Reheating breakfast: Age and multitasking on a computer-based and a non-computer-based task

Feinkohl, I. and Cress, U. and Kimmerle, J.

Computers in Human Behavior
2016 FEB ; 55(PtA): 432-438
2015 NOV 05 (first published online)
doi: 10.1016/j.chb.2015.09.040

Publisher: Elsevier

Reheating Breakfast: Age and Multitasking with a Computer-Based Task and a Non-Computer-Based Task

Insa Feinkohl[a*], Ulrike Cress[a,b], Joachim Kimmerle[a]

[a]Knowledge Media Research Center, Tuebingen, Germany

[b]Department of Psychology, University of Tuebingen, Germany

*Insa Feinkohl is now at Max-Delbrueck-Center for Molecular Medicine (MDC), Berlin, Germany

Corresponding author:

Insa Feinkohl
Max-Delbrueck-Center for Molecular Medicine (MDC)
Robert-Roessle-Str.10
D-13092 Berlin
Germany
Email: insa.feinkohl@mdc-berlin.de
Phone: +49 30 9406 4595

Abstract

Computer-based assessments are popular means to measure individual differences, including age differences, in cognitive ability, but are rarely tested for the extent to which they correspond to more realistic behavior. In the present study, we explored the extent to which performance on an existing computer-based task of multitasking ('cooking breakfast') may be generalizable by comparing it with a newly developed version of the same task that required interaction with physical objects. Twenty younger and 29 older adults performed both the computer-based multitasking task and its laboratory-based equivalent. In each task, two measures determined prospective memory and one measure ascertained speed of completion. The Prospective-Retrospective Memory Questionnaire determined self-reported memory failures. In both age groups, correlations between the computer-based and the laboratory version of the task were largely restricted to a single measure of prospective memory. Whereas performance on the laboratory task correlated with self-reported memory failures across the entire sample, there was no such relationship for the computer-based task. Finally, age group by task interactions suggested that younger and older adults were differentially affected by laboratory versus computer-based assessment. Our study illustrates the need for future evaluations of computer-based psychometric instruments on younger and older samples prior to their application.

Keywords: computer-based assessment, older adults, cognitive ageing, prospective memory, multitasking

Reheating Breakfast: Age and Multitasking Measured with a Computer-Based and a Non-Computer-Based Task

1. Introduction

Psychometric instruments are used as tools to quantify individual differences, including age differences, in cognitive abilities and aptitudes. Traditionally, testing has been based on paper-and-pencil tasks, but due to greater precision in data collection, ease and consistency of use, and immunity to biases, these methods were gradually replaced by tasks administered on computers (Deary, Liewald, & Nissan, 2011; Kush, Spring, & Barkand, 2012; Logie, Trawley, & Law, 2011). To justify their continued use, evidence of sufficient ecological validity—the extent to which performance is representative to that in a real-life setting—is imperative (Czaja & Sharit, 2003), but due to time-intensiveness and costs, studies of this kind remain sparse in many domains of psychology (Baumeister, Vohs, & Funder, 2007). For many computer-based instruments in particular, we currently do not have sufficient information on their generalizability beyond the specific task environment.

One rare study that reported data on associations between computer-based assessments and real-life performance produced very limited findings (Lawrence et al., 2004). Children with attention-deficit/hyperactivity disorder (ADHD) and healthy controls performed a real-life task set in a zoo, as well as a virtual environment and two paper-and-pencil tasks in the laboratory. All aimed to assess executive function. Across groups, performance on the real-life task was entirely unrelated to performance on the virtual environment task, which is suggestive of a limited ecological validity of the latter. Comparisons of computer-supported tests with paper-and-pencil tests of cognitive function (Coyne, Warszta, Beadle, & Sheehan, 2005; Ihme et al., 2009; Holt et al., 2011; Parsons & Courtney, 2014) and with psychometric assessments such as of personality (Naus, Phillip, & Samsi, 2009) also often indicate substantial task-dependent differences. In one study of healthy adults, for instance, scores on a paper-and-pencil Stroop test correlated only modestly (though significantly) with scores on a computer-based version of the same task (Parsons et al., 2013). The evidence was also weak for associations between the virtual and the two paper-and-pencil assessments of executive function in the study of children with ADHD and controls (Lawrence et al., 2004).

Thus, computer-based psychometric tests are often applied with little evidence of correlations with other tests thought to measure the same construct, let alone with its real-life applications. This may be particularly problematic in the assessment of cognitive abilities in older populations. Older adults perform cognitive tasks in different ways compared with younger people (Johnson, Logie, & Brockmole, 2010). Moreover, older adults are more commonly affected by a lack of experience with computers and by computer anxiety than are younger people (Broady, Chan, & Caputi, 2010; Slegers, van Boxtel, & Jolles, 2012). Age-related psychomotor slowing also influences the ability to interact with computer keyboard and mouse (Kallus, Schmitt, & Benton, 2005). Thus, in cross-sectional investigations of age effects on cognitive ability, a disproportionate disadvantage for older adults through the assessment on a computer as opposed to performance on a task that does not involve a computer is plausible.

The present study aimed to evaluate this claim for one such task that tapped multitasking abilities. Specifically, the present study compared age effects on performance on a computer-based task that has previously been used to study age effects on multitasking—the computer-based breakfast task (CBBT, Craik & Bialystok, 2006)—with age effects on a similar, though non-computer-based, task that was set in the laboratory (laboratory-based breakfast task, LBBT). The ability to multitask is vital for independent living at any age and has received increased attention in psychological research. It is characterized by the completion of a number of distinct tasks, which are dovetailed and performed in succession, as well as interruptions and the need to delay intentions (Burgess, 2000). Consequently, it necessitates intact planning, retention of plans in prospective memory, switching and executive control functions (Logie et al., 2010). The LBBT was designed to match these general cognitive requirements for multitasking in the CBBT. However, it was closer to a real-life breakfast situation due to taking place in the real (albeit laboratory-based) world and requiring several of the spatial interactions with physical objects that are involved in preparing breakfast. Separate analysis of groups of younger and of older adults allowed evaluation of age-related differences across the two versions of the task. This procedure provided a means to assess whether negative associations of age with cognitive performance reported in the original study by Craik and Bialystok (2006) may present differently when the same cognitive functions are assessed on a task that does not involve a computer. Associations of performance on the two tasks with self-reported frequency of memory failures in everyday life were additionally explored. If CBBT performance correlated with performance on the LBBT, this would provide a cross-validation of the two tasks. Additional correlation with self-reported memory failures would further indicate that individual differences in the abilities tapped by the tasks are recognized by people in their everyday lives.

In contrast, age group by task interactions demonstrating age-specific disadvantages on the CBBT but little or no age effect for the LBBT, in combination with lower correlations of prospective memory performance on the CBBT with self-reported memory failures in everyday life compared with the LBBT, would point to a biased estimation of multitasking abilities in older age by the CBBT.

2. Material and Methods

2.1 Participants

Twenty participants (14 women, six men, M_{age} = 24.9 years, age range: 19-30) were included in a younger group consisting predominantly of university students (55%) who were recruited through posters and flyers distributed in the Psychology Department at the University of Edinburgh, UK. An older group consisted of 30 individuals (20 women, 10 men, M_{age} = 72.0 years, age range: 65-87). The older group was recruited from a Psychology Department panel of volunteer members of the general public. Across age groups, a majority of participants were British, though native English speaking was not an inclusion criterion. Data from one older participant who did not understand instructions were excluded from the analyses, resulting in a total sample size of 49. The study was approved by the Institutional Review Board, and all participants gave full informed consent. Travel expenses were reimbursed for older adults.

2.2 Materials and Procedure

Testing sessions lasted approximately one hour. Initially, demographic information was collected and participants rated their previous experience with computers (*computer experience*) on a 7-point Likert scale (1 = *no experience* and 7 = *frequent use*). The Prospective-Retrospective Memory Questionnaire (PRMQ; Smith, Della Salla, Logie, & Maylor, 2000) was then administered to measure self-reported frequency of retrospective and prospective memory failures. Retrospective memory involves the remembering of information from the past (e.g., autobiographical information). Prospective memory is the remembering of intentions, which are stored either until detection of a cue for its retrieval, or until the appropriate time to carry out the intended action occurs. This form of memory involves a prospective (i.e., remembering to do something) as well as a retrospective component (i.e., remembering the specific intention; McDaniel & Einstein, 1992). On the PRMQ, responses to a total of 16 items were made on 5-point Likert scales (1 = *never* and 5 = *very often*), resulting in a score range of eight to 40 on each of two subscales. A sample item of the retrospective memory subscale is: *Do you fail to recall things that have happened to you in the last few days?* A sample item of the prospective memory subscale is: *Do you decide to do something* *in a few minutes' time and then forget to do it?* Higher scores on each of the subscales reflect a higher frequency of memory failures. The two components of the PRMQ typically correlate strongly with each other and with other self-report measures of episodic memory (Mäntylä, 2003). The instrument has previously been shown to exhibit high reliability (Crawford, Smith, Maylor, Della Salla, & Logie, 2003).

The experimental part of the study involved two breakfast tasks (the CBBT and the LBBT). Both were performed by each participant. Half of the participants in each age group performed the LBBT prior to the CBBT; the other half of each group began with the CBBT and went on to perform the LBBT. The CBBT was performed on a conventional computer monitor using a computer mouse (this task is the '6-screen' version that is described in detail in Craik & Bialystok, 2006). On a main screen, pictures of five breakfast foods each with a different cooking time (ranging from 30 seconds to five minutes) were shown. Each food was presented on a separate screen, with start and stop buttons, and a timer. Participants were instructed to 'cook' each food for the correct time and to ensure that all foods are ready at the same time. Thus, each food had an ideal start time. Participants could enter the screen for each of the respective food as often as they wished. The timer indicated the time left to reach target time for the respective food. As an open-ended distractor task, participants were to 'set a table' with four place settings by dragging images of plates and cutlery to a table presented on the main screen. When all four settings had been completed, the table was cleared on the computer screen for the participant to set four new place settings. Instructions specified that participants were to attempt to complete as many table settings as possible while 'cooking'. Performance of the task began after completion of a practice trial and confirmation by participants that task instructions were understood.

The LBBT resembled the virtual version in terms of design, cooking times and foods, but was set in the laboratory rather than on the computer. Five portable DVD players with nine-inch monitors represented the five foods. Other than these digital proxies for the five foods, all components of the LBBT were set in real life and required participants to interact with physical objects. Participants were informed that the aim, again, was to cook each of the foods for the correct time and to have all foods ready concurrently. In order to begin cooking a food, participants were instructed to press the play button on each DVD player. This resulted in previously recorded video footage of the specific real food being cooked and of a digital timer, which counted down the respective target cooking time, being shown (see Figure 1). The screen of each DVD player was then covered with a cloth by the experimenter, which could be lifted by participants to check on the progress of cooking, that is, to view the timer that indicated the time left before the target time would be reached. No auditory feedback was provided. Again, the open-ended distractor task was to set a table (here, a real table with paper plates and plastic cutlery), which was positioned across from the DVD players at a distance of around two meters. The DVD players were therefore out of sight while participants set the table. When all four settings had been completed, participants continued with the task by completing the next setting and laying plates and cutlery on top of those previously placed on the table, with the aim to complete as many settings as possible during ‘cooking’. During performance of the task, participants could move around freely in the space between the DVD players and the table. Performance again began after a practice session and confirmation that task instructions were understood. It was recorded via a fixed digital video camera for later coding of the number of times that the cloth was lifted for each respective food as well as for start times of each food. The number of table settings completed was counted by the experimenter upon task completion. The order of task performance is illustrated in Figure 2.

2.3 Outcome Measures

Outcome measures were equivalent in the two breakfast tasks. Prospective memory was determined by the participant's number of *clock checks* performed during cooking and by *mean deviation,* calculated as the mean difference (in seconds) between ideal and actual start times for cooking each of the five foods. A higher number of clock checks and a lower mean deviation represented higher prospective memory performance. Performance on the distractor task of setting a table was also recorded. Each complete table (four plates and four sets of cutlery dragged or put to the correct locations on the table) was counted as 'one' *table setting*. A high number of table settings represented high speed of task performance.

3. Results

3.1 Age Group Differences in Demographics and PRMQ

Two-tailed *t*-tests were performed to determine possible group differences in education, computer experience, and memory failures self-reported on the PRMQ. Years spent in education did not differ between the older (M = 15.07 years, SD = 3.36) and the younger group (M = 16.45 years, SD = 2.84; $t(47) = 1.50$, $p = .139$), but older adults reported less computer experience (M = 2.41, SD = 2.10) compared with the younger group (M = 3.70, SD = 1.38; $t(47) = 2.40$, $p = .020$). Self-reported frequency of prospective memory failure on the PRMQ was higher in the younger (M = 21.30, SD = 4.93) than in the older group (M = 17.55, SD = 4.01; $t(47) = 2.93$, $p = .005$). Self-reported frequency of retrospective memory failures on the PRMQ was similar for younger (M = 18.05, SD = 3.52) and for older adults (M = 16.52, SD = 3.31; $t(47) = 1.55$, $p = .127$).

3.2 Correlations among Breakfast Task Measures and between Breakfast Tasks and PRMQ

Two-tailed Pearson correlations across all measures are shown in Table 1 for the younger group and in Table 2 for the older group. In the younger group, there was a significant positive association between CBBT and LBBT clock checks and between CBBT table settings and LBBT mean deviation (see Table 1).

For the older group, associations between the two breakfast tasks were restricted to a positive relationship of CBBT and LBBT clock checks (see Table 2). When computer experience was entered as a covariate in partial correlations again performed separately for each age group, only the respective associations of CBBT and LBBT clock checks remained statistically significant (data not shown; both $p<.05$). A statistically non-significant trend for a positive association of CBBT mean deviation with LBBT clock checks in the younger group (see Table 1) reached statistical significance when adjustment was made for computer experience (r=.60; p=.012). Mean deviation and table settings performance were both unrelated between the two tasks in the unadjusted analysis and following adjustment for computer experience.

Further Pearson correlations revealed a positive association between the retrospective and the prospective memory components of the PRMQ in the younger group ($r = .78$, $p<.001$) and in the older group ($r = .76$, $p<.001$.). Across the entire sample, both types of self-reported memory failures correlated negatively with prospective memory performance in the LBBT in terms of clock checks ($r = -.28$, $p = .052$ for the prospective, and $r = -.34$, $p = .018$ for the retrospective memory component). In the CBBT, in contrast, the prospective memory performance measures did not correlate with the retrospective or the prospective memory components of the PRMQ (*p* range .24 to .72). Thus, participants with greater self-reported memory failure overall had lower prospective memory performance on the LBBT, but not on the CBBT. Only CBBT table settings correlated positively with the prospective component of the PRMQ ($r = .40$, $p = .004$) and was also marginally associated with the retrospective component of the PRMQ ($r = .27$, $p = .064$), showing that people with greater self-reported memory problems were faster during CBBT performance.

3.3 Age Group by Task Interaction in Breakfast Task Performance

In order to determine whether breakfast task performance levels in older and younger adults were differently affected by the two versions of tasks, mixed-design analyses of variance (ANOVA) with age group as between-subjects variable and task as within-subjects variable were performed. A statistically significant interaction for age group by task was found for clock checks ($F(1, 47) = 9.23$, $p = .004$, partial $\eta^2 = .16$; see Figure 3) and for table settings ($F(1, 47) = 78.33$, $p < .001$, partial $\eta^2 = .63$; see Figure 4). In these analyses, the main effect of task was significant for clock checks ($F(1,47) = 79.36$, $p<.001$, partial $\eta^2 = .63$) and for table settings ($F(1,47) = 173.83$, $p<.001$, partial $\eta^2 = .79$), as was the main effect of age group on table settings ($F(1,47) = 101.22$, $p<.001$, partial $\eta^2 = .68$). There was no main effect of age group on clock checks ($F(1,47) = .74$, $p=.394$, partial $\eta^2 = .02$).

From Figures 3 and 4, it is clear that these interaction effects are driven by the computer-based task generating a larger difference between the two age groups in clock checks and table settings relative to the laboratory-based version of the task. These interactions remained significant after computer experience was entered as a covariate into the analyses (for clock checks, $F(1, 46) = 4.90$, $p = .032$, partial $\eta^2 = .09$; for table settings, $F(1, 46) = 69.95$, $p < .001$, partial $\eta^2 = .60$). Despite a significant main effect of age group on mean deviation in unadjusted analyses ($F(1,45) = 12.90$, $p=.001$, partial $\eta^2 = .22$), there were no main effect of task ($F(1,45) = 2.37$, $p=.131$, partial $\eta^2 = .05$) and no interaction effect on that outcome ($F(1, 45) = 1.45$, $p = .234$, partial $\eta^2=0.03$; see Figure 5).

4. Discussion

4.1 Summary of Findings

The present article compared the performance of older and younger adults on a simulation of a task of multitasking that was set in the laboratory and had minimal involvement of technology (LBBT) with that on a computer-based simulation of this task (CBBT). Associations of performance on these tasks with self-reported memory failures in everyday life were also explored. We found a weak positive association between LBBT clock checks and CBBT mean deviation (both measuring prospective memory) in the younger group, which only became apparent following adjustment for computer experience. This speaks *against* measurement of the same construct by both tasks (high clock checks but lower mean deviation mean better prospective memory). The number of clock checks was the only measure to correlate between the two versions of the task in both age groups and irrespective of statistical adjustment.

Associations with self-reported frequency of memory failures in everyday life were limited to the LBBT. Together with the intuitive assumption of a higher ecological relevance in the LBBT than in the CBBT, we conclude on the basis of the correlations with self-reported memory failures for LBBT but not CBBT that the usefulness of the latter to measure real-life multitasking ability may be limited and requires further investigation. Additionally, age group by task interactions suggested differential age group effects on prospective memory (clock checks) and speed (table settings) in the CBBT compared with the LBBT. This cast further doubt on a successful cross-validation of both tasks.

4.2 Potential Mechanisms Underlying the Observations and Comparison with Previous Research

Although a lower previous exposure to computers in older than in younger adults (for which we also found evidence in our sample) could be assumed as a likely driving force behind the observation, statistical adjustment for self-reported computer experience did not change these findings. Our observation thereby contrasts with a previous investigation, which had identified computer experience as a mediator between age and performance on a computer-based task of memory (Laguna & Babcock, 2000). It may be the case that other factors, such as an age-related decline in processing speed, contributed to the disadvantage for the older adults of our sample on the computer-based task. Our measurement of computer experience on a single-item scale was imprecise, however, and so the contribution of computer experience to the present findings warrants further investigation. Future studies could make use of more precise instruments (e.g., the Computer Proficiency Questionnaire; Boot et al., 2015) for this purpose. Though implausible, we also cannot rule out the possibility that the observed interaction effect represented an unfair advantage of the older adults on the laboratory task (rather than a disadvantage on the computer-based task). We lack a 'yardstick' or 'gold-standard' which would allow ultimate determination of the ecological validity of either task.

The apparent positive association of CBBT mean deviation and LBBT clock checks in the younger group is puzzling, because both are measures of prospective memory but indicate better (clock checks) versus worse (mean deviation) ability. It is possible that participants who had a preference for the more realistic version of the task were less motivated when performing the task on the computer. We did not measure task preference and so the finding leaves scope for future research.

Finally, our study has shed light on some previous research of time monitoring abilities. In two early investigations, groups of younger and older adults were to press a key on a keyboard at certain time intervals, and older adults were consistently outperformed by younger adults in terms of accuracy and clock checking (Einstein et al., 1995; Park, Hertzog, Kidder, Morell, & Mayhorn, 1997). The interaction effect of age group by task on clock checks in combination with a lack of a main effect of age on this outcome in the present study has now shown that the use of a computer may have contributed to those early reports of age-related deficits in time monitoring. In one more recent investigation, older adults were indeed outperformed by younger adults in clock checking despite the task being set in real life (Mioni & Stablum, 2014), however, and so the overall disparity in observations of age effects on time monitoring requires clarification in future, ideally prospective, studies.

4.3 Implications

Our findings have implications first and foremost for future psychological research. It appears that poor performance by older adults on computer-based tasks may not always reflect true underlying deficits and so should not always be taken at face value by investigators. We have therefore highlighted a need for the evaluation of cognitive tests as to the degree that they are ecologically valid prior to their application, which is consistent with a recent review (Phillips, Henry, & Martin, 2012) concluding that in older adults ecological validity is a crucial factor in the assessment of prospective memory (which is an important component of multitasking). As an earlier paper on ecological validity aptly notes, “it is necessary to study real behavior sometimes” (Baumeister et al., 2007, p. 400; see also Neisser, 1976). Despite problems such as scaling aspects of real-world situations into experimental platforms (Czaja & Sharit, 2003), researchers are therefore encouraged to attempt a recreation of computer-based tasks in simulations with lesser digital involvement wherever possible. For tasks of prospective memory, Phillips et al. (2012) call for strategic investigations along a continuum of ecological validity ranging from tasks set in everyday life to artificial tasks in laboratory settings. With the creation of the LBBT in an ecologically relevant environment, we have provided an important step toward that goal.

One previous (unpublished) study applied a real-life cooking task to a sample of older adults to examine multitasking (Edwards & Ryan, 2004; cited in Craik & Bialystok, 2006). Participants also performed the CBBT in that study. Correlations of medium effect size were found between the number of sequencing steps (a measure of multitasking ability) in real-life and CBBT mean deviation. However, deviation was not measured in the real-life cooking task, and the task was subject to little experimental control and to potential effects of routine (Bergua et al., 2006). Our LBBT has supplemented this previous evidence by allowing a greater maintenance of experimental control. A similar approach to ours has since been described by another group. Here, performance on a realistic breakfast task ('Dresden Breakfast Task'), which was also based on Craik and Bialystok's (2006) CBBT, was associated with performance on other paper-and-pencil tests of cognitive ability. Yet, potential associations of that task with the CBBT itself were not explored (Altgassen et al., 2012). Direct comparison of the Dresden Breakfast Task with the CBBT as well as the LBBT in a single sample would now be a useful next step for future research in the field.

Research of age effects on time monitoring could also prove useful for health promotion in older adults. Should future assessments confirm age effects on time monitoring behavior, for which the evidence appears unclear at present, then older adults could benefit from support systems in form of alarms during important real-life tasks that rely on time monitoring, such as cooking. Implementation of such strategies could then ultimately help prevent domestic accidents which older adults are at particular risk of (Bhanderi & Choudhary, 2008).

4.3 Limitations and future research

A limitation of the present study is its relatively small sample size. Effect sizes and directions of associations suggest that further investigation using a larger sample may be useful. We also did not assess participants' personality and attitudes, which may play a role in the way in which people answer health-related questions (Cipolletta, Consolaro, & Hovarth, 2014) and so potentially influenced participants' self-reported memory failures on the PRMQ, or any other factors, such as participants' experience with cooking, that may have systematically influenced performance on the tasks.

Despite these shortcomings, our study suggests that an alteration of task designs of those that take place on a computer rather than in real life may be helpful in eliminating any spurious task-related age group differences in performance. After all, a majority of today's psychological research use computer-based tasks, and a continued use of this method is likely.

Recent developments of virtual environments appear to offer a promising alternative to traditional computer-based assessments such as the CBBT. In studies of neuropsychological patients, for instance, performance on virtual environment tasks involving the completion of multiple errands has been found to correlate with real-life performance of the same task (Grewe et al., 2014; McGeorge et al., 2001) or with the clinical descriptions of patients' deficits in this ability (Titov & Knight, 2005). It is plausible that older adults, too, could benefit from such task designs, particularly in view of evidence that computer-based assessment may reduce test-induced stress experienced by older adults (Collerton et al., 2007).

Recently, a virtual environment test of prospective memory was identified as a successful means for assessing working memory, planning, and prospective memory in young adults (Edinburgh Virtual Errands Task, EVET; Logie, Trawley, & Law, 2011; Trawley, Law, & Logie, 2011). However, subsequent unpublished pilot work with the present sample and with a separate group of participants showed that older adults had difficulty with the keyboard and mouse interface, and testing was terminated prematurely due to stress or excessively long testing sessions. Ongoing pilot work with older people using alternative interfaces to the virtual environment has proved more successful but is still under development. Therefore, if future studies use virtual reality environments to study older adults, these should be developed specifically to suit the target population and, as all computer-based tasks, should be tested both for ease of use by the participant sample and ecological validity prior to their use.

5. Conclusions

Taken together, we have highlighted the importance of evaluating computer-based tasks in terms of their real-life applicability by demonstrating performance disparity between one such task and an alternative with minimal digital involvement. Particularly in the cross-sectional comparison of older with younger adults, the use of a computer could hinder the accurate measurement of underlying cognitive abilities. The problem may well become less pronounced with time, as the current younger generation might sustain their level of computer proficiency into older age. People currently in the middle age range of 40-65 may also be much more familiar with the use of computers in their daily personal and working lives than are people who entered retirement before computers became commonplace. For now, our study has demonstrated that the issue requires further investigation as well as ongoing consideration in the design of psychological experiments. Scores on computer-based assessments should not always be taken at face value to reflect underlying ability without scrutiny, and so researchers are encouraged to 'think outside the box' in the evaluation of their tasks to determine whether or not research findings (particularly when using older samples) are transferable to real-life settings.

Acknowledgements

The work reported here was conducted in 2010 as a Masters research project of I.F. at the University of Edinburgh, UK, and was supported in part by the Centre for Cognitive Ageing and Cognitive Epidemiology at the University of Edinburgh. I.F. is grateful to Professor Robert H. Logie for guidance in the design, conduct of the research, and in the writing of the current manuscript.

References

Altgassen, M., Koban, N. & Kliegel, M. (2012). Do adults with autism spectrum disorders compensate in naturalistic prospective memory tasks? *Journal of Autism and Developmental Disorders, 42,* 2141-2151. doi: 10.1007/s10803-012-1466-3

Bailey, P. E., Henry, J. D., Rendell, P. G., Phillips, L. H., & Kliegel, M. (2010). Dismantling the "age–prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting. *Quarterly Journal of Experimental Psychology, 63*, 646–652. doi: 10.1080/17470210903521797

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403. doi:10.1111/j.1745-6916.2007.00051.x

Bergua, V., Fabrigoule, C., Barberger-Gateau, P., Dartigues, J.-F., Swendsen, J., & Bouisson, J. (2006). Preferences for routines in older people: Associations with cognitive and psychological vulnerability. *International Journal of Geriatric Psychiatry, 21*, 990–998. doi:10.1002/gps.1597

Bhanderi, D. J. & Choudhary, S. (2008). A study of occurrence of domestic accidents in semi-urban community. *Indian Journal of Community Medicine, 33,* 104-106. doi: 10.4103/0970-0218.40878

Boot, W. R., Charness, N., Czaja, S. J., Sharit, J., Rogers, W. A, Fisk, A. D., Mitzner, T, Lee, C. C., & Nair, S. (2015). Computer Proficiency Questionnaire: Assessing low and high computer proficient seniors. *The Gerontologist*, *55*, 404-411. doi:10.1093/geront/gnt117

Broady, T., Chan, A., & Caputi, P. (2010). Comparison of older and younger adults' attitudes towards and abilities with computers: Implications for training and learning. *British*

Journal of Educational Technology, *41*, 473–485. doi:10.1111/j.1467-8535.2008.00914.x

Burgess, P. W. (2000). Strategy application disorder: The role of the frontal lobes in human multitasking. *Psychological Research, 63,* 279-288. doi: 10.1007/s004269900006

Cipolletta, S., Consolaro, F., & Horvath, P. (2013). When health is an attitudinal matter. A qualitative research. *Journal of Humanistic Psychology, 54,* 391-413. doi: 10.1177/0022167813507630

Collerton, J., Collerton, D., Arai, Y., Barass, K., Eccles, M., Jagger, C., McKeith, I., Saxby, B. K., & Kirkwood, T. (2007). A comparison of computerized and pencil-and-paper tasks in assessing cognitive function in community-dwelling older people in the Newcastle 85+ Pilot Study. *Journal of the American Geriatrics Society, 55,* 1630-1635. doi:10.1111/j.1532-5415.2007.01379.x

Coyne, I., Warszta, T., Beadle, S., & Sheehan, N. (2005). The impact of mode of administration on the equivalence of a test battery: A quasi-experimental design. *International Journal of Selection and Assessment*, *13*, 220–224. doi:10.1111/j.1468-2389.2005.00318.x

Craik, F. I. M., & Bialystok, E. (2006). Planning and task management in older adults: Cooking breakfast. *Memory & Cognition*, *34*, 1236–1249. doi:10.3758/BF03193268.

Crawford, J. R., Smith, G., Maylor, E. A., Della Salla, S., & Logie, R. H. (2003). The Prospective and Retrospective Memory Questionnaire (PRMQ): Normative data and latent structure in a large non-clinical sample. *Memory*, *11*, 261–275. doi:10.1080/09658210244000027

Czaja, S. J., & Sharit, J. (2003). Practically relevant research: Capturing real world tasks, environments, and outcomes. *The Gerontologist*, *43*(Supplement I), 9–18. doi:10.1093/geront/43.suppl_1.9

Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: The Deary-Liewald reaction time task. *Behavior Research Methods*, *43*, 258–68. doi:10.3758/s13428-010-0024-1

Dubreuil, P., Adam, S., Bier, N., & Gagnon, L. (2007). The ecological validity of traditional memory evaluation in relation with controlled memory processes and routinization. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *22*, 979–989. doi:10.1016/j.acn.2007.08.002

Einstein, G. O, McDaniel, M. A., Richardson, S. L., Guynn, M. J., & Cunfer, A. R. (1995). Aging and prospective memory: examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 996-1007. doi: 10.1037/0278-7393.21.4.996

Grewe, P., Lahr, D., Kohsik, A., Dyck, E., Markowitsch, H. J., Bien, C. G., Botsch, M, & Piefke, M. (2014). Real-life memory and spatial navigation in patients with focal epilepsy: Ecological validity of a virtual reality supermarket task. *Epilepsy & Behavior, 31*, 57–66. doi:10.1016/j.yebeh.2013.11.014

Holt, D. V., Rodewald, K., Rentrop, M., Funke, J., Weisbrod, M., & Kaiser, S. (2011). The Plan-a-Day approach to measuring planning ability in patients with schizophrenia. *Journal of the International Neuropsychological Society, 17,* 327-335. doi:10.1017/S1355617710001712

Ihme, J. M., Lemke, F., Lieder, K., Martin, F., Müller, J. C., & Schmidt, S. (2009). Comparison of ability tests administered online and in the laboratory. *Behavior Research Methods*, *41*, 1183–1189. doi:10.3758/BRM.41.4.1183

Johnson, W., Logie, R. H., & Brockmole, J. R. (2010). Working memory tasks differ in factor structure across age cohorts: Implications for dedifferentiation. *Intelligence, 38, 513–528*. doi:10.1016/j.intell.2010.06.005

Kallus, K. W., Schmitt, J. A. J., & Benton, D. (2005). Attention, psychomotor functions and age. *European Journal of Nutrition*, *44*, 465–484. doi:10.1007/s00394-005-0584-5

Kush, J. C., Spring, M. B., & Barkand, J. (2012). Advances in the assessment of cognitive skills using computer-based measurement. *Behavior Research Methods*, *44*, 125–134. doi:10.3758/s13428-011-0136-2

Laguna, K. D., & Babcock, R. L. (2000). Computer testing of memory across the adult life span. *Experimental Aging Research*, *26*, 229–243. doi:10.1080/036107300404877

Lawrence, V., Houghton, S., Douglas, G., Durkin, K., Whiting, K., & Tannock, R. (2004). Executive function and ADHD: A comparison of children's performance during neuropsychological testing and real-world activities. *Journal of Attention Disorders, 7,* 137-149. doi:10.1177/108705470400700302

Logie, R. H., Law, A., Trawley, S. & Nissan, J. (2010). Multitasking, working memory and remembering intentions. *Psychologica Belgica, 50,* 309-326.

Logie, R. H., Trawley, S., & Law, A. S. (2011). Multitasking: Multiple, domain-specific cognitive functions in a virtual environment. *Memory and Cognition, 39*, 1561–1574. doi: 10.3758/s13421-011-0120-1.

Mäntylä, T. (2003). Assessing absentmindedness: Prospective memory complaint and impairment in middle-aged adults. *Memory and Cognition*, *31*, 15–25. doi:10.3758/BF03196078

McDaniel, M. A. & Einstein, G. O. (1992). Aging and prospective memory: Basic findings and practical applications. *Advances in Learning and Behavioral Disabilities, 7,* 87-105.

McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Della Sala, S., Milne, A. B., Hamilton, S., & Callender, J. S. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence, 10,* 375–383. doi:10.1162/1054746011470235

Mioni, G. & Stablum, F. (2014). Monitoring behaviour in a time-based prospective memory task: The involvement of executive functions and time perceptions. *Memory, 22,* 536-552. doi:10.1080/09658211.2013.801987

Naus, M. J., Philip, L. M., & Samsi, M. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior, 25,* 1-7. doi:10.1016/j.chb.2008.05.012

Neisser, U. (1976). *Cognition and Reality*. San Francisco: Freeman.

Park, D. C., Hertzog, C., Kidder, D. P., Morrell, R. W. & Mayhorn, C. B. (1997). Effect of age on event-based and time-based prospective memory. *Psychology and Aging, 12,* 314-329. doi: 10.1037/0882-7974.12.2.314

Parsons, T. D., & Courtney, C. (2014). An initial validation of the Virtual Reality Paced Auditory Serial Addition Test in a college sample. *Journal of Neuroscience Methods*, *222*, 15–23. doi:10.1016/j.jneumeth.2013.10.006

Phillips, L. H., Henry, J. D., & Martin, M. (2012). Adult aging and prospective memory: The importance of ecological validity. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives* (pp. 161–186). New York: Psychology Press.

Rendell, P. G., & Craik, F. I. M. (2000). Virtual week and actual week: Age-related differences in prospective memory. *Applied Cognitive Psychology*, *14*, S43–S62. doi:10.1002/acp.770

Roediger, H. L. I. (1996). Prospective memory and episodic memory. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective Memory: Theory and Applications* (pp. 149–155). Mahwah, NJ: Lawrence Erlbaum Associates.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*, 727–741. doi:10.1093/brain/114.2.727

Slegers, K., van Boxtel, M. P. J., & Jolles, J. (2012). Computer use in older adults: Determinants and the relationship with cognitive change over a 6 year episode. *Computers in Human Behavior, 28,* 1-10. doi:10.1016/j.chb.2011.08.003

Smith, G., Della Salla, S., Logie, R. H., & Maylor, E. A. (2000). Prospective and retrospective memory in normal ageing and dementia: A questionnaire study. *Memory*, *8*, 311–321. doi:10.1080/09658210050117735

Titov, N., & Knight, R. G. (2005). A computer-based procedure for assessing functional cognitive skills in patients with neurological injuries: The virtual street. *Brain Injury*, *19*, 315–322. doi:10.1080/02699050400013725

Trawley, S. L., Law, A. S., & Logie, R. H. (2011). Event-based prospective remembering in a virtual world. *Quarterly Journal of Experimental Psychology*, *64*, 2181–2193. doi:10.1080/17470218.2011.584976

Table 1

Associations among breakfast task outcome measures in the younger group.

	CBBT		
LBBT	**Mean deviation**	**Clock checks**	**Table settings**
Mean deviation	-.06 (.801)	-.05 (.822)	**.45 (.048)**
Clock checks	.43 (.075)	**.69 (.001)**	-.43 (.062)
Table settings	.13 (.611)	-.34 (.139	.03 (.908)

Values are Pearson correlation coefficients (*p*-values).

Table 2

Associations among breakfast task outcome measures in the older group.

	CBBT		
LBBT	**Mean deviation**	**Clock checks**	**Table settings**
Mean deviation	-.04 (.842)	-.21 (.283)	-.24 (.205)
Clock checks	.29 (.129)	**.73 (<.001)**	-.20 (.305)
Table settings	-.08 (.668)	-.02 (.928)	-.11 (.586)

Values are Pearson correlation coefficients (*p*-values).

Figure 1. Screen shot of DVD player representing ‘eggs’ in the LBBT.

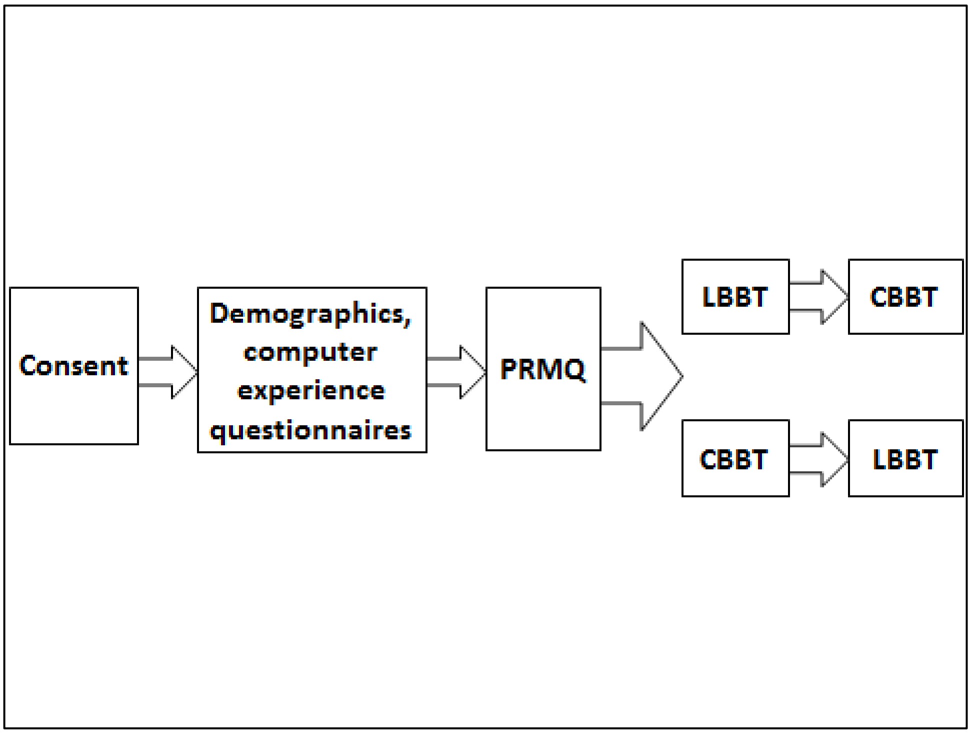

Figure 2. Order of task administration. All participants performed all tasks. One half of participants in each age group performed the LBBT first; the other half performed the CBBT first.

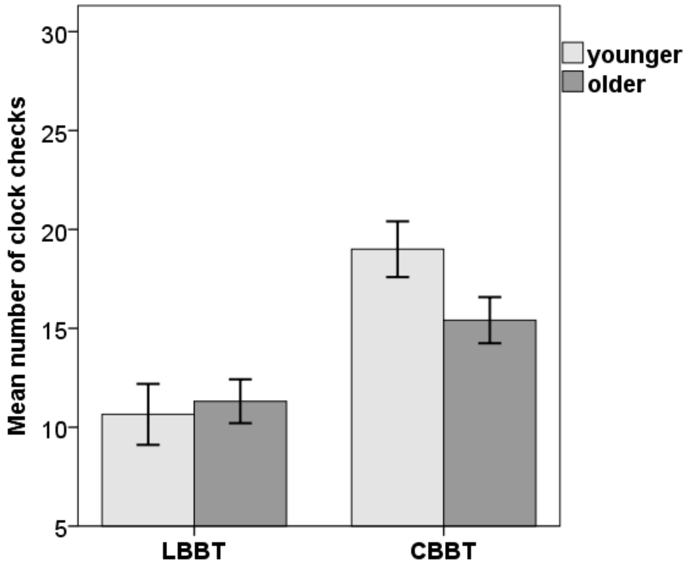

Figure 3. Mean number of clock checks performed by the younger and older group in the LBBT and the CBBT (error bars show standard errors).

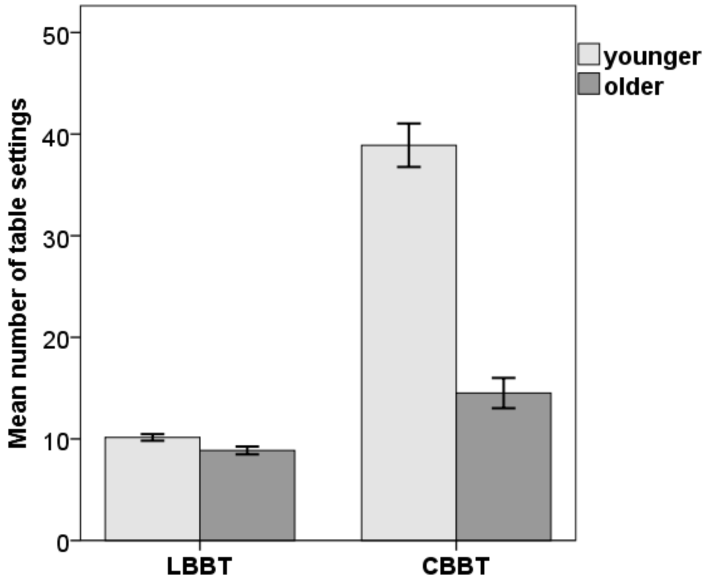

Figure 4. Mean number of table settings completed by the younger and older group in the LBBT and the CBBT (errors bars show standard errors).

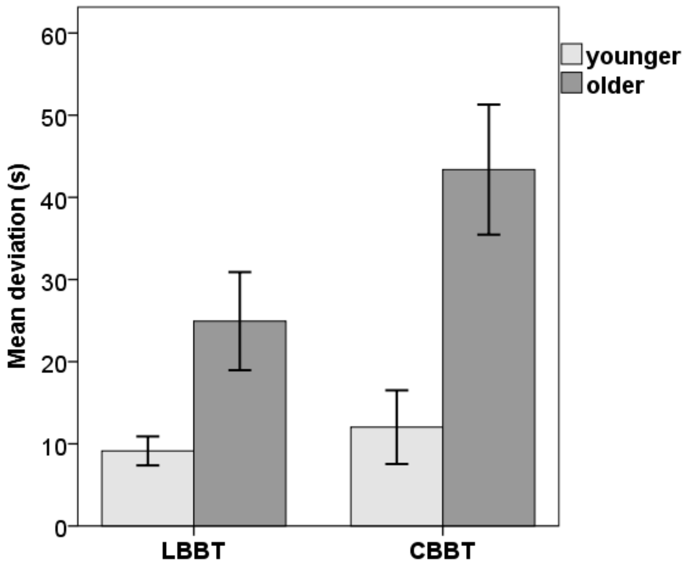

Figure 5. Mean deviation (s) in the younger and older group in the LBBT and the CBBT (error bars show standard errors).