subscale scores because participants will over score them.

b) Scoring accuracy will be related to participant experience using the ACE-R.

Method

Participants

Participants were recruited from two groups of professionals:

i) Community nurses (CNs) working in older adult community mental health teams, who use the ACE-R routinely in their clinical practice.

ii) Trainee Clinical Psychologists (TCPs) currently completing their Doctorate in Clinical Psychology who have used the ACE-R whilst on clinical placement.

Eligibility criteria

Participants were included if they had administered and scored the ACE-R independently in clinical practice on at least one older adult patient. Participants were excluded if they had not used the ACE-R in clinical practice; that is they had not completed it with an older adult patient. Some CN participants had previously received a training session on how to administer and score the ACE-R.

Justification of sample size

The sample size required was based on a power calculation for a one sample t-test, as this was the planned main method for statistical analysis. Since there have been no previously published studies on the ACE-R's rater reliability there was no data available to base the

power calculation upon. It was considered that a small effect size would translate into only a few points variation from the true score which would not threaten the accuracy of the overall conclusions drawn from the ACE-R in clinical practice. By contrast a medium effect size would translate into a difference significantly large enough that it could alter interpretation of the ACE-R results, which could in turn, influence a wider assessment process. Therefore the power calculation determined the number of participants necessary to detect a medium effect size, using the conservative values of: p-value 0.05 and power 0.8. When these values for a one sample t-test were entered into G*Power 3.010 (Faul et al., 2007, from www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/) the results indicated that a minimum of 34 participants would be required.

It was hypothesised that both participant sub-groups would over-score the ACE-R; no significant difference was expected between the participant groups. However, to ensure power would not be affected if the groups were significantly different, the study aimed to recruit a minimum of 34 participants from each of the participant groups.

Measures

The Addenbrooke's Cognitive Examination – Revised (ACE-R) was the primary outcome measure.

Design procedure

Three filmed vignettes of the principal researcher administering the ACE-R to older adult actors (mock patients) were devised (one for each version of the ACE-R). Vignettes had pre-determined 'true scores' to allow scoring consistency to be investigated separately from administration consistency. Scripts for these research vignettes were developed by an advisory group of Clinical Psychologists with extensive neuropsychological

experience within older adult services; the scripts were not based on individual clients. Each vignette script was designed to represent a cognitive and clinical profile typical of clinical presentation at a memory clinic. The script for Vignette 1 was designed to be reflective of early-stage Alzheimer's disease (with a true total score of 75); Vignette 2's script was representative of Fronto-temporal dementia, behavioural-variant (true total score 84); and Vignette 3 was scripted for affective disorder (true total score 85). There was 100% agreement and accuracy with the true scores, across all the vignettes, when two experienced Clinical Psychologists rated them. The University of Glasgow media services filmed and produced the vignettes. Professional actors were recruited through the NHS Role play coordinator; they were given information on the purpose of the study and how their performances would be used. Signed consent was obtained from each of the actors (Appendix 2.1). Each actor performed one vignette, having learned the script beforehand. The actors were paid for their time and were not patients of any relevant service. The researcher who was filmed administering the ACE-R was the same person who showed the vignettes to the study participants. However, this researcher was blind to the 'true scores' until data collection was completed.

Research procedure

Potential participants were invited to participate in the study either via email or verbally. Participants who opted in to the study attended one of several group sessions, which lasted on average 1hour30min. Each participant received an information sheet about the study (Appendix 2.2). Written consent was obtained from all participants prior to the study commencing (Appendix 2.3). Participation involved completing ACE-R scoring sheets in conjunction with watching the vignettes. Participants viewed each vignette in its entirety once and were not permitted to pause or rewind the vignettes, reflecting actual clinical practice. Vignettes were viewed on a projector screen and shown to participants

in groups (with a minimum of three participants in each group). All participants viewed the vignettes in the same order (1, 2, 3). After watching each vignette participants were permitted time to complete their scoring of the corresponding ACE-R form. The ACE-R administration and scoring guide was available to participants, on request. In addition to scoring the vignettes the participants completed an additional information form detailing their profession and ACE-R experience (Appendix 2.4). A standardised set of instructions was read aloud to each group of participants before they commenced participation in the study (Appendix 2.5). Participants were also given the correct orientation information (date and place) for each vignette (Appendix 2.6).

Ethical approval

Prior to the study commencing, ethical approval was gained from a Local Research Ethics Committee (Appendix 2.7) and practice was informed by The British Psychological Society Code of Ethics & Conduct (2009).

Data analysis

Statistical analyses were carried out using SPSS and Microsoft Excel.

Results

Fifty-seven participants took part in the study; 45 Community Nurses and 12 Trainee Clinical Psychologists. All participants completed the three vignettes, with the exception of one participant who only completed two due to clinical time constraints (this participant did not complete Vignette 3). Four participants did not complete the Additional Information sheet. Across the vignettes the majority of participants scored all items; however, some participants did not sum items to provide subscale and total scores

(information was missing for 10 participants in Vignettes 1 and 2, and 11 in Vignette 3). Although the principal researcher could have summed up missing total and subscale scores on participant ACE-R forms, given that the principle focus of the study was on scoring reliability, this was not done. If participants had not fully completed their ACE-R forms the missing information was treated as missing data and excluded from relevant analyses.

Preliminary analysis

Initial analysis was undertaken to investigate whether the data set was normally distributed. The Shapiro-Wilks test was undertaken for the participant total and subscale scores for each of the vignettes. The null hypothesis of the Shapiro-Wilks test is that data are normally distributed. The null hypothesis was rejected, across the vignettes, for total and subscale scores; this suggests that the data were not normally distributed. Furthermore, the skewness and kurtosis of the distribution of total and subscale scores for each of the three vignettes indicated that the data may not be normally distributed; the kurtosis results were all positive indicating a leptokurtic distribution (an acute peak around the mean). Since the tests for normality indicated that the data did not have a normal distribution, further analysis used non-parametric tests.

The Non-parametric Independent-Samples Mann-Whitney U Test was used to examine if a significant difference existed between CNs and TCPs total scores on each vignette. The null hypothesis of this test is that the distribution of scores is the same across different groups. The null hypothesis could not be rejected for Vignettes 2 and 3 ($U=248.50$, $p=0.96$ and $U=196.0$, $p=0.655$, respectively) but was rejected for Vignette 1 ($U=341.5$, $p=0.039$). When the TCP and CN groups were compared using Independent-Samples Mann-Whitney U Tests for each of the Vignette 1 subscales, the Memory subscale was the

only one where a significant difference was obtained between the two groups. Since the TCP sample size was small and there was not a significant group difference in Vignettes 2 and 3 it was decided to combine the two participant groups for further analysis across the three vignettes.

Summary of total score results

Examination of Vignette 1 showed 15% of participant total scores matched the total 'true' score (TS), while 57% of participant total scores deviated from the TS by 1-2 points. Only 9% deviated by 5 or more points. In Vignette 2, 19% of participant total scores matched the TS, 47% deviated by 1-2 points and 21% deviated by 5 or more points. For Vignette 3, 30% of participant total scores matched the TS, 60% deviated by 1-2 points and 6% of participant scores were 5 or more points from the TS.

Figure 1 displays the differences between participant total scores and the true total score for each vignette. Visual inspection indicates that the majority of participant scores were lower than the TS in Vignette 1 (M=-2.3, SD=4.1). There was no clear pattern of under or over-scoring in Vignette 2 (M=-1.22, SD=4.1) and deviation from the TS was less variable in Vignette 3 (M=-0.9, SD=1.9).

INSERT FIGURE 1 HERE

Primary hypothesis: participant ACE-R total scores will differ significantly from the ACE-R true scores (TS) because participants will over score the ACE-R.

To investigate the primary hypothesis, one-tailed, one sample Wilcoxon signed ranks tests were undertaken (the results of which are available in Table 1). The tests were one tailed because apriori, a specific directional hypothesis (that participants would over score the

vignettes) was made. The null hypothesis was that participant total scores would be the same as the TS for each vignette. The null hypothesis was rejected in Vignettes 1 and 3 but could not be rejected in Vignette 2. These results indicate that the primary hypothesis was partially supported in Vignettes 1 and 3, with participant scores differing significantly from the TS. The medians for the participant total scores in Vignettes 1 and 3 were below the TS indicating that over-scoring was not occurring.

INSERT TABLE 1 HERE

Secondary hypothesis a) participant scores on the ACE-R subscales which involve more subjective judgements (Visuospatial and Language) will significantly differ from 'true' subscale scores because participants will over score them.

One tailed, one sample Wilcoxon signed ranks tests were used to investigate this hypothesis (see Table 1). The tests were one tailed because apriori, a specific directional hypothesis (that participants would over score these subscales) was made. The null hypothesis was that participant scores would be the same as the TS for both the Language and Visuospatial subscales. The null hypothesis was retained for the Language subscale in Vignettes 1 and 2, but rejected in Vignette 3. Whilst the null hypothesis was rejected for the Visuospatial subscale in Vignettes 2 and 3, but not for Vignette 1. These results indicate partial support for secondary hypothesis a), with significant differences between participant scores and the TS found for the Language subscale in Vignette 3 and the Visuospatial subscale in Vignettes 2 and 3.

Figure 2 shows participant subscale scores in comparison to subscale TS for all three vignettes. Visual inspection indicates a trend of slight over-scoring on the Visuospatial subscale and no clear pattern of over or under-scoring on the Language subscale.

INSERT FIGURE 2 HERE

Additional analysis

Inspection of Figure 2 reveals that the majority of participant scores are lower than the TS on the Memory subscale and there is also a trend towards under-scoring on the Attention and orientation subscale. To investigate whether significant differences exist between participant scores and TS on the subscales of Attention and orientation, Memory and Fluency, two tailed one sample Wilcoxon signed ranks tests were undertaken (Table 1). These tests were two tailed because predictions had not been made about how participants would score these subscales. The null hypothesis of these tests was that participant subscale scores would be the same as the subscale true scores.

For the Attention and orientation subscale a significant difference was found between participant scores and TS in Vignettes 1 and 3 but not 2. For the Memory subscale the null hypothesis was rejected in Vignettes 1, 2 and 3; highlighting that participant scores significantly differ from the TS across all three Memory subscales. In comparison, for the Fluency subscale it was not possible to reject the null hypothesis in Vignettes 1, 2 and 3; this suggests there was no evidence of a significant difference found between participant scores and the TS for all three Fluency subscales.

Item level analysis

To examine how participant scores differed from the TS at the level of individual items, Table 2 was generated. The table shows, for every item and subscale in Vignettes 1, 2 and 3, the percentage of participant scores which were: the same as the TS; plus or minus 2 points from the TS; or plus or minus 3 or more points from the TS.

INSERT TABLE 2 HERE

Across all three vignettes the items where most participant scores deviated from the TS were:

- Orientation questions in the Attention and orientation subscale (23%).
- Naming item in the Language subscale (23%).
- Name and address recall (25%) and recognition (28%) questions in the Memory subscale.
- Clock item in the Visuospatial subscale (45%).

The percentages in brackets above represent the total deviation from the true score for each item across all three vignettes.

Secondary hypothesis b) scoring accuracy will be related to participant experience using the ACE-R.

The non-parametric Independent-Samples Mann-Whitney U Test was used to address this hypothesis. The null hypothesis was that for groups of participants with varying levels of ACE-R experience their total scores would be equal. The participants were divided into different groups as follows:

i) Participants who attended a one off training session on the ACE-R (n=31) and those who did not (n=21).

ii) Participants who utilised the ACE-R scoring and administration guide during their participation in the study (n=24) and those who did not (n=33).

iii) Participants who reported using the ACE-R in their clinical practice anywhere between once a week to once a month (n=25) and those who reported using the

ACE-R anywhere between once every three months to less than once every six months (n=27).

iv) Participants who reported their last use of the ACE-R as being up to a week to up to one month prior to the study (n=31) and those who reported their last use of the ACE-R as being more than a month to more than six months prior to the study (n=22).

The Independent-Samples Mann-Whitney U Test was completed separately for each of the above (i-iv), per vignette (output available in Table 3). The null hypotheses were retained for each of the above, across all three vignettes. This suggests that participant experience of the ACE-R did not lead to significant differences in participant performance. However, some of the sub-groups of participants in the analysis had small numbers, thus these conclusions should be interpreted with caution.

INSERT TABLE 3 HERE

Calculation errors

It was discovered that some of the subscale and overall total scores reported on participant ACE-R forms differed from the summed total of recorded items. This suggests that calculation errors were being made in the 'summing up' of items to obtain subscale and overall total scores. In Vignette 1, based on the total score, 26% of participants made a 'summing up' error; in Vignette 2, 40% of participants made this error; and in Vignette 3 it was 18%. In Vignette 2 the majority of participants made the same calculation error, they failed to include the Name and address recognition item in their total score.

<u>Discussion</u>

Despite the fact that some significant differences have been obtained in this study, overall the majority of participant scores were either the same as, or within one or two points of the true scores at both total and subscale levels. It is probable that such a discrepancy will be acceptable in clinical practice. Nonetheless, if emphasis is placed on fixed cut-off scores then even a few points of variation could have significant clinical implications; however, since the ACE-R is a screening tool its results should not be used in such a definitive manner.

Total scores

The majority of participants either rated vignette performance as the true score or within a few points of the true score for both the total and subscale scores across all three of the vignettes. In terms of the hypothesis that, due to participant over-scoring, participant total scores would be significantly different from true total scores, the results indicated that over-scoring did not systematically occur. In Vignettes 1 and 3 there was a significant difference between participant and true total scores, but results indicated that participants were under-scoring the ACE-R. In Vignette 2 participant variance from the true score did not occur in a systematic way. It may be that there was greater scope for under-scoring because there is a maximum amount of points per ACE-R item so when items are correct they cannot be over-scored, but may still be under-scored. In addition, participants may have been more cautious when scoring the vignettes in this research project than they are in clinical practice. They may also have been less likely to give the mock patients in the vignettes the 'benefit of the doubt' because they had not directly administered the ACE-R to them.

The significant differences between participant scores compared to true scores for the totals of Vignettes 1 and 3 had large effect sizes. The true total score for Vignette 1 was 75; in comparison the true total score for Vignette 3 was 85. Both vignettes indicated a degree of under-scoring; however, there was less variance in participant scores for Vignette 3 compared to Vignette 1. This suggests that scoring accuracy may be poorer when patient true scores involve more errors. This is concerning for clinical practice as the majority of participants who present to services for dementia screening are likely to present with some degree of cognitive impairment. If patient performance on the ACE-R is under-scored in clinical practice this may lead to false positive results on the tool, which may cause unnecessary distress and worry.

Subscales

Particular ACE-R subscales were associated with less scoring accuracy across the vignettes. Although the present study hypothesised that this would be found within more subjective subscales such as Language and Visuospatial, it was the Memory subscale that had the least accuracy. It was the only one where there was a significant difference between participant scores and the true score across all three vignettes. This is an interesting finding given that the Memory subscale contributes 26% to the weighted ACE-R total score and memory difficulties are commonly found in the intended patient population.

There was no significant difference found in the Fluency subscale across all of the vignettes. However in Vignette 2, 40% of participants did not identify the true Fluency scores, this may be reflective of the perseveration of words (e.g. pig and piggies) that occurred on the fluency items in Vignette 2, but not Vignettes 1 and 3. The scoring guidelines state that perseverations should not count toward the total of correct responses.

The Vignette 2 Fluency results indicate that notable numbers of participants were counting the perseveration responses; this suggests that participants were not aware of the finer details involved in scoring this subscale.

Individual items

There were certain ACE-R individual items found to have lower scoring accuracy across vignettes. In particular the Name and Address recall and recognition items of the Memory subscale were frequently scored inaccurately, this may indicate that the participants were not aware of the scoring instructions for these items. In Vignette 2 the filmed mock patient correctly recalled all the information in the recall item and so was not administered the recognition item; however, according to the ACE-R scoring guidelines, full marks for the recognition item should have been automatically rewarded. Several participants awarded no points for the recognition item; this may indicate that many participants were not aware of this scoring rule. Since the recognition item is weighted 5 points, omitting it can cause substantial deviation from the true score. Incorporating the guidelines for scoring the Name and address recall and recognition items into the ACE-R scoring sheet, instead of them being in the separate administration and scoring guidelines document, may reduce the likelihood of these scoring errors.

The clock item on the Visuospatial subscale was another item which had poorer scoring accuracy, this is one of the most subjective items and participants may not have been aware of the rules for assigning points according to the clock face, numbers and hands as outlined in the ACE-R scoring guide. The Naming item in the Language subscale was also identified as being scored with lower accuracy. This item has a maximum of ten points and involves the patient being required to name ten line drawings. It may be that participants differed in what they considered to be acceptable answers, which may also be

the case for the orientation questions which had a lower level of accuracy associated with them. It might be beneficial to have more detailed information on acceptable responses for these items available in the ACE-R administration and scoring guidelines.

In the current study 33 out of the 57 participants did not request a copy of the administration and scoring guidelines whilst scoring up their forms, indicating that guidelines may not be routinely referred to in clinical practice. Instead of increasing detail in the scoring guidelines, it may be more beneficial if the ACE-R scoring forms were adapted to incorporate more detailed scoring information for items identified as having poorer accuracy. This may minimise scoring errors and thus increase scoring accuracy.

The study identified that a significant number of participants made errors when summing up their item scores to give the total score, it may be that because the ACE-R is scored out of 100 this increases the likelihood of making such a human calculation error.

Limitations of current study

The study involved investigation of several hypotheses (and therefore multiple comparisons). Given that the majority of these were specific apriori hypotheses and that identifying significant differences between participant scores and true scores (or between experience variables and scoring accuracy) would potentially have important implications for clinical practice, it was felt that any correction for multiple comparisons (e.g. Bonferroni correction) would be too conservative and inflate the likelihood of a Type II error (Perneger, 1998). Nonetheless, not undertaking any correction for multiple comparisons may have increased the chance of a Type I error in the results.

The one sample Wilcoxon signed ranks tests aimed to identify if the participants were systematically scoring in a particular direction (i.e. over-scoring) which was significantly different from the true score. Using this test it is not possible to conclude from a non-significant result that the participant scores showed a high level of accuracy when compared to the true scores because a high scatter of differences may lead to a non-significant result when there is actually low accuracy (Altman, 1991). This is evidenced when Vignette 3 and Vignette 2 total scores are compared. It is clear that participant scores show greater variance from the true score in Vignette 2 than Vignette 3. In Vignette 3 many participant scores are slightly below the true score so the Wilcoxon reports a significant difference. In comparison, because the variance in Vignette 2 is on either side of the true score the Wilcoxon reports a non-significant result. It has been recommended that, when comparing two methods of measurement, (in this study this would translate to comparing the participant scores and the fixed true scores) the differences between them should be plotted in a graph and their mean differences and standard deviations obtained (Altman, 1991). However, the interpretation of these means and standard deviations must depend on clinical circumstances; it is not possible to use statistics to define acceptable agreement in such cases (Altman, 1991). Therefore, in this study, the limits of how many points deviation from the true score are clinically acceptable cannot be answered statistically. As a result, the conclusion that a few points deviation is likely to be clinically acceptable is based on expert opinion rather than statistical assessment.

The power calculation recommended a minimum sample of 34 participants. Whilst the study recruited 57 participants the participant group contained more CNs (45) than TCPs (12). It had been hoped that it would be possible to recruit 34 participants for each of the professional groups. Unfortunately several TCPs had only observed the ACE-R being

used in clinical practice and therefore did not meet inclusion criteria. There was a significant difference between the two groups in Vignette 1, but not in Vignettes 2 and 3; the decision was therefore made to combine the groups. However, there may be significant differences between these two professional groups in terms of their ACE-R scoring accuracy which, due to the small TCP sample size, this study was unable to reliably detect. In addition, the current study did not identify significant differences in scoring accuracy as a result of experience with the ACE-R. Again, the numbers in the participant groups involved in these analyses may have been too small to detect significant effects.

Furthermore, not all participants summed up the ACE-R items to obtain subscale and total scores; this meant there was some missing information which may have impacted on the analysis. The likelihood of this is reduced by the fact that the minimum sample size was never violated.

This study used a design which enabled scoring accuracy to be explored, independent of test administration, for multiple raters. Part of the design involved devising a set of 'true' scores for each of the vignettes. It is possible that the true scores included a degree of error; however, the likelihood of this was minimised by scripting the vignettes. In addition, the 100% accuracy and agreement that was obtained when the vignettes were scored by two experienced Clinical Psychologists suggests that the defined true scores were an accurate reflection of actual true scores.

Potential areas for future research

In future it would be useful if this study could be replicated using a larger participant sample. Future replications of the study would benefit from recruiting larger participant

sub-groups to enable more comprehensive assessment of potential scoring accuracy differences between professional groups and between individuals with varying levels of ACE-R experience.

This study exclusively investigated multiple rater scoring accuracy on the ACE-R, however scoring accuracy is only one element of inter-rater reliability, how the tool is administered by multiple raters should be addressed in future research. In addition research exploring the intra-rater reliability (test re-test) of the ACE-R would make a valuable contribution to the ACE-R literature.

It has been suggested that scoring accuracy on screening tools might be increased by using filmed vignettes as interactive training exercises (Queally et al., 2010). This suggests that the design used in this study could be adapted in future research. Using the vignettes of mock patients is an ethical means of generating a realistic clinical scenario which researchers are able to manipulate according to their objectives.

Conclusion

Results indicate that the ACE-R is a robust tool when scored by multiple raters. There were however significant differences between the participant and true scores across the vignettes at both the total and subscale level. Practitioners should take these findings into account in clinical practice and, as a result, be wary about using definitive cut off scores. Certain items on the ACE-R were identified across the vignettes as being associated with lower scoring accuracy; these items typically required a more subjective level of judgement or had a more complex scoring system. It may be that accuracy could be improved if raters familiarised themselves extensively with the scoring guide or more

detailed guidance on how to score these items was incorporated into the ACE-R scoring sheets. The vignettes design utilised in this study could be adapted to investigate similar issues related to scoring and administration in future.

References

Altman DG. 1991. *Practical statistics for medical research.* Chapman and Hall: London.

Bak TH, Mioshi E. 2007. A cognitive bedside assessment beyond the MMSE: The Addenbrooke's Cognitive Examination. *Pract Neurol* **7**(4): 245-249.

British Psychological Society. 2009. *Code of ethics and conduct.* [Internet] The British Psychological Society: Leicester. Available at: http://www.bps.org.uk/document-download-area/document-download$.cfm?file_uuid=E6917759-9799-434A-F313-9C35698E1864&ext=pdf [Accessed 9th June 2010].

Cullen B, O'Neill B, Evans JJ, et al. 2007. A review of screening tests for cognitive impairment. *J Neurol Neurosurg Psychiatry* **78**: 790-799.

Faul F, Erdfelder E, Lang AG, Buchner A. 2007 G*Power 3: A flexible statistical power analysis program for the social, behavioral & biomedical sciences. *Behav Res Methods* **39**: 175-191.

Gifford DR, Cummings JL. 1999. Evaluating Dementia Screening Tests: Methodologic standards to rate their performance. *Neurology* **52**(2): 224-227.

Hannay HJ, Howieson DB, Loring DW, et al. 2004. Neuropathology for Neuropsychologists. In *Neuropsychological Assessment 4th edition*, Lezak MD, Howieson B, Loring DW (eds). Oxford University Press Inc.: New York; 157-285.

Hannay HJ, Lezak MD. 2004. The neuropsychological examination: interpretation. In *Neuropsychological Assessment 4th edition*, Lezak MD, Howieson B, Loring DW (eds). Oxford University Press Inc.: New York; 133-156.

Mathuranath PS, Nestor PJ, Berrios GE, et al. 2000. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology* **55**(11): 1613–20.

McDowell I. 2006. *Measuring Health: a guide to rating scales and questionnaires.* [e-book] Oxford University Press: New York. Available at: http://books.google.co.uk/books?id=06t-63RaYk0C&dq=mcdowell+measuring+health&printsec=frontcover&source=bl&ots=ok5k80bBHN&sig=kde3L9sYcKmVrr9Yf3nAbuJv6Q8&hl=en&ei=wO8RSovfL8afjAf77-HkCA&sa=X&oi=book_result&ct=result&resnum=1#PPA8,M1 [Accessed 10th January 2009].

Mioshi E, Dawson K, Mitchell J, et al. 2006. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* **21**: 1078-1085.

National Audit Office. 2007. *Improving services and support for people with dementia*. [Internet] The Stationery Office: London. Available at: http://www.nao.org.uk/publications/0607/support_for_people_with_dement.aspx [Accessed 20th June 2010].

National Institute for Health and Clinical Excellence (NICE). 2006. *Clinical Guideline*

42, Dementia: Supporting people with dementia and their carers in health and social care. [Internet] NICE: London. Available at: http://www.nice.org.uk/nicemedia/pdf/CG042NICEGuideline.pdf [Accessed 5th December 2009].

Perneger TV. 1998. What's wrong with Bonferroni adjustments. *BMJ* **316**: 1236.

Queally VR, Evans JJ, McMillan TM. 2010. Accuracy in scoring vignettes using the mini mental state examination and the short orientation memory concentration test. *J Geriatr Psychiatry Neurol* **00**: 1-5.

Scottish Intercollegiate Guidelines Network (SIGN). 2006. *Management of patients with dementia, a National Clinical Guideline*, *(SIGN publication 86).* [Internet] SIGN: Edinburgh. Available at: http://www.sign.ac.uk/pdf/sign86.pdf [Accessed 5th December 2008].

Trochim WMK. 2006. *Research Methods Knowledge Base (Reliability Section).* [Internet] Web Center for Social Research Methods. Available at: http://www.socialresearchmethods.net/kb/reliable.php [Accessed 10th January 2009].

Woodford HJ, George J. 2007. Cognitive assessment in the elderly: a review of clinical methods. *QJM* **100**: 469-484.

Figure 1: Vignette Total Score

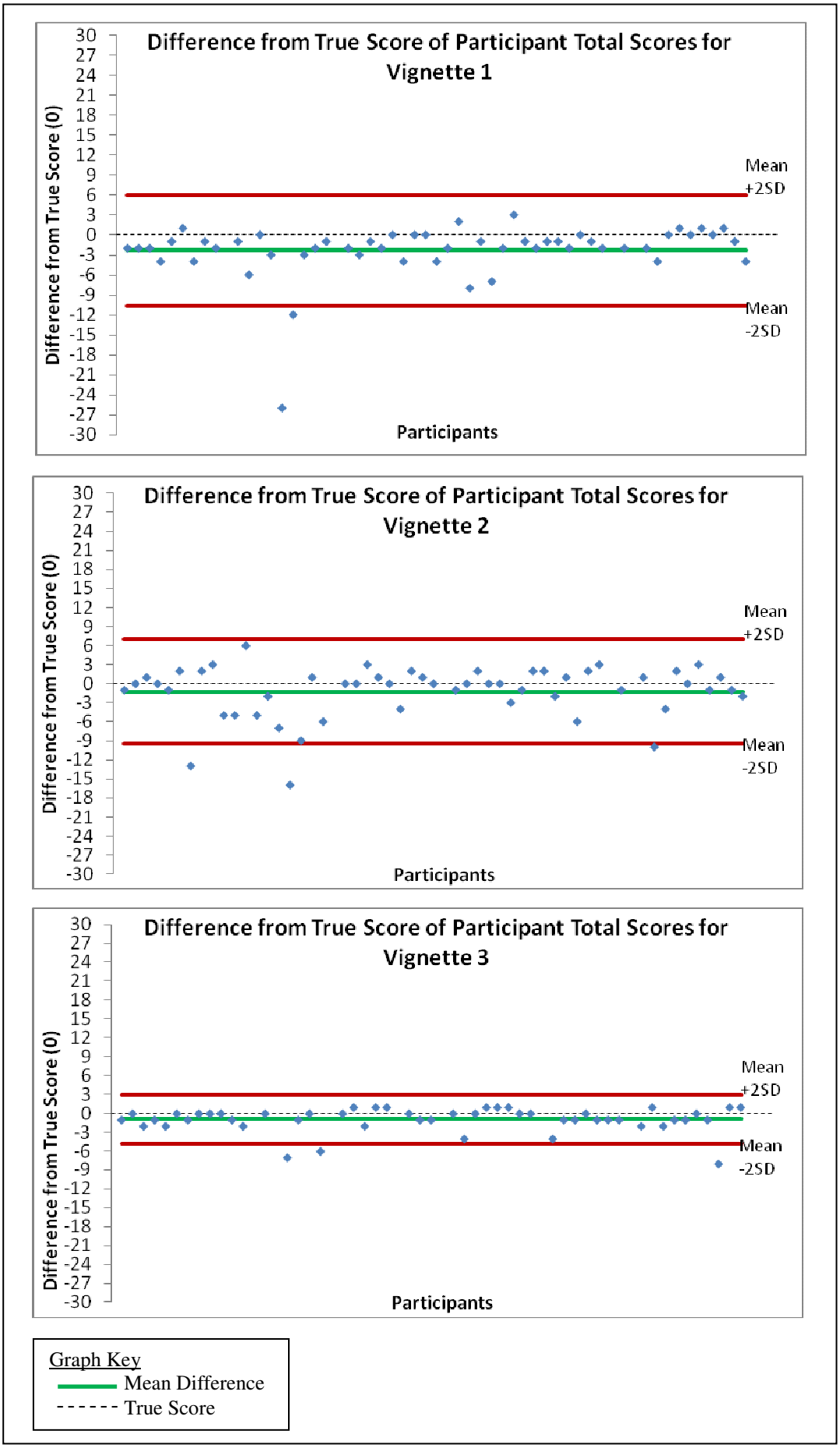

Table 1: Summary of one sample Wilcoxon signed rank tests

ACE-R scores	Vignette 1			Vignette 2			Vignette 3		
	True score	Median	One sample Wilcoxon Signed Ranks test	True score	Median	One sample Wilcoxon Signed Ranks test	True score	Median	One sample Wilcoxon Signed Ranks test
Total	75	73	W(52) = 87.5, z=-4.9, p=<0.001, r = 0.67	84	84	W(52) = 361.50, z=-1.36, p=0.09, r = 0.19	85	84	W(49) = 112.50, z=-3.46, p=<0.001, r = 0.49
Language	22	22	W(47) = 150.50, z=-0.643, p=0.26, r = 0.09	24	24	W(46) = 151.00, z=-0.327, p=0.37, r = 0.05	26	26	W(45) = 0.00, z=-3.71, p=<0.001, r = 0.55
Visuospatial	14	14	W(46) = 41.50, z=-1.40, p=0.08, r = 0.20	12	13	W(47) = 494.50, z=3.51, p=<0.001, r = 0.51	14	15	W(45) = 883.00, z=4.50, p=<0.001, r = 0.66
Attention & Orientation	16	16	W(46) = 100, z=-2.9, p=0.004, r = 0.42	17	17	W(46) = 39.00, z=-0.48, p=0.632, r = 0.07	18	17	W(44) = 0.00, z=-5.67, p=<0.001, r = 0.84
Memory	18	17	W(47) = 0.00, z=-5.5, p=<0.001, r = 0.79	26	26	W(46) = 1.00, z=-3.57, p=<0.001, r = 0.52	19	19	W(44) = 0.00, z=-3.30, p=0.001, r = 0.49
Fluency	5	5	W(46) = 23.50, z=-4.2, p=0.67, r = 0.61	5	5	W(46) = 139.00, z=1.91, p=0.057, r = 0.28	8	8	W(44) = 4.00, z=-0.38, p=0.705, r = 0.06

Figure 2: Participant differences from true subscale scores

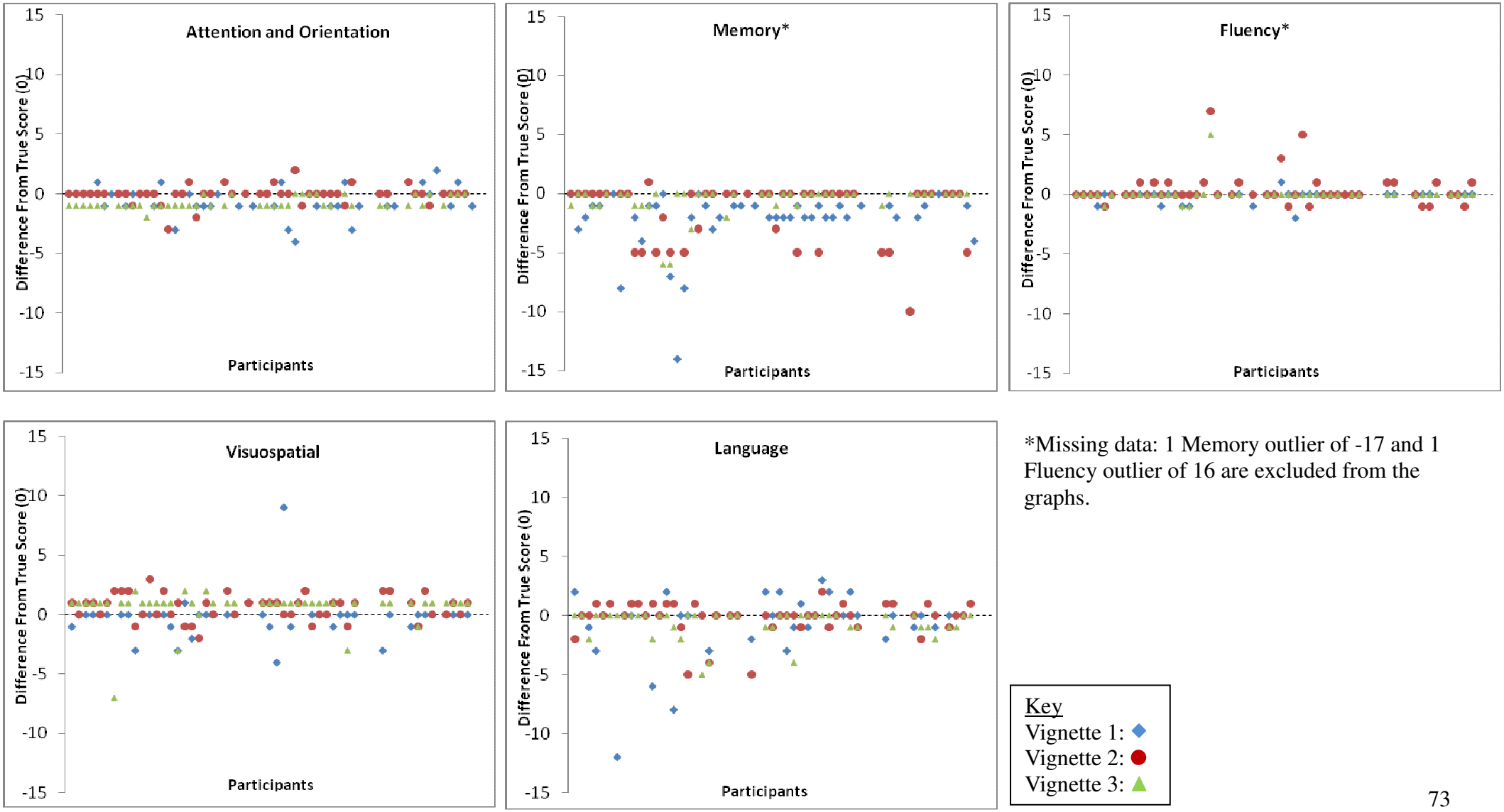

Table 2: Breakdown of participant scores per question for Vignettes 1, 2 and 3

Attention and Orientation	Vignette 1 (%)				Vignette 2 (%)				Vignette 3 (%)			
	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS
Orientation 1	3/5	68	32		5/5	100			5/5	100		
Orientation 2	5/5	89	11		4/5	79	21		5/5	25	75	
Registration	3/3	95	5		3/3	100			3/3	100		
Attention & Conc.	5/5	81	17	2	5/5	86	12	2	5/5	98	2	
Subscale Total	**16/18**	**36**	**53**	**11**	**17/18**	**72**	**26**	**2**	**18/18**	**27**	**73**	

Memory	Vignette 1 (%)				Vignette 2 (%)				Vignette 3 (%)			
	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS
Recall	2/3	100			3/3	98	2		3/3	100		
Anterograde memory	6/7	100			7/7	98	2		6/7	100		
Retrograde memory	2/4	88	12		4/4	98	2		4/4	98	2	
Name&AddressRecall	4/7	28	70	2	7/7	100			1/7	89	11	
Name&AddressRecog	4/5	53	47		5/5	75		25	5/5	95	3	2
Subscale Total	**18/26**	**21**	**62**	**17**	**26/26**	**66**	**2**	**32**	**19/26**	**71**	**22**	**7**

Fluency	Vignette 1 (%)				Vignette 2 (%)				Vignette 3 (%)			
	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS
Letter fluency	4/7	91	7	2	2/7	60	37	3	3/7	98	2	
Animal fluency	1/7	86	12	2	3/7	72	26	2	5/7	91	9	
Subscale Total	**5/14**	**79**	**19**	**2**	**5/14**	**60**	**34**	**6**	**8/14**	**91**	**7**	**2**

Language	Vignette 1 (%)				Vignette 2 (%)				Vignette 3 (%)			
	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS
Comprehension 1	1/1	100			1/1	100			1/1	100		
Comprehension 2	3/3	100			3/3	93	7		3/3	98	2	
Writing	1/1	98	2		1/1	98	2		1/1	100		
Repetition 1	2/2	100			2/2	98	2		2/2	100		
Repetition 2	1/1	100			1/1	100			1/1	100		
Repetition 3	0/1	81	19		0/1	96	4		1/1	100		
Pencil/watch naming	2/2	98	2		2/2	100			2/2	100		
Naming	7/10	56	42	2	10/10	96	4		10/10	76	24	
Comprehension 3	4/4	96	4		4/4	91	9		4/4	100		
Reading	1/1	100			0/1	74	26		1/1	100		
Subscale Total	**22/26**	**46**	**39**	**14**	**24/26**	**47**	**47**	**6**	**26/26**	**63**	**30**	**7**

Visuospatial	Vignette 1 (%)				Vignette 2 (%)				Vignette 3 (%)			
	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS	True Score (TS)	Same as TS	+/- 2 points from TS	More than +/-3 points from TS
Pentagon	1/1	93	7		1/1	100			1/1	98	2	
Cube	0/2	98	2		0/2	91	9		2/2	98	2	
Clock	5/5	93	2	5	3/5	33	67		4/5	2	98	
Dots	4/4	100			4/4	98	2		3/4	100		
Letters	4/4	96	4		4/4	96	4		4/4	100		
Subscale Total	**14/16**	**66**	**23**	**11**	**12**	**29**	**69**	**2**	**14/16**	**2**	**91**	**7**

Table 3: Output for Independent Samples Mann Whitney U Tests

Participant groups	Vignette 1	Vignette 2	Vignette 3
Guide used	U=252.5, p=0.083	U=421.0, p=0.169	U=250.5, p=0.312
Training	U=291.0, p=0.983	U=279.0, p=0.821	U=212.5, p=0.231
How often	U=292.5, p=0.895	U=274.5, p=0.607	U=186.0, p=0.077
Last use	U=301.5, p=0.952	U=294.0, p=0.905	U=193.0, p=0.102

Chapter 3: Advanced Clinical Practice 1, Reflective Critical Account (**Abstract only**)

The person vs. the disease: a reflective account of issues encountered when working with patients with dementia

Stephanie Crawford[1*]

[1]Section of Psychological Medicine, University of Glasgow

[*]Address for Correspondence
Section of Psychological Medicine
Division of Community Based Sciences
University of Glasgow
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XY
E-mail: scrawford@nhs.net

Word Count: 4315

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology (D.Clin.Psy)

Abstract

In this reflective account I apply the Atkins and Murphy (1994) model of reflection to an incident which I experienced within a Cognitive Stimulation Therapy Group for people with dementia. Using the model, I identify my initial emotional response within the group to a man with dementia, who presented as acutely distressed due to his inability to recall important information from his past. I then go on to consider my emotions in more detail, alongside their accompanying thoughts. From this process I produce some key questions about the incident and I draw upon my own knowledge, social beliefs and the writings of Terry Pratchett to explore these questions further. I consider the impact of the answers I generate within the wider context of services for people with dementia. Finally, I conclude by reflecting on the model used and my experience of writing the account.

Chapter 4: Advanced Clinical Practice 2, Reflective Critical Account

(Abstract only)

Inequality in health care: a hernia in need of removal

Stephanie Crawford[1*]

[1]Section of Psychological Medicine, University of Glasgow

[*]Address for Correspondence
Section of Psychological Medicine
Division of Community Based Sciences
University of Glasgow
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
G12 0XY
E-mail: scrawford@nhs.net

Word Count: 4786

Submitted in partial fulfilment of the requirements for the degree of Doctorate in Clinical Psychology (D.Clin.Psy)

Abstract

In this reflective account I apply the Rolfe et al. (2001) Framework of Reflexive Practice as a model to guide my reflections triggered by an encounter with a patient while on a joint home visit with a learning disability nurse. Using the Rolfe et al. (2001) model I initially detail the situation wherein a patient showed the nurse and I a hernia, which was causing him significant levels of discomfort. I consider my immediate and subsequent emotional responses to the incident, before going on to contemplate my thoughts and emotions in more detail, drawing upon wider sources of knowledge to assist my reflective processes as appropriate. I then consider the wider service level issues related to my reflections. Finally, I conclude by reflecting on the model used and my experience of writing the account.

Appendices

Appendix 1.1: Publication guidelines

Author Guidelines

1. AIMS & SCOPE

The rapidly increasing world population of aged people has led to a growing need to focus attention on the problems of mental disorder in late life. The aim of the *International Journal of Geriatric Psychiatry* is to communicate the results of original research in the causes, treatment and care of all forms of mental disorder which affect the elderly. The Journal is of interest to psychiatrists, psychologists, social scientists, nurses and others engaged in therapeutic professions, together with general neurobiological researchers.

The Journal provides an international perspective on the important issue of geriatric psychiatry, and contributions are published from countries throughout the world. Topics covered include epidemiology of mental disorders in old age, clinical aetiological research, post-mortem pathological and neurochemical studies, treatment trials and evaluation of geriatric psychiatry services.

Further information about the Journal, including links to the online sample copy and contents pages, can be found on the Journal homepage .

2. MANUSCRIPT CATEGORIES

The *International Journal of Geriatric Psychiatry* invites the following types of submission:

Research Articles

Research Articles are the Journal's primary mode of scientific communication. Peer-review of Research Articles will be handled by the most appropriate Editor . Research Articles must not exceed **3500** words of body text, and are limited to 6 figures/tables.

Review Articles

Review Articles will typically be solicited by the Editors. Authors who wish to submit an unsolicited review should first contact one of the Editors to determine its suitability for publication in the Journal. All reviews will be peer-reviewed. Reviews must not exceed **4500** words of body text, and are limited to 6 figures/tables and 150 references.

Letters to the Editor

Letters to the Editor, or Correspondence, may be in response to issues arising from recently published articles, or short, free-standing pieces expressing an opinion, but should not exceed **700** words of body text, and are limited to 1 figure/table and 5 references. Letters are not subject to external peer-review.

3. MANUSCRIPT SUBMISSION

All submissions should be made online at the *International Journal of Geriatric Psychiatry* **ScholarOne Manuscripts** site— http://mc.manuscriptcentral.com/gps . New users should first create an account. Once a user is logged onto the site, submissions should be made via the Author Centre.

Authors must also supply to the Editorial Office by fax (+44 1243 770460) the following:

- a Copyright Transfer Agreement with original signature(s)—without this we are unable to accept the submission, and
- permission grants—if the manuscript contains extracts, including illustrations, from other copyright works (including material from on-line or intranet sources) it is the author's responsibility to obtain written permission from the owners of the publishing rights to reproduce such extracts using the Wiley Permission Request Form .

4. MANUSCRIPT PREPARATION

Manuscripts must be written in English.

Text should be supplied in a format compatible with Microsoft Word for Windows (PC). Charts and tables are considered textual and should also be supplied in a format compatible with Word. All figures (illustrations, diagrams, photographs) should be supplied in jpg, tiff or eps format.

All manuscripts must be typed in 12pt font and in double space with margins of at least 2.5 cm.

Manuscripts must comply with the word limits defined in section 2, and include:

Title Page

The first page of the manuscript should contain the following information:

- the title of the paper
- a running head not exceeding 50 characters
- 2–6 article keywords **and** up to 4 key points
- names of authors
- names of the institutions at which the research was conducted
- name, address, telephone and fax number, and email address of corresponding author
- the name(s) of any sponsor(s) of the research contained in the paper, along with grant number(s)
- the word count of the body text

Structured Abstracts

Authors submitting Research and Review Articles should note that structured abstracts (maximum 250 words) are required. The structured abstract should adopt the format: Objective, Methods, Results, Conclusions. (Authors of Reviews may use Design instead of Method.) Abstracts should contain no citation to other published work.

Letters to the Editor do not require abstracts.

Text

This should in general, but not necessarily, be divided into sections with the headings: Introduction, Methods, Results, Discussion, Conclusion.

Research Letters and Correspondence should be formatted in one continuous section.

Tables and Figures

Tables and figures should not be inserted in the appropriate place in the text but should be included at the end of the paper, each on a separate page.

Tables and figures should be referred to in text as follows: Figure 1, Figure 2; Table 1, Table 2. The place at which a table or figure is to be inserted in the printed text should be indicated clearly on a manuscript. Each table and/or figure must have a legend that explains its purpose without reference to the text.

Any figure submitted as a colour original will appear in colour in the Journal's online edition free of charge. Colour figures will be printed in the Journal on the condition that authors contribute to the associated costs: **£350** for the first page; **£150** for each subsequent page thereafter. Corresponding authors will be invoiced post-publication.

References

References should be in 'Harvard' format, i.e, names and dates in brackets in the text (Jones, 2000; Smith and Jones, 2001; Jones *et al* ., 2002), and the full reference listed at the end of the paper, in alphabetical order by first author, as follows:

Porteus SD. 1959. *The Maze Tests and Clinical Psychology* . Pacific Books: Palo Alto.

Rabbitt PMA. 1982. How do old people know what to do next? In *Aging and Cognitive Processes* , Craik FIM, Trehub S (eds). Plenum Press: New York; 79–98.

Chou K-L, Chi I. 2004. Combined effect of vision and hearing impairment on depression in elderly Chinese. *Int J Geriatr Psychiatry* **19** : 825–832. DOI: 10.1002/gps.1174

(Titles of periodicals should be abbreviated according to the style used in Index Medicus.)

We recommend the use of a tool such as EndNote for reference management and formatting. Click here to download the most up-to-date EndNote reference style for *International Journal of Geriatric Psychiatry*.

5. DECLARATION

Original Publication

Submission of a manuscript will be held to imply that it contains original unpublished work and is not being submitted for publication elsewhere at the same time. The author must supply a full statement to the Editor-in-Chief about all submissions and previous reports that might be regarded as redundant or duplicate publication of the same or very similar work.

Conflict of Interest

Authors are responsible for disclosing all financial and personal relationships between themselves and others that might bias their work. To prevent ambiguity, authors must state explicitly whether potential conflicts do or do not exist. Investigators should disclose potential conflicts to study participants and should state in the manuscript whether they have done so. Authors should describe the role of the study sponsor(s), if any, in study design, in the collection, analysis and interpretation of data, in the writing of the report and in the decision to submit the report for publication. If the supporting source had no such involvement, the authors should so state.

Ethics

When reporting experiments on human subjects, indicate whether the procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional or regional) and with the Helsinki Declaration of 1975, as revised in 1983. Do not use patients' names, initials or hospital numbers, especially in illustrative material. When reporting experiments on animals, indicate whether the institution's or a national research council's guide for, or any national law on, the care and use of laboratory animals was followed. A statement describing explicitly the ethical background to the studies being reported should be included in all manuscripts in the Materials and Methods section. Ethics committee or institutional review board approval should be stated.

Patients have a right to privacy that should not be infringed without informed consent. Identifying information should not be published in written descriptions, photographs and pedigrees unless the information is essential for scientific purposes and the patient (or parent or guardian) gives written informed consent for publication. Informed consent for this purpose requires that the patient be shown the manuscript to be published. Identifying details should be omitted if they are not essential but patient data should never be altered or falsified in an attempt to attain anonymity. Complete anonymity is difficult to achieve and informed consent should be obtained if there is any doubt. For example, masking the eye region in photographs of patients is inadequate protection of anonymity.

Authorship

All persons designated as authors should qualify for authorship and all those who qualify should be listed. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content. One or more authors should take responsibility for the integrity of the work as a whole, from inception to published article. Authorship credit should be based only on 1) substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; 2) drafting the article or revising it critically for important intellectual content; 3) final approval of the version to be published. Conditions 1, 2 and 3 must all be met. Acquisition of funding, the collection of data or general supervision of the research group, by themselves, do not justify authorship. All others who contributed to the work who are not authors should be named in the Acknowledgements section.

Committee on Publication Ethics (COPE)

As a member of the Committee on Publication Ethics (COPE), adherence to these submission criteria is considered essential for publication in *International Journal of Geriatric Psychiatry* ; mandatory fields are included in the online submission process to ensure this. If, at a later stage in the submission process or even after publication, a manuscript or authors are found to have disregarded these criteria, it is the duty of the Editor-in-Chief

to report this to COPE. COPE may recommend that action may be taken, including but not exclusive to, informing the authors' professional regulatory body and/or institution of such a dereliction.

The website for COPE may be accessed at: http://www.publicationethics.org.uk

6. ADDITIONAL INFORMATION ON ACCEPTANCE

Copyright

To enable the publisher to disseminate authors' work to the fullest extent, all authors must sign a Copyright Transfer Agreement, transferring copyright in the article to John Wiley & Sons. An agreement with original signatures will be requested if a paper is accepted for publication. A copy of the agreement to be used can be downloaded from this link: Copyright Transfer Agreement .

Proofs

Proofs of accepted articles will be sent to the author for checking. This stage is to be used only to correct errors that may have been introduced during the production process. Prompt return of the corrected proofs, preferably within two days of receipt, will minimise the risk of the paper being held over to a later issue.

Offprints

Free access to the final PDF offprint or your article will be available via Author Services. Please therefore sign up for Author Services if you would like to access your article PDF offprint and enjoy the many other benefits the service offers.

Early View

Early View is Wiley's exclusive service presenting individual articles online as soon as they are ready before the release of the compiled print issue. Early View articles are complete, citable and are published in an average time of 6 weeks from acceptance.

OnlineOpen

OnlineOpen is available to authors of primary research articles who wish to make their article available to non-subscribers on publication, or whose funding agency requires grantees to archive the final version of their article. With OnlineOpen the author, the author's funding agency, or the author's institution pays a fee to ensure that the article is made available to non-subscribers upon publication via Wiley InterScience, as well as deposited in the funding agency's preferred archive.

Note to NIH grantees

Pursuant to NIH mandate, Wiley-Blackwell will post the accepted version of contributions authored by NIH grant holders to PubMedCentral upon acceptance. This accepted version will be made publicly available 12 months after publication. For further information, click here

Appendix 1.2: Detailed search strategy

All of the text word searches outlined below were conducted on the databases: Medline, EMBASE, PsychINFO, All EBM Reviews. Each search was limited to the time span 2000-April 2010 and duplicates were removed.

Search 1	((ACE?R or Addenbrooke*) and (dement* or alzheimer* or cognitive impair*)).tw.
Search 2	(ACE and (dement* or alzheimer* or cognitive impair*) and (scale* or test* or tool* or screen* or assess* or battery or questionnaire*)).tw.
Search 3	(ACE?R or Addenbrooke*) were combined with the (depressi* or head injur* or stroke or brain injur* or cerebrovascular accident* or cerebrovascular incident* or TIA or transient isch?emic or scan* or MRI or magnetic resonance or CT or compute* tomograph*)).tw.
Search 4	Addenbrooke*.tw.
Search 5	(ACE and (depressi* or head injur* or stroke or brain injur* or cerebrovascular accident* or cerebrovascular incident* or TIA or transient isch?emic or scan* or MRI or magnetic resonance or CT or compute* tomograph*) and (scale* or test* or tool* or screen* or assess* or battery or questionnaire*)).tw
Search 6	((scale* or test* or tool* or screen* or assess* or questionnaire) adj (dement* or Alzheimer* or cognitive impair*)).tw
Search 7	Combined searches 3, 4, 5, 6, 7 using 'OR'

Appendix 1.3: Quality rating checklist

Checklist for Diagnostic Accuracy

Study Reference: _____

SCORING:
2 – information well presented and detailed
1 – information present but lacks adequate detail
0 – information absent

Items	Score
Abstract and Introduction	
Abstract provides structured summary of trial design, methods, results, and conclusions and identifies article as a study of diagnostic accuracy	
Introduction clearly states the research question or study aims	
Total for abstract and introduction	**/4**
Methodology	
The spectrum of participants is representative of the participants who will receive the test in practice	
Selection criteria are clearly described (inclusion and exclusion)	
Participant recruitment explained (ie was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?)	
Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard.	
Is the reference standard likely to classify the condition correctly	
The period between reference standard and index test is short enough to be reasonably sure that the target condition did not change between the two tests.	
The reference standard was independent of the index test (i.e. the index test did not form part of the reference standard).	
The execution of the index test was described in sufficient detail to permit replication of the test	
The execution of the reference standard was described in sufficient detail to permit replication of the test.	
Index test results were interpreted without knowledge of the results of the reference standard	
Total for methodology	**/20**
Results and Discussion	
An explanation is provided for withdrawals from the study	
Report when study was done, including beginning and ending dates of recruitment.	
Report clinical and demographic characteristics of the study population.	
Describe methods for calculating or comparing measures of diagnostic accuracy, and other statistical methods used seem reasonable.	
Discussion: Trial limitations are acknowledged and the clinical applicability of the study findings are discussed.	
Total for Results and Discussion	**/10**
Grand Total	**/34**

Appendix 1.4: References of excluded studies

Ahmed S, Mitchell J, Arnold R, et al. 2008. Predicting rapid clinical progression in amnestic mild cognitive impairment. *Dement Geriatr Cogn Disord* **25**(2): 170-177.

Alexopoulos P, Greim B, Nadler K, et al. 2006. Validation of the Addenbrooke's cognitive examination for detecting early Alzheimer's disease and mild vascular dementia in a German population. *Dement Geriatr Cogn Disord* **22**(5-6): 385-391.

Bak TH, Rogers TT, Crawford LM, et al. 2005. Cognitive bedside assessment in atypical parkinsonian syndromes. *J Neurol Neurosurg Psychiatry*, **76**(3): 420-422.

Bak TH, Mioshi E, 2007. A cognitive bedside assessment beyond the MMSE: The Addenbrooke's Cognitive Examination. *Pract Neurol* **7**(4): 245-249.

Bier JC, Donckels V, Van Eyll E, et al. 2005. The French Addenbrooke's cognitive examination is effective in detecting dementia in a french-speaking population. *Dement Geriatr Cogn Disord* **19**(1): 15-17.

Bier JC, Ventura M, Donckels V, et al. 2004. Is the Addenbrooke's Cognitive Examination effective to detect frontotemporal dementia? *J Neurol* **251**(4): 428-431.

Carvalho VA, Barbosa MT, Caramelli P. 2010. Brazilian version of the addenbrooke cognitive examination-revised in the diagnosis of mild Alzheimer disease. *Cogn Behav Neurol* **23**(1): 8-13.

Gaber TA. 2008. Evaluation of the Addenbrooke's Cognitive Examination's validity in a brain injury rehabilitation setting. *Brain Inj* **22**(7-8): 589-593.

Garcia-Caballero A, Garcia-Lado I, Gonzalez-Hermida J. 2006. Validation of the Spanish version of the Addenbrooke's Cognitive Examination in a rural community in Spain. *Int J Geriatr Psychiatry* **21**(3): 239-245.

Hancock P, Larner AJ. 2009. Diagnostic utility of the Informant Questionnaire on Cognitive Decline in the Elderly (IQCODE) and its combination with the Addenbrooke's Cognitive Examination-Revised (ACE-R) in a memory clinic-based population. *Int Psychogeriatr* **21**(3): 526-530.

Hornberger M, Shelley BP, Kipps CM. 2009. Can progressive and non-progressive behavioural variant frontotemporal dementia be distinguished at presentation? *J Neurol Neurosurg Psychiatry* **80**(6): 591-593.

Hummelova-Fanfrdlova Z, Rektorova I, Sheardova K. 2009. Czech adaptation of Addenbrooke's Cognitive Examination. *Cesk Psychol* **53**(4): 376-388.

Jeyapaul P, Kerwick S. 2008. Addenbrooke's Cognitive Examination as a better discriminator of cognitive impairment than the Mini-Mental State Examination in patients with dementia. *Int Psychogeriatr* **20**(3): 642-643.

Kipps CM, Nestor PJ, Dawson CE, et al. 2008. Measuring progression in frontotemporal dementia: implications for therapeutic interventions. *Neurology* **70**(22): 2046-52.

Larner AJ. 2008. Reply. *Age Ageing* **37**(3): 350-351.

Larner AJ. 2006. Audit of the Addenbrooke's Cognitive Examination (ACE) in clinical practice: 2. longitudinal change. *Int J Geriatr Psychiatry* **21**(7): 698-699.

Larner AJ. 2005. An audit of the Addenbrooke's Cognitive Examination (ACE) in clinical practice. *Int J Geriatr Psychiatry* **20**(6): 593-594.

Stacho L, Dudas R, Ivady R, et al. 2003. Addenbrooke's Cognitive Examination: Developing the Hungarian Version. *Psychiatr Hung* **18**(4): 226-240.

Mathuranath PS, Hodges JR, Mathew R, et al. 2004. Adaptation of the ACE for a Malayalam speaking population in southern India. *Int J Geriatr Psychiatry* **19**(12): 1188-1194.

Mathuranath PS, Cherian PJ, Mathew R, et al. 2010. Dementia in Kerala, South India: Prevalence and influence of age, education and gender. *Int J Geriatr Psychiatry* **25**(3): 290-297.

Mathuranath PS, Cherian J, Mathew R, et al. 2007. Mini Mental State Examination and the Addenbrooke's Cognitive Examination: Effect of education and norms for a multicultural population. *Neurol India* **55**(2): 106-110.

Mioshi E, Kipps CM, Dawson K, et al. 2007. Activities of daily living in frontotemporal dementia and Alzheimer disease. *Neurology* **68**(24): 2077-2084.

Mioshi E, Kipps CM, Hodges JR. 2009. Activities of daily living in behavioral variant frontotemporal dementia: Differences in caregiver and performance-based assessments. *Alzheimer Dis Assoc Disord* **23**(1): 70-76.

Mitchell J, Arnold R, Dawson K, et al. 2009. Outcome in subgroups of mild cognitive impairment (MCI) is highly predictable using a simple algorithm. *J Neurol* **256**(9):1500-1509.

Newman JP. 2005. Brief assessment of cognitive mental status in Hebrew: Addenbrooke's Cognitive Examination. *Isr Med Assoc J* **7**(7): 451-452.

Newman JP. 2007. Invited commentary. Mini mental status examination and the Addenbrooke's cognitive examination: effect of education and norms for a multicultural population. *Neurol India* **55**(2): 99.

Nyatsanza S, Shetty T, Gregory C, et al. 2003. A study of stereotypic behaviours in Alzheimer's disease and frontal and temporal variant frontotemporal dementia. *J Neurol Neurosurg Psychiatry* **74**(10): 1398-1402.

Pouretemad HR, Khatibi A, Ganjavi A. 2009. Validation of addenbrooke's cognitive Examination (ACE) in a persian-speaking population. *Dement Geriatr Cogn Disord* **28**(4): 343-347.

Robben SHM, Sleegers MJM, Dautzenberg PLJ, et al. 2010. Pilot study of a three-step diagnostic pathway for young and old patients with Parkinson's disease dementia: Screen, test and then diagnose. *Int J Geriatr Psychiatry* **25**(3): 258-265.

Roca M, Torralva T, Lopez P, et al. 2008. Differentiating early dementia from major depression with the Spanish version of the Addenbrooke's Cognitive Examination. *Rev Neurol* **46**(6): 340-343.

Sarasola D, De Lujan-Calcagno M, Sabe L, Crivelli L, et al. 2005. Validity of the Spanish version of the Addenbrooke's Cognitive Examination for the diagnosis of dementia and to differentiate Alzheimer's disease and frontotemporal dementia. *Rev Neurol* **41**(12): 717-721.

Shelley BP, Hodges JR, Kipps CM, et al. 2009. Is the pathology of corticobasal syndrome predictable in life? *Mov Disord* **24**(11): 1593-9.

Stokholm J, Vogel A, Johannsen P, Waldemar G. 2009. Validation of the danish addenbrooke's

cognitive examination as a screening test in a memory clinic. *Dement Geriatr Cogn Disord* **27**(4): 361-365.

Torralva T, Roca M, Gleichgerrcht E, et al. 2009. A neuropsychological battery to detect specific executive and social cognitive impairments in early frontotemporal dementia. *Brain* **132**(5): 1299-1309.

Woodford HJ, George J. 2008. Addenbrooke's Cognitive Examination-Revised in day-to-day clinical practice. *Age Ageing* **37**(3): 350.

Zarei M, Pouretemad HR, Bak T, Hodges JR. 2010. Autobiographical memory in progressive supranuclear palsy. *Eur J Neurol* **17**: 238-241.

Appendix 2.1: Actors consent form

Section of Psychological Medicine
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow

Consent Form

I consent to the DVD of my performance as an older adult actor completing the Addenbrooke's Cognitive Examination – Revised (ACE-R) being used in the following ways:

- To aid the completion of Stephanie Crawford's (Trainee Clinical Psychologist) Major Research Project: A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination - Revised (ACE-R)

- As part of an ACE-R training package developed by and administered within NHS Greater Glasgow and Clyde Healthboard.

- As a means of assisting NHS Greater Glasgow and Clyde professionals to gain a better understanding of dementia screening tools.

| --- | --- | --- |
| Name of Participant | Date | Signature |

Appendix 2.2: Participant information sheet

Section of Psychological Medicine
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow

A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination (ACE-R)

Information Sheet
10/09/09 (Version 2)

I would like to invite you to take part in a research study. Before you decide you need to understand why the research is being done and what it would involve for you. Please take time to read the following information carefully. If there is anything that is not clear or if you would like more information please contact me.

Who is conducting the research?

The research is being carried out by Stephanie Crawford (Final year Trainee Clinical Psychologist), from the Section of Psychological Medicine.

What is the purpose of the study?

The study aims to explore how different clinical presentations are scored on the Addenbrooke's Cognitive Examination - Revised (ACE-R) by professionals working within the NHS.

Why have I been invited?

You have been invited to take part in this study because you have had experience administering the Addenbrooke's Cognitive Examination – Revised (ACE-R) clinically.

Do I have to take part?

It is up to you to decide. You will be asked to sign a consent form to show you have agreed to take part. You are free to withdraw at any time, without giving reason.

What does taking part involve?

Taking part involves attending a one off session in which you will be shown three filmed vignettes of older adult actors being administered the ACE-R. You will be required to watch each vignette in its entirety whilst concurrently scoring an accompanying ACE-R form. You will also be asked to complete an Additional Information sheet detailing your occupation and your experience to

date with the ACE-R. It is anticipated that the session will last no more than 2 hours. Once the study is completed you will receive feedback on the overall group results. Depending on the overall results there may be further one off sessions offered which would discuss in more detail the results of the study; you will be invited to attend such follow-up sessions if you wished to.

What happens to the information?

Your identity and personal information will be completely confidential and known only to the researcher and her academic supervisor (Professor Jonathan Evans). The information obtained will remain confidential and stored within a locked filing cabinet. The data are held in accordance with the Data Protection Act, which means they are kept safely and cannot be revealed to other people, without your permission.

What are the possible benefits of taking part?

It is hoped that by taking part in this research, you will be providing valuable information regarding how professionals score the ACE-R in clinical situations. It is hoped that this information will influence further research into how dementia screening tools are utilised by high volumes of practitioners.

Who has reviewed the study?

This study has been reviewed by the NHS Greater Glasgow and Clyde Research Ethics Committee 3 and the University of Glasgow.

If you have any further questions?

You will have a copy of the information sheet and signed consent form to keep. If you would like further information about this research project please contact Stephanie Crawford or her clinical supervisors Dr Leigh Whitnall and Dr Joanne Robertson. If you wish to seek general advice about participating in this study from someone **not** closely linked to the study, please contact Dr Susan Cross (Consultant Clinical Psychologist). Please find all contact details overleaf.

Contacts:

Miss Stephanie Crawford
Trainee Clinical Psychologist
Department of Psychological Medicine
Gartnavel Royal Hospital
1055 Great Western Road
G12 0XH
Tel: 0141 211 0607
Email: scrawford@nhs.net

Dr Joanne Robertson
Chartered Principal Clinical Psychologist
Clinical Psychology Stroke Services – Clyde
Royal Alexandra Hospital
Corsebar Road
Paisley
Tel: 0141 314 6893
Email: Joanne.Robertson2@nhs.net

Dr Leigh Whitnall
Chartered Principal Clinical Psychologist
Cardiac Rehabilitation – Clyde
Health at Heart Centre
Royal Alexandra Hospital
Corsebar Road
Paisley
Tel: 0141 314 6852
Email: leigh.whitnall@nhs.net

Dr Susan Cross
Consultant Clinical Psychologist
Unit 8C, The Quadrangle
59 Ruchill Street
Glasgow
G20 9PX
Tel: 0141 232 0060
Email: Susan.Cross@ggc.scot.nhs.uk

If you have a complaint about any aspect of the study?

If you are unhappy about any aspect of the study and wish to make a complaint, please contact the researcher in the first instance. The normal NHS complaint mechanisms are also available to you.

Thank-you for your time and co-operation

Appendix 2.3: Participant consent form

Section of Psychological Medicine
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow

Subject number:

A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination - Revised (ACE-R)

Consent Form

Please initial box

I confirm that I have read and understand the information sheet dated 10/09/09 (Version 2) for the above study and have had the opportunity to ask questions. ☐

I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason. ☐

I agree to take part in the above study

----------	----------	----------
Name of Participant	Date	Signature

1 copy to the participant, 1 copy to the researcher

Appendix 2.4 Additional information sheet

Section of Psychological Medicine
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow
Subject number:

A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination (ACE-R)

Additional Information Sheet

Please answer the following questions.

1) What is your occupation and how long have you been in this post?

_____

2) How often do you use the Addenbrooke's Cognitive Examination – Revised (ACE-R) in your clinical work? (please circle the most appropriate response).

once a week once a fortnight once a month once every 3 months

once every 6 months less than once every 6 months

3 a) Have you attended an ACE-R training course? (please circle)

Yes No

b) If you answered 'No' to question 3 a) please briefly state how you learned to use the ACE-R.

_____

_____

4) Please give an estimate of how long you have used the ACE-R in your clinical practice:

_____ years and _____ months

5) When was the last time you administered the ACE-R in clinical practice? (please circle the most appropriate response)

Up to 1 week ago Up to a fortnight ago Up to 1 month ago

Over a month but less than 6 months ago Over 6 months ago

Appendix 2.5: Instructions given to participants

Having now read the information sheet, does anyone have any questions?

If you are happy to proceed and participate in this study could you please complete the consent forms and the Additional Information sheets that are in your research packs.

During the study you will be shown three vignettes. In each a version of the ACE-R is administered to a mock patient. Each vignette will be shown in its entirety only once; you are to score each ACE-R sheet as you watch the vignettes, just as you would in clinical practice. There should be no conferring throughout the study. There will be time to total up each of the ACE-R forms between vignettes. Scoring guides are available on request should you wish to refer to them when scoring the forms up between vignettes.

In your pack there is a sheet detailing which version of the ACE-R (A,B or C) should be used for each vignette. That sheet also contains what the correct answers for the orientation section of the ACE-R are for each vignette, please refer to it when scoring this section.

Appendix 2.6: Orientation information

Vignette 1

Version B of the ACE-R

Orientation information:

Monday 16th June, 2009
Govan Health Centre
1st Floor

Vignette 2

Version A of the ACE-R

Orientation information:

Monday 16th June, 2009
Parkview Resource Centre
Ground Floor

Vignette 3

Version C of the ACE-R

Orientation information:

Monday 16th June, 2009
Elderpark Clinic
1st Floor

Appendix 2.7: Research ethics committee letter

West of Scotland Research Ethics Service
West of Scotland REC 3
Ground Floor, The Tennent Institute
Western Infirmary
38 Church Street
Glasgow G11 6NT

Telephone: 0141 211 2123
Facsimile: 0141 211 1847
08 October 2009

Miss Stephanie Crawford
Trainee Clinical Psychologist
NHS Greater Glasgow and Clyde
Section of Psychological Medicine
Gartnavel Royal Hospital
1055 Great Western Road
Glasgow G12 0XH

Dear Miss Crawford

Study Title:	**A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination - Revised (ACE-R)**
REC reference number:	**09/S0701/92**
Protocol number:	**Version 7**

The Research Ethics Committee reviewed the above application at the meeting held on 01 October 2009. Thank you for attending to discuss the study.

Ethical opinion

The members of the Committee present gave a favourable ethical opinion of the above research on the basis described in the application form, protocol and supporting documentation, subject to the conditions specified below.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

For NHS research sites only, management permission for research ("R&D approval") should be obtained from the relevant care organisation(s) in accordance with NHS research governance arrangements. Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk. Where the only involvement of the NHS organisation is as a

Participant Identification Centre, management permission for research is not required but the R&D office should be notified of the study. Guidance should be sought from the R&D office where necessary.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

Approved documents

The documents reviewed and approved at the meeting were:

Document	*Version*	*Date*
Covering Letter		11 September 2009
REC application		11 September 2009
Protocol	Version 7	27 July 2007
Investigator CV		11 September 2009
Participant Information Sheet	Version 2	10 September 2009
Participant Information Sheet: Additional	Version 2	10 September 2009
Participant Consent Form	Version 1	10 September 2009
Letter of invitation to participant	Version 1	10 September 2009
Questionnaire: Vignette 1	Version 2	10 September 2009
Questionnaire: Vignette 2	Version 2	10 September 2009
Questionnaire: Vignette 3	Version 2	10 September 2009
Examinations Officer Letter		30 June 2009
Project Proposal Assessment Mark Sheet		29 June 2009
Letter - Research Director		28 July 2009
Scoring and Instructions Guide		

Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

The attached document “After ethical review – guidance for researchers” gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk.

09/S0701/92	**Please quote this number on all correspondence**

With the Committee's best wishes for the success of this project

Yours sincerely

Liz Jamieson
Committee Co-ordinator on behalf of Dr Robert McNeill – Acting Chair

Email: Liz.Jamieson@ggc.scot.nhs.uk

Enclosures: List of names and professions of members who were present at the meeting and those who submitted written comments
"After ethical review – guidance for researchers"

Copy to: Darren Gibson, Research and Development

Appendix 2.8: Major research project proposal

Major Research Project Proposal

A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination – Revised (ACE-R)

Matriculation Number: 0206932

Date of Submission: 27th July, 2009

Version No. 7

Word Count: 3903

Title of Project

A Novel Approach to Investigating the Reliability of the Addenbrooke's Cognitive Examination – Revised (ACE-R)

Abstract

Background

Dementia rates are expected to markedly increase in the future. There is a consensus that early diagnosis is an important element in the management of the Dementia population. Cognitive screening tools which aid dementia diagnosis are commonly used in clinical practice. Dementia screening tools need to be both valid and reliable. The Addenbrooke's Cognitive Examination – Revised (ACE-R) is showing promise as a dementia screening tool. The ACE-R is becoming increasingly used in clinical practice; however there is as yet, no study which solely and explicitly investigates its reliability. This study plans to investigate the reliability of the ACE-R using a novel approach involving multiple raters and filmed vignettes.

Aims

This study aims to explore how accurately the ACE-R is scored by multiple raters; both in terms of its total and sub-category scores. It will also consider whether scoring accuracy is effected by participant experience in using the ACE-R.

Methods

Scoring consistency will be investigated using three filmed vignettes. These vignettes will be of the ACE-R being administered to older adult actors (mock patients). These vignettes will have a pre-determined 'true score'. Study participants will be required to

complete ACE-R scoring sheets for each vignettes. The participants will be Trainee Clinical Psychologists and Community Psychiatric Nurses (CPNs) working in Older Adult Community Mental Health Teams (CMHT).

Applications

This study will develop an original approach to investigating reliability in a clinically relevant way. It will also address a gap in the literature on the ACE-R by providing information on its reliability.

Introduction

Dementia is a disorder involving progressive, global, cognitive decline (Lezak et al., 2004); Alzheimer's disease is the most common type of dementia. In 2005 a group of international experts estimated that worldwide there were 24.3 million people with dementia (Ferri et al., 2005). Ferri et al. (2005) further estimated that this prevalence rate would double every 20 years; meaning that by 2020 an estimated 42 million people worldwide will have dementia. Wimo et al. (2007) estimated that in 2005 the societal cost of dementia was 315.4 billion U.S. Dollars worldwide. These studies make it clear that Dementia is a disorder which requires increasing attention on a global scale.

From a Scottish perspective, in April 2008 the Scottish Government announced that Dementia had been made a National Health Priority and that each NHS board was to deliver improvements in early diagnosis and management of dementia by 2011 (Scottish Government Health and Community Care, 2008).

There is a general consensus that early diagnosis of dementia is desirable. The Comptroller and Auditor General's report (2007) stated that early detection enables for earlier intervention and gives: "people the opportunity to make choices and plan for the future while they are still relatively well" (p.43). Shulman and Feinstein (2003) also identified several potential benefits of early detection, including: providing the opportunity for patients to participate in clinical research at a relatively early stage; highlighting a need to monitor risks for driving and allowing cognitive enhancer drug treatments to be commenced early.

The Role of Cognitive Screening in Dementia Diagnosis

According to the National Institute for Health and Clinical Excellence (NICE), Clinical Guideline 42, Dementia: Supporting people with dementia and their carers in health and social care (2006) a Dementia diagnosis should be made following a comprehensive assessment involving: history taking, cognitive, mental and physical examination and medication review. Cognitive assessment tools are well established aids in this diagnostic process. Woodford and George (2007) state that cognitive assessment in the elderly is typically used for three reasons: i) screening for cognitive impairment; ii) differential diagnosis and iii) rating disease severity or monitoring disease progression. In clinical practice, where time constraints are omnipresent, the possibility of achieving an objective, quantitative measure of cognitive functioning following a short administration period is desirable. This may explain why, in the arena of cognitive assessment, recent decades have witnessed the development of various dementia screening tools to aid the diagnostic process.

It has been suggested that the fundamental aim of screening tools is to infer from the patient's score, compared to reference norms, the likelihood of a genuine cognitive

impairment (Cullen et al., 2007). Cullen et al. (2007) emphasise that the success of screening tools developed to meet this aim will be dependent on their statistical robustness. Clinical practice requires such statistical robustness to be achieved in the minimum time possible using an instrument that is easy to administer (Cullen et al., 2007). Furthermore Cullen et al. (2007) noted that the ideal screen should also provide rich qualitative data; thus enabling a symptom orientated approach to assessment and preventing over emphasis on cut off scores. Cut off scores distinguish between patients in terms of the presence or absence of the condition under study (Lezak et al., 2004). It is unlikely there will ever be a cognitive screening tool which will always accurately make this distinction (Lezak et al., 2004). Therefore professionals using cognitive screening tools should know their cut off scores; understand their limitations and extrapolate relevant qualitative information from the tool.

Currently, the use of one established, standardized screening test remains elusive (Gifford and Cummings, 1999; Mathuranath et al., 2000). Opinions also differ regarding how dementia screening tests should be applied and whose responsibility they should be. Cullen et al. (2007) suggested three main screening purposes: i) brief assessment in the doctor's office; ii) large scale screening programmes in the community and iii) domain specific screening to guide further assessment (tertiary care clinicians). Cullen et al.'s (2007) review of screening tools concluded that those most likely to be useful:
"..have good sensitivity and specificity for all dementia types in unselected populations, and which elicit information about key cognitive abilities, which can then be compared with neuropsychological profiles in different types of dementia" (p.9).
The Addenbrooke's Cognitive Examination – Revised (ACE-R) was identified as one of the tools which best matched this criteria.

Measurement Issues in Cognitive Screening Tools

It is generally agreed that the value obtained from any measurement is a combination of the 'true score' and a degree of measurement error (McDowell, 2006). The fundamental difficulty across all cognitive measures is that the true score is never precisely known; it can only be inferred from the value obtained (Trochim, 2006). It is essential cognitive tools minimise the degree of measurement error they create: thus making the value they obtain as accurate a reflection of the underlying true score as possible. Measurement errors can be categorised into random errors and systematic errors. Random errors increase group variability but do not affect the average performance (overall mean) of the group. Systematic errors tend to be consistently either positive or negative and thus are considered a source of bias (Trochim, 2006).

There are numerous sources of potential measurement error, which can affect the reliability of a cognitive tool, these may be broadly conceptualised into three categories: i) client factors e.g. client motivation, attention, and mood; ii) contextual factors e.g setting in time and place; iii) rater factors e.g. administration and scoring issues.

Reliability refers to a measurements consistency. Inter-rater reliability assesses whether the tool is administered and scored consistently between raters (Trochim, 2006). If a tool is not consistently administered and scored by different raters then its clinical utility is limited. Traditionally inter-rater reliability has been investigated by assessing agreement between a few (expert) raters across numerous trials. Such designs assume that high agreement indicates strong inter-rater consistency for both administration and scoring. However by assessing agreement between a few raters these designs fail to investigate administration and scoring reliability across multiple raters in a clinical setting. Furthermore by considering administration and scoring collectively these designs may fail

to detect important aspects of each. Inter rater reliability studies may benefit from using greater numbers of raters and investigating administration and scoring separately.

Once initial research has identified a cognitive screening tool as promising the primary question usually becomes whether or not its usefulness can be maintained when it is applied to a more varied, clinical population. Another important, yet less frequently studied, aspect which affects the generalisability of a tool is the consistency with which it is administered and scored by a range of professionals other than the few experts who initially developed it. An originally designed rater reliability study could address such an issue.

The ACE-R as a Dementia Cognitive Screening Tool

Cullen et al. (2007) identified the ACE-R as a cognitive screening tool currently used in domain specific screening (tertiary care settings, e.g. Older Adult CMHTs) which may be potentially useful in primary care settings. However until it is validated in non-specialist settings this suggestion remains purely speculative (Cullen et al., 2007).

The Addenbrooke's Cognitive Examination (ACE) was originally designed to detect mild dementia and differentiate Alzheimer's disease from Fronto-temporal Dementia (Mathuranath et al. 2000). In 2006 Mioshi et al. published an article introducing the revised version: the ACE-R. The ACE-R was designed as a dementia screening test, sensitive to early indicators of cognitive deterioration.

The ACE-R has five sub-categories, each representing a cognitive domain, these are: Attention and orientation (18 points), Language (26 points), Fluency (14 points), Memory (26 points) and Visuo-spatial (16 points). The five sub-category scores are summed

together to produce an overall total score (maximum 100 points). Mioshi et al. (2006) identified two ACE-R cut off scores (88 – sensitivity: 0.94, specificity: 0.89 and 82 – sensitivity: 0.84, specificity: 1.00).

In 2006 the ACE-R was recommended by the Scottish Intercollegiate Guidelines Network, (SIGN 86), as a means of improving initial cognitive testing for dementia. It therefore seems likely that its use will become increasingly pervasive.

ACE-R Reliability

In their original validation study of the ACE Mathuranath et al. (2000) did not assess its inter-rater reliability but acknowledged that this would be a useful focus for future research. Mathuranath et al. (2000) theorised that rater related bias should be low since the ACE assesses cognitive function in an objective manner. Mioshi et al. (2006, p.1078) stated that the ACE-R had underwent "design changes to make the test easier to administer". This suggests that the rater reliability of the ACE-R should be high.

Molloy and Standish (1997) highlighted that, without provision of clear instructions, multiple raters inevitably administer and score a test in their own way and different groups may develop their own unique standardization procedures. This raises reliability issues both between professionals and between different health groups. The ACE-R Administration and Scoring guide aims to prevent such issues by providing descriptions of acceptable responses for each item. However in a busy clinical setting this guide may not be regularly referred to. In addition the guide does not provide advice on how to use the ACE-R as a tool for gathering qualitative information or for differentiating between different types of dementia.

As far as this proposal is aware there is no published study looking explicitly at the ACE-R's rater reliability. Current circumstances suggest that the ACE-R is increasing in popularity and thus is being used by increasing numbers of professionals. It therefore seems timely to investigate the ACE-R's rater reliability. Conclusions from such a study could also provide information relevant to future cognitive screening research.

The Current Study

Cognitive screening tools which aid dementia diagnosis are commonly used in clinical practice. The ACE-R has been identified as one such tool. The reliability of the ACE-R has not been explicitly studied and given the likelihood that it will be increasingly used in clinical settings this seems a potentially useful research area.

This study plans to focus on the rater reliability of the ACE-R using a novel approach wherein several raters participate and scoring consistency is investigated whilst administration is kept constant. Several items in the ACE-R require a degree of rater interpretation. Based on the clinical experiences of Older Adult Clinical Psychologist's it is hypothesised that professionals are more likely in clinical practice to give clients "the benefit of the doubt", which may in turn lead to the 'test score' being inflated from the 'true score'. This may become troublesome if over scoring means the score obtained is above the cut off whilst the underlying 'true score' is below the cut off. In other words, if over scoring causes conclusions to be drawn which do not accurately reflect the underlying true score. Such over scoring by several professionals may lead to systematic measurement errors. It seems probable that over scoring will be more likely on the visuo-spatial and language sub-categories because they require more subjective judgements when scoring.

Furthermore, it may be hypothesised that if individuals do not frequently refer to the ACE-R Administration and Scoring guide they may develop their own unique scoring methods. This in turn is likely to impact on the inter rater reliability of the measure. Between rater scoring discrepancies may also be affected by rater experience of using the ACE-R. Logically one would expect that scoring accuracy would be better for raters who regularly use the ACE-R.

Aims

The primary aim is to investigate how accurately the ACE-R is scored by multiple raters; particularly in terms of whether participants consistently over score the ACE-R. In addition there will be exploration of whether scoring consistency varies across ACE-R sub-categories and whether scoring accuracy is effected by experience of using the ACE-R. Any sub-categories identified as having partcipant scores which significantly differ from their true scores will be further investigated to identify which items within them are causing the greatest discrepancies. Finally there will be general considerations regarding participant ability to correctly identify ACE-R cut off scores and participant reports of their next steps in clinical practice following ACE-R testing.

Primary Hypothesis

The participant ACE-R total scores will differ significantly from the ACE-R true scores.

Secondary Hypothesis

a) Participant scores on the ACE-R sub-categories which involve more subjective judgements (visuo-spatial and language) will significantly differ from the 'true' subtest scores.

b) Scoring accuracy may be related to participant experience using the ACE-R

Plan of Investigations

Participants

There will be two groups of participants:

i) Older Adult CMHT community psychiatric nurses (CPNs) who use the ACE-R routinely in their clinical practice.

ii) Trainee Clinical Psychologists currently completing the University of Glasgow's Doctorate in Clinical Psychology course who have used the ACE-R whilst on placement in the NHS.

Participants will be included if they have administered and scored the ACE-R independently in clinical practice on at least one older adult client. Participants will be excluded if they have not used the ACE-R in clinical practice; that is they have not completed it with an older adult patient. If participants have only observed the ACE-R being used or practiced its use with another colleague this will not be considered sufficient use and they will be excluded from the study. Some of the CPN participants will have received a one off Clinical psychology led training session on how to use the ACE-R during 2008. Although participants will not be excluded if they did not participate in such training all participants will be required to state whether they did or did not participate in this training. Participants will be invited by letter to take part in the study. The researcher will visit each of the Older Adult CMHT team meetings and the three Trainee Clinical Psychology year groups to provide a brief overview of the study.

Measures

The Addenbrooke's Cognitive Examination – Revised (ACE-R) is the primary outcome

measure.

Design and Research Procedure

This study has developed a novel, clinically relevant method for investigating inter-rater reliability. The study plans to use filmed vignettes of the researcher administering the ACE-R to older adult actors (mock patients). These vignettes will have pre-determined 'true scores' and will allow scoring consistency to be investigated separately from administration consistency. Scripts for the research vignettes will be developed by an advisory group of Older Adult Clinical Psychologists. These scripts will reflect performance on the ACE-R suggestive of different Dementia profiles. The agreed vignettes will have 'true' overall and sub-category scores. There will be three vignettes in total. The University of Glasgow media services will film and produce the vignettes. The actors will be professional actors recruited through the NHS Role play coordinator. The actors will be given information on the purpose of the study and how their performances will be used. Each actor will perform one vignette, having learned it beforehand and will be paid for their time.

Study participants will be required to complete ACE-R scoring sheets in conjunction with watching the vignettes. Participants will view each vignette in its entirety once and will not be permitted to pause or rewind any part of it. The researcher will ensure that participants are able to hear and see the vignettes clearly. The vignettes will be viewed either on a laptop or projector screen. The vignettes will be shown to participants in groups (with a minimum of three participants in each group). After watching each vignette participants will be permitted time to complete their scoring of that particular ACE-R form. The ACE-R Administration and Scoring Guide will be available to view at this time. In addition to scoring the vignettes the participants will be asked to complete

an additional form detailing: their profession and details of their ACE-R experience.

Settings and Equipment

A lap top, and at times projector will be required to show the vignettes. The NHS Greater Glasgow Older Adult Psychology service is able to provide such equipment. The study will be conducted in clinic rooms or lecture/study rooms; the researcher will be able to book time in such settings through the Older Adult Psychology Department and the University of Glasgow. ACE-R scoring sheets, ACE-R Administration and Scoring guides and additional information sheets will also be required.

Justification of Sample Size

The power calculation was based on the one sample t-test because it will be the main method for statistical analysis. Since there have been no previously published studies on the ACE-R's inter-rater reliability there is no data available to base the power calculation on. It was felt that a small effect size would translate into only a few points variation from the true score which would not threaten the overall conclusions drawn from the ACE-R. Whereas a medium effect size would translate into a difference significantly large enough that it could alter interpretation of the ACE-R results, which could in turn influence a wider assessment process. Therefore the power calculation determined the number of participants necessary for a medium effect size used the conservative values of: a p-value of 0.05 and a power of 0.8. When these values for a one sample t-test were entered into G*Power 3.010 (Faul, Erdfelder et al., 2008, from www.psycho.uni-duesseldorf.de/abteilungen/aap/gpower3/) the results indicated that a minimum of 34 participants would be required.

It is hypothesised that both participant groups will over score the ACE-R and therefore no

significant difference is expected between the participant groups. However at analyses the groups will be initially compared to ensure there are no significant differences between them. If there are significant differences, the groups will be analysed separately and comparisons of their sub-category scores conducted to detect where the differences lie. To ensure power is not affected if the groups are significantly different a minimum of 34 participants will be recruited from each of the participant groups.

Data Analysis

The details of data analysis outlined below will be carried separately out for each vignette. Descriptive statistics will be used to gain an overall understanding of the data. In particular proportions and standard deviations will be used to analyse the variation of test scores from the 'true score'. Percentages will also be used to determine how many participant correctly identified the relevant cut off scores; total and sub-category true scores. The primary hypothesis will be analysed using a one sample t-test to compare whether the ACE-R scores obtained significantly differ from the predetermined ACE-R 'true scores'. Secondary hypothesis a) will also be assessed using one sample t-tests to investigate whether the results obtained differ significantly from the true subtest scores. Secondary hypothesis b) will be analysed using correlational methods.

Health and Safety issues

Researcher and Participant Safety Issues

No significant health and safety issues are expected. Participants will all be NHS staff and the study will be conducted in either NHS or University of Glasgow premises.

Ethical Issues

The study will not progress until ethical approval is obtained from the West of Scotland Research Ethics Committee 3 and Greater Glasgow and Clyde Research and Development approval is also obtained. The participants in the study will be NHS staff. Participants will be required to give informed consent prior to their participation. Following participant completion of the vignettes they will be debriefed. This debrief will include information on how the data will be used and details of the follow up process detailed below. The debrief will also highlight that performance on the vignettes is not a reflection of overall professional competency and that individual results will not be revealed. All data gathered will be anonymous. It is not anticipated that the study will cause participants distress. However, if the results indicate that as a group the participants have made significant scoring errors this may threaten their feeling of professional competency. To overcome this, the researcher will provide written feedback to all participants on the overall findings of the study and also offer every participant the opportunity to participate in one training session. Depending on participant uptake on the offer for a training session the researcher will host as many training sessions as required to ensure all participant who would like to take part in a training session have the opportunity do so. Training session content will focus on addressing any areas that the study identified as being significantly mis-scored by the participant group. The actors involved will not be CMHT patients and will not have cognitive difficulties. The actors will be fully briefed prior to their participation. The vignettes used shall be developed by a small panel of Clinical Psychologists working within Older Adult services and therefore will not be based on individual clients.

Financial issues

The University of Glasgow Media Services will be available free of charge with the exception of the recoup of digital tape stock and DVDs, which will come to £35. Pene Herman-Smith, NHS role coordinator, will provide professional older adult actors to perform the vignettes. Each actor requires a fee of £125 for their time. Other financial costs will be minimal.

Practical Applications

This study will develop an original approach to investigating inter-rater reliability in a clinically relevant way. It will also address a gap in the literature on the ACE-R by providing information on its inter-rater reliability.

Timetable

Obtain Course Approval	by August 2009
Complete scripts for vignettes	by September 2009
Obtain ethical approval	by October 2009
Film vignettes	by November 2009
Data collection	Nov 2009 – Jan 2010
Data analysis and write up	Completed by Jun 2010
Written summary of results sent to all participants	Completed by Aug 2010
Follow up training sessions	Offered throughout Aug

References

Cullen B., O'Neill B., Evans J.J., Coen R.F., Lawler B.A., (2007). A review of screening tests for cognitive impairment. *Journal of Neurology, Neurosurgery, and Psychiatry*, 78, pp.790-799.

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

Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, M., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P.R., Rimmer, E., Scazufca, M., 2005. Global prevalence of dementia: a Delphi consensus study. *Lancet*, 366, pp.2112-2117.

Gifford D.R., Cummings J.L., (1999). Evaluating Dementia Screening Tests: Methodologic standards to rate their performance. *American Academy of Neurology*, 52(2), pp.224-227.

Lezak M.D., Howieson B., Loring D.W., Hannay J., (2004), *Neuropsychological Assessment*, 4th edition, New York, Oxford University Press, Inc.

Mathuranath, P. S., Nestor, P. J., Berrios, G. E., Rakowicz, W., and Hodges, J. R., (2000). A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology*, 55(11), pp.1613–20.

McDowell, I., 2006. *Measuring Health: a guide to rating scales and questionnaires.* [e-book] New York: Oxford University Press.
Available at: http://books.google.co.uk/books?id=06t-63RaYk0C&dq=mcdowell+measuring+health&printsec=frontcover&source=bl&ots=ok5k80bBHN&sig=kde3L9sYcKmVrr9Yf3nAbuJv6Q8&hl=en&ei=wO8RSovfL8afjAf77-HkCA&sa=X&oi=book_result&ct=result&resnum=1#PPA8,M1
[Accessed 10 January 2009]

Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., Hodges, J.R., 2006. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International Journal of Geriatric Psychiatry*, 21, pp.1078-1085.

Molloy, D.W., and Standish, T.I.M., 1997. Mental status and neuropsychological assessment a guide to the standardized Mini-Mental State Examination. *International Psychogeriatrics*, 9(Suppl.1), pp.87-94.

National Audit Office, 2007. *Improving services and support for people with dementia*. (Report by the Comptroller and Auditor General) (HC 604 session 2006-2007) London: the Stationery Office.

National Institute for Health and Clinical Excellence (NICE), 2006. *Clinical Guideline 42, Dementia: Supporting people with dementia and their carers in health and social care* [Internet] London: NICE (Published 2006)
Available at: http://www.nice.org.uk/nicemedia/pdf/CG042NICEGuideline.pdf
[Accessed 5 December 2008]

Scottish Intercollegiate Guidelines Network, 2006, Management of patients with dementia, a National Clinical Guideline, (SIGN publication 86) [Internet] Edinburgh: SIGN, (Published 2006)

Available at: http://www.sign.ac.uk/pdf/sign86.pdf [Accessed 5 December 2008]

Shulman, K.I., and Feinstein, A., 2003. *Quick cognitive screening for clinicians.* [e-book] Informa Healthcare.
Available at:
http://books.google.co.uk/books?id=oud4rX2EdswC&dq=%22Quick+Cognitive+Screening+for+Clinicians.%22&printsec=frontcover&source=bl&ots=SLjy8F_rY0&sig=v3SzOJPLnPGbwhowCONts2-brAs&hl=en&ei=bX4QSsbgIdm2jAejxrW3Bg&sa=X&oi=book_result&ct=result&resnum=3#PPP1,M1 [Accessed 21 April 2009]

The Scottish Government Health and Community Care, 2008. Dementia. [Online] (Updated 21 July 2008)
Available at: http://www.scotland.gov.uk/Topics/Health/health/mental-health/servicespolicy/DFMH/dementia [Accessed 10 December 2008]

Trochim, W.M.K., Web Center for Social Research Methods, 2006. *Research Methods Knowledge Base (Reliability Section).* [Online] (Updated 20 October 2006)
Available at: http://www.socialresearchmethods.net/kb/reliable.php
[Accessed 10 January 2009]

Wimo, A., Winbald, B., Jönsson, L., 2007. An estimate of the total worldwide societal costs of dementia in 2005. *Alzheimer's and Dementia*, 3, pp.81-91.

Woodford H.J. and George J., 2007. Cognitive assessment in the elderly: a review of clinical methods. *QJM*, 100, pp.469-484.