ELSEVIER

Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring ■ (2017) 1-6

Alzheimer's & Dementia

Cognitive & Behavioral Assessment

Improving the quality of cognitive screening assessments: ACEmobile, an iPad-based version of the Addenbrooke's Cognitive Examination III

C. G. J. Newman[a,*], A. Bevins[b], J. Zajicek[c], J. Hodges[d], E. Vuillermoz[e], J. Dickenson[f], D. Kelly[f], S. Brown[g], R. Noad[h]

[a]Plymouth University Peninsula Schools of Medicine and Dentistry (PU PSMD), Plymouth, United Kingdom
[b]Devon Partnership NHS Foundation Trust, Exeter, United Kingdom
[c]School of Medicine, Medical & Biological Sciences, St Andrews, United Kingdom
[d]Brain Sciences UNSW, UNSW Australia, UNSW Sydney, Australia
[e]School of Psychology, Plymouth University, Plymouth, United Kingdom
[f]Plymouth University, Doctorate in Clinical Psychology, School of Psychology, Plymouth, United Kingdom
[g]Devon Partnership NHS Trust, Exeter, United Kingdom
[h]Plymouth Hospitals NHS Trust, Department of Neuropsychology, Level 7, Derriford Hospital, Plymouth, United Kingdom

Abstract

Introduction: Ensuring reliable administration and reporting of cognitive screening tests are fundamental in establishing good clinical practice and research. This study captured the rate and type of errors in clinical practice, using the Addenbrooke's Cognitive Examination-III (ACE-III), and then the reduction in error rate using a computerized alternative, the ACEmobile app.

Methods: In study 1, we evaluated ACE-III assessments completed in National Health Service (NHS) clinics (n = 87) for administrator error. In study 2, ACEmobile and ACE-III were then evaluated for their ability to capture accurate measurement.

Results: In study 1, 78% of clinically administered ACE-IIIs were either scored incorrectly or had arithmetical errors. In study 2, error rates seen in the ACE-III were reduced by 85%–93% using ACEmobile.

Discussion: Error rates are ubiquitous in routine clinical use of cognitive screening tests and the ACE-III. ACEmobile provides a framework for supporting reduced administration, scoring, and arithmetical error during cognitive screening.

Keywords: Screening assessment; Cognitive assessment; Alzheimer's; Dementia; Computerized; App; Usability; Validity; Administrator error

1. Introduction

The psychometric properties of cognitive screening tools for dementia are routinely reported, yet far less is known about the clinician's ability to administer and score these tests accurately. Evidence suggests that users make many more errors than expected [1–3]. There is surprisingly little detail in the literature on how well the cognitive screening tests perform in the hands of the clinicians for whom they are designed.

Despite the brevity and perceived simplicity of two of the most commonly used cognitive assessment instruments in the United Kingdom—the Mini–Mental State Examination [4] and the Addenbrooke's Cognitive Examination–Revised (ACE-R) [5]—test scoring simulation studies have revealed high rates of errors on both measures [6,7]. Both the Mini–Mental State Examination and ACE-R use cutoffs for determining caseness, and this influences subsequent diagnostic/treatment pathways, highlighting the importance of accurate assessment.

The authors have declared that no conflict of interest exists.

*Corresponding author. Tel.: (+44)1752 315264; Fax: 01752 588072.
E-mail address: craig.newman@plymouth.ac.uk

https://doi.org/10.1016/j.dadm.2017.12.003

Computerized approaches to cognitive assessment have the potential to improve the standards of administration, scoring, and reporting, as by automating processes, the possibility of human error is reduced. For example, it is possible to more closely control administration instructions, thus reducing the chance of intrarater and interrater variation. The scoring of a test can also be supported or automated, reducing the chance of arithmetical errors.

The use of a product in real-world settings cannot be presumed and must be tested. Such *usability* research is a worldwide standard in medical device evaluation [8–10] but until recently has been largely neglected in the validation of cognitive screening assessments. Furthermore, there are few studies that have evaluated the use of these tools in routine clinical practice.

The present study set out to explore the nature of scoring errors using the ACE-III in routine National Health Service (NHS) practice. This was followed by a comparison of the ACE-III and ACEmobile, a new iPad-based version developed by the authors. The aim was to assess the ability of each tool to support a clinician to capture accurate measurement, that is, the hypothetical score that the patient would receive with zero measurement error. ACEmobile was designed to support users of the ACE-III by guiding and automating the administration, rule adherence, scoring, and reporting.

1.1. Aims and hypotheses

Aim 1: To establish the presence, nature, and extent of scoring errors on the ACE-III in standard clinical practice, via the analysis of completed ACE-III assessments from NHS memory assessments (study 1).
Hypothesis 1: High rates of administration and arithmetical errors will be observed in ACE-III assessments from NHS memory clinics.
Aim 2: To compare the measurement accuracy of ACE-III and ACEmobile (study 2).
Hypothesis 2: Administration and reporting errors will be significantly less for ACEmobile than for ACE-III.

2. Study 1: Identification of scoring errors on the ACE-III in standard NHS clinical practice

2.1. Sample

ACE-III scoresheets ($N = 132$) were extracted from the electronic patient records of four Older People's Community Mental Health teams in Devon, UK (with NHS ethical approval). Of these, 45 (34%) were subsequently excluded from the analysis because they were not suitable for further analysis (i.e., incomplete assessments, scores omitted, illegible, and older version of ACE used [i.e., ACE-R]). A total of 87 ACE-III scoresheets were subsequently analyzed. The ACE-IIIs were administered by the community's psychiatric nurses (63%, $n = 55$), psychiatrists (26%, $n = 23$), and occupational therapists (10%, $n = 9$). Details of specific training undertaken by each administrator were not collected but were assumed to be the standard required for that clinical service. This was deemed to be representative of standard NHS clinical practice.

2.2. Measures

The ACE-III is a cognitive screening tool to detect mild dementia and distinguish between Alzheimer's disease and frontotemporal dementia [11]. It contains 24 individual test items contributing to five subdomains—attention (18 points), memory (26 points), fluency (14 points), language (26 points), and visuospatial functioning (16 points), with a total score of 100. The ACE-III shows high sensitivity and specificity for dementia using a cutoff of 88 or 82, respectively [12].

2.3. Procedure

Two anonymized copies of each ACE-III were produced. Rescoring was conducted by two raters, strictly following the published scoring guidelines. The two data sets were compared for consistency using an Excel formula. There were 121 discrepancies between raters, equating to an error rate of 2.58%. The second author adjudicated on the discrepancies to reconcile the differences and produce a single data set with an accurate score at the individual item level, subdomain level, and ACE-III total score.

Data were double entered, and any discrepancies were adjudicated by the lead researcher. Finalized ACE-IIIs were then compared back to the clinician-scored ACE-IIIs. *Scoring errors* (points deducted or added in error by each clinician, for each subtest), *arithmetic errors* (mental arithmetic errors made in adding the scores together), and *total error* (scoring and arithmetic errors combined) were calculated.

2.4. Results

The range of clinician ACE-III total scores in the sample was from 30 to 88 points ($\mu_x = 64.80$, SD = 13.24).

Scoring errors were observed in 68% of the ACE-IIIs. Arithmetic errors were observed in 24% of ACE-IIIs, with a range of −10 to 10. Only 22% of ACE-IIIs had no errors at all. The total error rate ranged from 0 to 22, with a mean of 3.3 (SD = 4.2). In 22% of the sample, the total error rate was 5 or more points (Fig. 1).

At the subdomain level, 46% and 44.8% of clinicians made at least one error on the visuospatial and language domains, respectively. Errors were present but observed less frequently for the memory (20%), fluency (15%), and attention and orientation (12%) domains. At the individual item level, 39% of clinicians made at least one error on sentences, 34% on clock drawing, and 11% on animal fluency.

2.5. Summary

In NHS settings, clinician errors in scoring, mental arithmetic, and reporting the ACE-III were commonplace.