Does telephone testing of long-term memory retention and forgetting influence performance in young and older adults? An examination using the Crimes Test.

Richard J. Allen[1], Stephen Kemp[2], Suzannah Morson[3], Christine Wells[4], Katherine Grindheim[1], & Alan D. Baddeley[5]

[1] School of Psychology, University of Leeds

[2] St James' University Hospital, Leeds

[3] School of Psychology, University of Aberdeen

[4] Department of Social & Behavioural Sciences, Leeds Trinity University

[5] Department of Psychology, University of York

Correspondence should be addressed to: Dr Richard Allen, School of Psychology, University of Leeds, Leeds, UK. LS2 9JT. Email: r.allen@leeds.ac.uk

Abstract

Measuring memory over long delays requires multiple sessions, often administered remotely (e.g. by telephone) to maximise convenience and participant access. However, the efficacy of testing delayed memory via telephone has not previously been examined. We administered the Crimes Test to young and older adults, with a one-week delay test either in person or over the telephone. Testing via telephone had no detrimental effect, indicating this to be an appropriate method of examining delayed episodic memory.

Keywords: Telephone testing; Long-term forgetting; Episodic memory; Ageing; Crimes Test.

<u>Does telephone testing of long-term memory retention and forgetting influence performance in</u>
<u>young and older adults? An examination using the Crimes Test.</u>

While most patient groups who show impaired learning show normal rates of forgetting, certain groups of individuals may display relatively intact performance on measures of memory administered within single testing sessions but show greater forgetting over longer delays of days or weeks. This phenomenon, known as 'accelerated long-term forgetting' or ALF (Blake, Wroe, Breen, & McCarthy, 2000), was first described by De Renzi and Lucchelli (1993) and has since been observed in a number of case and group studies, particularly in relation to individuals with temporal lobe epilepsy (TLE; see Elliott, Isaac, & Mulhert, 2014, and Butler & Zeman, 2008). ALF may also be apparent in some patients with mild cognitive impairment (MCI) and raises the possibility of memory consolidation being a risk factor for conversion to dementia (Walsh, Wilkins, Bettcher et al, 2013; Manes, Serrano, Calcagno et al, 2008).

These findings clearly underline the need for development and validation of testing materials and methodologies able to detect forgetting over the longer term. A range of clinical tests of memory exist that are designed to measure different aspects of memory functioning and detect specific and generalized problems. Such tests are typically administered in the clinic and assess memory performance over periods of 30-45 minutes. While this is sufficient to detect many forms of memory problem, they do not provide measures of forgetting over longer intervals (i.e. days, weeks, or months). This is potentially problematic for detecting ALF and may be one reason why the subjective memory complaints of TLE patients sometimes appear to be underestimated based on clinical tests (Piazzini, Canevini, Maggiori, & Canger, 2001).

One practical issue that arises when assessing memory for materials over longer retention intervals is the requirement for participants, researchers, or clinicians to travel for each testing session. This can increase cost and inconvenience, and hinder recruitment and assessment, particularly for population groups with difficulties in mobility and independence. One solution often adapted in studies of long-term retention and ALF is the use of telephone testing. Thus,

participants are presented with and tested on materials in an initial 'in person' session, with follow-up long-term tests for these materials then administered over the telephone. For example, Kemp, et al. (2012) observed ALF for story and picture recall using telephone testing over delays up to 30 days, while Ricci, Wong, Nikpour, and Miller (2019) used telephone testing to examine impacts of rehearsal and retesting on patients with established ALF across delays of up to 4 weeks. Similarly, Gascoigne and colleagues have observed ALF on word recall at 7-day delays in children with idiopathic generalized epilepsy (Gascoigne, Barton, Webster, Gill, Antony, & Lah, 2012). Telephone testing has also been used to observe increased rates of ALF in pre-symptomatic individuals at enhanced genetic risk of Alzheimer's disease (Weston et al., 2018). In all cases, these studies used telephone testing but did not directly compare against face to face assessment.

This approach has also been applied in healthy populations. Baddeley and colleagues (Baddeley, Rawlings, and Hayes (2014; Baddeley, Allen, Atkinson, & Kemp, 2019) developed the 'Crimes Test', a novel test of long-term forgetting involving four short fictional prose descriptions of relatively minor crimes. Baddeley et al. (2014) presented these materials to groups of young and older adults and assessed immediate and delayed memory using cued recall. Initial presentation and immediate recall were carried out in person, before further telephone-based testing at an intermediate point (between 24 hours and 24 days) and 6-weeks after initial presentation. Both age groups showed forgetting over time, with the older adults demonstrating an accelerated loss relative to the younger group. While Baddeley et al. suggested that participants appeared to cope with testing by telephone, this could only be indirectly inferred as the impact of telephone vs. in person testing was not directly examined at the same time point. Baddeley et al. (2019) followed this up using both the Crimes test and Four-Doors test (a test of memory for visual information), running all follow-up sessions (at delays of 24-hours, one week, and one month) over the telephone. For these studies and for work with clinical populations, it is not clear what effect this variation in test method might have, particularly when comparing immediate and short delays (typically

administered in person) with telephone-based long-term tests. The present study was designed to test this key practical issue.

The utility of telephone-based cognitive test administration has previously been examined, particular in relation to dementia screening, with findings generally illustrating this to be an appropriate approach (Brandt, Spencer, & Folstein, 1988; Duff, Tometich, & Dennett, 2015; Lipton et al., 2003; see Castanho, Amorim, Zihl, Palha, Sousa, & Santos, 2014, for review). Quality of data gathered in semi-structured interviews also appears to be comparable for over the telephone versus face-to-face administration (Sturgess & Hanrahan, 2004). Similarly, cognitive behavioral therapy for the treatment of depression has been demonstrated to have comparable efficacy (at least over the short-term) when carried out over the telephone versus face-to-face, while also reducing attrition rates (Mohr et al., 2012).

Crucially however, these studies all focus on administration of screening and interview procedures carried out within individual sessions; no previous studies (to our knowledge) have explored the efficacy of telephone-based testing of long-term memory and forgetting over multiple sessions administered across delays of days or more. The present study is therefore the first to directly compare delayed recall accuracy in face-to-face vs. telephone testing of materials. This should have important implications both for experimental work examining the mechanisms of forgetting in general, and for development of measures for use with individuals with TLE (Butler & Zeman, 2008; Elliott et al., 2014), MCI (Walsh et al., 2013; Manes et al., 2008), and other cognitive impairments. This is an important issue to address for the development and administration of ALF tests.

It cannot necessarily be assumed that in-person vs. telephone-based testing of memory retention will produce equivalent performance levels. Telephone-based assessment is likely to reduce availability of nonverbal or environmental cues, as well as affecting the quality of the auditory-verbal input. This may also influence memory performance over multiple sessions if telephone testing follows an initial in-person encoding phase, as a number of studies have shown

environmental-context dependent memory effects (e.g. Smith & Vela, 2001). However, while this may provide some a priori justification for the prediction of a performance decrement associated with telephone testing, the successful use of this method by Baddeley et al. (2014, 2019) means that we predicted only minimal difference at the final test between the in-person and telephone tested groups.

Method

Participants

Eighty participants took part in this experiment, all drawn from School of Psychology participant databases. This sample consisted of forty younger adults (7 male, 33 female) aged 18-23 years (mean 19.45, SD = 1.06), and forty older adults (9 male, 31 female) aged 55-86 years (mean 73.03, SD = 7.67). All older adults scored 27 or above on the Mini-Mental State Examination (Folstein et al., 1975).

This study was approved by the School of Psychology, University of Leeds ethics committee (reference no. 12-0176).

Materials

We used the version of the Crimes test as described by Baddeley et al. (2019). This involves four short descriptions of fictional, relatively minor crimes, each involving a different perpetrator; victim, crime, and location. Memory for these items is then tested via sets of cued recall questions, with the current version of the task allowing 80 possible questions; we used three sets of 20 questions from the materials used by Baddeley et al. (2014, 2019). Each set of questions differs, and is designed to assess various aspects of each of the vignettes (e.g. *Who committed the hit and run crime? What was the nationality of the victim in the pub? What was the location of the crime committed against the young man?*).

Design and Procedure

A 3x2x2 mixed design was implemented, examining test session as a within-subjects factor (Immediate, 20-minute, and 1-week delay tests), along with age group (young vs. older adults), and final session (in person vs. telephone-based) as between-subjects factors. There were 20 younger adults in each of the two final session conditions, while 17 older adults performed the 'in person' final session and 23 older adults the 'telephone' condition. Based on the age group effect of $d = .86$ in Baddeley et al. (2014), power analysis (calculated using G*Power 3.1.9.3; Faul, Erdfelder, Lang, & Buchner, 2009) indicated 95% power to detect such an effect at $\alpha = .05$.

The study was run over two sessions. All participants attended the School of Psychology, University of Leeds, for the first session, while only those participants in the 'in person' condition attended for the second session. The first session lasted around 60 minutes, and proceeded as follows. Following administration of the MMSE (for older adults), the Crimes task was introduced and participants had the opportunity to ask questions. The Crimes passages were then verbally presented to the participant by a researcher (a female native English speaker). The three sentences within each crime vignette were read aloud at an even pace, with a 2s pause between each sentence, and a 5s pause at the end of each vignette. The immediate cued recall test was then administered, with the 20 questions from the first set read out by the researcher and participants attempting to provide a verbal response to each. Following a delay of approximately 20 minutes (filled with an adapted version of the Doors visual recognition task – Baddeley, Emslie, & Nimmo-Smith, 1994), the 20-min delay test was then administered using the same methodology (and a second set of questions).

The second session was 15 minutes in duration and took place 1 week later, comprising the final 20 cued recall questions from the Crimes Test.

Results

Performance was scored as the mean number of cued recall questions correctly answered per test session (max score = 20). Descriptive statistics showing mean recall at each test stage, for each age group and final test method, are reported in Table 1.

All data were analysed using both frequentist and Bayes Factor (BF) analysis (e.g. Dienes & Mclatchie, 2018; van Doorn et al., 2019), using JASP (JASP Team, 2018). BF analysis provides a continuous measure of the strength of evidence for the alternative hypothesis against the null hypothesis and provides a test of equivalence between groups/conditions (Jarosz & Wiley, 2014; Mulder & Wagenmaker, 2016). The Bayes Factor (BF_{10}) describes how many times more likely the alternative hypothesis is than the null hypothesis (of no effect/interaction), with a BF<1 indicating support for the null hypothesis, and BF>1 indicating support for the alternative hypothesis. BFs for the inclusion of each main effect and interaction in the model are reported alongside frequentist outcomes.

Table 1 here

The primary focus of the current work was on whether recall varies when assessed in person or via telephone. Therefore, final test recall accuracy was analysed, within a 2x2 (age group x testing method) ANOVA. This indicated a significant effect of age group, $F(1,76) = 6.61$, $MSE = 106.99$, $p = .012$, $np^2 = .08$, $BF = 4.79$, with younger adults (M=13.78, SE=.60) more accurate than older adults (11.38, .66). There was no effect of testing method, $F(1,76) = .72$, $MSE = 11.64$, $p = .4$, $np^2 = .01$, $BF = .32$, and no interaction between these factors, $F(1,76) = .06$, $MSE = .93$, $p = .81$, $np^2 = .00$, $BF = .34$.

As recall at the final (1-week delay) test is not independent of earlier memory retention, we calculated final test accuracy as a proportion of recall at the immediate and 20-minute tests (see Figure 1). Separate 2x2 frequentist and Bayesian ANOVA were carried out in each case. For the comparison with the immediate test, this indicated no effect of age group, $F(1,76) = .76$, $MSE = .01$, $p = .38$, $np^2 = .01$, $BF = .32$, testing method, $F(1,76) = .02$, $MSE = .00$ $p = .88$, $np^2 = .00$, $BF = .24$, or the interaction, $F(1,76) = .02$, $MSE = .00$, $p = .89$, $np^2 = .00$, $BF = .33$. Similarly, for the comparison with the 20-minute test, this indicated no effect of age group, $F(1,76) = .53$, $MSE =$

.01, $p = .82$, $np^2 = .00$, $BF = .24$, testing method, $F(1,76) = .81$, $MSE = .18$ $p = .37$, $np^2 = .01$, $BF = .34$, or the interaction, $F(1,76) = .04$, $MSE = .01$, $p = .85$, $np^2 = .00$, $BF = .29$.

Figure 1 here

Discussion

The current study explored whether the use of telephone testing has any major impact on cued recall performance in the Crimes Test (Baddeley et al., 2014, 2019) by younger and older adults. The outcomes overall were clear; there was no impact of telephone-based testing, relative to in-person testing, at the final session carried out one week after initial presentation and assessment.

These findings indicate that the Crimes Test can be implemented over the telephone at delayed follow-up, thus simplifying its application as a possible assessment tool. While acknowledging the caution required when extending outcomes to other tasks, delayed recall tests of story, word, or picture memory as used to measure forgetting over time (e.g. Baddeley et al., 2018; Gascoigne et al., 2012; Kemp et al., 2012; Ricci et al., 2019; Weston et al., 2018) should also be suitable for use in telephone-based testing. The absence of an interaction with age also indicates that older adults did not find telephone testing to be relatively more challenging, compared to the younger age group. Though the present study was focused on healthy young and older adults with no known cognitive or sensory impairments, we would cautiously suggest the present findings to be applicable to clinical groups for whom forgetting over time may be problematic, provided of course that such individuals do not possess auditory impairments that preclude telephone use. It may therefore be appropriate for future work to incorporate this method into study designs and development of test materials to minimise expense, inconvenience, and participant attrition. However, the efficacy of telephone-based assessment of long-term forgetting should ideally also be explored in relevant clinical samples before its widespread use can be confidently adopted. It is also worth noting that this form of remote assessment may reduce the ability of the tester to assess

symptom validity, non-verbal behaviour, or the degree to which the participant is being influenced by the environmental context.

More broadly, an effect of age was observed on the final session, with older adults performing less accurately overall, relative to the younger group. This replicates the overall age difference on Crimes recall reported by Baddeley et al. (2014) and is in line with well-established age-related deficits in episodic learning and long-term recall (e.g. Craik, 1994). The older group did not show faster forgetting over the one-week retention interval, as there was no effect of age on the final test recall score when calculated as a proportion of either the immediate or 20-minute test. Indeed, forgetting was not apparent in either group over this time period (as illustrated by Figure 1). The one-week delay used in the present study was shorter than the delays implemented by Baddeley et al. (2014), who studied intermediate intervals of 24 hours to 24 days, and a longer delay of 6 weeks. Meaningful long-term forgetting may only start to emerge in healthy individuals on the Crimes test over intervals substantially beyond one week. Furthermore, the 20-minute delay test may have served to strengthen the memory representation and slow rate of forgetting. Baddeley et al. (2019) recently demonstrated that forgetting in young adults on the Crimes test, and also on a visual equivalent (the Four Doors Test), was substantially reduced over a 1-month period if intervening tests were administered at 1-day and 1-week delays following encoding. Indeed, at least on the Crimes Test, Baddeley et al. (2019) observed that participants showed a slight numerical improvement between the immediate and 1-week tests when an additional 1-day was included. The current findings might indicate that a 20-minute test serves a similar purpose and extends this to an older adult group.

In conclusion, this study provides evidence that a telephone method of follow-up for measuring episodic memory retention and forgetting over delays is equivalent to face-to-face testing and acceptable to adult participants. We present these findings as an important contribution to the emerging literature and debate regarding methodology, in measuring memory consolidation in neurological disorders.

<center>References</center>

Baddeley, A.D., Allen, R.J., Atkinson, A.L., & Kemp, S. (2019). The problem of detecting long-term forgetting: Evidence from the Crimes Test and the Four Doors Test. *Cortex, 110,* 69-79.

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

Baddeley, A.D., Rawlings, B., & Hayes, A. (2014). Constrained prose recall and the assessment of long-term forgetting: The case of ageing and the Crimes Test. *Memory*, *22*(8), 1052-1059.

Blake, R. V., Wroe, S. J., Breen, E. K., & McCarthy, R. A. (2000). Accelerated forgetting in patients with epilepsy Evidence for an impairment in memory consolidation. *Brain*, *123*(3), 472-483.

Brandt, J., Spencer, M., & Folstein, M. (1988). The telephone interview for cognitive status. *Cognitive and Behavioral Neurology*, *1*(2), 111-118.

Butler, C. R., & Zeman, A. Z. (2008). Recent insights into the impairment of memory in epilepsy: Transient epileptic amnesia, accelerated long-term forgetting and remote memory impairment. Brain, 4, 2243– 2263. doi:10.1093/brain/awn127

Castanho, T. C., Amorim, L., Zihl, J., Palha, J. A., Sousa, N., & Santos, N. C. (2014). Telephone-based screening tools for mild cognitive impairment and dementia in aging studies: a review of validated instruments. *Frontiers in aging neuroscience*, *6*.

Craik, F. I. (1994). Memory changes in normal aging. *Current directions in psychological science*, 155-158.

De Renzi, E., & Lucchelli, F. (1993). Dense retrograde amnesia, intact learning capability and abnormal forgetting rate: A consolidation deficit? *Cortex, 29*, 449–466. doi:10.1016/S0010-9452(13)80253-5

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic bulletin & review*, *25*(1), 207-218.

Duff, K., Tometich, D., & Dennett, K. (2015). The Modified Telephone Interview for Cognitive

    Status is More Predictive of Memory Abilities Than the Mini-Mental State Examination.

    Journal of Geriatric Psychiatry and Neurology, 0891988715573532.

Elliott, G., Isaac, C. L., & Muhlert, N. (2014). Measuring forgetting: A critical review of

    accelerated long-term forgetting studies. *Cortex*, *54*, 16-32.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*

    Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, *41*(4),

    1149-1160.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method

    for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*,

    12(3), 189-198.

Gascoigne, M. B., Barton, B., Webster, R., Gill, D., Antony, J., & Lah, S. S. (2012). Accelerated

    long-term forgetting in children with idiopathic generalized epilepsy. *Epilepsia*, *53*(12),

    2135-2140.

JASP Team (2018). JASP (Version 0.10.0) [Computer software].

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting

    Bayes factors. *The Journal of Problem Solving*, *7*(1), 2-9.

Kausler, D. H. (1991). Experimental psychology, cognition and human aging (2nd ed.). New York,

    NY: Springer-Verlag.

Kemp, S., Illman, N. A., Moulin, C. J., & Baddeley, A. D. (2012). Accelerated long-term forgetting

    (ALF) and transient epileptic amnesia (TEA): Two cases of epilepsy-related memory

    disorder. *Epilepsy & Behavior*, *24*(3), 382-388.

Lipton, R. B., Katz, M. J., Kuslansky, G., Sliwinski, M. J., Stewart, W. F., Verghese, J., ... &

    Buschke, H. (2003). Screening for dementia by telephone using the memory impairment

    screen. *Journal of the American Geriatrics Society*, *51*(10), 1382-1390.

Manes, F., Serrano, C., Calcagno, M. L., Cardozo, J., & Hodges, J. (2008). Accelerated forgetting
in subjects with memory complaints: A new form of mild cognitive impairment? *Journal of
Neurology, 255*, 1067-1070. doi: 10.1007/s00415-008-0850-6

Mary, A., Schreiner, S., & Peigneux, P. (2013). Accelerated long-term forgetting in aging and intra-
sleep awakenings. *Frontiers in Psychology, 4, 750*. doi:10.3389/fpsyg.2013.00750

Mohr, D. C., Ho, J., Duffecy, J., Reifler, D., Sokol, L., Burns, M. N., Jin, L., & Siddique, J. (2012).
Effect of telephone-administered vs face-to-face cognitive behavioral therapy on adherence
to therapy and depression outcomes among primary care patients: a randomized trial. *JAMA,
307*(21), 2278-2285.

Mulder, J., & Wagenmakers, E. J. (2016). Editors' introduction to the special issue "Bayes factors
for testing hypotheses in psychological research: Practical relevance and new
developments". *Journal of Mathematical Psychology*, *72*, 1-5.

Piazzini, A., Canevini, M. P., Maggiori, G., & Canger, R. (2001). The perception of memory
failures in patients with epilepsy. *European Journal of Neurology*, *8*(6), 613-620.

Ricci, M., Wong, T., Nikpour, A., & Miller, L. A. (2019). Testing the effectiveness of cognitive
interventions in alleviating accelerated long term forgetting (ALF). *Cortex, 110,* 37-46.

Salthouse, T. A. (1991). Theoretical perspectives on cognitive aging. Hillsdale, NJ: Erlbaum.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-
analysis. *Psychonomic Bulletin & Review*, *8*(2), 203-220.

Sturgess, J. E., & Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative
interviewing: a research note. *Qualitative Research*, *4*(1), 107-118.

Tombaugh, T. N., & Hubley, A. M. (2001). Rates of forgetting on three measures of verbal learning
using retention intervals ranging from 20 minutes to 62 days. *Journal of the International
Neuropsychological Society, 7(1)*, 79–91. doi:10.1017/S1355617701711083

van Doorn, J., van den Bergh, D., Bohm, U., Dablander, F., Derks, K., Draws, T., ... & Ly, A. (2019). The JASP Guidelines for Conducting and Reporting a Bayesian Analysis.

Walsh, C. M., Wilkins, S., Bettcher, B. M., Butler, C. R., Miller, B. L., & Kramer, J. H. (2014). Memory consolidation in aging and MCI after 1 week. *Neuropsychology*, *28*(2), 273.

Weston, P. S., Nicholas, J. M., Henley, S. M., Liang, Y., Macpherson, K., Donnachie, E., ... & Zeman, A. Z. (2018). Accelerated long-term forgetting in presymptomatic autosomal dominant Alzheimer's disease: a cross-sectional study. *The Lancet Neurology*, *17*(2), 123-132.

List of authors:

Dr Richard J. Allen, Associate Professor, University of Leeds. r.allen@leeds.ac.uk

Dr Steven Kemp, Consultant Clinical Neuropsychologist, Leeds Teaching Hospitals NHS Trust, stevenkemp@nhs.net

Dr Suzannah Morson, Research Fellow, University of Aberdeen. s.morson@abdn.ac.uk

Dr Christine Wells, Senior Lecturer, Leeds Trinity University. ch.wells@leedstrinity.ac.uk

Katherine Grindheim, Psychologist in Clinical Training, University of Leeds. ps09kas@leeds.ac.uk

Professor Alan D. Baddeley, University of York. a.baddeley@york.ac.uk

Table 1. Mean number of correct responses at each test point (with standard error), as a function of age group and final session testing method

	Immediate test	20-minute test	Final (1-week) test
Younger adults			
In person (N=20)	13.85 (.69)	13.35 (.94)	14.05 (.81)
Telephone (N=20)	13.30 (.61)	13.40 (.81)	13.50 (.90)
Older adults			
In person (N=17)	11.76 (1.08)	12.94 (1.22)	11.94 (.93)
Telephone (N=23)	11.30 (.83)	12.35 (1.10)	10.96 (.93)

Figure 1. Recall accuracy at the final test session as a proportion of accuracy in the earlier sessions.

Error bars show standard error.

Figure 1